Astronomy & Astrophysics manuscript no. Planck2011-1.6
January 12, 2011

Planck Early Results: The Low Frequency Instrument data processing

A. Zacchei[33] ⋆, D. Maino[20,35], C. Baccigalupi[50], M. Bersanelli[20,35], A. Bonaldi[31], L. Bonavera[50,5], C. Burigana[34], R. C. Butler[34], F. Cuttaia[34], G. de Zotti[31,50], J. Dick[50], M. Frailis[33], S. Galeotta[33], J. González-Nuevo[50], K. M. Górski[44,55], A. Gregorio[21], E. Keihänen[13], R. Keskitalo[44,13], J. Knoche[48], H. Kurki-Suonio[13,28], C. R. Lawrence[44], S. Leach[50], J. P. Leahy[45], M. López-Caniego[43], N. Mandolesi[34], M. Maris[33], F. Matthai[48], P. R. Meinhold[16], A. Mennella[20,33], G. Morgante[34], N. Morisset[37], P. Natoli[22,2,34], F. Pasian[33], F. Perrotta[50], G. Polenta[2,32], T. Poutanen[28,13,1], M. Reinecke[48], S. Ricciardi[34], R. Rohlfs[37], M. Sandri[34], A.-S. Suur-Uski[13,28], J. A. Tauber[26], D. Tavagnacco[33], L. Terenzi[34], M. Tomasi[20,35], J. Valiviita[41], F. Villa[34], A. Zonca[16], A. J. Banday[53,7,48], R. B. Barreiro[43], J. G. Bartlett[4,44], N. Bartolo[18], L. Bedini[6], K. Bennett[26], P. Binko[37], J. Borrill[47,51], F. R. Bouchet[40], M. Bremer[26], P. Cabella[22], B. Cappellini[35], X. Chen[38], L. Colombo[12,44], M. Cruz[43], A. Curto[43], L. Danese[50], R. D. Davies[45], R. J. Davis[45], G. de Gasperis[22], A. de Rosa[34], G. de Troia[22], C. Dickinson[45], J. M. Diego[43], S. Donzelli[35,41], U. Dörl[48], G. Efstathiou[54], T. A. Enßlin[48], H. K. Eriksen[41], M. C. Falvella[3], F. Finelli[34], E. Franceschi[34], T. C. Gaier[44], F. Gasparo[33], R. T. Génova-Santos[42,24], G. Giardino[26], F. Gómez[42], A. Gruppuso[34], F. K. Hansen[41], R. Hell[48], D. Herranz[43], W. Hovest[48], J. Jewell[44], M. Juvela[13], T. S. Kisner[47], L. Knox[15], A. Lähteenmäki[1,28], J.-M. Lamarre[46], R. Leonardi[25,26,16], J. León-Tavares[1], P. B. Lilje[41,9], P. M. Lubin[16], G. Maggio[33], D. Marinucci[23], E. Martínez-González[43], M. Massardi[31], S. Matarrese[18], M. T. Meharga[37], A. Melchiorri[19], M. Migliaccio[22], S. Mitra[44], A. Moss[11], H. U. Nørgaard-Nielsen[10], L. Pagano[44], R. Paladini[52,8], D. Paoletti[34], B. Partridge[27], D. Pearson[44], V. Pettorino[50], D. Pietrobon[44], G. Prézeau[8,44], P. Procopio[34], J.-L. Puget[39], C. Quercellini[22], J. P. Rachen[48], R. Rebolo[42,24], G. Robbers[48], G. Rocha[44,8], J. A. Rubiño-Martín[42,24], E. Salerno[6], M. Savelainen[13,28], D. Scott[11], M. D. Seiffert[44,8], J. I. Silk[17], G. F. Smoot[14,47,4], J. Sternberg[26], F. Stivoli[36], R. Stompor[4], G. Tofani[29], J. Tuovinen[49], M. Türler[37], G. Umana[30], P. Vielva[43], N. Vittorio[22], C. Vuerli[33], L. A. Wade[44], R. Watson[45], S. D. M. White[48], and A. Wilkinson[45]

(Affiliations can be found after the references)

Preprint online version: January 12, 2011

ABSTRACT

We describe the data processing pipeline employed by the Low Frequency Instrument (LFI) Data Processing Centre (DPC) to create and characterize the frequency maps used by the ERCSC (Early Release Compact Source Catalogue) first product of *Planck* to become public. In particular, we discuss the various steps involved in reducing the data, starting from telemetry (TM) packets through to the production of cleaned calibrated timelines and calibrated frequency maps. Data are continuously calibrated using the modulation induced on the mean temperature of the Cosmic Microwave Background Radiation by the proper motion of the spacecraft. The sky signals other than the dipole are removed by an iterative procedure based on simultaneous fitting of calibration parameters and sky maps. Noise properties are estimated from time-ordered data where the sky signal is removed using a Generalized Least Square map-making algorithm. The measured $1/f$ noise knee-frequencies range from ~ 100 mHz at 30 GHz to a few tens of mHz at 70 GHz. A destriping code (`Madam`) is employed to combine radiometric data and pointing information into sky maps, minimizing the variance of correlated noise. Noise covariance matrices required to compute statistical uncertainties on LFI and *Planck* products are also produced. Main beams are estimated down to the ≈ -10 dB level using Jupiter transits which are also used for the geometrical calibration of the focal plane.

Key words. methods:data analysis; cosmology: cosmic microwave background

1. Introduction

Planck [1] (Tauber et al. 2010a; Planck Collaboration 2011a) is a third generation space mission to measure the anisotropy of the cosmic microwave background (CMB). It observes the sky in nine frequency bands covering 30–857 GHz with high sensitivity and angular resolution from 31′to 5′. The Low Frequency Instrument (LFI) (Mandolesi et al. 2010; Bersanelli et al. 2010; Mennella et al. 2011) covers the 30, 44, and 70 GHz bands with amplifiers cooled to 20 K. The High Frequency Instrument (HFI) (Lamarre et al. 2010; Planck HFI Core Team 2011a) covers the 100, 143, 217, 353, 545, and 857 GHz bands with bolometers cooled to 0.1 K. Polarization is measured in all but the highest two bands (Leahy et al. 2010; Rosset et al. 2010). A combination of radiative cooling and three mechanical coolers produces the temperatures needed for the detectors and optics (Planck Collaboration 2011b). Two Data Processing Centres (DPCs), conceived as interacting and complementary since the earliest design of Planck Scientific Ground Segment (Pasian & Gispert 2000), check and calibrate the

⋆ Corresponding author; email: zacchei@oats.inaf.it.

[1] *Planck* (http://www.esa.int/*Planck*) is a project of the European Space Agency (ESA) with instruments provided by two scientific consortia funded by ESA member states (in particular the lead countries France and Italy), with contributions from NASA (USA) and telescope reflectors provided by a collaboration between ESA and a scientific consortium led and funded by Denmark.

data and make maps of the sky (Planck HFI Core Team 2011b; Zacchei et al. 2011). *Planck*'s sensitivity, angular resolution, and frequency coverage make it a powerful instrument for galactic and extragalactic astrophysics, as well as cosmology. Early astrophysics results are given in Planck Collaboration, 2011h–z.

The Low Frequency Instrument LFI on-board the *Planck* satellite is a set of 11 Radiometer Chain Assemblies (RCA), each of them composed of two independent pseudo-correlation radiometers. A total of two RCA are available at 30 GHz, three at 44 and six at 70 GHz. Each of the radiometers has two independent diodes for the detection, integration and conversion from RF signal into DC Voltages. The LFI is cryogenically cooled to 20 K to improve noise figures while the pseudo-correlation design with a reference target at ≈ 4 K ensures a good suppression of $1/f$ noise (see Mennella et al. 2011, for further details).

LFI produces full-sky maps centered at 30, 44 and 70 GHz, with significant improvements with respect to current CMB data in the same frequency range. The complexity of the *Planck* satellite, and specifically of the LFI instrument, demands careful data processing to ensure that systematic effects are limited to a few μK per resolution element. Fidelity of the maps at this level is required to achieve the ambitious science goals of the *Planck* project.

In this paper we will describe the processing steps implemented to create LFI (and therefore *Planck*) data products with particular attention to the needs of the first set of astrophysics results.

The structure of this paper follows the data flow for the nominal data analysis pipeline implemented at the LFI DPC. The first basic step consists in the creation of Time Ordered Information (TOI) from Telemetry (TM) packets delivered by the satellite, as well as time stamping, pointing reconstruction and data flagging (see Section 2).

In Section 3 we describe the main operations performed on the TOI which comprise removal of frequency spikes, the creation of differenced data, computing an optimal Gain Modulation Factor to reduce $1/f$ noise (see Mennella et al. 2011, for further details) and diode combination to mitigate the effects of imperfect gain balance between the two radiometers legs and to produce independent noise time streams which makes the treatment of error propagation easier in subsequent steps of the pipeline.

The reconstruction of the main beams via Jupiter transits and the calibration of the focal plane geometry are discussed in Section 4. The photometric calibration pipeline is discussed in detail in Section 5.

Knowledge of instrument noise properties is essential for quantitative scientific analysis of the LFI data: the white noise level allows us to create noise covariance matrices used for component separation and power spectrum estimation; residual $1/f$ noise is used for the creation of noise covariance matrices at low resolution, where a maximum likelihood approach to estimating the power spectrum is feasible. Section 6 reports results on TOI noise properties. Both noise and beam figures are reported and discussed in detail in Mennella et al. (2011).

Map-making, the creation of white noise covariance matrices and various tests based on jackknife analysis and Monte Carlo simulations, are presented in detail in Section 7.

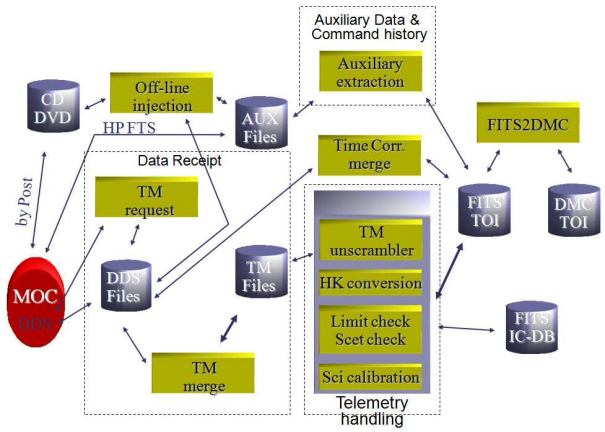

Fig. 1. Level 1 pipeline structure.

Section 8 reports the computation of Colour correction done on LFI maps, Section 9 describes the process applied to estimate and remove the CMB from LFI and HFI maps.

Finally Section 10 give an overview of the software infrastructure used at LFI DPC.

2. Creation of Time Ordered Information (TOI)

The principal task of the first level of the LFI DPC pipeline is the creation of scientific TOI and House-Keeping (H/K) data. The scope is to retrieve all the necessary information from packets received each day from the Mission Operation Centre (MOC) and transform it into a form that is *manageable* by the scientific pipeline.

During the Daily Tele-Communication Period (DTCP), lasting ≈ 3 hours per day, the MOC receives from the *Planck* satellite a dump of the telemetry from the entire previous day, archived in the on-board mass memory, together with real-time telemetry. Additional auxiliary files, such as the attitude history of the satellite (AHF), are produced.

Before delivering the dumped telemetry, the MOC consolidates the data for each day, checking for gaps or corrupted TM packets. The MOC provides three interfaces to distribute the data. The consolidated telemetry, together with additional auxiliary data, is provided through a client/server application called the Data Disposition System (DDS).

For the real-time telemetry, the MOC has defined a dedicated protocol based on the CORBA middleware(Real-time Science Interface - RTSI): when a new TM packet arrives, the server notifies the packet stream to the client. The third interface, called the File Transfer System (FTS), is based on the FTP protocol and is used, via a dedicated line, to exchange auxiliary data (e.g. the attitude history) between the DPCs, the MOC and the Planck Science Office (PSO). The data received at the DPC are in the form of a stream of packets which are handled automatically, as shown in Figure 1.

The Level 1 pipeline consists of four main pipelines: Data Receipt, Telemetry Handling, Auxiliary Data, and Command History.

The Data Receipt pipeline implements the client side of the interface with the MOC DDS. It requests a subset of data provided through this interface. A finite state machine model has been used in the design of the main program

pipeline (TM request) for better formalization of the actions (transactions) required to handle all cases (statutes) in the interaction with the DDS server.

The Telemetry Handling pipeline is triggered when a new segment of TM data is received. The first task (TM Unscrambler) discriminates between scientific (SCI) and HouseKeeping (H/K) telemetry packets. SCI packets are grouped according to radiometer, detector source and processing type applied. For each group, scientific data are uncompressed and decoded (see next paragraph) and the On-Board Time (OBT) of each sample is computed based on the packet OBT and the detector sampling frequency. H/K TM packets are also grouped, according to packet type, and each H/K parameter within the packet is extracted and saved into TOI. The subsequent tasks of the pipeline perform calibration of H/K and SCI TOIs together with additional quality checks (e.g. out of limits, time correlation). The last task, FITS2DMC, ingests the TOIs into the Data Management Component (DMC) making them available to the Level 2 and Level 3 pipelines.

The main purpose of the Auxiliary data pipeline is to ingest into the DMC the Attitude History Files (AHF) provided by the Flight Dynamics at the MOC. Finally the Command History pipeline requests and stores the list of telecommands sent to the satellite during the DTCP.

The four pipelines are implemented as `perl` scripts, scheduled every 5 minutes. Trigger files are created to activate the processing in the Auxiliary and Command history pipelines and a Pipeline Monitoring facility displays information about the status of each pipeline The entire Level 1 pipeline was heavily tested and validated before the start of *Planck* operations (see Frailis et al. 2009, for more details).

2.1. Scientific data Processing

When creating the TOIs, the Level 1 pipeline has to recover accurately the values of the original (averaged) sky and load samples acquired on-board. The instrument has the possibility to acquire scientific data in several modes (called PTypes) but we describe here only the nominal one (PType 5) (see Zacchei et al. 2009). The key feature is that two independent differenced time streams are created from the sky and load signals with two different known Gain Modulation Factors.

Data acquired with PType 5 are first uncompressed. The loss-less compression applied on-board is simply inverted and the number of samples obtained is checked with auxiliary packet information. Decompressed data Q_i ($i = 1, 2$) are then subject to a de-quantized step to recover the original signals P_i according to

$$P_i = \frac{Q_i}{\mathrm{SECOND\_QUANT}} - \mathrm{OFFSET\_ADJUST}\,, \qquad (1)$$

where SECOND_QUANT and OFFSET_ADJUST are parameters of the Read Out Electronic Box Assembly (REBA), the calibration of which is described in detail by Maris et al. (2009).

After dequantization, data are demixed to obtain $\overline{S}_{sky}$ and $\overline{S}_{load}$ using as inputs the two know Gain Modulation Factors:

$$\overline{S}_{sky} = \frac{R_2 \cdot P_1 - R_1 \cdot P_2}{R_2 - R_1}, \qquad (2)$$

$$\overline{S}_{load} = \frac{P_1 - P_2}{R_2 - R_1}\,, \qquad (3)$$

where R_1 and R_2 are the two Gain Modulation Factors (GMF) that are REBA parameters determined during the REBA calibration (Maris et al. 2009).

The digitized data obtained (two timelines for each diode $\overline{S}_{sky}$ and $\overline{S}_{load}$) are not in physical units but in ADU (Analog to Digital Units). Conversion of $\overline{S}_{sky}$ and $\overline{S}_{load}$ to Volts requires auxiliary packet information (e.g. Blanking Time). Finally, voltages for sky and load are obtained from

$$\overline{V}_i = \frac{\overline{S}_i - Z_{\mathrm{DAE}}}{G_{\mathrm{DAE}}} - O_{\mathrm{DAE}}\,, \qquad (4)$$

where G_{DAE}, O_{DAE} and Z_{DAE} are DAE (Data Acquisition Electronic) Gain, Offset and a small tunable offset, with optimal values determined during the ground-test campaigns (Maris et al. 2009).

2.2. On-Board Time (OBT) reconstruction

To make proper use of scientific data, a time stamp has to be assigned to each data sample. This requires knowledge of the phase switch status (see Mennella et al. 2011, for details of the LFI radiometer scheme): if it is off (not switching) the packet contains consecutive values of either sky or load samples. In this case, denoting with $i_{\mathrm{smp}} \geq 0$ the sample index within the packet, we have:

$$t^{\mathrm{obt}}_{i_{\mathrm{smp}}} = t^{\mathrm{obt}}_0 + i_{\mathrm{smp}} \frac{N_{\mathrm{aver}}}{f_{\mathrm{samp}}}\,, \qquad (5)$$

where t^{obt}_0 is the time stamp of the packet and $i_{\mathrm{smp}} = 0$ denotes the first sample in the packet. The quantity f_{samp} is the detector sampling frequency and N_{aver} denotes the total number of fast samples (at $\simeq 4$ kHz) averaged together to obtain a single detector sample. However if the phase switch status is on (nominal case), either consecutive pairs of sky−load (or load−sky) samples are stored in the packet. Hence, consecutive pairs of samples have the same time stamp and

$$t^{\mathrm{obt}}_{i_{\mathrm{smp}}} = t^{\mathrm{obt}}_0 + 2\ \mathrm{trunc}\left[\frac{i_{\mathrm{smp}}}{2}\right] \frac{N_{\mathrm{aver}}}{f_{\mathrm{samp}}}. \qquad (6)$$

At the end, OBT information is stored in the form of TOI and directly linked to its scientific sample.

2.3. Data Flagging

For each sample we also define a 32 bit flag (from 0 to 31) mask to identify potential inconsistencies in the data and to enable the pipeline to skip over a sample during the data analysis. Flags are related to different kinds of situations that may occur during the observations (either normally or in special cases) which might impact the data.

In the current analysis, the flags that are checked are: those identifying the stable pointing period (as derived from the AHF); science data that cannot be recovered (e.g. because of saturation); samples artificially created to fill data gaps; samples affected by planet transits and moving objects within the Solar System.

3. TOI processing

3.1. Handling of Electronic Spikes

The first step in processing the LFI data is removal of electronic contamination in the data. This contamination appears in the frequency domain as a series of "spikes", with the most prominent ones at multiples of 1 Hz (Meinhold et al. 2009). Fig. 2 shows this signal in the time domain obtained by binning the data (from Operation Day – OD – 91 to 389) into a span of one second: the signal appears as a square wave in most detectors. The signal is synchronous with the on-board time, with no change in phase over the entire survey. This square wave signal is caused by the clock of the housekeeping electronics, which was not properly shielded from the data lines.

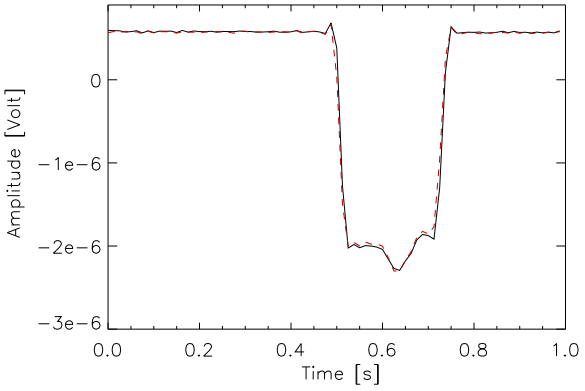

Fig. 2. A square wave template for both sky (black) and load (red/dashed line) for one of the 44 GHz detectors, computed from data between OD 91 and OD 389. Individual templates are directly subtracted from the un-differenced data

Because this signal is synchronous in one second periods, we have opted for a time-domain subtraction of the signal. Our signal model is a one-second piecewise-continuous template, which is used to better resolve the leading and trailing edges of the square wave. To estimate each template, we sum the data over the entire survey span into the one second period. There are some variations in the level of the square wave signal over the survey. We do not use this in the subtraction, but instead simulate this amplitude variation in an attempt to estimate the residual of the square wave signal after subtraction. Since the true modulation is likely to be less than the measured modulation, as the measured modulation includes both noise and the true modulation, this is a conservative estimate of the residual.

We would like to estimate the level of this systematic effect in the maps to decide whether or not to correct the data for it. To do this we generated three simulated maps: the first contains pure white noise, which can be regarded as a best case scenario and therefore the lowest possible noise level in a real map. The second map, which represent a worst case scenario, takes the square wave template for each detector measured over the entire survey, and then modulates it by a time-varying amplitude measured from the data. The final map removes a constant square wave template from the amplitude-modulated simulation, providing an indication of the residual effects of electronic spikes after their subtraction. The 30 GHz maps were produced at HEALPix (Górski et al. 2005) resolution parameter $N_{\rm side} = 512$, while the 44 GHz and 70 GHz maps were produced at $N_{\rm side} = 1024$. We compare the largest deviation from zero of both the square wave and residual maps and compare these with the pure white noise map rms, as shown in the left hand panel of Fig. 3. In all cases, the maximum effect of the square wave signal at the pixel level is significantly below the noise. Therefore the only correction we applied to the data is via the square-wave template and only for the 44 GHz detectors, as in the 70 GHz detectors the signal is already extremely small, and in the 30 GHz detectors it is not clear that the square wave templates are well-resolved even after months of integration time.

We also evaluated the impact of spike signals on the angular power spectra derived from the maps described above, and compared these to the power spectra of the white noise maps. Fig. 3 (right panel) shows effect at 44 GHz. It is clear that subtraction of a constant amplitude square-wave template is able to reduce unwanted artefacts by almost two orders of magnitude. Remaining artefacts are reduced further during the map-making process.

3.2. Gain Modulation Factor and Differenced data

Each LFI detector (diode) provides two voltage outputs carrying the signals from the sky, $V_{\rm sky}$, and from the 4 K reference load, $V_{\rm load}$, alternating at 4096 Hz (Mennella et al. 2010). Such undifferenced signals are strongly dominated by $1/f$ noise, which is characterized by knee frequencies of the order of a few tens of Hertz. However the pseudo-correlation design of the instrument (Bersanelli et al. 2010) results in highly correlated noise within the two streams, hence the knee frequencies can be reduced considerably by appropriate differencing. Since the two arms of the radiometer are slightly unbalanced (one looks at the sky at 2.7 K while the other looks at the reference load at ~ 4.5 K), the difference is taken after multiplying the load signal by the GMF, R, that nulls the mean of the difference between sky and load signals. This parameter can be computed in several ways (Mennella et al. 2003). The simplest method (which is the one implemented in the processing pipeline) is to take the ratio of DC levels from sky and load outputs obtained by averaging the two time streams, and to compute the differenced signal as:

$$\Delta V(t) = V_{\rm sky}(t) - \frac{\langle V_{\rm sky}\rangle}{\langle V_{\rm load}\rangle} V_{\rm load}(t)\,. \qquad (7)$$

We proceed with computing $R = \langle V_{\rm sky}\rangle/\langle V_{\rm load}\rangle$ for each pointing period (identified from the AHF information) and for unflagged data. The R factor is very stable over the timescale of the mission so far, with typical overall variations of $\sim 0.03 - 0.04\%$ for all the detectors. We therefore keep the pipeline as simple as possible and apply the single pointing period value of the GMF to create differenced data.

Fig. 4 shows the effect of applying the GMF to real flight data. The high correlation between noise in sky and load streams (evident in the two upper plots of the figure) is clearly reduced by the application of the R factor and differencing of $V_{\rm sky}$ and $V_{\rm load}$.

The reduction of $1/f$ noise is substantial, with a final noise spectrum characterized by small residual $1/f$ noise

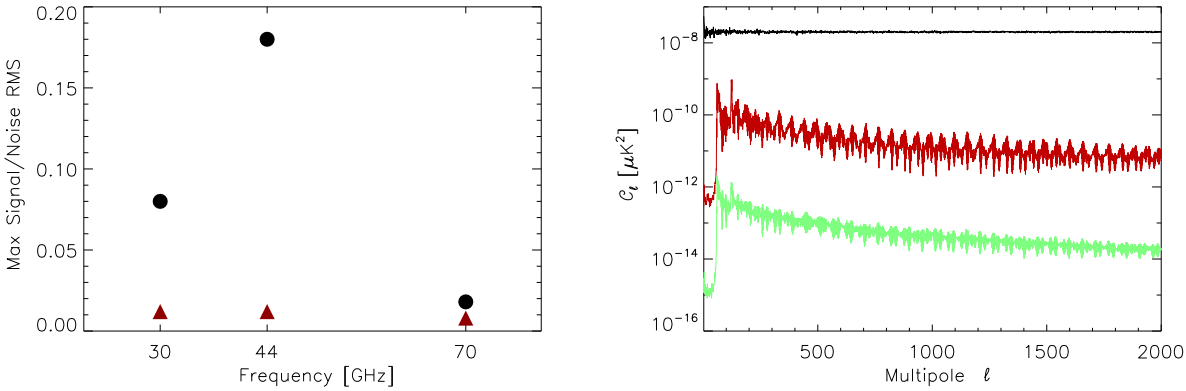

Fig. 3. Left – Estimated levels of square wave systematics (black circles) for each frequency. Plotted is the maximum amplitude of the systematic at the pixel level divided by the white noise rms for each frequency. Note that in our data processing, the square wave signal in only subtracted in 44 GHz detectors. The black circles therefore represent the estimated signal level in the 30 GHz and 70 GHz maps, while the red triangle represents the estimated remaining signal level in the 44 GHz map. Right – Angular power spectra of the 44 GHz simulations: the red (middle) line shows the power spectrum of the simulation, with a square wave with varying amplitude and the green (bottom) line shows the power spectrum after subtraction of a constant amplitude square wave. Subtraction reduces the amplitude by a factor of about 100, reducing it to a small fraction of the white noise level, shown by the black (top) line. The effects of electronic spikes are reduced further during the map-making process.

with knee frequencies of about 25 mHz. The impact of this residual $1/f$ noise on the maps is reduced further by the map-making pipeline described in Section 7.

3.3. The Diode Combination

Each LFI correlation radiometer has two diode outputs, arranged so that when one diode is responding to the signal from the sky horn the other is measuring the reference horn and vice-versa. The primary purpose of this is to maximize integration time. Inclusion of the second diode allows observations with a duty cycle of almost 100% and (in the ideal case of perfect gain balance between the two radiometer legs) a noise reduction of $\sqrt{2}$ when the diodes are averaged. Separate analysis of these two timelines can be useful in investigating gain drifts, temperature correlations and similar systematic effects. In the nominal science pipeline it is critical to propagate noise and correlations of the various time-streams contributing to the maps as accurately as possible.

The diode-to-diode "isolation" due to gain imbalances has been measured for each radiometer in a series of ground tests, and has been verified in flight data using the CMB dipole, planet and Galactic plane crossings and other signal events. Typical values of this isolation range from −13 dB to −20 dB or more. While the isolation is within specifications, and does not compromise LFI sensitivity, it does produce a significant anti-correlation of the white noise of the two diodes of a given radiometer. When data from the two diodes are averaged, the white noise of the resulting TOI is lower than would be expected if they were statistically independent. This behaviour is expected, and the level of anti-correlation is in line with the measured diode-to-diode isolation for each radiometer. A complete mathematical description of this behaviour is given in Mennella et al. (2011).

While the effects of this correlated noise component and the proper propagation through the pipeline to maps could in principle be calculated and corrected, it is more effective to combine the diodes in the time domain (see Mennella et al. 2011), performing calibration and further processing on the combined time stream. To do this, we assign relative weights to the two uncalibrated diode time-streams (labelled 0 and 1) based on their calibrated noise. We make a first order calibration of the timelines, G_0 and G_1, subtract a signal estimate, and calculate the calibrated white noise levels, σ_0 and σ_1 for the two diodes. The diode weights are found, for the 0, 1-diode, as:

$$W_{0,1} = \frac{\sigma_{1,0}^2}{G_{01}} \frac{1}{\sigma_0^2 + \sigma_1^2} \tag{8}$$

where we have calibrated by the weighted calibration constant G_{01} as

$$G_{01} = \frac{1}{\sigma_0^2 + \sigma_1^2} \left[G_0 \sigma_1^2 + G_1 \sigma_0^2 \right] , \tag{9}$$

(which is the same for each diode).

Fixing the weights to a single value per radiometer for the entire mission simplifies the noise analysis. Since all calibrations, noise estimation and other tests are done on these combined data streams, small errors in the weights will cause only inconsequential losses in sensitivity, and no other systematic errors.

3.4. Detector Pointing computation

Detector pointing information is a fundamental ingredient in the data processing pipeline and this involves knowledge of the Space Craft (S/C) attitude and the geometry of the focal plane (i.e. the actual physical locations of the horns). Attitude information is provided through the Attitude History File, where complete attitude of the S/C is encoded in quaternions sampled at a rate of 8 Hz. In

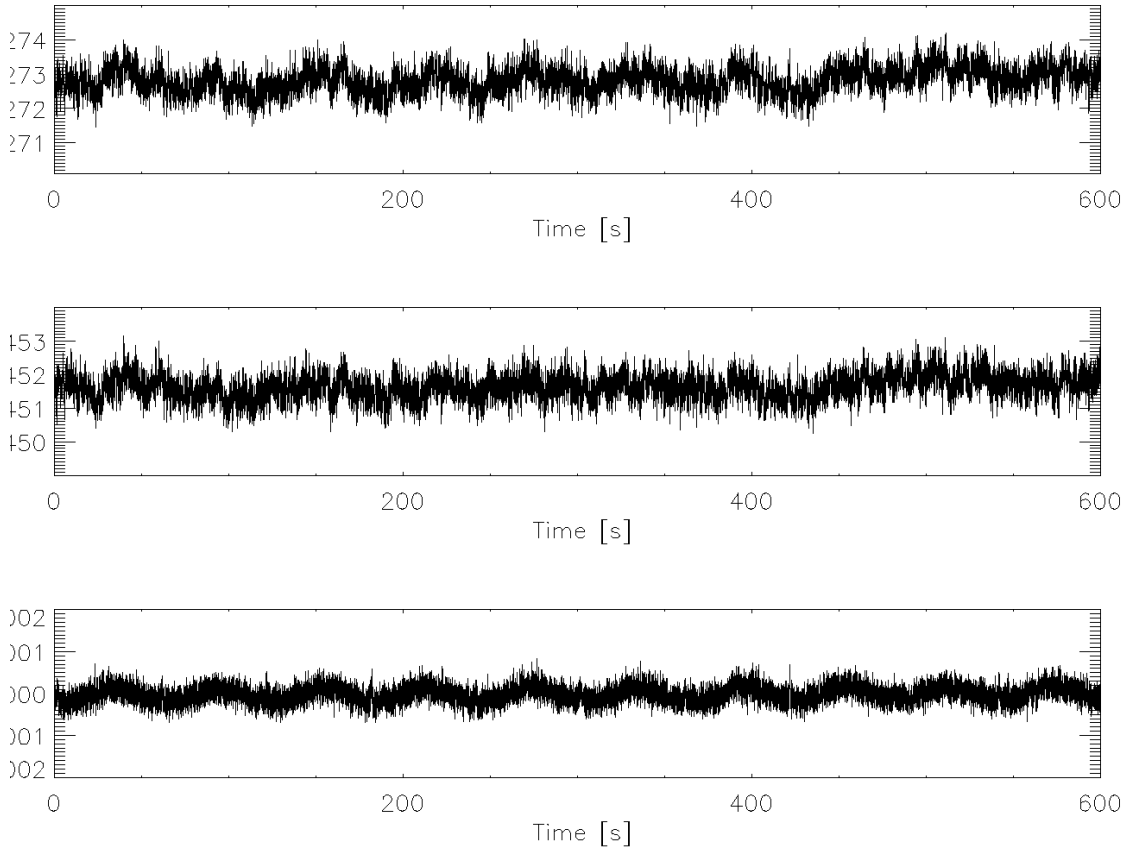

Fig. 4. Effect of the Gain Modulation factor on Sky and Load signals for flight data. The upper and middle panels show about 10 minutes of Sky and Load signals of the `LFI27S-11` detector: they are highly correlated and clearly show signatures of low-frequency noise. After application of the GMF, such fluctuations are considerably reduced revealing the presence of a sky signal dominated by the CMB dipole (lower panel). Note also the change in the y-axis scale.

addition the AHF gives times for the start/end of a single pointing period and identifies (with appropriate flags) when wobbling of the spin-axis is finished and stable pointing begins. Horn locations within the focal plane are stored in the Instrumental Model, derived from ground measurements and from planet transits.

As soon as differenced data are produced (see Section 3.2) we proceed with quaternion interpolation at the time of each data sample via linear spherical interpolation. From the attitude of S/C at times of data samples, we compute detector pointings by means of simple rotations: from the spin-axis reference frame to the telescope optical axis and then to the horn location reference frame, with an additional rotation accounting for the orientation of the horn on the focal plane.

In some cases (basically at the end of each pointing period) we have to perform extrapolation instead of interpolation. This has been verified, using simulations, to be robust and does not introduce any significant degradation of the pointing accuracy.

4. Main Beams and the geometrical calibration of the Focal Plane

Knowledge of the beam profile is of paramount importance in CMB experiments aiming at the extraction of the angular power spectrum of anisotropies. Indeed it is at the power spectrum level that the effect of the beam convolution is corrected for in the data. Therefore uncertainties in the knowledge of the main beam (but also its complex non-Gaussian shape) directly affect the determination of cosmological parameters.

This part of the pipeline is aimed at the reconstruction of main beam parameters, as well as performing a geometrical calibration of the focal plane, i.e. to measure the actual position of each horn within the focal plane. This is done during planet transits, of which Jupiter provides the best signal for this purpose (we have also tried with other external planets and celestial bright point sources, but these have proved less accurate). Input to this pipeline are: TOI from each radiometer spanning the range in time of the planet transit, the AHF for the same period and planet ephemerides. These are provided by the JPL Horizons system, which returns planet positions as seen by *Planck* ,taking into account both S/C and planet motion, which is not negligible.

4.1. The algorithm

The algorithm implemented takes the AHF for the whole planet transit time and interpolates quaternions at the sample time, as performed by the detector pointing creation

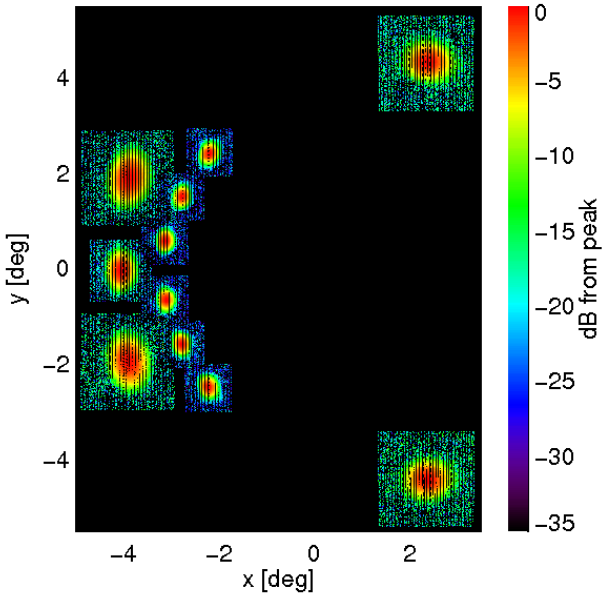

Fig. 5. LFI focal plane from first season of Jupiter observations from 24 October to 1 November 2009. Contour levels are in dB from the peak. All beams are well approximated by an elliptical Gaussian down to the ≈ -10 dB level.

code. Here the main differences come from the rotations applied to interpolated quaternions. Since we would like to reconstruct the actual position of the horn, we perform a rotation from the spin-axis reference to the telescope optical axis reference frame, i.e. the so-called Line Of Sight (LOS) reference frame defined in Sandri et al. (2010) and Tauber et al. (2010b).

We then create a simple 2D map of the footprint of the focal plane on the sky by selecting data which are within a given angle (usually 10°) from the LOS direction which comprises the whole extension of the LFI focal plane. We have verified that, in general, $1/f$ noise has no large impact on the final results in terms of horn position on the focal plane and the main beam parameters. However there are cases for which this is not true, e.g. weak sources. Therefore we use TOI where offsets per ring as derived from the `Madam` destriper (see Section 7) have already been removed.

On these data we perform a χ^2 fit of our beam model (Burigana et al. 2001). This is a bivariate Gaussian model

$$B(x_i, y_i) = \frac{A}{d^2}\exp\left\{-\frac{1}{2}\left[\frac{(\Delta x_i \cos\alpha + \Delta y_i \sin\alpha)^2}{\sigma_x^2} + \frac{(-\Delta x_i \sin\alpha + \Delta y_i \cos\alpha)^2}{\sigma_y^2}\right]\right\}, \quad (10)$$

where A is an overall amplitude, x_i and y_i are the 2D Cartesian coordinates, σ_x and σ_y are the beamwidth parameters of the elliptical approximation of the beam shape and $\Delta x_i \equiv x_i - x_c$ and $\Delta y_i \equiv y_i - y_c$, with x_c, y_c the position of the centre of the beam. Finally the angle α is the reconstructed orientation of the beam in the focal plane and d is the actual distance (in astronomical units) of the planet.

We tested our technique for beam and focal plane geometry reconstruction via optical simulations using the measured beam patterns together with a detailed model of the *Planck* telescope. The simulations included the nominal main and far beam patterns and the effects of smearing, caused by the motion of the satellite and pointing uncertainties. Using realistic simulations of Jupiter crossings (including instrumental noise and complete sky signal) we are able to reconstruct the main beam shape down to -20 dB and to recover the main beam properties with high accuracy (at the percent level or better) for all of the LFI beams. Table 1 reports results for the main beam properties for a sample of the LFI beams located in various places within the *Planck* focal plane. These figures are representative of our expected accuracy for in-flight beam and focal plane reconstruction.

Table 1. Results of simulations of the reconstructions of the beams and focal plane geometry using a Jupiter transit. Simulations include realistic models of the beams, instrument noise, beam smearing and uncertainties in the pointing reconstruction of the star-trackers.

Beam and Focal Plane Geometry Reconstruction			
Parameter	Input	Reconstruction	δ [%]
`LFI19S` – 70 GHz			
FWHM	12.83′	12.97′	1.12
ellipticity	1.280	1.276	0.98
x_0[°]	−2.8715	−2.8704	0.36
y_0[°]	−1.5678	−1.5829	0.96
`LFI25S` – 44 GHz			
FWHM	29.33′	30.23′	3.07
ellipticity	1.170	1.230	5.12
x_0[°]	−2.8227	−2.8293	0.23
y_0[°]	−5.1369	−5.0844	1.02
`LFI27M` – 30 GHz			
FWHM	32.42′	32.89′	1.45
ellipticity	1.380	1.384	0.32
x_0[°]	−4.7788	−4.7798	0.02
y_0[°]	2.4903	2.3958	3.79

Fig. 5 shows the actual footprint of the LFI focal plane obtained during the first season of Jupiter observations between 24 October 2009 and 1 November 2009. Fig. 6 shows beam images for `LFI28M`, `LFI25M` and `LFI21M` from the same Jupiter transit. As expected all beams are asymmetric, though they are close to elliptical and no significant departures from an elliptical shape are visible down to the ~ -10 dB level. For lower levels, aberration starts to distort the beam response, creating non-elliptical shapes.

We have also constructed a planet mask, including Jupiter, Mars and Saturn, which are the most luminous planets at LFI frequencies. The planet mask is radiometer dependent, since each horn observes a planet at different times. The planet masking algorithm assigns an appropriate flag to those sample that lie within an ellipse, centred at the position of the planet and with an orientation that matches the beam orientation, with axes ~ 3 times larger than the beam widths derived from beam fitting. These flags are used in the map-making and ensuing data analysis to discard samples affected by planet transits.

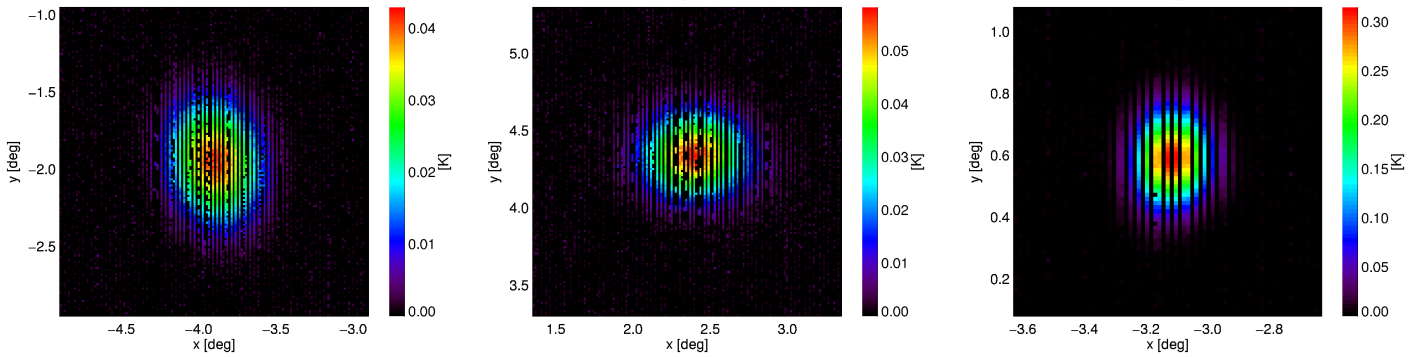

Fig. 6. Beam map of LFI data around the Jupiter transit (24 October/1 November 2009) for three LFI horns: `LFI28M` (left), `LFI25M` (middle) and `LFI21M` (right).

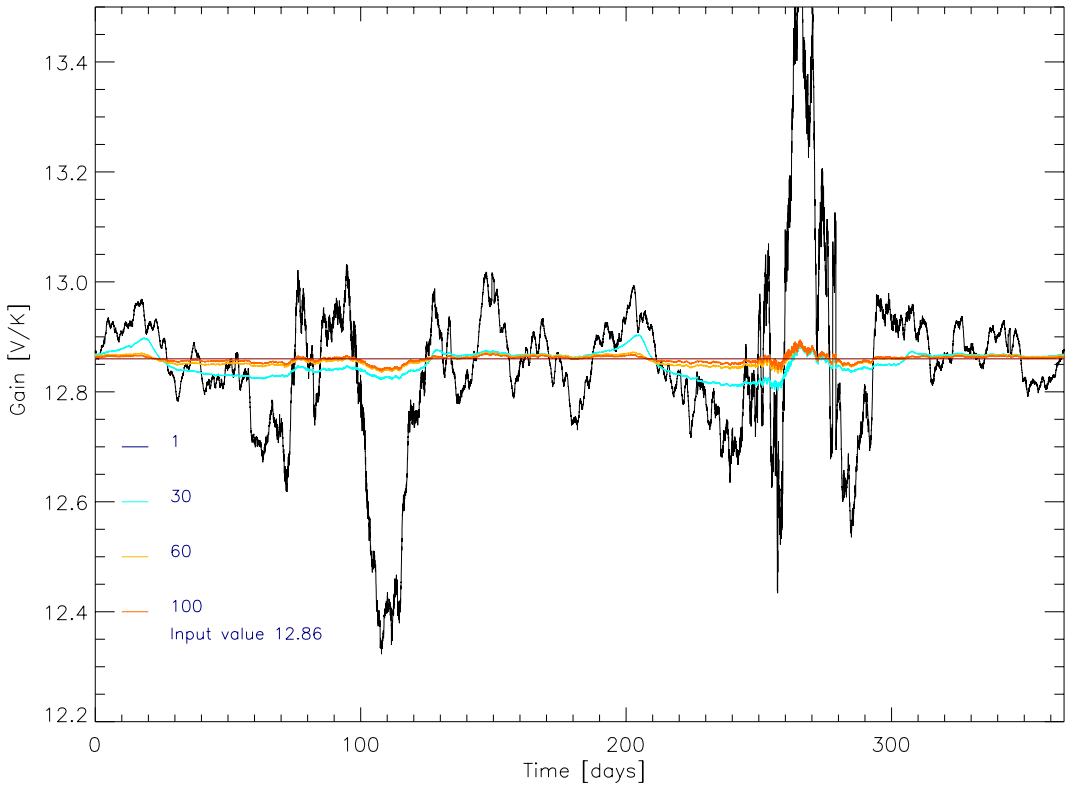

Fig. 7. Simulation results of the convergence of the gain solution for one year of observations of one of the four 30 GHz detectors. Simulations include CMB dipole(s), CMB anisotropies and Galactic emission. The input gain was 12.86 V/K and the sky included all diffuse components as well as nominal instrumental noise. The first iteration shows large errors caused by Galactic and CMB anisotropy emissions. However after one hundred iterations, convergence is achieved with an overall deviation from the input value of less than 0.01%. The various curves show the solution after 1, 30, 60 and 100 iterations.

5. Photometric Calibration

After the removal of frequency spikes, differentiation of $V_{\rm sky}$ and $V_{\rm load}$ using the derived Gain Modulation Factors, and diode combination into a single radiometer data stream, we proceed with the photometric calibration, i.e. the conversion from Volts to Kelvin. Ideally, a calibration source should always be present during the observations and have a well known, stable, spectral shape, with a spectrum similar to that of the signal that we are interested in. In the frequency range of the LFI (from 30 to 70 GHz) a suitable celestial source for calibration is the CMB dipole, caused by the motion of the Solar system (Sun) with respect to the CMB reference frame and by the modulation induced on the CMB dipole by the space-craft orbital motion. The calibration procedure employs an iterative algorithm, in which the calibration solution (instrument gain and baseline) is computed for a sufficiently short time period (usually a pointing period $\sim$ 1 hour). This solution is used as a trend-analysis tool to monitor the behaviour and health of the instrument.

We follow essentially the same calibration procedure, described in Hinshaw et al. (2003), which was implemented for the *WMAP* first year data. For a given pointing period, k, the signal from each detector can be modelled as

$$\Delta V_k = g_k(\Delta T_{\rm sky} + n) + b_k\,, \qquad (11)$$

where $\Delta T_{\rm sky}$ is the actual sky signal and n is noise contribution, while g_k and b_k are the gain and baseline solution for the kth pointing period. The sky source dominating the signal on short time scales is the CMB dipole (plus contribution from the Galaxy during crossings of the Galactic Plane) and this is modeled as

$$\Delta V_{\rm m}(g_k, b_k) = g_k(\Delta T_d + \Delta T_{\rm v}) + b_k \,, \qquad (12)$$

where we have considered both the cosmological dipole ΔT_d and the additional modulation from space-craft motion $\Delta T_{\rm v}$. We therefore fit for g_k and b_k for each pointing period k by minimising

$$\chi^2 = \sum_{i \in k} \frac{[\mathbf{\Delta V}(t_i) - \mathbf{\Delta V}_m(t_i|g_k, b_k)]^2}{rms_i^2} \,, \qquad (13)$$

The sum is extended to those unflagged samples within a given pointing period k and with coordinates that lie outside a Galactic mask. The Galactic mask is created from realistic simulations of microwave emissions at the *Planck* frequencies, provided by the Planck Sky Model[2]. Specifically the 30 GHz channel is treated as the LFI frequency that is most strongly contaminated by diffuse foreground emission. The mask excludes all pixels with a Galactic signal at 30 GHz that is larger than 5×10^{-4} times the expected CMB rms. Point sources brighter than 1 Jy, from a compilation of all radio catalogues available at high frequencies (the so-called *Planck* Input Catalogue, see Massardi (2006)) are also masked. The Galactic and point source masks preserve $\sim 82\%$ of the sky.

The *Planck* scanning strategy is such that the instrument field of view very nearly describes great circles on the sky. The signal mean is therefore almost zero and close to constant from one circle to the next. This helps to reduce the correlation between the gain and baseline solutions, as has already been observed by *WMAP* (Hinshaw et al. 2003) where the same situation (approximately zero mean signal on one calibration period) occurred, although *WMAP* scanned the sky in a different way.

As pointed out by Hinshaw et al. (2003) and by Cappellini et al. (2003), the largest source of error in the present procedure arises from unmodelled sky signal $\mathbf{\Delta T}_a$ from CMB anisotropy and emission from the Galaxy. To solve this problem, we perform an iterative approach of the basic calibration algorithm. If g'_k is the solution at a certain iteration step, the next solution is derived using the same expression as in Eq. (13), but with

$$\Delta V' = \Delta V - g'_k \Delta T'_{\rm a} \,, \qquad (14)$$

where $\Delta T'_{\rm a}$ is the sky signal (minus the dipole components) estimated from a sky map built from the previous iteration step. This is repeated until convergence is achieved, which usually occurs after a few tens of iterations. Fig. 7 shows the results from a simulation for one of the four 30 GHz detectors with nominal noise, illustrating the convergence of the calibration procedure.

The algorithm therefore alternates therefore dipole fitting and map-making, where the map-making is a simplified version of the `Madam` destriper used in the map-making step of our pipeline (see Section 7) ignoring polarisation and with no noise prior and baseline length equal to the pointing period length. To improve calibration and reduce noise, calibration is performed simultaneously for both radiometers of each single horn. We are also developing an iterative calibration approach, in which dipole fitting and map-making are done simultaneously. Preliminary tests show that both approaches give the same results, though the simultaneous technique converges much more quickly.

[2] The *Planck* Sky Model is available at: `http://www.apc.univ-paris7.fr/APC_CS/Recherche/Adamis/PSM/psky-en.html`

Table 2. Summary of Dipole based Gain statistics

	Main Arm		Side Arm	
Det ID	$\langle g_k \rangle$ [V/K]	σ_{g_k} [%]	$\langle g_k \rangle$ [V/K]	σ_{g_k} [%]
	70 GHz			
LFI 18	14.935	0.279	22.932	0.243
LFI 19	27.434	0.141	41.843	0.228
LFI 20	25.572	0.253	29.581	0.261
LFI 21	41.629	0.367	41.999	1.038
LFI 22	64.275	0.367	62.504	0.185
LFI 23	36.492	0.290	54.121	0.382
	44 GHz			
LFI 24	282.295	0.349	175.728	0.306
LFI 25	123.141	0.358	123.958	0.279
LFI 26	167.364	0.398	142.061	0.411
	30 GHz			
LFI 27	12.875	0.314	15.320	0.349
LFI 28	15.802	0.225	19.225	0.379

In the presence of real noise, the actual scatter from one gain solution to the other is quite large. Fig. 8 shows an example of the hourly gain solution (grey line) derived from the iterative scheme described above for `LFI18M`, one of the 70 GHz radiometers. Note that apart from instrumental induced noise scatter, the gain solution is quite stable along the observation period: during dipole maxima typical noise induced variations are $\sim 0.8\%$ (rms). Nonetheless, this stability is poor compared to the stability of the instrument which can be inferred by other means (e.g. the trend of the uncalibrated white noise level of both differenced and undifferenced data, and trend of the $\Delta V/V$ of the undifferenced data from both sky and load signals) and this is particularly evident during minima of the dipole signal Mennella et al. (see 2011, for further details). Furthermore there are also other effects that can mimic a gain variation. For example an ADC non-linearity, discovered during data analysis, produced a multiplicative effect on the data which is recovered by the calibration pipeline as a gain variation. In addition, from OD 259 the transponder was set permanently in "ON" status (instead of switching ON/OFF every day) since switching caused a temperature variation in the back-end unit (BEM) of the LFI radiometers which translated directly into a real gain variation. These are examples of problematic situations that affect the calibration and need to be dealt with. To tackle the ADC we applied two independent, complementary methods. In the first, we calibrated the data using the gain solution that follows the induced ADC gain variation. In the second, we were able to model the ADC effect and to test that the correction at the raw TOI level worked as expected, producing a gain solution that did not include the ADC gain variation.

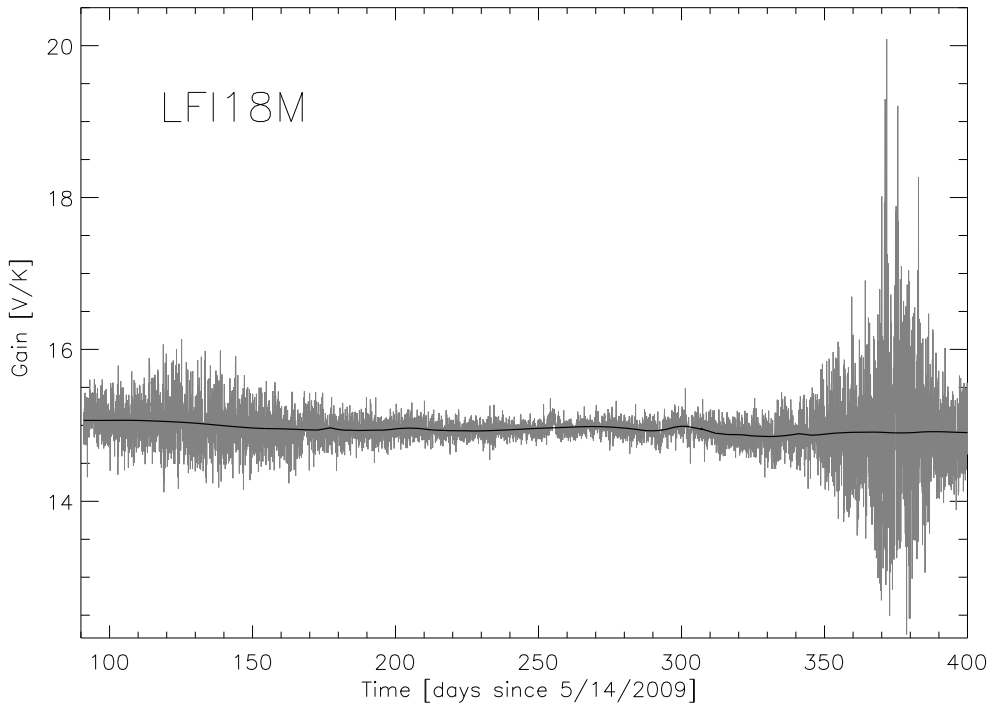

Fig. 8. Hourly gain solution (gray line) from flight data for `LFI18M`, as derived from our iterative calibration algorithm. The gain is quite stable over the observing period, although there is a lot of scatter due to the noise, especially during dipole minima. The thick black line is the refined gain solution (see text) applied to create calibrated TOI and sky maps.

5.1. Improving Calibration Accuracy

As shown in Fig. 8, the hourly gain solutions are noisy and their application to raw data would lead to stripes in the sky maps. Therefore, prior to generating calibrated time lines a refined calibration pipeline is applied to remove most of the gain solution noise and to produce gain tables closer to the actual instrument stability. We processed the hourly gain solution according to the following pipeline:

- create less noisy gain solutions by constructing running averages of 5 and 30 OD lengths respectively (the first creates artifacts during dipole minima, while the latter does not accurately follow real rapid gain changes);
- further smooth the 5 and 30 OD curves with wavelets;
- use the 30 ODs wavelet smoothed curve during dipole minima;
- use the 5 ODs unsmoothed curve around OD 259 to trace real gain variations;
- use the 5 OD wavelet smoothed curve elsewhere.

A typical gain solution applied to create calibrated data is plotted in Figure 8 as the solid black line.

From the 5 and 30 OD gain curves we infer information on the actual gain stability of the instrument as the mission progresses and also of the overall uncertainty in the gain reconstruction. A summary of the gain statistics from flight data is reported in Table 2, which lists the mean gain values and the rms of the gain solution for all LFI radiometers (both Main and Side Arms) after computing the running average (worst case reported comparing the 5 and the 30 ODs windows). Again although daily gains show peak-to-peak variations, sometimes at the $\sim 10\%$ level (with mean $\sim 7\%$), the rms of the smoothed gain solution is, in most cases, at the $\sim 0.3 - 0.4\%$ level, which can be considered an indication of the LFI calibration accuracy.

Although the current pipeline provides results close to those expected from the stability of the instrument, we are working to improve it as much as possible. In particular we would like to trace gain variations on time scales shorter than the pointing period. To achieve this, we are developing a detailed gain model (currently under test) based on calibration constants estimated from the pipeline and instrument parameters (temperature sensors, total power data Mennella et al. (see 2011, for further information)).

6. Noise estimation

Once data are calibrated, we proceed to the evaluation of noise properties for each radiometer. We would like to use noise properties to check also for possible instabilities and changes in the radiometer output along the mission lifetime. We therefore select data in chunks of 5 ODs each and then compute noise properties. This is done using the `roma` Iterative Generalized Least Square (IGLS) map-making algorithm (Natoli et al. 2001; de Gasperis et al. 2005) which includes a noise estimation tool based on the iterative approach described in (Prunet et al. 2001). IGLS map-making is time and resource intensive and cannot be run over the whole data set within the current DPC system. However since the TOD length considered here is only 5 ODs, it is possible to use the `roma` implementation of this algorithm which has a noise estimator built-in. (The `Madam` destriper code does not have this capability, although it could be modified to provide noise estimates.) The method implemented here is summarized as follows. Let us model the

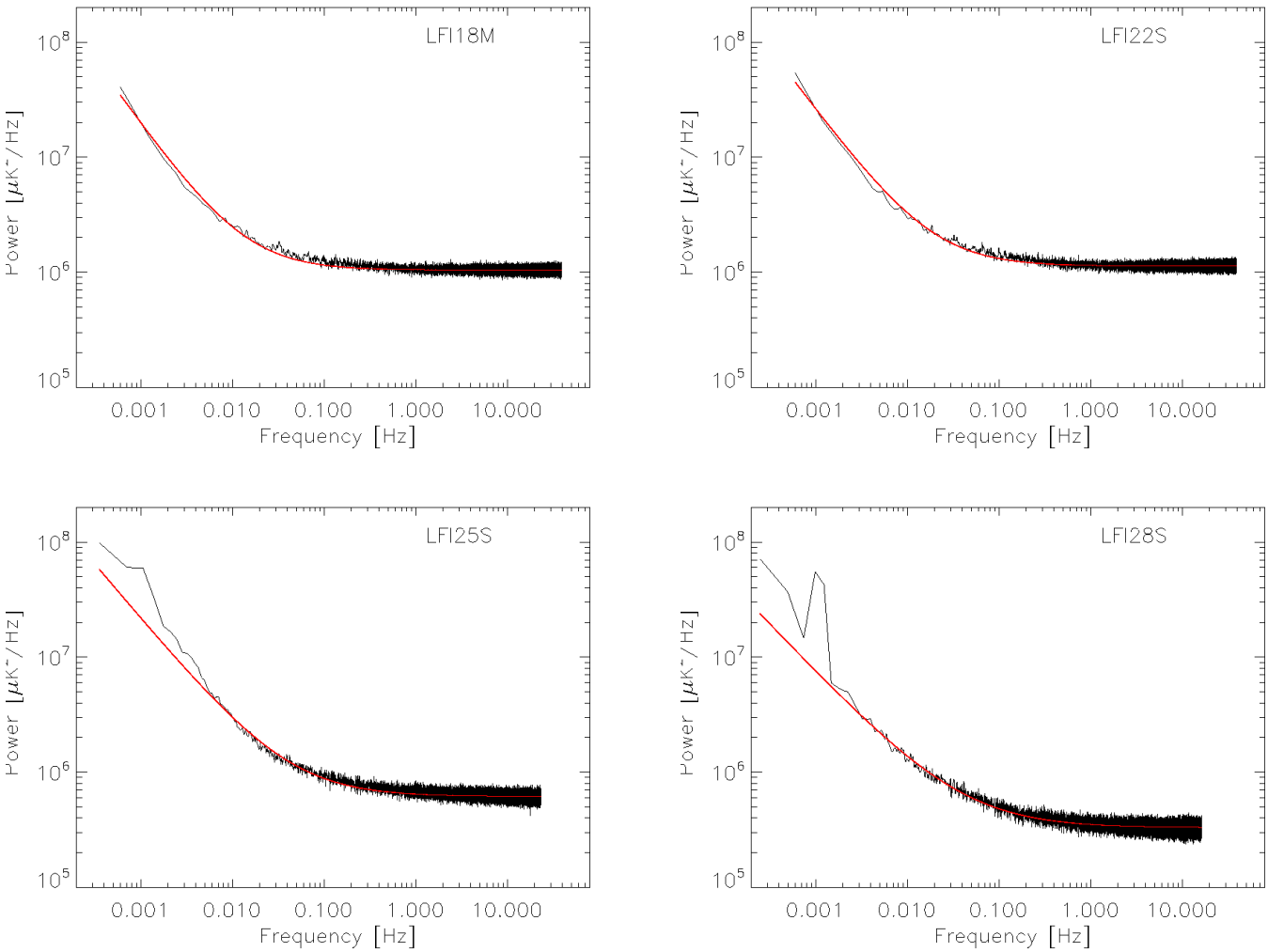

Fig. 9. Noise spectra of flight data from a subset of LFI radiometers (`LFI18M`,`LFI22S`, `LFI25S` and `LFI28S`) estimated by the noise pipeline (see text for details) and the noise spectra from fitted parameters (red continuous lines). All spectra are well described by a single knee-frequency and slope. A peak at around 1mHz is visible in `LFI25S` spectrum and more evident for `LFI28S`: this is a typical sorption cooler frequency and the different slope in `LFI28S` and `LFI25S` on the low-frequency side of the spectrum are possibly indications of thermal effects on radiometer output.

calibrated TOD as

$$\mathbf{\Delta T} = \mathbf{P m} + \mathbf{n}, \tag{15}$$

where $\mathbf{n}$ is the noise vector, and $\mathbf{P}$ is a projection matrix that relates a map pixel $\mathbf{m}$ to a TOD measurement $\mathbf{\Delta T}$. We obtain a zeroth order estimate of the signal through a rebinned map and then iterate noise and signal estimation:

$$\mathbf{\hat{n}_i} = \mathbf{\Delta T} - \mathbf{P \hat{m}_i}, \tag{16}$$

$$\mathbf{\hat{m}_{i+1}} = (\mathbf{P^T \hat{N}_i^{-1} P})^{-1} \mathbf{P^T \hat{N}_i^{-1} \Delta T}, \tag{17}$$

where $\mathbf{\hat{N}_i}$ is the noise covariance matrix in time domain estimated at iteration i. We have verified that convergence is reached in a few, usually three, iterations.

We then perform a Fast Fourier Transform over the noise time stream and fit the resulting spectrum for the three main noise parameters, namely the white noise level, the knee-frequency and the slope of the $1/f$ noise part:

$$P(f) = \sigma^2_{\rm WN} \left[1 + \left(\frac{f}{f_{\rm k}} \right)^{\beta} \right] . \tag{18}$$

The white noise level is simply computed taking the average of the last, few percent, of frequency bins. A linear fit on the log-log spectrum low frequency tail is done to extract the slope of the $1/f$ noise part and the knee frequency, $f_{\rm k}$, is estimated by the intersection of these two straight lines. The actual accuracy, verified with dedicated simulations, is around the percent level and to improve the situation a simple Monte-Carlo Markov Chain algorithm is implemented and currently under test. Examples of noise spectra and corresponding fits are shown in Fig. 9.

6.1. Noise constrained realizations and gap filling

This FFT-based power spectrum estimation method requires continuity of the noise time stream. As discussed

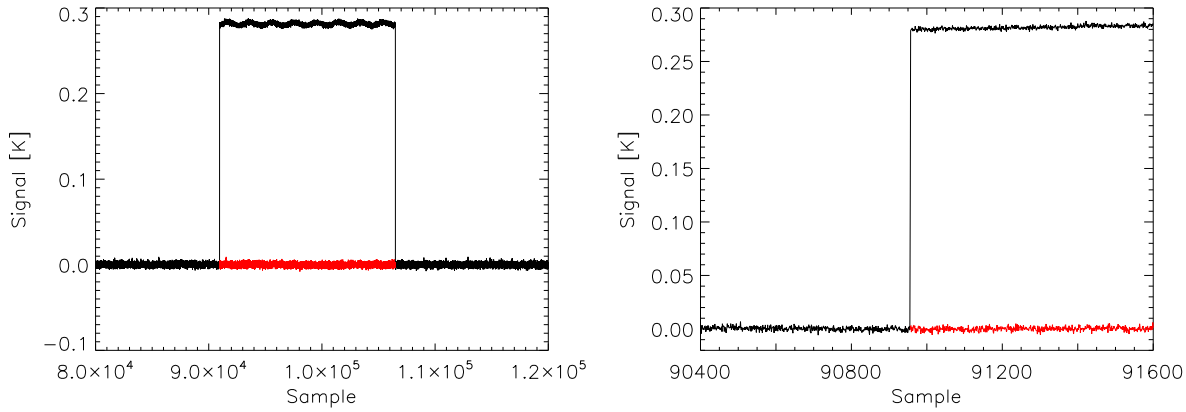

Fig. 10. Gap filling procedure applied on `LFI28M` time ordered data for OD 239. The upper panel shows the original TOI (black) where a step is evident caused by a DAE gain change which produces saturated data. The (red) lines show the constrained noise realization used to replace those data. The lower panel shows a zoom around the position of the step to highlight the consistency of the gap filling data with the unflagged part of the TOI.

in Section 3, we identify bad data (e.g. unstable space-craft pointing, data saturation effects) as well as gaps with missing data with appropriate flags. To preserve the noise correlations over different time scales, we need to implement a gap filling procedure to replace flagged, or missing data, with a Gaussian noise realization constrained using data outside the gaps as described in (Hoffman & Ribak 1991). Although the method described by these authors requires a pure noise time stream outside the gap to constrain the simulated noise inside the gap, we have verified that given the low signal-to-noise ratio in the LFI TOD, the procedure is not affected by the signal present in the time streams. To generate simulated noise whose properties are consistent with the actual data, we produce a Gaussian noise time stream starting from a noise power spectrum computed over the OD immediately before the one with flagged data. An example of the gap filling procedure applied to the LFI data is shown in Fig. 10.

7. The Map-Making pipeline

7.1. Frequency maps

Map-making is the most computationally demanding step in the data processing chain presented here. The aim of map-making is to produce sky maps of temperature and polarisation for each frequency channel. The map-making pipeline takes as an input the calibrated timelines and pointing information in the form of three angles (θ, ϕ, ψ) describing the orientation of the feed horns for each single sample considered. All of the pointing information need to be held in computer memory simultaneously, placing considerable demands on the computational resources required.

An essential part of the map-making process is the reduction of correlated $1/f$ noise in the TOD. A large part of the noise can be removed by exploiting redundancies in the scanning strategy: while the underlying sky signal remains the same, the observed signal varies due to noise. A statistical analysis of the signal variations allows one to distinguish between true sky signals and noise.

A number of map-making codes have been tested with simulated *Planck* data. For a review of the comparison see Ashdown et al. (2007a,b, 2009). The LFI baseline (Mandolesi et al. 2010) is to use the `Madam` map-making code, which is based on the destriping method (Maino et al. 2002). The basic idea is to model the correlated noise component by a sequence of constant offsets, called baselines. One of the key parameters in the code is the length of the baseline to be fitted to the data. `Madam` allows the use of an optional noise prior, if the noise spectrum can be reliably estimated, which further improves the accuracy of the output map. Without the noise prior, the optimal baseline length is of the order of the satellite spin period ($\approx$ 1 minute). With an accurate noise prior, much shorter baselines can be used. The shorter the baselines, the closer the `Madam` solution will be to the optimal Generalized Least Square solution.

The algorithm and the underlying theory are described in detail in the literature (see Keihänen et al. 2010; Kurki-Suonio et al. 2009; Keihänen et al. 2005, for further details). We are continually improving our knowledge of the instrument and its noise characteristics and this information will eventually be used in the `Madam` algorithm. However, at this stage in the processing we decided to make two simplifications when running our map-making pipeline: no noise prior was used and all radiometers were weighted equally. These choices lead to a simpler and faster map-making algorithm, which is sufficiently accurate for the *Planck* Early Results and avoids using detailed parameters describing the instrument which are under continual revision.

With these simplifications, the map-making equations can be written in a concise form. Technically, we are neglecting the baseline covariance, $\mathbf{C_a}$, and setting the white noise covariance $\mathbf{C_n}$ to unity.

The basic model behind the algorithm is the following. Let us write the signal in TOD as

$$\mathbf{\Delta T} = \mathbf{P m} + \mathbf{n}', \tag{19}$$

where $\mathbf{\Delta T}$ is the calibrated TOD, $\mathbf{P}$ and $\mathbf{m}$ are the pointing matrix and the pixelized sky map respectively and $\mathbf{n}'$ is the instrumental noise. This last term is further modeled as

$$\mathbf{n}' = \mathbf{F b} + \mathbf{n}, \tag{20}$$

where $\mathbf{b}$ is the vector of unknown base function amplitudes and the matrix $\mathbf{F}$ projects these amplitudes into the TOD. Since `Madam` uses uniform baselines, the matrix $\mathbf{F}$ consists of ones and zeros, indicating which TOD sample belongs to which baseline. Finally $\mathbf{n}$ is a pure white noise stream assumed to be statistically independent of the baselines.

The maximum likelihood solution is obtained by minimizing

$$\chi^2 = (\mathbf{\Delta T} - \mathbf{F}\mathbf{b} - \mathbf{P}\mathbf{m})^T(\mathbf{\Delta T} - \mathbf{F}\mathbf{b} - \mathbf{P}\mathbf{m}), \quad (21)$$

with respect to the quantities $\mathbf{b}$ and $\mathbf{m}$. The baseline amplitudes $\mathbf{b}$ are detemined by solving

$$(\mathbf{F}^T\mathbf{Z}\mathbf{F})\mathbf{b} = \mathbf{F}^T\mathbf{Z}\mathbf{y}\,, \quad (22)$$

where

$$\mathbf{Z} \equiv \mathbf{I} - \mathbf{P}(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\,. \quad (23)$$

`Madam` uses an iterative conjugate-gradient method to solve Eq. (22). An estimate for the map is finally obtained as

$$\mathbf{m} = (\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T(\mathbf{\Delta T} - \mathbf{F}\mathbf{b})\,. \quad (24)$$

The map $\mathbf{m}$ has as many elements as pixels in the sky. A single element is a Stokes parameter triplet (I, Q, U) at each pixel p. The matrix $\mathbf{P}^T\mathbf{P}$ is a 3×3 block diagonal matrix that operates on map space. There is a block for each pixel p of the map. A block can only be inverted if the pixel p is sampled with sufficiently different polarisation directions so that the three Stokes parameters can be determined for that pixel. This is gauged by the condition number of the block. If the inverse condition number *rcond* (ratio of the smallest to largest eigenvalue) is below a predetermined limit, the pixel p is excluded from (I, Q, U) of the map (as are pixels with no hits). For the present analysis pixels with *rcond* $\leq$ 0.01 were excluded from the maps.

The $\mathbf{P}^T\mathbf{P}$ blocks need to be inverted when the baselines are solved from Eqs. (22) and (23). These invertions were computed by eigenvalue decomposition. Eigenvalues whose magnitudes were smaller than 10^{-6} from the largest eigenvalue were discarded and only the remaining part was inverted.

For the purposes of the present analysis, we provide, from the full (I, Q, U) analysis, the I-component maps at the three LFI nominal frequencies (30, 44 and 70 GHz) combining observations of all radiometers at a given frequency. Figure 11 shows (left column) the hit count maps obtained combining all radiometers at a given frequency. In addition we also produce maps from horn pairs scanning the same path in the sky (see Mennella et al. 2011, for details on LFI focal plane arrangement). We have produced maps at a HEALPix resolution of $N_{\rm side} = 1024$ at 44 and 70 GHz and $N_{\rm side} = 512$ at 30 GHz. All maps are in the NESTED scheme, in Galactic coordinates, with units of thermodynamic Kelvins. The baseline length in the `Madam` destriping was one minute[3].

7.2. White noise covariance matrices

If we bin the pure white noise stream $\mathbf{n}$ to a map using the pointing $\mathbf{P}$, we obtain a binned white noise map,

$$\mathbf{w} = (\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\mathbf{n}\,. \quad (25)$$

[3] One minute baselines for 30 GHz, 44 GHz, and 70 GHz are 1950, 2792, and 4726 samples respectively.

This map is a theoretical concept because we do not have access to the radiometer white noise streams. Its covariance matrix is, however, of practical importance because it provides an estimate of both the white noise power in each pixel and white noise correlations between Stokes parameters at a given pixel.

This *white noise covariance matrix* (WNC) is computed as (Eq. 25)

$$\mathbf{C}_w = \langle \mathbf{w}\mathbf{w}^T \rangle = (\mathbf{P}^T\mathbf{P})^{-1}(\mathbf{P}^T\mathbf{C}_n\mathbf{P})(\mathbf{P}^T\mathbf{P})^{-1}\,. \quad (26)$$

Here angular brackets denote the ensemble mean and $\mathbf{C}_n \equiv \langle \mathbf{n}\mathbf{n}^T \rangle$. $\mathbf{C}_n$ is a matrix that operates in TOD domain. Because the radiometers have independent white noise, $\mathbf{C}_n$ is diagonal. We assumed that a radiometer TOD has a uniform white noise variance $\sigma^2_{\rm WN}$ (see Section 6) but that each radiometer has its own variance. The radiometer $\sigma_{\rm WN}$ values that we used in the WNC computation are reported in Mennella et al. (2011).

7.3. Half-ring jackknife noise maps

For noise estimation purposes we divided the time ordered data into two halves and produced jackknife maps as follows:

Each pointing period lasts typically ≈ 44 minutes (median 43.5 min, standard deviation 10 min). Typically, during the first 4 minutes the pointing is unstable, so these data are not used for science. During the remaining stable 40 minutes, each horn scans a ring on the sky. This ring consists of scan circles. One full scan circle takes 1 minute. Therefore, each ring has about 40 scan circles. We made half-ring jackknife maps $\mathbf{j_1}$ (and $\mathbf{j_2}$) with the same pipeline as described in Sect. 7.1, but using stable data only from the first (second) half of each pointing period. Technically this is implemented by marking the other half of each ring as a gap in the data. This way the destriper `Madam` knows that for any given pointing period the first used scan circle/sample of the half-ring is far apart in time (typically 25 minutes) from the last used scan circle/sample of the previous pointing period.

At each pixel p, the jackknife maps $\mathbf{j_1}$ and $\mathbf{j_2}$ contain the same sky signal (if there are no time-varying sources or moving objects crossing p at the time of observation), since they result from the same scanning pattern on the sky. However, because of instrumental noise the maps $\mathbf{j_1}$ and $\mathbf{j_2}$ are not identical.

We can estimate the sky signal (+noise) as

$$\mathbf{m_{1+2}}(p) = [\mathbf{j_1}(p) + \mathbf{j_2}(p)]/2, \quad (27)$$

and a realization of noise in map $\mathbf{m_{1+2}}$ as

$$\mathbf{n_{1+2}}(p) = [\mathbf{j_1}(p) - \mathbf{j_2}(p)]/2. \quad (28)$$

This noise map includes noise that is not correlated on timescales longer than 20 minutes. In particular, $\mathbf{n_{1+2}}$ gives a good estimate of the white noise in $\mathbf{m_{1+2}}$.

However, we are interested in the noise level in the full map $\mathbf{m}$, (see Eq. 24). To estimate this, we construct another noise map

$$\mathbf{n_m}(p) = \frac{\mathbf{j_1}(p) - \mathbf{j_2}(p)}{\mathbf{w}_{\rm hit}(p)}\,, \quad (29)$$

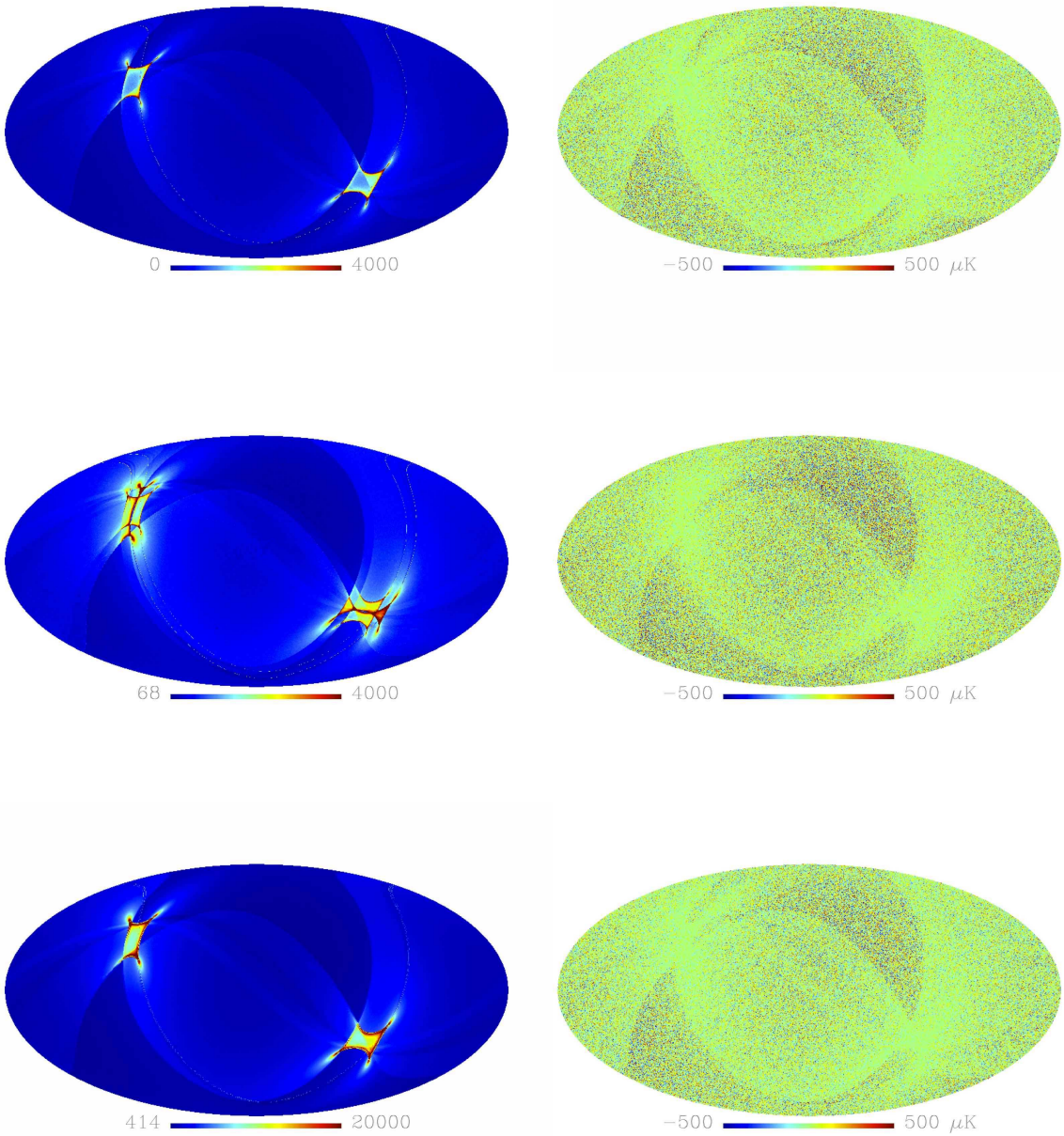

Fig. 11. Hit count (on the left) and noise maps (on the right) at the 30 (top), 44 (middle) and 70 GHz (bottom) LFI frequency channels. Note that the complex distribution around the Ecliptic Poles in the 44 GHz hit map is caused by the actual location of 44 GHz horns on the focal plane. The noise maps are derived from half-ring jackknife tests described in the text.

where the hit count weight is

$$\mathbf{w}_{\rm hit}(p) = \sqrt{\mathbf{hit}_{\rm full}(p)\left[\frac{1}{\mathbf{hit_1}(p)} + \frac{1}{\mathbf{hit_2}(p)}\right]}. \qquad (30)$$

Here $\mathbf{hit}_{\rm full}(p) = \mathbf{hit_1}(p) + \mathbf{hit_2}(p)$ is the hit count at pixel p in the full map $\mathbf{m}$, while $\mathbf{hit_1}$ and $\mathbf{hit_2}$ are the hit counts of $\mathbf{j_1}$ and $\mathbf{j_2}$, respectively. The weight factor $\mathbf{w}_{\rm hit}(p)$ is equal to 2 only in those pixels where $\mathbf{hit_1}(p) = \mathbf{hit_2}(p)$. In a typical pixel, $\mathbf{hit_1}(p)$ will differ slightly from $\mathbf{hit_2}(p)$ and hence the weight factor is $\mathbf{w}_{\rm hit}(p) > 2$.

Noise maps from half-ring jackknifes are shown in the right-hand column of Fig. 11. A detailed comparison of the jackknife noise estimates and other noise estimates (WNC, noise Monte-Carlo - see next section) are presented in the LFI instrument paper (Mennella et al. 2011).

7.4. Noise Monte-Carlo simulations

To check the noise analysis, we produced 101 noise Monte-Carlo realizations on the "Louhi" supercomputer at "CSC-IT Center for Science" in Finland. The noise Monte-Carlo

simulation takes as an input estimates of the white noise $\sigma_{\rm WN}$, the knee frequency, and the slope of the $1/f$ noise estimated from the TOD for each radiometer (see Mennella et al. 2011) as well as satellite pointing information.

The flight pointing was reconstructed on the Louhi supercomputer to machine accuracy using Planck Level-S simulation software Reinecke et al. (2006). For each frequency channel, we generated 101 noise Monte-Carlo realizations and simulated the white noise and correlated noise $(1/f)$ streams separately. As with the real *Planck* frequency maps, the maps from these noise streams were produced with the same map-making pipeline as described in Sect. 7.1. In addition to 101 noise Monte-Carlo maps, for each simulated map we computed the corresponding binned white noise maps defined in Eq. (25). The production of the binned white noise maps allows us to study the residual correlated noise, i.e. the difference between the total and binned white noise maps Kurki-Suonio et al. (2009).

Noise Monte-Carlo simulations are used to test and validate several approaches to noise estimation described in detail in the LFI instrument paper (Mennella et al. 2011).

8. Colour Correction

The power measured by LFI, converted into tension by the detectors, is expressed as

$$P = \frac{G}{2}\int g(\nu)\Delta T_{RJ}(\nu)d\nu\,, \tag{31}$$

where $g(\nu)$ is the bandpass and G is the overall gain. Here ΔT_{RJ} is the Rayleigh-Jeans brightness temperature signal, in the case of LFI calibration procedure, due to the CMB dipole. Therefore the overall gain G is equal to $2P/\Delta T_{RJ}(\nu_0)$ at a given frequency ν_0. For small fluctuations around the mean CMB temperature, T_0, the relation between intensity, I, brightness temperature and thermodynamic temperature T is

$$\Delta I(\nu_0) = \frac{2k_B\nu_0^2}{c^2}\Delta T_{RJ}(\nu_0) = \left(\frac{\partial B(\nu_0,T)}{\partial T}\right)_{T_0}\Delta T\,, \tag{32}$$

where the differential black-body spectrum is

$$\left(\frac{\partial B(\nu,T)}{\partial T}\right)_{T_0} = \frac{2k_B\nu^2}{c^2}e^{k\nu/k_BT_0}\left(\frac{h\nu/k_BT_0}{e^{h\nu/k_BT_0}-1}\right)^2, \tag{33}$$

$$\equiv \frac{2k_B\nu^2}{c^2}\eta_{\Delta T}(\nu)\,. \tag{34}$$

The function $\eta_{\Delta T}(\nu)$ is the differential black-body spectrum in Rayleigh-Jeans units. With our definition of the overall gain G, the bandpasses are normalised such that

$$\int g(\nu)\eta_{\Delta T}(\nu)d\nu = \eta_{\Delta T}(\nu_0)\,. \tag{35}$$

Calibration data provide a nominal brightness temperature $\Delta\tilde{T}_{RJ} = (2/G)P$. However, this is only exact for a monochromatic response. For a finite bandwidth, a colour correction $C(\alpha)$ is required to convert the brightness temperature for emission with a particular spectral index α to that of the map:

$$C(\alpha)\Delta T_{RJ}(\nu_0) = \Delta\tilde{T}_{RJ} = \eta_{\Delta T}(\nu_0)\Delta\tilde{T}\,. \tag{36}$$

By definition, the colour correction is unity when the source observed has a CMB spectrum. Within each finite LFI band $g(\nu)$ is well approximated by a power law with spectral index $\alpha = 2-(h\nu_0/k_{\rm B}T)^2/6$.

The general expression for the colour correction is

$$C(\alpha) = \left[\frac{\eta_{\Delta T}(\nu_0)}{\int g(\nu)\eta_{\Delta T}(\nu)d\nu}\right]\int g(\nu)(\nu/\nu_0)^\beta d\nu\,, \tag{37}$$

where we assumed a power-law spectrum with temperature spectral index $\beta = \alpha - 2$. The term in square brackets is unity with our normalisation for $g(\nu)$ but has been included to show that $C(\alpha)$ depends only on the shape and not the amplitude of the bandpass. Thus $C(\alpha)$ is independent of G.

Each LFI detector has a different bandpass and hence its own colour correction. We derive approximate colour corrections for band-averaged sky maps using bandpasses averaged over: (i) the two detectors in each radiometer; (ii) the two orthogonally-polarised radiometers behind each feed horn; (iii) the several feed horns in each frequency band.

In addition, although the bandpass is mainly defined by the front-end (Bersanelli et al. 2010), differences between back-end bandpasses on a single radiometer are measurable, e.g., in the form of β-dependent residual in difference images.

Since the current sky maps have been produced, both for pairs of horns and for several horns in each band, with calibrated data combined with equal weights, we have used an unweighted average of all the contributing bandpasses for our band-averaged corrections. Using the bandpass models given in Zonca et al. (2009) which were derived from the pre-launch calibration campaign we evaluate the integrals in Eq.(37) analytically for several spectral indices. The results are given in Table 3.

At the current stage of the mission and data analysis, uncertainties in the colour corrections are much smaller than those of the gains G. However we aim to reduce the calibration error (using the orbital dipole) to below 0.2%. Two primary sources of error in $C(\alpha)$ will then need to be considered. The first is related to uncertainties in the bandpass model (Leahy et al. 2010; Zonca et al. 2009). The second arises from the uneven sampling of individual sky pixels by the full set of detectors, which causes pixel-to-pixel variations in the colour correction.

9. CMB removal

This section was developed in common with HFI (Planck HFI Core Team 2011b) and for this reason is reported identically in both papers.

In order to facilitate some foreground studies with the frequency maps, a set of maps were made available with the best estimate of the CMB subtracted from them. The steps undertaken jointly by HFI and LFI DPC teams in determining the best CMB map, subtracting it from the frequency maps and characterising the errors in the subtraction are described below.

9.1. Masks

Point source masks were constructed for each of the LFI frequency channels from the source catalogues produced by the LFI pipeline (Mexican Hat Wavelet 2 code, MHW2). All

Table 3. Colour Correction for LFI band for different input power-law spectral indices

Colour Correction for LFI maps							
Det ID	spectral index α						
	-2.0	-1.0	0.0	1.0	2.0	3.0	4.0
				70 GHz			
LFI18	1.054	1.028	1.011	1.003	1.003	1.010	1.026
LFI19	1.170	1.113	1.066	1.026	0.994	0.969	0.949
LFI20	1.122	1.079	1.044	1.017	0.997	0.983	0.975
LFI21	1.087	1.053	1.028	1.010	1.000	0.996	0.998
LFI22	0.973	0.971	0.976	0.988	1.007	1.033	1.066
LFI23	1.015	1.014	0.999	0.998	1.003	1.012	1.026
$\langle C \rangle_{70}$	1.070	1.041	1.021	1.007	1.001	1.001	1.007
				44 GHz			
LFI24	1.028	1.015	1.007	1.002	1.000	1.003	1.009
LFI25	1.039	1.024	1.013	1.005	1.000	0.999	1.000
LFI26	1.050	1.032	1.017	1.007	1.000	0.997	0.997
$\langle C \rangle_{44}$	1.039	1.024	1.012	1.004	1.000	0.999	1.002
				30 GHz			
LFI27	1.078	1.049	1.026	1.010	1.000	0.996	0.998
LFI28	1.079	1.049	1.026	1.009	1.000	0.997	1.002
$\langle C \rangle_{30}$	1.079	1.049	1.026	1.010	1.000	0.997	1.000

sources detected with a signal-to-noise (SNR) ratio greater than 5 were masked with a cut of radius $3\,\sigma \approx 1.27\,\mathrm{FWHM}$ of the effective beam. A similar process was applied to the HFI frequency maps (see Planck HFI Core Team 2011b).

A series of Galactic masks were constructed from the 30 GHz and 353 GHz frequency channel maps. An estimate of the CMB was subtracted from the maps in order not to bias the construction of the masks. The maps were smoothed to a common resolution of 5 degrees. The pixels within each mask were chosen to be those with values above a threshold value. The threshold values were chosen to produce masks with the desired fraction of the sky remaining. Five masks were made with $f_{\mathrm{sky}} = 0.99$, 0.97, 0.90, 0.80 and 0.60.

The point source masks and Galactic masks were provided as additional inputs to the component separation algorithms.

9.2. Selection of the CMB Template

Six component separation or foreground removal algorithms were applied to the HFI and LFI frequency channel maps to produce CMB maps. They are, in alphabetical order:

- AltICA: Internal linear combination (ILC) in map domain;
- CCA: Bayesian component separation in the map domain;
- FastMEM: Bayesian component separation in the harmonic domain;
- Needlet ILC: ILC in the needlet (wavelet) domain;
- SEVEM: Template fitting in map or wavelet domain;
- Wi-fit: Template fitting in wavelet domain.

see Leach et al. (2008) for further details. These six algorithms make different assumptions about the data and the combination of frequency channels used as input by each one was not necessarily the same. Comparing results from these methods demonstrated the consistency of the CMB template and provided estimates of the uncertainties in the reconstruction. A detailed comparison of the output of these methods, largely based on the CMB angular power spectrum, was used to select the CMB template that was removed from the frequency channel maps. The comparison was quantified using a jackknife procedure: each algorithm was applied to two additional sets of frequency maps made from the first half and second half of each pointing period. A residual map consisting of the half of the difference between the two reconstructed CMB maps was considered to be indicative of the noise level in the reconstruction from the full data set. The Needlet ILC (hereafter NILC) map was chosen as the CMB template because it contained the lowest noise level at small scales.

The CMB template was removed from the frequency channel maps after application of a filter in the spherical harmonic domain. The filter has a transfer function made of two factors. The first factor corresponds to the Gaussian beam of the channel to be cleaned; the second factor is a transfer function attenuating the multipoles of the CMB template which have low signal-to-noise ratio. It is designed in Wiener-like fashion, being close to unity at large scales up to multipoles around $\ell = 1000$ and then dropping smoothly to zero with a cut-off frequency around $\ell = 1700$. All angular frequencies above $\ell = 3900$ are completely suppressed. See Fig. 12 for the Wiener-like transfer function. This procedure was adopted to avoid doing more harm than good to the small scales of the frequency channel maps where the signal-to-noise ratio of the CMB is low.

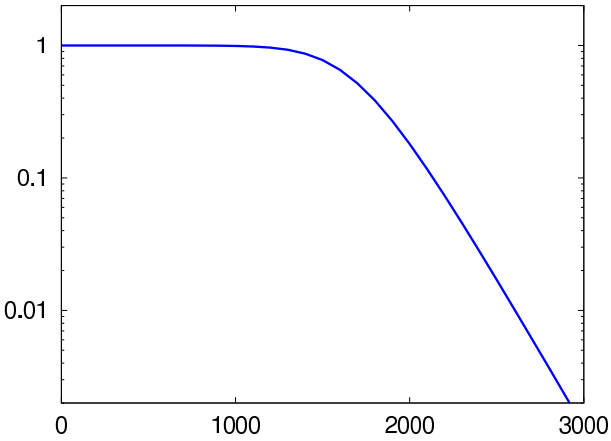

Fig. 12. Wiener-like transfer function applied for CMB removal.

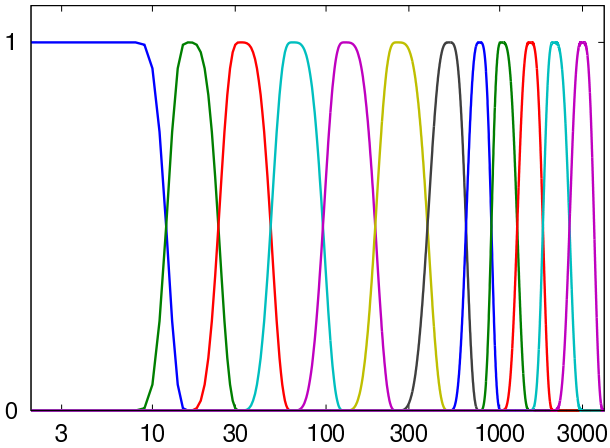

Fig. 13. The bandpass filters defining the spectral domains used in the Needlet ILC.

9.3. Description of Needlet ILC

Principle: The Needlet ILC map was produced using the ILC method in the "needlet" domain. Needlets, which are spherical wavelets, allow localizing both in multipole and sky direction. The input maps are decomposed into twelve overlapping multipole domains (called "scales"), using band-pass filters, shown on Fig. 13, and further decomposed into regions of the sky. Independent ILCs are applied in each sky region of each needlet scale. At large scale, large regions are used while smaller regions are used at fine scales.

Processing details: The Needlet ILC template was produced from all 6 HFI channels, using a tight Galactic mask shown on Fig. 14. It covers 99.36% of the sky but additional areas are excluded on a per-channel basis for point source masking.

Before applying Needlet ILC, the missing pixels due to point source masking and Galactic masking are filled in by a simple "diffusive inpainting" technique (it consists in replacing each missing pixel by the average of its neighbours and iterating this procedure until convergence, similar to solving the heat diffusion equation in the masked areas with boundary conditions given by the available pixels at the borders of the mask). Finally, all maps are "re-beamed" to a common resolution of 5 arc-minutes. Re-beaming induces

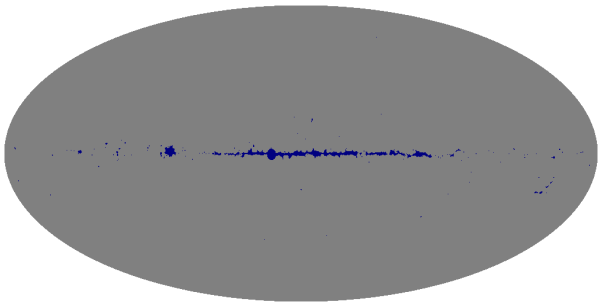

Fig. 14. Galactic mask used with Needlet ILC.

a blow up of the noise in the less resolved channels but that effect is of course automatically taken into account by the ILC filter. The CMB template obtained after needlet ILC processing, is then re-beamed to have the 'Wiener beam' shown in Fig. 12. The ILC coefficients are saved to be later applied to the jackknife maps for performance evaluation as described in Section 9.4.2. Fig. 15 shows LFI frequency maps where the CMB template estimated has been subtracted.

9.4. Uncertainties in the CMB Removal

The uncertainties in the CMB removal have been been estimated in two ways, first by comparing the CMB maps produced by the different algorithms and second by an internal estimation in NILC algorithm.

9.4.1. External estimate by comparing the CMB maps produced by the various algorithms.

The methods that were used to produce the estimates of the CMB, are diverse. They work by applying different algorithms, ILC, template fitting, or Bayesian parameter estimation, in a variety of domains, pixel space, Needlet/wavelets space, or spherical harmonic coefficients. Each method does its optimisation in a different way and thus will respond to the foregrounds differently. If we assume that the resulting CMB maps span the range of uncertainties induced by the foregrounds, we can provide an estimate of the uncertainties in the determination of the CMB, and thus in the subtraction process. The rms difference between the NILC map and the other CMB estimates is shown in Fig. 16. As expected, the uncertainties are largest in the Galactic plane, and smallest around the Ecliptic poles where the noise levels are lowest.

9.4.2. Internal Needlet ILC estimate of the uncertainties

The cleanliness of the CMB template produced by the Needlet ILC filter can be estimated using jackknives. We apply the Needlet ILC filter to the first and last halves of the ring set. The power distribution of the half-difference of the results provides us with a reliable estimation of the power of the noise in the Needlet ILC CMB template.

We use jackknives to estimate the contributions of sky signal and noise to the total data power. The data is in the form $X = S+N$ where, roughly speaking, S is the sky signal and N is the noise, independent of S. The total data power $\mathrm{Var}(X)$ decomposes as $\mathrm{Var}(X) = \mathrm{Var}(S) + \mathrm{Var}(N)$. The data power $\mathrm{Var}(X)$ is readily available and jackknives can

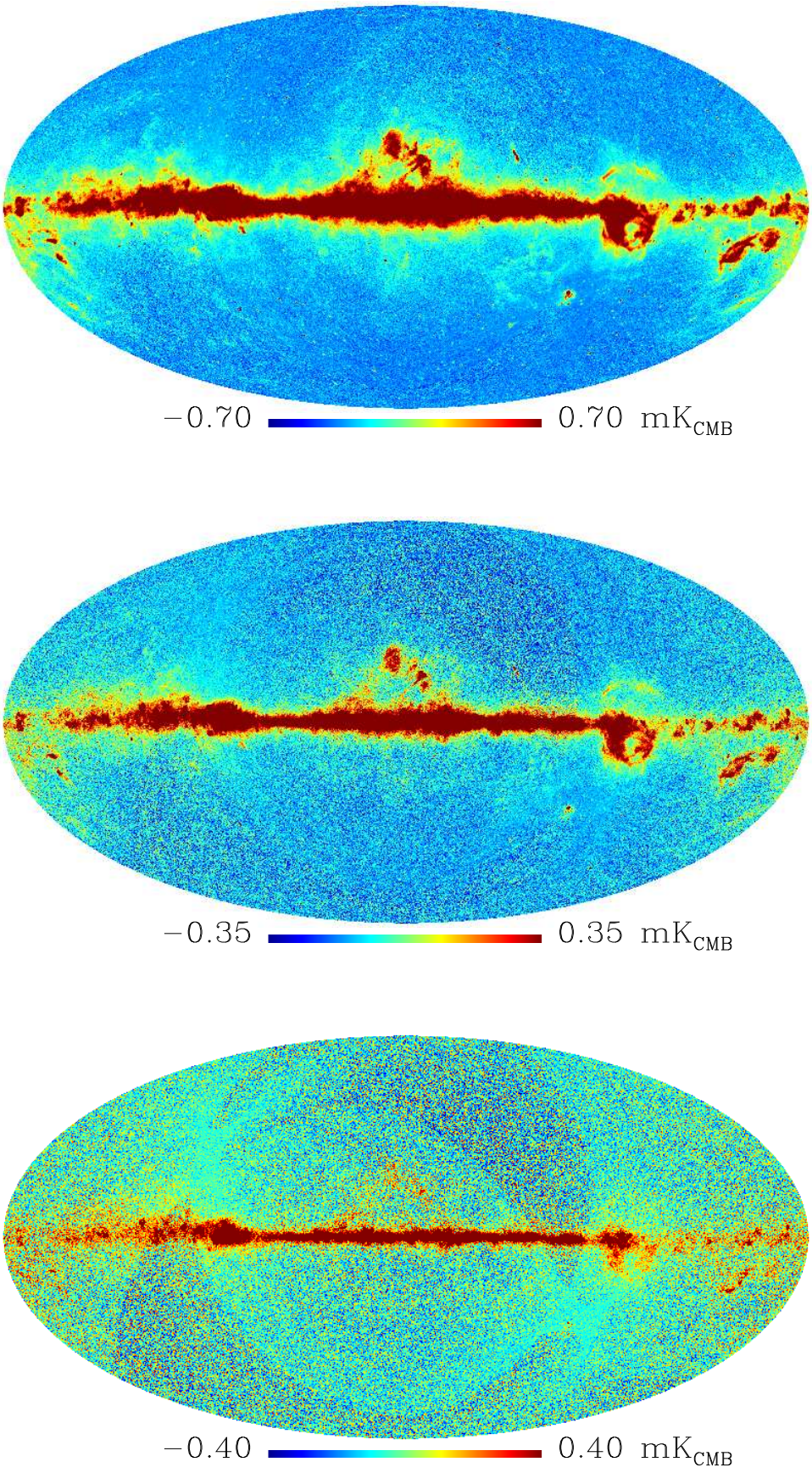

Fig. 15. LFI frequency maps where needlet ILC CMB template has been remove. Main galactic structures are clearly visible, as well as scanning strategy signatures at 44 and 70 GHz.

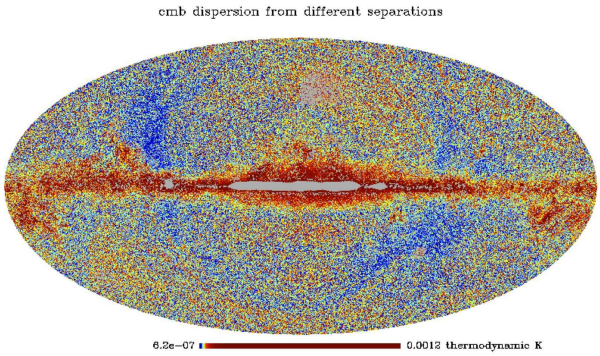

Fig. 16. Estimate of the rms error in the CMB subtraction. The map is plotted on a histogram-equalised scale to bring out the details.

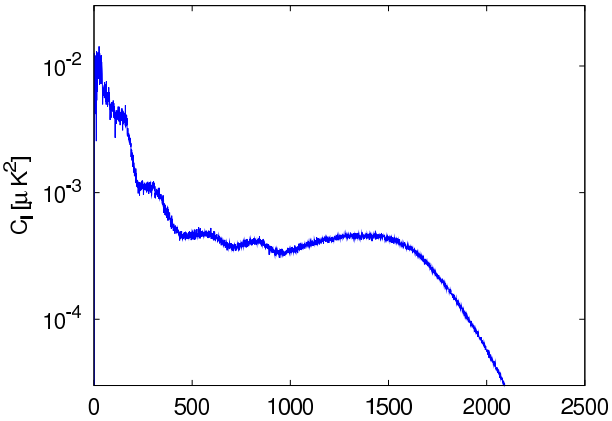

Fig. 18. Angular spectrum in μK^2 of the noise (estimated by jackknife) in the NILC CMB map. It corresponds to 11 μK per pixel.

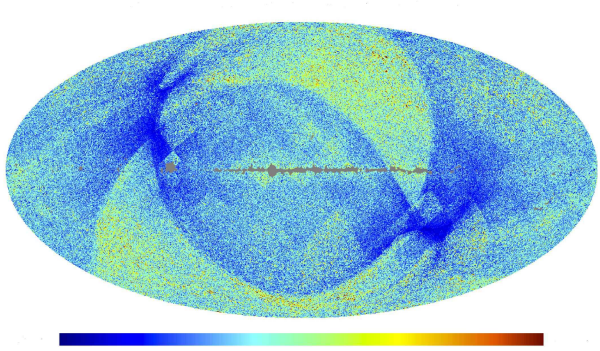

Fig. 17. Local rms of the noise (estimated by jackknife) in the NILC CMB map. The colour scale is from 0 to 30 μK per pixel at resolution $N_{\rm side} = 2048$.

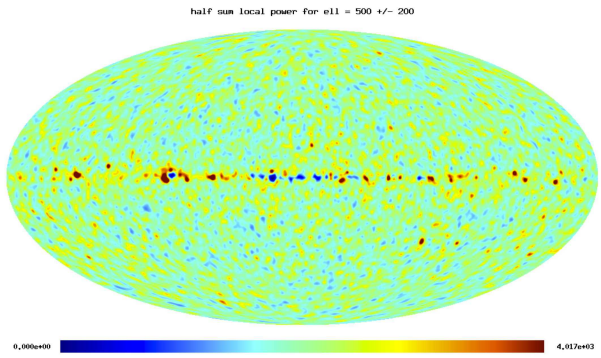

Fig. 19. Local power of the Needlet ILC CMB template in the range $\ell = 500 \pm 200$.

be used to obtain its two components Var(S) and Var(N). Specifically, Var(N) is obtained by by applying the NILC filter to half difference maps and Var(S) is obtained as Var(X) − Var(N). This procedure can be applied in pixel space, in harmonic space, or in pixel space *after* the maps have been bandpass-filtered, as described next.

We first used pixel space jackknifing to estimate the spatial distribution of noise. Fig. 17 shows a map of the local rms of the noise. We applied the NILC filter to a half-difference map and we display the square root of its smoothed squared values, effectively resulting in an estimate of the local noise rms. By the same device, we obtain an estimate of the angular spectrum of the noise in the NILC map, shown on Fig. 18. That spectrum corresponds to a rms $(1/4\pi \sum_\ell (2\ell+1) C_\ell)^{1/2}$ of 11 μK per pixel. The 'features' in the shape of the noise angular spectrum at large scale are a consequence of the needlet-based filtering (such features would not appear in a pixel-based ILC map). Recall that the coefficients of an ILC map are adjusted to minimize the total contamination by both foregrounds *and* noise. The strength of foregrounds relative to noise being larger at coarse scales, the needlet-based ILC tends to let more noise in, at the benefit of a better foreground rejection.

The half-difference maps offer a simple access to the power distribution of the residual noise in the estimated CMB template, but it is more difficult to evaluate other residual contamination since all fixed sky emissions cancel in half difference maps. All such large scale contamination are barely visible in the CMB template since they are dominated by the CMB itself. However, contamination is more conspicuous if one looks at intermediate scales. Fig. 19 shows the local power of the CMB template after it is band-passed to retain only the multipoles in the range $\ell = 500 \pm 200$.

10. Infrastructure Overview

To organize the large number of data processing codes and data products, the DPC employs the *Planck* Integrated Data and Information System (IDIS). The aim of this infrastructure is to allow flexible development of the processing pipeline, while ensuring complete traceability and reproducibility of data products.

In this respect, the most relevant components of IDIS are the Data Management Component (DMC) and the Process Coordinator (ProC), developed at the MPA Planck Analysis Centre (MPAC) at the Max-Planck-Institute for Astrophysics in Garching. Access to both components as well as to the *Planck* document and software management system is controlled via another IDIS component: the Federation Layer developed and maintained at the ESTEC RSSD.

Here we describe the essential features of the IDIS data processing components and their use at the DPC. A more

detailed description of these components and their capabilities can be found in (to be prepared).

10.1. Data Management Component - DMC

The IDIS DMC organizes the storage and access to DPC data products. To combine optimal performance in data I/O with the data management capabilities of modern databases, scientific data are stored in files, while metadata identifying them are stored in a database. The data files can only be modified in synchronization with the database, preventing concurrent access to data objects via locking mechanisms. The DMC software supports several database management systems of various complexity; the LFI DPC operates an Oracle 10g database which ensures good performance and stability.

The DMC provides a uniform application programming interface (API) for several programming languages (Fortran, C, C++, and Java), hiding all specific database operations from the user, who is therefore not required to have database experience. The DMC data types are defined in a Data Definition Layer (DDL) with description of both data and metadata structure. The DDL supports inheritance of data types (e.g., a data type `polarized_map` can be inherited from a data type `map`), and also association of data types (i.e. one data type containing a reference to another).

In addition to the API, the DMC provides a graphical user interface (GUI), which supports user queries of the database and retrieval of information on the data. The GUI offers the user the ability to list the (meta-)data of specific objects and also to visualize the data in a simple way (although data can also be exported to other powerful visualizing tools). The DMC GUI also allows the display of history information on data objects, permitting the user to browse intermediate data products which have been used in the data generation, and the controlled deletion of data, observing dependencies of data types and maintaining the history information for all remaining data. For this, the DMC relies on additional metadata on the processing history of the data objects, which is generated by the Process Coordinator workflow engine (ProC).

10.2. Pipeline Management - the ProC workflow engine

The ProC is a generic engine to construct, verify and execute computational workflows, comprised of computing modules and data flows between them. The modules can be written in any programming language, provided they conform to simple I/O format requirements which are described in a XML module description file. These interface files specify the input and output objects as well as the parameters of the individual programs in terms of DMC data types as described above.

The ProC provides a *Pipeline Editor* to support graphical construction of data processing workflows. It allows users to arrange and connect computing modules of a workflow in a clearly structured manner, and at the same time to configure the parameters of the algorithms used. It provides control structures for data flow, for data object I/O and consistent parameter definition (see Fig. 20 for an example).

The execution of workflows is controlled by a forward chaining algorithm, which ensures that modules are executed as soon as all necessary data products and parameters are known. If the same version of a module has been executed with identical inputs and parameters, the ProC will skip the execution and use the data product from the earlier execution for further processing. The ProC maintains control of pipeline execution also on massive parallel computing environments. In the LFI implementation, the ProC communicates with the PBS scheduling system to send jobs to the DPC cluster and to log their execution status.

The ProC logs workflow executions on log files, which can also contain logging messages of the executed modules. Additionally, it creates so-called *Pipeline-Run* and *Module-Run* objects in the DMC, which are used to recover the generation history of data products (including versions of processing modules via MD5-sums). Besides the GUI, the ProC can also be executed from the command line.

At the LFI DPC, the ProC is used to execute the official pipeline producing *Planck* data products. The current pipeline for the generation of calibrated TOI can be seen in Fig. 20.

11. Discussion and Conclusions

We have described the status of the pipeline as it stands at the time of the ERCSC release and the series of *Planck* early papers. All the algorithms run during this process have been verified, validated and tested before launch and the start of operations using realistic simulations. This allowed us to begin analyzing the data as soon as they were acquired from the first day of operations. The entire Level 1 pipeline did not suffer any significant problems and all of the data were transformed efficiently from telemetry packets to timelines. At present, the Level 2 pipeline is capable of providing relative calibration to an overall statistical accuracy in the range 0.05-0.1% and absolute calibration at around $\sim 1\%$ level. The beams are accurately characterised down to -10 dB level. We expect to improve many aspects in the near future. Concerning the calibration, our intention is to reach the levels determined by the stability of the instrument. For the beam reconstruction, our aim is to improve the characterisation of the far side lobes and to refine the entire beam reconstruction pipeline with particular attention to polarization measurements.

Acknowledgements. Planck is too large a project to allow full acknowledgement of all contributions by individuals, institutions, industries, and funding agencies. The main entities involved in the mission operations are as follows. The European Space Agency operates the satellite via its Mission Operations Centre located at ESOC (Darmstadt, Germany) and coordinates scientific operations via the Planck Science Office located at ESAC (Madrid, Spain). Two Consortia, comprising around 50 scientific institutes within Europe, the USA, and Canada, and funded by agencies from the participating countries, developed the scientific instruments LFI and HFI, and continue to operate them via Instrument Operations Teams located in Trieste (Italy) and Orsay (France). The Consortia are also responsible for scientific processing of the acquired data. The Consortia are led by the Principal Investigators: J.L. Puget in France for HFI (funded principally by CNES and CNRS/INSU-IN2P3) and N. Mandolesi in Italy for LFI(funded principally via ASI). NASA US Planck Project, based at JPL and involving scientists at many US institutions, contributes significantly to the efforts of these two Consortia. The author list for this paper has been selected by the Planck Science Team, and is composed of individuals from all of the above entities who have made multi-year contributions to the development of the mission. It does not pretend to be inclusive of all contributions. The

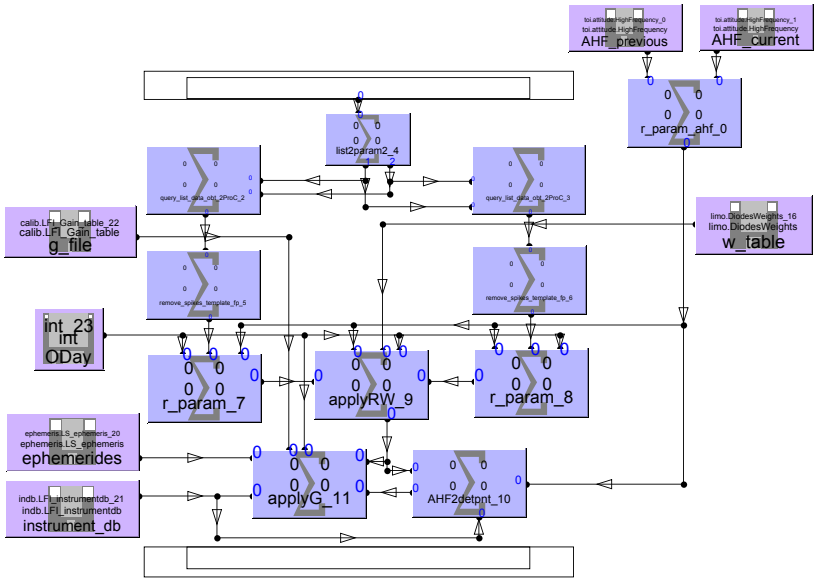

Fig. 20. The implemented part of the LFI DPC calibration pipeline in ProC editor view. The pipeline shows computing modules (blue) and I/O structures (purple), data and parameter flows are represented by arrows. A parameter element is used to ensure that the same parameter for operational day are used by all modules. The bars on top and bottom represent a ProC control element for parallel execution of the calibration pipeline over all horns and radiometers.

Planck-LFI project is developed by an International Consortium lead by Italy and involving Canada, Finland, Germany, Norway, Spain, Switzerland, UK, USA. The Italian contribution to *Planck* is supported by the Italian Space Agency (ASI) and INAF. This work was supported by the Academy of Finland grants 121703 and 121962. We thank the DEISA Consortium (www.deisa.eu), co-funded through the EU FP6 project RI-031513 and the FP7 project RI-222919, for support within the DEISA Virtual Community Support Initiative. We thank CSC – IT Center for Science Ltd (Finland) for computational resources. We acknowledge financial support provided by the Spanish Ministerio de Ciencia e Innovaciõn through the Plan Nacional del Espacio y Plan Nacional de Astronomia y Astrofisica. We acknowledge The Max Planck Institute for Astrophysics Planck Analysis Centre (MPAC) is funded by the Space Agency of the German Aerospace Center (DLR) under grant 50OP0901 with resources of the German Federal Ministry of Economics and Technology, and by the Max Planck Society. This work has made use of the Planck satellite simulation package (Level-S), which is assembled by the Max Planck Institute for Astrophysics Planck Analysis Centre (MPAC) Reinecke et al. (2006). We acknowledge financial support provided by the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Some of the results in this paper have been derived using the HEALPix package Górski et al. (2005). A description of the Planck Collaboration and a list of its members, indicating which technical or scientific activities they have been involved in, can be found at http://www.rssd.esa.int/index.php?project=PLANCK&page=Planck_Collaboration.

References

Ashdown, M. A. J., Baccigalupi, C., Balbi, A., et al. 2007a, A&A, 471, 361
Ashdown, M. A. J., Baccigalupi, C., Balbi, A., et al. 2007b, A&A, 467, 761
Ashdown, M. A. J., Baccigalupi, C., Bartlett, J. G., et al. 2009, A&A, 493, 753
Bersanelli, M., Mandolesi, N., Butler, R. C., et al. 2010, A&A, 520, A4+
Burigana, C., Natoli, P., Vittorio, N., Mandolesi, N., & Bersanelli, M. 2001, Experimental Astronomy, 12, 87, 10.1023/A:1016338603042
Cappellini, B., Maino, D., Albetti, G., et al. 2003, A&A, 409, 375
de Gasperis, G., Balbi, A., Cabella, P., Natoli, P., & Vittorio, N. 2005, A&A, 436, 1159
Frailis, M., Maris, M., Zacchei, A., et al. 2009, Journal of Instrumentation, 4, 2021
Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759
Hinshaw, G., Spergel, D. N., Verde, L., et al. 2003, ApJS, 148, 135
Hoffman, Y. & Ribak, E. 1991, ApJ, 380, L5
Keihänen, E., Keskitalo, R., Kurki-Suonio, H., Poutanen, T., & Sirviö, A. 2010, A&A, 510, A57+
Keihänen, E., Kurki-Suonio, H., & Poutanen, T. 2005, MNRAS, 360, 390
Kurki-Suonio, H., Keihänen, E., Keskitalo, R., et al. 2009, A&A, 506, 1511
Lamarre, J., Puget, J., Ade, P. A. R., et al. 2010, A&A, 520, A9+
Leach, S. M., Cardoso, J., Baccigalupi, C., et al. 2008, A&A, 491, 597
Leahy, J. P., Bersanelli, M., D'Arcangelo, O., et al. 2010, A&A, 520, A8+
Maino, D., Burigana, C., Górski, K. M., Mandolesi, N., & Bersanelli, M. 2002, A&A, 387, 356
Mandolesi, N., Bersanelli, M., Butler, R. C., et al. 2010, A&A, 520, A3+
Maris, M., Tomasi, M., Galeotta, S., et al. 2009, Journal of Instrumentation, 4, 2018
Massardi, M. 2006, in CMB and Physics of the Early Universe
Meinhold, P., Leonardi, R., Aja, B., et al. 2009, Journal of Instrumentation, 4, 2009
Mennella, A., Bersanelli, M., Butler, R. C., et al. 2010, A&A, 520, A5+
Mennella, A., Bersanelli, M., Seiffert, M., et al. 2003, A&A, 410, 1089
Mennella et al. 2011, Planck early results 03: First assessment of the Low Frequency Instrument in-flight performance (Submitted to A&A)
Natoli, P., de Gasperis, G., Gheller, C., & Vittorio, N. 2001, A&A, 372, 346
Pasian, F. & Gispert, R. 2000, Astrophysical Letters Communications, 37, 247
Planck Collaboration. 2011a, Planck early results 01: The Planck mission (Submitted to A&A)

Planck Collaboration. 2011b, Planck early results 02: The thermal performance of Planck (Submitted to A&A)
Planck Collaboration. 2011c, Planck early results 07: The Early Release Compact Source Catalogue (Submitted to A&A)
Planck Collaboration. 2011d, Planck early results 08: The all-sky early Sunyaev-Zeldovich cluster sample (Submitted to A&A)
Planck Collaboration. 2011e, Planck early results 09: XMM-Newton follow-up for validation of Planck cluster candidates (Submitted to A&A)
Planck Collaboration. 2011f, Planck early results 10: Statistical analysis of Sunyaev-Zeldovich scaling relations for X-ray galaxy clusters (Submitted to A&A)
Planck Collaboration. 2011g, Planck early results 11: Calibration of the local galaxy cluster Sunyaev-Zeldovich scaling relations (Submitted to A&A)
Planck Collaboration. 2011h, Planck early results 12: Cluster Sunyaev-Zeldovich optical Scaling relations (Submitted to A&A)
Planck Collaboration. 2011i, Planck early results 13: Statistical properties of extragalactic radio sources in the Planck Early Release Compact Source Catalogue (Submitted to A&A)
Planck Collaboration. 2011j, Planck early results 14: Early Release Compact Source Catalogue validation and extreme radio sources (Submitted to A&A)
Planck Collaboration. 2011k, Planck early results 15: Spectral energy distributions and radio continuum spectra of northern extragalactic radio sources (Submitted to A&A)
Planck Collaboration. 2011l, Planck early results 16: The Planck view of nearby galaxies (Submitted to A&A)
Planck Collaboration. 2011m, Planck early results 17: Origin of the submillimetre excess dust emission in the Magellanic Clouds (Submitted to A&A)
Planck Collaboration. 2011n, Planck early results 18: The power spectrum of cosmic infrared background anisotropies (Submitted to A&A)
Planck Collaboration. 2011o, Planck early results 19: All-sky temperature and dust optical depth from Planck and IRAS — constraints on the "dark gas" in our Galaxy (Submitted to A&A)
Planck Collaboration. 2011p, Planck early results 20: New light on anomalous microwave emission from spinning dust grains (Submitted to A&A)
Planck Collaboration. 2011q, Planck early results 21: Properties of the interstellar medium in the Galactic plane (Submitted to A&A)
Planck Collaboration. 2011r, Planck early results 22: The submillimetre properties of a sample of Galactic cold clumps (Submitted to A&A)
Planck Collaboration. 2011s, Planck early results 23: The Galactic cold core population revealed by the first all-sky survey (Submitted to A&A)
Planck Collaboration. 2011t, Planck early results 24: Dust in the diffuse interstellar medium and the Galactic halo (Submitted to A&A)
Planck Collaboration. 2011u, Planck early results 25: Thermal dust in nearby molecular clouds (Submitted to A&A)
Planck Collaboration. 2011v, The Explanatory Supplement to the Planck Early Release Compact Source Catalogue (ESA)
Planck HFI Core Team. 2011a, Planck early results 04: First assessment of the High Frequency Instrument in-flight performance (Submitted to A&A)
Planck HFI Core Team. 2011b, Planck early results 06: The High Frequency Instrument data processing (Submitted to A&A)
Prunet, S., Ade, P. A. R., Bock, J. J., et al. 2001, ArXiv Astrophysics e-prints
Reinecke, M., Dolag, K., Hell, R., Bartelmann, M., & Enßlin, T. A. 2006, A&A, 445, 373
Rosset, C., Tristram, M., Ponthieu, N., et al. 2010, A&A, 520, A13+
Sandri, M., Villa, F., Bersanelli, M., et al. 2010, A&A, 520, A7+
Tauber, J. A., Mandolesi, N., Puget, J., et al. 2010a, A&A, 520, A1+
Tauber, J. A., Norgaard-Nielsen, H. U., Ade, P. A. R., et al. 2010b, A&A, 520, A2+
Zacchei, A., Frailis, M., Maris, M., et al. 2009, Journal of Instrumentation, 4, 2019
Zacchei et al. 2011, Planck early results 05: The Low Frequency Instrument data processing (Submitted to A&A)
Zonca, A., Franceschet, C., Battaglia, P., et al. 2009, Journal of Instrumentation, 4, 2010

[1] Aalto University Metsähovi Radio Observatory, Metsähovintie 114, FIN-02540 Kylmälä, Finland

[2] Agenzia Spaziale Italiana Science Data Center, c/o ESRIN, via Galileo Galilei, Frascati, Italy

[3] Agenzia Spaziale Italiana, Viale Liegi 26, Roma, Italy

[4] Astroparticule et Cosmologie, CNRS (UMR7164), Université Denis Diderot Paris 7, Bâtiment Condorcet, 10 rue A. Domon et Léonie Duquet, Paris, France

[5] Australia Telescope National Facility, CSIRO, P.O. Box 76, Epping, NSW 1710, Australia

[6] CNR - ISTI, Area della Ricerca, via G. Moruzzi 1, Pisa, Italy

[7] CNRS, IRAP, 9 Av. colonel Roche, BP 44346, F-31028 Toulouse cedex 4, France

[8] California Institute of Technology, Pasadena, California, U.S.A.

[9] Centre of Mathematics for Applications, University of Oslo, Blindern, Oslo, Norway

[10] DTU Space, National Space Institute, Juliane Mariesvej 30, Copenhagen, Denmark

[11] Department of Physics & Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, British

Columbia, Canada

[12] Department of Physics and Astronomy, University of Southern California, Los Angeles, California, U.S.A.

[13] Department of Physics, Gustaf Hällströmin katu 2a, University of Helsinki, Helsinki, Finland

[14] Department of Physics, University of California, Berkeley, California, U.S.A.

[15] Department of Physics, University of California, One Shields Avenue, Davis, California, U.S.A.

[16] Department of Physics, University of California, Santa Barbara, California, U.S.A.

[17] Department of Physics, University of Oxford, 1 Keble Road, Oxford, U.K.

[18] Dipartimento di Fisica G. Galilei, Università degli Studi di Padova, via Marzolo 8, 35131 Padova, Italy

[19] Dipartimento di Fisica, Università La Sapienza, P. le A. Moro 2, Roma, Italy

[20] Dipartimento di Fisica, Università degli Studi di Milano, Via Celoria, 16, Milano, Italy

[21] Dipartimento di Fisica, Università degli Studi di Trieste, via A. Valerio 2, Trieste, Italy

[22] Dipartimento di Fisica, Università di Roma Tor Vergata, Via della Ricerca Scientifica, 1, Roma, Italy

[23] Dipartimento di Matematica, Università di Roma Tor Vergata, Via della Ricerca Scientifica, 1, Roma, Italy

[24] Dpto. Astrofísica, Universidad de La Laguna (ULL), E-38206 La Laguna, Tenerife, Spain

[25] European Space Agency, ESAC, Planck Science Office, Camino bajo del Castillo, s/n, Urbanización Villafranca del Castillo, Villanueva de la Cañada, Madrid, Spain

[26] European Space Agency, ESTEC, Keplerlaan 1, 2201 AZ Noordwijk, The Netherlands

[27] Haverford College Astronomy Department, 370 Lancaster Avenue, Haverford, Pennsylvania, U.S.A.

[28] Helsinki Institute of Physics, Gustaf Hällströmin katu 2, University of Helsinki, Helsinki, Finland

[29] INAF - Osservatorio Astrofisico di Arcetri, Largo Enrico Fermi 5, Firenze, Italy

[30] INAF - Osservatorio Astrofisico di Catania, Via S. Sofia 78, Catania, Italy

[31] INAF - Osservatorio Astronomico di Padova, Vicolo dell'Osservatorio 5, Padova, Italy

[32] INAF - Osservatorio Astronomico di Roma, via di Frascati 33, Monte Porzio Catone, Italy

[33] INAF - Osservatorio Astronomico di Trieste, Via G.B. Tiepolo 11, Trieste, Italy

[34] INAF/IASF Bologna, Via Gobetti 101, Bologna, Italy

[35] INAF/IASF Milano, Via E. Bassini 15, Milano, Italy

[36] INRIA, Laboratoire de Recherche en Informatique, Université Paris-Sud 11, Bâtiment 490, 91405 Orsay Cedex, France

[37] ISDC Data Centre for Astrophysics, University of Geneva, ch. d'Ecogia 16, Versoix, Switzerland

[38] Infrared Processing and Analysis Center, California Institute of Technology, Pasadena, CA 91125, U.S.A.

[39] Institut d'Astrophysique Spatiale, CNRS (UMR8617) Université Paris-Sud 11, Bâtiment 121, Orsay, France

[40] Institut d'Astrophysique de Paris, CNRS UMR7095, Université Pierre & Marie Curie, 98 bis boulevard Arago, Paris, France

[41] Institute of Theoretical Astrophysics, University of Oslo, Blindern, Oslo, Norway

[42] Instituto de Astrofísica de Canarias, C/Vía Láctea s/n, La Laguna, Tenerife, Spain

[43] Instituto de Física de Cantabria (CSIC-Universidad de Cantabria), Avda. de los Castros s/n, Santander, Spain

[44] Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, California, U.S.A.

[45] Jodrell Bank Centre for Astrophysics, Alan Turing Building, School of Physics and Astronomy, The University of

Manchester, Oxford Road, Manchester, M13 9PL, U.K.

[46] LERMA, CNRS, Observatoire de Paris, 61 Avenue de l'Observatoire, Paris, France

[47] Lawrence Berkeley National Laboratory, Berkeley, California, U.S.A.

[48] Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 85741 Garching, Germany

[49] MilliLab, VTT Technical Research Centre of Finland, Tietotie 3, Espoo, Finland

[50] SISSA, Astrophysics Sector, via Bonomea 265, 34136, Trieste, Italy

[51] Space Sciences Laboratory, University of California, Berkeley, California, U.S.A.

[52] Spitzer Science Center, 1200 E. California Blvd., Pasadena, California, U.S.A.

[53] Université de Toulouse, UPS-OMP, IRAP, F-31028 Toulouse cedex 4, France

[54] University of Cambridge, Institute of Astronomy, Madingley Road, Cambridge, U.K.

[55] Warsaw University Observatory, Aleje Ujazdowskie 4, 00-478 Warszawa, Poland