Lawrence Berkeley National Laboratory

Title

BEYONDPLANCK

Permalink

https://escholarship.org/uc/item/403251qj

Authors

Andersen, KJ

Aurlien, R

Banerji, R

et al.

Publication Date

2023-07-01

DOI

10.1051/0004-6361/202244953

Peer reviewed

BeyondPlanck I. Global Bayesian analysis of the *Planck* Low Frequency Instrument data

BeyondPlanck Collaboration[*]: K. J. Andersen[1], R. Aurlien[1], R. Banerji[1], A. Basyrov[1], M. Bersanelli[2,3,4], S. Bertocco[5], M. Brilenkov[1], M. Carbone[6], L. P. L. Colombo[2,4], H. K. Eriksen[1], J. R. Eskilt[1], M. K. Foss[1], C. Franceschet[2,4], U. Fuskeland[1], S. Galeotta[5], M. Galloway[1], S. Gerakakis[6], E. Gjerløw[1], B. Hensley[7], D. Herman[1], M. Iacobellis[6], M. Ieronymaki[6], H. T. Ihle[1], J. B. Jewell[8], A. Karakci[1], E. Keihänen[9,10], R. Keskitalo[11], J. G. S. Lunde[1], G. Maggio[5], D. Maino[2,3,4], M. Maris[5], A. Mennella[2,3,4], S. Paradiso[2,4], B. Partridge[12], M. Reinecke[13], M. San[1], N.-O. Stutzer[1], A.-S. Suur-Uski[9,10], T. L. Svalheim[1], D. Tavagnacco[5,14], H. Thommesen[1], D. J. Watts[1], I. K. Wehus[1], and A. Zacchei[5]

[1] Institute of Theoretical Astrophysics, University of Oslo, Blindern, Oslo, Norway
[2] Dipartimento di Fisica, Università degli Studi di Milano, Via Celoria, 16, Milano, Italy
[3] INAF-IASF Milano, Via E. Bassini 15, Milano, Italy
[4] INFN, Sezione di Milano, Via Celoria 16, Milano, Italy
[5] INAF - Osservatorio Astronomico di Trieste, Via G.B. Tiepolo 11, Trieste, Italy
[6] Planetek Hellas, Leoforos Kifisias 44, Marousi 151 25, Greece
[7] Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, U.S.A.
[8] Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, California, U.S.A.
[9] Department of Physics, Gustaf Hällströmin katu 2, University of Helsinki, Helsinki, Finland
[10] Helsinki Institute of Physics, Gustaf Hällströmin katu 2, University of Helsinki, Helsinki, Finland
[11] Computational Cosmology Center, Lawrence Berkeley National Laboratory, Berkeley, California, U.S.A.
[12] Haverford College Astronomy Department, 370 Lancaster Avenue, Haverford, Pennsylvania, U.S.A.
[13] Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 85741 Garching, Germany
[14] Dipartimento di Fisica, Università degli Studi di Trieste, via A. Valerio 2, Trieste, Italy

September 13, 2022

ABSTRACT

We describe the BeyondPlanck project in terms of motivation, methodology and main products, and provide a guide to a set of companion papers that describe each result in fuller detail. Building directly on experience from ESA's *Planck* mission, we implement a complete end-to-end Bayesian analysis framework for the *Planck* Low Frequency Instrument (LFI) observations. The primary product is a full joint posterior distribution $P(\omega \mid d)$, where ω represents the set of all free instrumental (gain, correlated noise, bandpass etc.), astrophysical (synchrotron, free-free, thermal dust emission etc.), and cosmological (CMB map, power spectrum etc.) parameters. Some notable advantages of this approach compared to a traditional pipeline procedure are seamless end-to-end propagation of uncertainties; accurate modeling of both astrophysical and instrumental effects in the most natural basis for each uncertain quantity; optimized computational costs with little or no need for intermediate human interaction between various analysis steps; and a complete overview of the entire analysis process within one single framework. As a practical demonstration of this framework, we focus in particular on low-ℓ CMB polarization reconstruction with *Planck* LFI. In this process, we identify several important new effects that have not been accounted for in previous pipelines, including gain over-smoothing and time-variable and non-$1/f$ correlated noise in the 30 and 44 GHz channels. Modelling and mitigating both previously known and newly discovered systematic effects, we find that all results are consistent with the ΛCDM model, and we constrain the reionization optical depth to $\tau = 0.066 \pm 0.013$, with a low-resolution CMB-based χ^2 probability-to-exceed of 32 %. This uncertainty is about 30 % larger than the official pipelines, arising from taking into account a more complete instrumental model. The marginal CMB Solar dipole amplitude is $3362.7 \pm 1.4 \mu$K, where the error bar is derived directly from the posterior distribution without the need of any *ad-hoc* instrumental corrections. We are currently not aware of any significant unmodelled systematic effects remaining in the *Planck* LFI data, and, for the first time, the 44 GHz channel is fully exploited in the current analysis. We argue that this framework can play a central role in the analysis of many current and future high-sensitivity CMB experiments, including *LiteBIRD*, and it will serve as the computational foundation of the emerging community-wide CosmoGlobe effort, which aims to combine state-of-the-art radio, microwave, and submillimeter data sets into one global astrophysical model. All software is made publicly available under an OpenSource license, and both codes and products may be obtained through http://beyondplanck.science.

Key words. ISM: general – Cosmology: observations, polarization, cosmic microwave background, diffuse radiation – Galaxy: general

Contents

1. Introduction

1.1. CMB cosmology

According to the current cosmological concordance model, the observable universe came into existence some 13.8 billion years ago in a process often referred to as the Big Bang. While the physical laws underpinning this singular event remain unknown, it is a testament to the success of modern cosmology that physicists today are able to measure and model the evolution and energy content of the universe to exquisite precision, starting only a fraction of a second after the Big Bang.

Among the most important cosmological observables is the cosmic microwave background (CMB), first detected by Penzias & Wilson (1965). This leftover heat from the Big Bang fills the entire universe, and may today be observed as a nearly isotropic blackbody signal with a temperature of 2.7255 K (Fixsen 2009). Initially, CMB photons were trapped locally within a dense electron–proton plasma by Thomson scattering. However, once the mean plasma temperature fell below 3000 K as the Universe expanded, electrons and protons combined into neutral hydrogen atoms, and the photons were free to move throughout the entire observable universe, with almost no further scattering. This event took place some 380 000 years after the Big Bang, at a time often referred to as the *recombination epoch*. To any observer, the resulting photons appear to come from a so-called
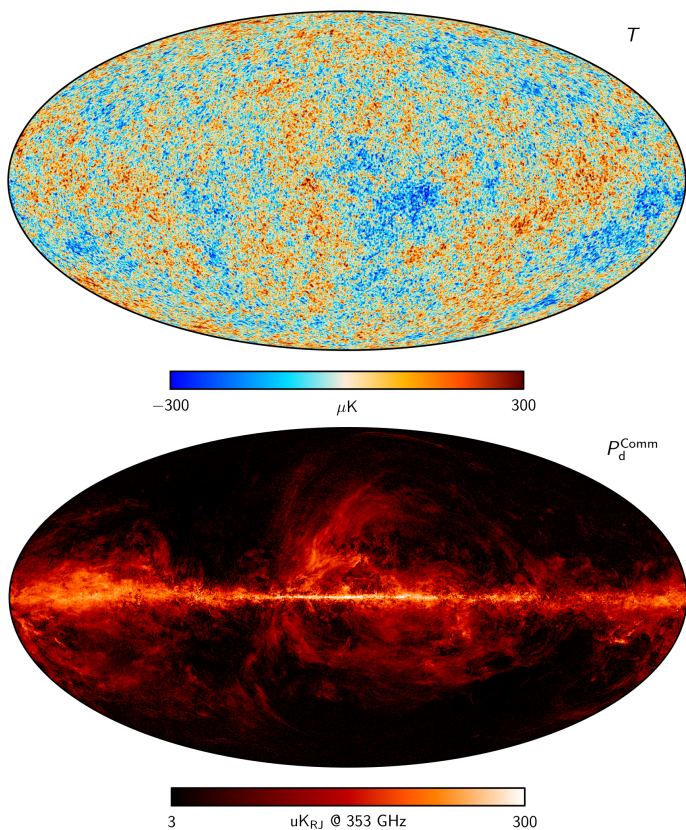
Fig. 1. *Planck* 2018 `Commander` CMB temperature (*top panel*) and thermal dust polarization amplitude (*bottom panel*) maps. Reproductions from Planck Collaboration IV (2018).

last-scattering surface, a sphere corresponding to a light travel distance of just under the entire history of the universe.

While the CMB field is very nearly isotropic, it does exhibit small spatial variations at the $O(30\,\mu\text{K})$ level (e.g., Hu & Dodelson 2002, and references therein). These fluctuations are produced primarily by variations in the local gravitational potential, temperature, density, and velocity at the last-scattering surface. Smaller amplitude fluctuations arise from various secondary interactions taking place after the photons leave the last-scattering surface, for instance through gravitational lensing or Thomson scattering in the hot, ionized medium in clusters of galaxies. It is precisely by measuring and modelling all these small variations that cosmologists are able to decipher the history of the universe in ever greater detail. The current best-fit cosmological model derived from this work is often referred to as *the ΛCDM model*, which posits that the universe is isotropic and homogeneous on large scales; that it started in a hot Big Bang; that it underwent a brief period of exponential expansion called inflation that seeded the universe with Gaussian random density fluctuations drawn from a scale-invariant power spectrum; and that the energy contents of the universe comprise 4.9 % baryonic matter, 26.5 % cold dark matter, and 68.5 % dark energy (Planck Collaboration VI 2020). Flat spatial curvature is also frequently assumed.

The rich cosmological information embedded in the CMB is not, however, easy to extract. Even the most dominant physical effects produce only temperature fluctuations at the 10^{-5}–10^{-4} level in the CMB. A primary goal for next-generation CMB experiments is the detection of primordial gravitational waves through the subtle polarization they imprint on the CMB (e.g., Kamionkowski & Kovetz 2016, and references therein). These

so-called *B*-modes are likely to have an amplitude no larger than 30 nK, or a relative amplitude smaller than $O(10^{-8})$.

The fact that current CMB observations reach the μK level in the face of instrument noise and systematics is a testament to the effort of many scientists and engineers in this field, and to the time and money they have spent. Here, we list only a few of the results of five decades of observational milestones. NASA's *COBE* mission produced the first highly accurate measurement of the thermal spectrum of the CMB (Mather et al. 1994) and the first detection of large scale fluctuations in the CMB (Smoot et al. 1992). The first high-fidelity map of smaller scale CMB fluctuations was made by the BOOMERanG team (de Bernardis et al. 2000), and the first detection of polarized fluctuations by DASI (Kovac et al. 2002). These are among the more than 50 past and present CMB projects, all of which have contributed to technological innovations or scientific breakthroughs.

Two space missions, however, are primarily responsible for today's cosmological concordance model. They are NASA's *Wilkinson Microwave Anisotropy Probe* (*WMAP*; Bennett et al. 2013) and ESA's *Planck* (Planck Collaboration I 2020) satellite missions. *WMAP* was the first experiment to take full advantage of the exquisite thermal stability at Earth's second Lagrange point (L2), and observed the CMB sky for nine years (2001–2010) in five frequency bands (23–94 GHz) with precision unprecedented at the time.

1.2. Planck

The state-of-the-art in all-sky CMB observations as of 2022 is defined by ESA's cosmology flagship mission called *Planck* (Planck Collaboration I 2020). *Planck* observed the CMB sky for four years (2009–2013) in nine frequency bands (30–857 GHz), with three times higher angular resolution and ten times higher sensitivity than *WMAP*. Its original design goal was to measure the primary CMB temperature fluctuations with a precision limited only by fundamental physical processes, including cosmic variance, not by instrumental sensitivity (Planck Collaboration 2005).

Planck comprised two separate instruments within a common focal plane. One was the *Low Frequency Instrument* (LFI; Planck Collaboration II 2020), which employed coherent High Electron Mobility Transistor (HEMT) radiometers with center frequencies near 30, 44 and 70 GHz, each with a fractional bandwidth of roughly 20 %. The other was the *High Frequency Instrument* (HFI; Planck Collaboration III 2020), which employed spider-web and polarization sensitive bolometers with center frequencies of 100, 143, 217, 353, 545 and 857 GHz, each with a fractional bandwidth of 25 %.

Two different detector technologies were required to span *Planck*'s frequency range. The use of two very different detector technologies also provided crucial cross-checks against subtle instrumental errors. *Planck*'s wide frequency range fully covered most of the spectrum of a 2.7255 K blackbody, but more crucially allowed for the removal of contaminating foreground signals (e.g., Leach et al. 2008). These arise from synchrotron emission from relativistic electrons moving in the magnetic field of the Galaxy, thermal emission from warm Galactic dust and bremsstrahlung emission from ionized gas, as well as microwave emission from extra-galactic sources. This list is not exhaustive; but each mechanism for foreground emission has a unique spatial distribution on the sky and a unique, non-blackbody spectrum which allows it to be distinguished from the CMB. The preferred method for separating cosmological fluctuations in the CMB from astrophysical foreground signals is to map the sky

at multiple frequencies, and then perform a joint fit to this set of maps while taking into account the particular spatial and spectral behaviour of each foreground. These considerations drove the design of *Planck* (Planck Collaboration 2005). The capability to detect polarized signals was added at the seven lowest frequency bands, from 30 to 353 GHz. Figure 1 shows the CMB temperature fluctuation and the polarized thermal dust emission maps as derived from *Planck* observations, which rank among the most iconic results from the experiment.

The *Planck* satellite was launched to L2 on May 14th 2009 and deactivated on October 23rd 2013; it thus completed in total almost 4.5 years of observations (Planck Collaboration I 2020). Unlike the case for *WMAP*, both *Planck* instruments were cryogenically cooled. The last 18 months of operation included only LFI science measurements, as HFI exhausted its cooling capacity in January 2012.

The first *Planck* data release (denoted either "PR1" or 2013 here; Planck Collaboration I 2014) took place in March 2013, and was based on the first 15.5 months of data, covering the full sky twice. By and large, these measurements confirmed the cosmological model presented by *WMAP* and other previous experiments, but with significantly higher signal-to-noise ratio. This higher sensitivity also supported several truly groundbreaking results, two of which were a 25σ detection of gravitational lensing of CMB anisotropies (Planck Collaboration XVII 2014), and a revolutionary new image of polarized thermal dust emission in the Milky Way (Planck Collaboration XI 2014).

The 2013 release, however, did not include any CMB polarization results. In addition, the initial angular power spectrum of CMB anisotropies exhibited a ~2 % shift in amplitude compared to the earlier *WMAP* power spectrum (Planck Collaboration XV 2014). Both of these issues had a common origin, namely incompletely controlled systematic errors arising from instrumental effects (Planck Collaboration I 2016). As noted earlier, CMB observations are not easy: even small errors in assumptions made about foregrounds or instrumental behaviour can have dramatic effects on the recovered CMB signal. Examples of instrumental effects include: uncertainties in the beam shape and far sidelobes; mis-estimation of the frequency response of detectors, which can introduce temperature to polarization leakage; unaccounted-for non-linearity in the analog-to-digital converters (ADCs) used in each detector chain; and uncertainties in the polarization properties of detectors.

The *Planck* team grappled with all of these, as well as uncertainties in foreground contamination, in the years between 2013 and the release of the final *Planck* results in 2020 (Planck Collaboration I 2020). Very substantial investments of time and money were made to develop increasingly accurate models of the two *Planck* instruments; these allowed for more precise and robust science results. We emphasize that the official LFI and HFI pipelines evolved step-by-step in the post-launch period as instrument-specific effects emerged due to increased calibration accuracy. BEYONDPLANCK builds on all this accumulated experience in implementing a global approach to the data analysis problem.

A major milestone in this iterative process was the second *Planck* data release ("PR2" or 2015; Planck Collaboration I 2016), which for the first time included the full set of *Planck* observations (50 months of LFI data and 27 months of HFI data). At this point, the polarization properties of both the LFI and HFI instruments were sufficiently well understood to allow for a direct measurement of CMB polarization on intermediate and small angular scales (Planck Collaboration XI 2016). For HFI, however, accurate large-scale polarization was still out of reach

due to systematic errors, and only LFI provided such constraints. The original power spectrum discrepancy relative to *WMAP* was tracked down to inaccuracies in the calibration procedure and reference dipole values used for the *Planck* 2013 analysis, and these were subsequently corrected in the 2015 release. With this second data release, *Planck* finally fulfilled its promise of measuring the primary CMB temperature fluctuations to the limits set by astrophysical and cosmological effects (Planck Collaboration I 2016).

1.3. Large-scale CMB polarization, the reionization optical depth, and systematic errors

Planck analysis continued beyond 2015, with a particular emphasis on reducing large-scale polarization systematics (Planck Collaboration I 2020). Both the importance and difficulty of this specific issue may be summarized in terms of the reionization optical depth, τ (e.g., Planck Collaboration Int. XLVII 2016). This parameter is directly related to the epoch during which the first stars were born, often called the *epoch of reionization* (e.g., Loeb & Barkana 2001, and references therein). According to detailed measurements of the abundance of neutral hydrogen in the universe from quasar spectra (the so-called "Lyman alpha forest"; Gunn & Peterson 1965), this event cannot have happened later than about 1 billion years after the Big Bang, corresponding to an optical depth of $\tau \gtrsim 0.048$. However, an independent measurement of τ may also be derived through CMB observations, by noting that the first stars or galaxies ionized their surrounding medium, and thereby released large numbers of free electrons off which CMB photons could scatter. Detailed models predict a CMB polarization signal with an amplitude of about 0.5 μK on angular scales larger than 10° (e.g., Alvarez et al. 2006, and references therein).

While the scientific potential in establishing robust large-scale polarization measurements is very high, potentially pinpointing a critical epoch in the history of the universe, the technical challenges are massive. The expected curl-free E-mode polarization signal is only about 1 % of the corresponding CMB temperature fluctuations, and the signal is only clearly visible on large angular scales. Among all parameters in the cosmological concordance model, the reionization optical depth is the most susceptible to systematic errors, and for this reason it is often adopted as a monitor for residual errors.

To illustrate the difficulties associated with measuring τ, it is interesting to consider its value as reported in the literature as a function of time. The first CMB constraint was reported in the first-year *WMAP* release, which claimed $\tau = 0.17 \pm 0.04$ corresponding to a reionization epoch of $t_{\rm r} = 180^{+220}_{-80}$ Myr (Kogut et al. 2003). Such an early reionization epoch imposed strong limits on galaxy formation processes, and was not immediately compatible with standard theories. However, this preliminary measurement was based on the cross-correlation between temperature and polarization fluctuations for which uncertainties and degeneracies are large. Furthermore, it also did not account for bias introduced by foreground emission.

After adding more data, and, critically, allowing more time for understanding the data and controlling systematic errors, the 3-year *WMAP* data release resulted in a significantly revised estimate of $\tau = 0.089 \pm 0.03$, nearly doubling the time allowed for structure formation (Page et al. 2007). This estimate was derived directly from polarization-only measurements, and included proper foreground corrections. Based on further improvements and additional data, the reported 5-year *WMAP* posterior mean value was $\tau = 0.085 \pm 0.016$ (Komatsu et al. 2009), while

in the 7-year release it was $\tau = 0.088 \pm 0.015$ (Larson et al. 2011), before finally settling on $\tau = 0.089 \pm 0.014$ in the 9-year release (Hinshaw et al. 2013). This represented the state-of-the-art before *Planck* in terms of large-scale CMB polarization measurements.

As already mentioned, no CMB polarization measurements were included in the first *Planck* 2013 release (Planck Collaboration I 2014). However, from temperature measurements alone, the best-fit optical depth was constrained to $\tau = 0.097 \pm 0.038$, in seemingly excellent agreement with the final *WMAP* polarization results (Planck Collaboration XVI 2014). Then, in the *Planck* 2015 release, the LFI data allowed the first independent constraint on large-scale CMB polarization since *WMAP* (Planck Collaboration XI 2016). At the same time, the HFI polarization observations provided new and powerful constraints on Galactic polarized thermal dust (Planck Collaboration X 2016), to which the *WMAP* experiment was only marginally sensitive. The combination of LFI CMB and HFI thermal dust polarization measurements alone resulted in $\tau = 0.064^{+0.022}_{-0.023}$, or 1.1σ lower than the 9-year *WMAP* value. Furthermore, when combining the *WMAP* large-scale polarization CMB data with the same HFI polarization foreground data, the best-fit value was $\tau = 0.067 \pm 0.013$, in full agreement with LFI.

The HFI large-scale CMB polarization data were not considered sufficiently mature for scientific analysis until 2016, when new calibration, mapmaking, and simulation procedures had been implemented in a code called SROLL (Planck Collaboration Int. XLVIII 2016). Taking advantage of these new developments, and leveraging the higher statistical power of the HFI data, the reported estimate of the reionization optical depth was adjusted further down by HFI to $\tau = 0.055 \pm 0.009$. In parallel, the LFI procedure was improved by merging calibration and component separation into one framework. Combined, these new analysis procedures formed the basis for the third official *Planck* release (Planck Collaboration I 2020), for which a final value of $\tau = 0.053 \pm 0.009$ was reported. The good agreement with the lower limit imposed by quasar measurements, $\tau > 0.048$, implies both that reionization by the first generation of stars occurred relatively late, and that we can pin down the epoch of reionization with precision.

While a stable and internally consistent ΛCDM model, including τ, had emerged by the official end of the *Planck* consortium in 2018, one could still see clear signatures of residual systematics present in various subsets of the data. For HFI, several internal cross-correlations did not agree with each other to statistical precision (Planck Collaboration III 2020). For LFI the 44 GHz channel failed internal null tests (Planck Collaboration II 2020), and there were clear discrepancies between the raw frequency maps as seen by LFI and *WMAP* (Planck Collaboration IV 2018), indicating that there were still issues to be resolved within either LFI or *WMAP*, or both.

The last effort of the *Planck* collaboration to resolve these questions was organized within *Planck* Release 4 (PR4; also sometimes referred to as "Data Release 4" — DR4 — or simply NPIPE[1]; Planck Collaboration Int. LVII 2020). One unique feature of this pipeline was its ability to analyze both LFI and HFI jointly within the same framework. Combining some of the most powerful features from each of the instrument analysis pipelines, this approach led to further reduction of systematic errors in both data sets. The resulting best-fit estimate of the reionization op-

tical depth from DR4 reads $\tau = 0.058 \pm 0.006$ (Tristram et al. 2021).

An independent initiative to improve the *Planck* processing was SROLL2 (Delouis et al. 2019), which was a direct continuation of the HFI SROLL effort (Planck Collaboration III 2020). A defining feature of this approach is improved ADC corrections, which in particular leads to more robust large-scale polarization estimates. From the SROLL2 polarization analysis alone, the current best-fit estimate of the reionization optical depth is $\tau = 0.0566^{+0.0053}_{-0.0062}$ (Pagano et al. 2020).

A second independent initiative is called BeyondPlanck, and this is the primary focus of the current paper and suite of companion papers. The scope of this project is significantly different than the previous efforts, as BeyondPlanck aims at building a complete integrated end-to-end analysis pipeline for current and future CMB experiments. The current work focuses in particular on the *Planck* LFI data set, although significant effort is spent ensuring that the tools are generalizable to other experiments. Indeed, one example of this is already presented within the current project in the form of a preliminary application to *WMAP* (Bennett et al. 2013; Watts et al. 2022).

Because instrumental systematics and residual foreground contamination have such a dramatic impact on the large-scale CMB polarization estimates, in this paper we will use the reionization optical depth as a direct demonstration of the BeyondPlanck framework, and our ultimate scientific goal is to estimate the posterior distribution $P(\tau \mid \boldsymbol{d})$ from *Planck* LFI and *WMAP* observations, $\boldsymbol{d}$. The posterior summarizes our knowledge about τ in the form of a probability distribution, and we will estimate $P(\tau \mid \boldsymbol{d})$ within a strict Bayesian framework, with as few approximations and little data selection as possible. We will avoid the use of cross-spectrum techniques, which frequently are used to reduce the sensitivity of the final products to instrumental systematics (e.g., Planck Collaboration V 2020). In this project, we aim to do the opposite, and *highlight* the impact of residual systematics, such that, if needed, they can be addressed at a lower level of the analysis. As such, internal consistency, goodness-of-fit and χ^2 tests will play critical roles.

1.4. Lessons learned from Planck

To understand the background, historical context, and motivation for the BeyondPlanck program, it is useful to revisit the "Lessons learned from *Planck*,"[2] as compiled by the *Planck* consortium in 2016. In Section 9.6 ("Understanding the data") one can read the following:

> *In a project like Planck, "understanding the data" is certainly the most significant driver of the quality of the final products and science it can produce. This activity must be at the core of the data processing. It covers a lot of ground – photometry, optical response, time response, calibration, systematic effects, etc. – all interlinked issues that can be diagnosed at many different levels in the data processing pipelines, from raw data streams to finished maps and scientific products.*
>
> *(. . .) In the early phases of Planck, much of the strategy was based on separating the various elements of the problem into independent parts. This was adequate for a first treatment of the data. However, as the quality of the data improved, it became harder to find and analyse subtler non-ideal effects, and to do so required a more in-*

---

[1] This name is short for "NERSC pipeline", a name deriving from the computer facilities at which it is executed, namely the National Energy Research Scientific Computing Center (NERSC).

tegrated approach, where a variety of effects were treated simultaneously.

(. . .) An example is the influence of foregrounds on calibration: initially model foreground templates were used to isolate the CMB dipole signal (the calibrator), but in later stages the template had to be iterated within the calibration pipeline to include and self-consistently reduce the effects of polarization, sidelobes, dipoles, etc.

(. . .) As understanding of the data progresses, analysis – and the teams doing it – need to become more and more integrated, pulling in parts of the pipeline which initially could be separated out.

As described in these paragraphs, the analysis approach adopted by *Planck* became gradually more and more integrated as the effective sensitivity of the data set improved through more refined analysis, and new systematic effects were uncovered. Indeed, only toward the end of the *Planck* mission period did it become evident that the single most limiting factor for the overall analysis was neither instrumental systematics nor astrophysical foregrounds as such, but rather the *interplay* between the two. Intuitively speaking, the problem may be summarized as follows: *One cannot robustly characterize the astrophysical sky without knowing the properties of the instrument, and one cannot characterize the instrument without knowing the properties of the astrophysical sky.* The calibration and component separation procedures are intimately tied together. By the time this issue was fully understood, there were neither sufficient resources nor time to redesign a complete *Planck* analysis pipeline from bottom-up. An important organizational goal of the BeyondPlanck program has therefore been to provide a financial structure that allows the team to consolidate this experience into practical computer code, and make this publicly available to the general community.

1.5. The next frontier: Primordial gravitational waves

While a statistically coherent analysis of existing data is undoubtedly both interesting and useful in its own right, it is important to emphasize that none of the developments detailed in this work are likely to impact the overall cosmological concordance model to any significant degree. Indeed, looking at the big picture, the cosmological model has been remarkably stable even before *WMAP* and *Planck* provided their high-precision measurements; see, e.g., Wang et al. (2003) for a summary of pre-*WMAP* measurements and constraints. The main achievement of *WMAP* and *Planck* has been to refine this model to the level at which cosmology now is a high-precision science within which competing theoretical models can be tested and rejected at high statistical significance.

Planck has for all practical purposes completed the study of primary CMB temperature fluctuations. Currently, however, another frontier is driving the CMB field, namely the search for primordial gravitational waves created during inflation. These are predicted to exist by most inflationary theories, although their predicted amplitudes can vary by many orders of magnitude, depending on the precise details of the assumed inflationary model (e.g., Kamionkowski & Kovetz 2016). Typically, this amplitude is quantified in terms of the tensor-to-scalar ratio, r, which measures the ratio in fluctuation power attributable to gravitational waves and ordinary density perturbations, respectively.

If such gravitational waves do exist, one generically expects a specific imprint in the CMB polarization field in the form of a large-scale "divergence-free" or B-mode polarization signal. The observational challenges associated with gravitational wave

detection are essentially the same as those for measuring τ. However, the state-of-the-art upper limits on the tensor-to-scalar ratio are $r < 0.036$ (Ade et al. 2021) and $r < 0.032$ (Tristram et al. 2022), both at 95 % confidence, which immediately implies that the B-mode signal must be more than one order of magnitude smaller than the E-mode signal, and thus no more than a few tens of nK in amplitude.

With such a small target amplitude, it is safe to assume that an integrated analysis approach will no longer be optional for future CMB missions, but rather a strict prerequisite. Establishing both the experience and appropriate code required to implement such an approach for future CMB missions is a main long-term scientific motivation for the BeyondPlanck program; current experiments such as *Planck* and *WMAP* provide real-world testbeds that ensure that the BeyondPlanck approach is both realistic and practical.

1.6. The BeyondPlanck program

We are now ready to formulate the main goal of the BeyondPlanck program:

BeyondPlanck *aims to implement and apply a single statistically coherent analysis pipeline to Planck and other CMB data sets, processing raw uncalibrated time-ordered data into final astrophysical component maps, angular power spectra, and cosmological parameters within one single code.*

Important secondary goals include

1. to model and propagate instrumental uncertainties from raw time-ordered data into final high-level *Planck* LFI scientific results;
2. to provide a computationally convenient interface to the raw *Planck* LFI data that can be accessed and extended by external users;
3. to develop a framework that allows joint analysis of *Planck* LFI with other data sets; and
4. to prepare for next-generation CMB experiments, in particular those aiming to detect primordial gravitational waves through their imprint on large-scale polarization of the CMB.

The "BeyondPlanck" name serves as a reminder that this work builds directly on several decades of *Planck* efforts and experience, while at the same time highlights the fact that it aims to apply the *Planck* methodology to data sets beyond *Planck*, both archival and future.

Clearly, this is a very ambitious program that will require long-term and dedicated support. The first stage of the program, which is reported in the current suite of papers, has been funded within an EU-based Horizon 2020 action called "Leadership in Enabling and Industrial Technologies" (LEIT), as well as through various individual grants. This funding only covers end-to-end analysis of the *Planck* LFI data, which is smaller in volume than HFI data, and therefore serves as a convenient real-world test case for development purposes, while still representing a very important scientific data set in its own right.

As detailed in the H2020 LEIT contract, the BeyondPlanck program started on March 1st 2018, and ended formally on November 30th 2020; the total duration of the funded program was thus strictly limited to two years and nine months. During this period, large amounts of software, products and documentation had to be written from scratch. Indeed, a first fully operational pipeline was completed as late as June 2020, and the first

data release took place during an online conference on November 20-22 2020; the conference was remote due to COVID-19 travel restrictions. However, as discussed during that conference, the first BeyondPlanck data products were significantly affected by important unresolved systematic effects, including a notable power excess in the Southern Galactic hemisphere. Understanding those effects has been the main focus of the collaboration since that time, and, as will be detailed in the following and in a suite of companion papers, we now finally believe that all previously reported issues have been successfully understood and resolved.

1.7. From BeyondPlanck to Cosmoglobe

The amount of funding and human effort spent on mapping the sky at radio, microwave, and submillimeter frequencies during the last few decades has truly been massive (Gerakakis et al. 2022), and an experiment like *Planck* alone represents an investment of about €700 M for the spacecraft, payload, launch, and operations. The cost of *WMAP* was about $150 M, while a typical ground-based or sub-orbital CMB experiment costs about 10 M euros or dollars; LiteBIRD's cost cap is $300 M, while the future CMB-S4 project is likely to cost about $500 M.

Optimally exploiting these expensive data is clearly of the essence for the field as whole. In particular, it is important to realize that each experiment is unique in its own way, and teaches us something new about the astrophysical sky. Some experiments have very broad frequency coverage, while others have very high angular resolution; some provide absolutely calibrated data, while others have very low amounts of correlated noise. Some are designed for intensity observations, while others are optimized for polarization measurements. At the same time, each experiment invariably has its own "blind-spots" or degeneracies. For instance, both *Planck* and *WMAP* have their own so-called "poorly measured modes", which correspond to sky modes that are not well measured by their respective scanning strategies (Planck Collaboration I 2020; Bennett et al. 2013; Watts et al. 2022).

The ultimate solution to breaking such degeneracies is through joint analysis of complementary data sets; *WMAP* can be used to constrain the largest polarization modes and thereby improve *Planck*'s gain model (Planck Collaboration II 2020), while *Planck* can be used to constrain *WMAP*'s transmission imbalance parameters (Jarosik et al. 2007). Similarly, for future LiteBIRD or CMB-S4 analyses, *Planck* and *WMAP* must be used as models of the temperature sky, while the new experiments will provide an entirely new view of the polarization sky. And once those new polarization measurements become available, it will be essential to re-analyze the *Planck* and *WMAP* measurements, taking into account the new polarization information, to further reduce their systematic uncertainties.

In general, the most cost-efficient and productive way for the field as a whole to make progress in the future is through joint and global analysis of all available experiments. Cosmoglobe[3] represents an Open Science platform for this work, which main goal is to establish a global model of the radio, microwave, and submillimeter sky through joint analysis of all available state-of-the-art data sets. This project builds directly on the current BeyondPlanck efforts, and will use the computational machinery developed in the following as the starting point for its multi-experiment analysis. All interested parties and collaborations are highly encouraged to participate in the Cosmoglobe effort.

---

[3] https://cosmoglobe.uio.no

2. BeyondPlanck analysis strategy and organization

2.1. End-to-end Bayesian CMB analysis

Recognizing the lessons learned from *Planck* as summarized in Sect. 1.4, the defining design philosophy of BeyondPlanck is tight integration of all steps from raw time-ordered data processing to high-level cosmological parameter estimation. Traditionally, this process has been carried out in a series of weakly connected steps, pipelining independent executables with or without human intervention. Some steps have mostly relied on frequentist statistics, employing forward simulations to propagate uncertainties, while other steps have adopted a Bayesian approach, using the posterior distribution to quantify uncertainties. For instance, traditional mapmaking is a typical example of the former (e.g., Ashdown et al. 2007b), while cosmological parameter estimation is a typical example of the latter (e.g., Lewis & Bridle 2002); for component separation purposes, both approaches have been explored in the literature (e.g., Planck Collaboration Int. XLVI 2016).

BeyondPlanck is the first real-world CMB analysis pipeline to adopt an end-to-end Bayesian approach. This solution was in fact first proposed by Jewell et al. (2004). However, it took more than 15 years of computational and algorithmic developments to actually make it feasible.

Perhaps the single most important advantage of a uniform Bayesian approach is that it allows seamless propagation of uncertainties within a well-established statistical framework. This aspect will become critically important for future experiments, as demonstrated by *Planck*. For most CMB experiments prior to *Planck*, the dominant source of uncertainty was noise; for most CMB experiments after *Planck*, the dominant source of uncertainty will be instrumental systematics, foreground contamination, and the interplay between the two. As a logical consequence of this fact, BeyondPlanck adopts a consistent statistical framework that integrates detailed error propagation as a foundational feature.

The Bayesian approach also has several notable advantages in terms of intuition and transparency. In particular, the most critical step for any Bayesian analysis is the definition of the data model. This may often be described in terms of a handful of equations, and these equations subsequently serve as a road-map for the entire analysis. While the complexity of the numerical implementation may vary from model to model, the posterior distribution itself has a very intuitive and direct interpretation.

At a practical level, integrating the entire pipeline into a single computational code also has significant advantages in terms of net computational speed and resources. Not only are slow disk operations reduced to a minimum by performing all operations within one single code, but more importantly, all intermediate human interactions are eliminated from the process. This both saves significant amounts of human time required for "code shepherding" and file transfers, and it significantly reduces the risk of human errors. Thus human resources are saved that can be better spent on fundamental modelling aspects.

A fourth significant advantage of end-to-end integration is increased transparency of implicit and explicit priors. For a distributed analysis process, it is critically important to communicate all assumptions made in each step to avoid errors, while in an integrated approach internal inconsistencies become much more visible; there are simply fewer opportunities for misunderstandings to propagate undetected throughout an integrated analysis pipeline.

Table 1. Overview of BEYONDPLANCK and preliminary COSMOGLOBE papers.

REFERENCE	TITLE
Pipeline	
BeyondPlanck (2022)	I. Global Bayesian analysis of the *Planck* Low Frequency Instrument data
Keihänen et al. (2022)	II. CMB mapmaking through Gibbs sampling
Galloway et al. (2022a)	III. Commander3
Brilenkov et al. (2022)	IV. On end-to-end simulations in CMB analysis — Bayesian versus frequentist statistics
Instrument characterization	
Herman et al. (2022a)	V. Minimal ADC Corrections for Planck LFI
Ihle et al. (2022)	VI. Noise characterization and modelling
Gjerløw et al. (2022)	VII. Bayesian estimation of gain and absolute calibration for CMB experiments
Galloway et al. (2022b)	VIII. Efficient Sidelobe Convolution and Correction through Spin Harmonics
Svalheim et al. (2022a)	IX. Bandpass and beam leakage corrections
Cosmological and astrophysical results	
Basyrov et al. (2022)	X. *Planck* LFI frequency maps with sample-based error propagation
Colombo et al. (2022)	XI. Bayesian CMB analysis with sample-based end-to-end error propagation
Paradiso et al. (2022)	XII. Cosmological parameter estimation with end-to-end error propagation
Andersen et al. (2022)	XIII. Intensity foregrounds, degeneracies and priors
Svalheim et al. (2022b)	XIV. Polarized foreground emission between 30 and 70 GHz
Herman et al. (2022b)	XV. Limits on Large-Scale Polarized Anomalous Microwave Emission from *Planck* LFI and *WMAP*
COSMOGLOBE	
Gerakakis et al. (2022)	From BEYONDPLANCK to COSMOGLOBE: Open Science, reproducibility, and data longevity
Watts et al. (2022)	From BEYONDPLANCK to COSMOGLOBE: Preliminary WMAP Q-band analysis

2.2. Commander

We adopt Commander2 (Eriksen et al. 2004, 2008; Seljebotn et al. 2019), a well-established Bayesian CMB Gibbs sampler developed for *Planck*, as the starting point of our pipeline. As demonstrated in Planck Collaboration IV (2018), this code already supports Bayesian multi-resolution component separation, which is precisely the operation that connects low-level mapmaking to high-level cosmological parameter estimation. A main implementational goal for BEYONDPLANCK is thus to extend this framework to incorporate Bayesian calibration and mapmaking, as well as to connect component separation and cosmological parameter estimation.

We will refer to three different versions of the Commander code in the following. Commander1 refers to the original implementation described by Eriksen et al. (2004, 2008), which at the beginning of the BEYONDPLANCK project represented the most mature version in terms of foreground spectral parameter fitting. However, a major limitation of that code is a requirement of common angular resolution among all data sets. Commander2 removes this limitation through explicit beam convolution for each frequency map during component separation, as detailed by Seljebotn et al. (2019), and thereby allows for full resolution analysis of the *Planck* data. Due to the much higher computational cost associated with increased angular resolution, the development of Commander2 required a re-implementation of the original algebra from scratch, adopting a much more fine-grained parallelization strategy than Commander1.

Commander3 (Galloway et al. 2022a) refers to the time-domain version of the algorithm, as developed in BEYONDPLANCK, and is a direct generalization and extension of Commander2 in terms of code implementation. As a result, Commander2 is no longer an independent code, but we will still refer to it in cases where it might be convenient for pedagogical purposes to distinguish between multi-resolution component separation in the pixel-domain versus the time-domain. All Commander source codes are available under a GNU Public Library (GPL) Open-Source license.[4]

2.3. Paper organization

The BEYONDPLANCK methodology and results are described in a suite of companion papers, as listed in Table 1. The present paper provides a broad overview in terms of motivation, algorithms, and main results. However, it is not intended to be comprehensive, and specific details are deferred to the relevant companion papers.

The remaining papers may be divided into four main categories, namely 1) pipeline papers; 2) instrument characterization papers; 3) cosmological and astrophysical results papers; and 4) joint BEYONDPLANCK–Cosmoglobe papers. The first category of papers provides a comprehensive overview of the current implementation of the BEYONDPLANCK pipeline, at a level that is hopefully sufficiently detailed to allow external users to understand intuitively its statistical and computational basis, what assumptions it relies on, and what its limitations are. The ultimate goal of these papers is that external users should be able to repeat and extend the work that is presented here.

The second category of papers address the various relevant instrumental parameters required to process the raw time-ordered data into sky maps. These include noise characterization, gain estimation, sidelobe corrections, and bandpass and beam mismatch modelling. Each paper aims both to provide an intuitive understanding of the effect in question, and to show how it impacts the final results. These papers also demonstrate how to quantitatively model each instrumental effect, and how to propagate uncertainties into other parameters. Particular emphasis is placed on building intuition regarding leading internal parameter degeneracies, both among the various instrumental parameters and with astrophysical and cosmological parameters.

---

[4] https://github.com/Cosmoglobe/Commander

The third category of papers present the main scientific results in terms of frequency and component maps, as well as angular power spectra and cosmological parameters. Consistency between the BeyondPlanck products and non-*Planck* sets is also considered in this category of papers.

The fourth category includes papers that aim to generalize the BeyondPlanck data model to other data sets within the Cosmoglobe framework. One worked example is provided in the form of a preliminary *WMAP Q*-band reanalysis (Watts et al. 2022).

We note that, in the spirit of reproducibility and accessibility (Gerakakis et al. 2022), a significant emphasis is put on intuition and background throughout the BeyondPlanck papers. The paper suite is intended to be largely self-contained, and detailed knowledge of the *Planck* publication list is not an assumed prerequisite. As such, a substantial amount of review material is included, both in terms of general background material and algorithmic details. The style of the papers is consciously tuned toward Ph.D. students and early postdoctoral fellows, rather than seasoned CMB experts.

3. Parameterizing the microwave sky

As already noted, the single most important component in any Bayesian analysis is the parametric model that is fitted to the data. In our case, this model consists of both astrophysical and instrumental components. In this section we consider the cosmological and astrophysical parameters, before introducing the instrumental parameters in the next section.

3.1. Conventions: Stokes parameters, pixelization, spherical harmonics, and units

In order to characterize each astrophysical component quantitatively, we need to introduce some general notation and conventions. First, each astrophysical component will be described in terms of three Stokes parameters, namely intensity (denoted either I or T) and two linear polarizations (denoted Q and U). We will ignore circular polarization (V) for now, but we note that this may be added in future work.

To discretize the Stokes parameters on the sphere, we adopt the `HEALPix` pixelization[5] (Górski et al. 2005). This pixelization has several highly desirable properties, including equal-area pixels and support for fast spherical harmonics transforms, and is now effectively a standard in modern CMB analysis. The `HEALPix` pixel resolution is controlled through a parameter called N_{side}, and the total number of pixels on the sky is $N_{\text{pix}} = 12N_{\text{side}}^2$. We organize the Stokes parameters into vectors of length $3N_{\text{pix}}$, simply by stacking $\{T, Q, U\}$ into a map vector $s(\hat{n})$, where $\hat{n}$ is a unit direction vector.

Unless otherwise noted, we define the Stokes parameters with respect to Galactic coordinates. We adopt the cosmological convention for the polarization angle, γ, in which $\gamma = 0$ for vectors pointing north and increases westward. This is opposite to the IAU convention used in most other fields of astronomy, in which γ increases eastward. To convert from one convention to the other, one must multiply Stokes U by -1.

The Stokes polarization parameters Q and U form a spin-2 field, which intuitively may be interpreted as a "headless vector field". In contrast, the intensity T is a spin-0 field, and does not change under rotations. Thus, when rotating Stokes parameters

---

by an angle α, the transformed Stokes parameters are

$$\begin{bmatrix} T' \\ Q' \\ U' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos 2\alpha & -\sin 2\alpha \\ 0 & \sin 2\alpha & \cos 2\alpha \end{bmatrix} \begin{bmatrix} T \\ Q \\ U \end{bmatrix}. \tag{1}$$

As described by Zaldarriaga & Seljak (1997), the polarization Stokes parameters may be expanded into spherical harmonics through the following relations,

$$T(\hat{n}) = \sum_{\ell=0}^{\ell_{\max}} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(\hat{n}) \tag{2}$$

$$(Q \pm iU)(\hat{n}) = \sum_{\ell=2}^{\ell_{\max}} \sum_{m=-\ell}^{\ell} {}_{\pm 2}a_{\ell m} {}_{\pm 2}Y_{\ell m}(\hat{n}), \tag{3}$$

where ${}_k a_{\ell m}$ are called (spin-k) spherical harmonic coefficients. The polarization coefficients are often combined algebraically into E and B coefficients,

$$a_{\ell m}^E = -\frac{1}{2} \left({}_2 a_{\ell m} + {}_{-2} a_{\ell m} \right) \tag{4}$$

$$a_{\ell m}^B = \frac{i}{2} \left({}_2 a_{\ell m} - {}_{-2} a_{\ell m} \right), \tag{5}$$

which each form a spin-0 field, fully analogous to the intensity T.

From the spherical harmonic coefficients we may compute the observed angular power spectrum as

$$\sigma_\ell^{XY} = \frac{1}{2\ell + 1} \sum_{m=-\ell}^{\ell} \left(a_\ell^X \right)^* a_{\ell m}^Y, \tag{6}$$

where $\{X, Y\} \in \{T, E, B\}$. These quantify the strength of fluctuations at a given multipole ℓ as directly measured from some sky map. In addition, we define the ensemble-averaged power spectrum as

$$C_\ell^{XY} \equiv \left\langle \left(a_\ell^X \right)^* a_{\ell m}^Y \right\rangle = \left\langle \sigma_\ell^{XY} \right\rangle, \tag{7}$$

where brackets indicate an average over statistical realizations. This function is thus independent of the observed sky, and only depends on the model that describes the field in question.

Finally, each sky map s must be quantified in terms of a physical unit. In the following work, we will encounter many different conventions for this, depending on the particular application in question. However, three conventions are more common than others, and we limit our discussion here to these special cases.

The first measure is *surface brightness per solid angle*, which simply measures the amount of energy emitted by some source per surface area, per frequency interval, per sky solid angle. This is often measured in units of $\text{MJy}\,\text{sr}^{-1} \equiv 10^{-20}\,\text{W}\,\text{m}^{-2}\,\text{Hz}^{-1}\,\text{sr}^{-1}$, and it quantifies the specific intensity I_ν of a given source as a function of wavelength, ν.

The second measure we will use is *thermodynamic temperature*. In this case, we identify the intensity with that emitted by a blackbody source with temperature T,

$$I_\nu = B_\nu(T) = \frac{2h\nu^3}{c^2} \frac{1}{e^{\frac{h\nu}{kT}} - 1}, \tag{8}$$

where h is Planck's constant, c is the speed of light, and k is the Boltzmann constant. This measure is particularly useful for CMB applications, because the CMB is itself a near-perfect blackbody, and a single temperature $T(\hat{n})$ therefore uniquely

specifies its intensity at any wavelength at a given position. The unit for thermodynamic temperature is denoted K_{CMB} or simply K.

Our third and final measure is the *brightness temperature* or *Rayleigh-Jeans temperature*, T_{RJ}. This is defined by the long wavelength limit ($h\nu \ll kT$) of Eq. (8), such that

$$I_\nu = \frac{2\nu^2 k T_{RJ}}{c^2}. \tag{9}$$

While the thermodynamic temperature is convenient to describe the CMB, most astrophysical foreground signals have a non-blackbody nature, and are more naturally quantified in terms of brightness temperature. In particular, while the spectral energy density of many foregrounds can span many tens of orders of magnitude when expressed in K_{CMB}, they are usually limited to a few orders of magnitude when expressed in either MJy sr^{-1} or K_{RJ}. To avoid numerical problems, all astrophysical components are therefore expressed in units of K_{RJ} internally in `Commander`, and only converted to the respective natural unit before outputting results to disk. Monochromatic conversion between K_{RJ} and MJy sr^{-1} is performed through Eq. (9), while monochromatic conversion between K_{RJ} and K_{CMB} is given by

$$\Delta T_{CMB} = \frac{(e^x - 1)^2}{x^2 e^x} T_{RJ}, \tag{10}$$

where $x = h\nu/kT_0$, and $T_0 = 2.7255$ K is the mean CMB temperature (Fixsen 2009). Note that this conversion applies only to small temperature variations around the CMB mean value, $\Delta T \equiv T - T_0$, which is precisely the form of most CMB temperature maps in common use today.

We are now ready to write down parametric models for each of the main astrophysical components that are relevant for the *Planck* frequency range. Each component will be described in terms of a spectral energy density (SED) in brightness temperature units, and, in some cases, in terms of an angular power spectrum or some other similar spatial coherence measure.

3.2. Cosmic microwave background anisotropies

We start our survey with the CMB component, which is the scientifically most important one for *Planck*. For this, we first define s^{CMB} to be a $3N_{pix}$ sky vector of CMB Stokes parameters as described above. Second, we assume that the CMB SED may be approximated as a blackbody. As such, its brightness temperature SED is given by Eq. (10),

$$s_{RJ}^{CMB}(\nu) \propto \frac{x^2 e^x}{(e^x - 1)^2} s^{CMB}, \tag{11}$$

where $x = h\nu/kT_0$. (Note that we define the effective SED only up to a normalization constant, as we will typically parameterize each component in terms of an amplitude map at a given reference frequency times the SED normalized to unity at the reference; any normalization factor is therefore accounted for in the amplitude coefficient.)

In addition to the pure cosmological blackbody SED, the CMB component exhibits a relativistic correction due to our motion with respect to the CMB monopole. This effect is frequency dependent, and is primarily dominated by quadrupole correction (Notari & Quartin 2015). Unlike the official *Planck* sky maps (Planck Collaboration II 2020; Planck Collaboration LVII 2020), we do *not* subtract this component from the frequency maps (Basyrov et al. 2022), but rather include it as part of the signal model (Andersen et al. 2022).

For component separation purposes, these are the *only* assumption we make regarding the CMB. However, for cosmological parameter estimation purposes, we make two important additional assumptions, namely that the CMB temperature fluctuations are both Gaussian distributed and statistically isotropic. The assumption of Gaussianity determines the conditional probability distribution for the CMB signal,

$$P(\mathbf{s} \mid C_\ell) \propto \frac{e^{-\frac{1}{2}\mathbf{s}^t \mathsf{S}^{-1}\mathbf{s}}}{\sqrt{|\mathsf{S}|}}, \tag{12}$$

where S is the covariance matrix of the CMB fluctuation field, and we have dropped the "CMB" superscript for convenience. The assumption of statistical isotropy implies that S is fully specified in terms of the angular power spectrum,

$$\mathsf{S}_{\ell m, \ell' m'}^{XY} \equiv \left\langle \left(a_\ell^X\right)^* a_{\ell' m'}^Y \right\rangle = C_{\ell m}^{XY} \delta_{\ell \ell'} \delta_{mm'}. \tag{13}$$

For practical parameter estimation purposes, both of these assumptions have been shown to be excellent approximations to the true CMB sky (see, e.g., Planck Collaboration VII 2020; Planck Collaboration IX 2020, and references therein).

The connection to cosmological parameters, such as the Hubble constant H_0 or the reionization optical depth τ, is made through cosmological Boltzmann codes, such as `CMBfast` (Seljak & Zaldarriaga 1996) or `CAMB` (Lewis et al. 2000). These deterministically calculate the ensemble-averaged CMB power spectrum based on well-understood physics given some specific set of cosmological parameters, ξ. However, this calculation is only straightforward going from ξ to C_ℓ; it is highly nontrivial to go directly from C_ℓ to ξ. Instead, Markov Chain Monte Carlo (MCMC) methods such as `CosmoMC` (Lewis & Bridle 2002) are typically employed to perform the inversion, in which a series of parameter combinations are proposed and rejected or accepted, ultimately resulting in a set of parameter samples that jointly represents the final parameter posterior distribution. As described in Sect. 1.6, the goal of the BEYONDPLANCK program is to implement a similar MCMC method that accounts for the entire process from raw time-ordered data to final cosmological parameters with full Bayesian end-to-end error propagation.

3.3. Galactic foreground emission

The second most important class of sky emission components consists of diffuse Galactic foregrounds. These all originate from within the Milky Way, and are due to particles (electrons, ions, dust, etc.) associated with various processes such as star formation or supernova explosions. Furthermore, these particles all interact with the same magnetic field, and as a result they produce correlated polarized emission. In this section, we provide a brief survey of each of the main physical emission mechanisms, with a particular focus on parametric models.

3.3.1. Synchrotron emission

At low microwave frequencies, synchrotron emission dominates the radio sky. This emission is mostly due to relativistic electrons ejected from supernova, spiralling in the magnetic field of the Milky Way. CMB observations are typically made at frequencies in the range of tens or hundreds of GHz, and at these frequencies, the synchrotron SED falls rapidly with increasing frequency. Indeed, detailed models and observations both suggest that the effective spectrum may be closely approximated by a power-law at frequencies higher than a few gigahertz, with some evidence

for possible curvature. In this work, we therefore follow Kogut (2012), and adopt a general SED model of the form

$$s_{\rm RJ}^{\rm s}(\nu) \propto \left(\frac{\nu}{\nu_{0,\rm s}}\right)^{\beta + C \ln \nu/\nu_{0,\rm s}}, \tag{14}$$

where $\nu_{0,\rm s}$ is a reference frequency, β is a power-law index, and C is a curvature parameter. However, in most cases we set $C = 0$, as the signal-to-noise ratio for this parameter is very low with the limited data set considered in this work (Svalheim et al. 2022b).

When the local magnetic field is highly structured, synchrotron emission can be highly polarized, with a theoretical maximum polarization fraction of $p = 75\,\%$. In practice, this value is decreased due to line-of-sight and volume integration effects, and according to *Planck* and *WMAP*, more typical values are $\lesssim 15\,\%$ at high Galactic latitudes, with extreme cases reaching 30–50 % only in a few large-scale supernova remnants that, when projected on the sky, take the form of so-called "Galactic spurs" (Planck Collaboration XXV 2016).

At low frequencies, polarized synchrotron emission is also significantly affected by Faraday rotation (e.g., Beck et al. 2013, and references therein). This effect is caused by circular bire-fringence, i.e., left- and right-handed circular polarized emission travel at different speeds through a magnetic field embedded in an ionized medium, resulting in a net rotation of the polarization angle of linearly polarized emission. The polarization angle rotation is proportional to the magnetic field strength as well as to the square of the wavelength of the emission. Numerically, the rotation angle is typically a few degrees at 23 GHz at low Galactic latitudes (Carretti et al. 2019; Fuskeland et al. 2021), and we neglect it for the higher-frequency *Planck* and *WMAP* surveys considered here.

3.3.2. Free-free emission

Free-free emission (or *bremsstrahlung*) arises primarily from free electrons scattering off protons without being captured, and emitting a photon in the process. Since free electrons only exist in appreciable amounts when the temperature of the medium is comparable to the hydrogen binding energy, corresponding to $10^3 - 10^4$ K, free-free emission predominantly traces hot H II regions and, as such, active star forming regions. Free-free emission is particularly important for CMB experiments because it is the only foreground component that is non-negligible at all frequencies between 1 and 1000 GHz, and it is therefore particularly sensitive to degeneracies with respect to both the CMB and other foreground components.

The free-free SED depends primarily on the number of free protons and electrons along the line of sight, which typically is quantified in terms of the *emission measure* (EM), i.e., the integral of the square electron density along the line of sight,

$$\mathrm{EM} \equiv \int_0^\infty n_e^2 \, \mathrm{d}l, \tag{15}$$

where the number densities of free protons and electrons are assumed to be equal. The conventional unit adopted for the EM is pc cm^{-6}, and typical values for the Milky Way range between 0 and 1000 (Planck Collaboration X 2016).

Assuming local thermodynamic equilibrium and first considering an optically thick medium, the free-free SED is determined by a blackbody spectrum given its electron temperature, $T_{\rm e}$, alone. Since the optical depth drops rapidly with increasing frequency, however, free-free emission in astrophysical contexts

and at CMB frequencies is optically thin. Hence, the effective SED can be expressed as

$$s_{\rm RJ}^{\rm ff}(\nu) = T_{\rm e} \, (1 - e^{-\tau}). \tag{16}$$

As shown by Dickinson et al. (2003) and Draine (2011), τ may be very well approximated by

$$\tau = 0.05468 \cdot T_{\rm e}^{-3/2} \cdot \nu_9^{-2} \cdot \mathrm{EM} \cdot g_{\rm ff}, \tag{17}$$

where

$$g_{\rm ff} = \log \left\{ \exp \left[5.960 - \sqrt{3}/\pi \log(\nu_9 \cdot T_4^{-3/2}) \right] + e \right\} \tag{18}$$

is called the Gaunt factor, and ν_9 and T_4 are the frequency and the electron temperature measured in units of GHz and 10^4 K, respectively.

This SED is a nonlinear function of EM and T_e. A complete free-free model therefore corresponds to a complicated probability distribution with expensive special-purpose sampling algorithms, as for instance employed in Planck Collaboration IX (2016). In this work, we instead adopt a simpler linearized version of Eq. (16) that is only strictly valid in the optically thin case, $\tau \ll 1$, namely

$$s_{\rm RJ}^{\rm ff}(\nu) \propto \frac{g_{\rm ff}(T_e)}{\nu^2}, \tag{19}$$

and we correspondingly quantify the free-free amplitude in terms of the observed signal at a given reference frequency in μK$_{\rm RJ}$, as opposed to the full nonlinear EM parameter described above.

There is essentially no effective alignment mechanism for thermal electrons in a hot medium, and large-scale free-free emission is therefore expected to be nearly unpolarized. The main exception to this are sharp edges around hot H II regions, which do introduce a preferred direction in the emission geometry. However, even these are only expected to be mildly polarized, and over large angular scales, the net polarization fraction is expected to be well below 1 % (see discussion in Keating et al. 1998). In the following, we thus assume that free-free emission is completely unpolarized.

3.3.3. Thermal dust emission

The interstellar medium (ISM) is filled not only with hydrogen and electrons, but also with tiny dust grains ranging in diameter from less than a nanometer (i.e., a few atoms across) to roughly a micron (i.e., thousands of atoms across). Dust grains typically condense from stellar outflows and ejecta, and so dust abundance is correlated with star formation. Newly-formed dust is rapidly mixed in the dynamic, turbulent ISM, where it undergoes significant processing. Dust is therefore ubiquitous in the Galaxy, found wherever there is interstellar gas.

It is known from spectroscopic features that dust is made from, at minimum, silicate and carbonaceous materials. However, the precise grain composition is likely to vary with local environment. Dust grains are heated by ambient stellar radiation, and large grains accounting for the bulk of the dust mass equilibrate to a steady-state temperature ranging between 10 and 30 K. This energy is thermally re-emitted with a peak wavelength in the sub-mm frequency range, typically between 1000 and 3000 GHz. Since these grains are inefficient radiators at longer wavelengths, the thermal dust SED falls rapidly at frequencies below the peak, where CMB observations are typically carried out. The varied composition and geometry of ISM dust

particles makes the thermal dust SED significantly more complicated to model from first principles, when compared to the free-free emission described above; for recent examples of such modelling efforts, see, e.g., Guillet et al. (2018) and Hensley & Draine (2022).

In practice, simpler fitting formulae are therefore usually adopted for practical analyses, and one particularly popular class of models is the so-called modified blackbody spectrum, which in intensity units reads

$$I_\nu^{\mathrm{d}} \propto \tau \nu^{\beta_{\mathrm{d}}} B_\nu(T_{\mathrm{d}}). \qquad (20)$$

This function is simply a blackbody spectrum with temperature T_{d}, modulated by a power-law having index β_{d}. In physical terms, this corresponds to dust having an opacity that scales as ν_{d}^β, a reasonable approximation for wavelengths longer than $\sim 20\,\mu$m (Hensley & Draine 2020).

The amplitude is, as for free-free emission, given by the optical depth, τ, which depends directly on the surface density of particles along the line of sight. Typical numerical values for these three parameters are $\tau \sim 10^{-6}, \beta_{\mathrm{d}} \sim 1.6$, and $T_{\mathrm{d}} \sim 20$ K. Intuitively speaking, β_{d} determines the slope (or first derivative in log-log space) of the SED below 200 GHz, while T_{d} determines the SED peak position, and second derivative at lower frequencies. However, we will model thermal dust emission in terms of brightness temperature, and in these units the effective SED may be written in the form

$$s_{\mathrm{RJ}}^{\mathrm{d}}(\nu) \propto \frac{\nu^{\beta_{\mathrm{d}}+1}}{e^{h\nu/kT_{\mathrm{d}}} - 1}. \qquad (21)$$

Interaction with gas and radiation torques up grains, and they tend to rotate about their axis of greatest moment of inertia, i.e., their short axis. Dust grains having unpaired electrons can develop a non-zero magnetic moment anti-parallel to their angular velocity through the Barnett effect (Dolginov & Mitrofanov 1976). Dissipative processes act to align the rotation axis with the local magnetic field. For a more detailed discussion of grain alignment, see Andersson et al. (2015).

The preferential alignment of the short axes of grains with the local magnetic field leads to significant net polarization from the ensemble of grains. Thermal dust polarization fractions as large as 20 % are found using the high frequency *Planck* polarization measurements (Planck Collaboration XI 2020). We therefore include all three Stokes parameters in our thermal dust model. At the same time, we note that the highest polarization-sensitive *Planck* frequency channel is 353 GHz, and this does not provide sufficient frequency range to allow an independent estimate of the thermal dust temperature in polarization. We therefore assume the same T_{d} for intensity and polarization, while β_{d} is allowed to be different.

3.3.4. Spinning dust (or anomalous microwave) emission

Dust grains rotate with rotational kinetic energy of order the thermal energy in the ambient gas. Consequently, sub-nanometer grains can achieve rotational frequencies of tens of GHz. If these grains possess an electric dipole moment, as generally expected for particles of this size (Macià Escatllar & Bromley 2020), this rotation produces emission in the microwave frequency range, as first predicted theoretically by Erickson (1957), and described quantitatively by Draine & Lazarian (1998). The spinning dust mechanism currently provides the most popular theoretical explanation for so-called "anomalous microwave emission" (AME) observed around 20 GHz in CMB surveys, as first identified and named by Leitch et al. (1997).

In this work, we will adopt a spinning dust model for this component, starting from an SED template, $s_0^{\mathrm{sd}}(\nu)$, computed with the SpDust2 code (Ali-Haïmoud et al. 2009; Ali-Haïmoud 2010; Silsbee et al. 2011) for environmental parameters typifying the Cold Neutral Medium. This spectrum is intrinsically computed in intensity units, in which it peaks at 30 GHz. After converting to brightness temperature by scaling with ν^{-2}, as given by Eq. (9), the peak shifts to 17.4 GHz, and the overall spectrum is less than 1 % of its peak value at frequencies below 1.3 GHz or above 66 GHz. To fit this SED model to the data, we follow Bennett et al. (2013), and introduce a peak position parameter, ν_{p}, that shifts the spectrum rigidly in log ν–log s space,

$$s_{\mathrm{RJ}}^{\mathrm{sd}}(\nu) \propto \nu^{-2}\, s_0^{\mathrm{sd}}\left(\nu \cdot \frac{30.0\,\mathrm{GHz}}{\nu_p}\right) \qquad (22)$$

We note, however, that this emission component is associated with large uncertainties, both in terms of the physical mechanism that is actually responsible for the observed emission, and in terms of detailed modelling within the chosen paradigm. In general, we assume this component to be unpolarized, and we adopt the spinning dust model in Eq. (22) for the AME component.

Despite sensitive searches in individual objects (Génova-Santos et al. 2015, 2017) and over large sky areas (Macellari et al. 2011), polarization has not been detected in the AME. In principle, AME could be highly polarized if small spinning grains are efficiently aligned. Theoretical estimates of the alignment efficiency of ultrasmall grains vary widely, with predicted AME polarization fractions ranging from $\lesssim 1\%$ (Hoang et al. 2013) to completely negligible (Draine & Hensley 2016). We perform a detailed study of AME polarization in Herman et al. (2022b), but assume it to be unpolarized in all other analysis.

3.3.5. Carbon monoxide emission

In the same way that rotating dust particles can emit radio emission, so can molecules with a non-zero electric dipole moment. One particularly important example of such molecules is carbon monoxide (CO), which resides primarily in dense clouds where it is shielded from destruction by UV radiation. The most common isotopologue of CO is $^{12}C^{16}O$ (abbreviated ^{12}CO), which is typically 10–100 times more abundant than $^{13}C^{16}O$ (abbreviated ^{13}CO) (Szűcs et al. 2014).

For a simple system such as CO, quantum mechanical effects are highly significant. In particular, only very specific rotational states are allowed by quantization of angular momentum. Let us denote the masses of the two atoms by m_{C} and m_{O}, respectively, and the corresponding atomic distances from their center of mass by r_{C} and r_{O}. We also define $r_{\mathrm{CO}} = r_{\mathrm{C}} + r_{\mathrm{O}}$ to be the effective atom size and $m_{\mathrm{CO}} = m_{\mathrm{C}}m_{\mathrm{O}}/(m_{\mathrm{C}} + m_{\mathrm{O}})$ its reduced mass.

With this notation, the moment of inertia of the CO molecule is $I = m_{\mathrm{C}}r_{\mathrm{C}}^2 + m_{\mathrm{O}}r_{\mathrm{O}}^2$. The corresponding eigenvalues of the Schrödinger equation are given by

$$E_{\mathrm{rot}} = \frac{J(J+1)\hbar^2}{2I}, \qquad (23)$$

where $J = 0, 1, \ldots$ is the angular momentum quantum number. Quantum mechanically allowed energy changes are given by $\Delta J = \pm 1$, and each such transition either absorbs or emits a photon with wavelength

$$\nu_0 = \frac{\Delta E_{\mathrm{rot}}}{h} = \frac{\hbar J}{2\pi I} = \frac{\hbar J}{2\pi m_{\mathrm{CO}} r_{\mathrm{CO}}^2}, \quad J = 1, 2, \ldots \qquad (24)$$

For the ^{12}CO $J=1\leftarrow0$ transition, one finds $\nu_0 = 115.27$ GHz, while for the ^{13}CO $J=1\leftarrow0$ transition, it is $\nu_0 = 110.20$ GHz. Higher-order transitions, such as $J=2\leftarrow1$, are very nearly multiples of these frequencies.

The width of CO lines is small compared to the broad *Planck* bandpasses, and so we model the corresponding SED by a delta function at the respective frequency,

$$s_{\mathrm{RJ}}^{\mathrm{CO}}(\nu) \propto \delta(\nu - \nu_0). \tag{25}$$

We note that specific intensity units are not appropriate for CO emission since all of the energy is being emitted in a narrow spectral range. Instead, CO emission is conventionally quantified in terms of $\mathrm{K\,km\,s^{-1}}$. Because the central frequency is known from theory, emission away from line center can be attributed to radial motion with a definite mapping between frequency and velocity. The line intensity in brightness temperature units is integrated over all velocities, yielding $\mathrm{K\,km\,s^{-1}}$.

CO emission is expected to be only weakly polarized, with a polarization fraction around 1 % in molecular clouds (Greaves et al. 1999). Detecting such low levels of true polarization is challenging with the currently available *Planck* data, primarily due to instrumental temperature-to-polarization leakage. For now, we assume CO line emission to be fully unpolarized, but note that this is likely to change in future analysis.

Finally, we note that although the base CO frequencies do not lie within the *Planck* LFI frequency bands themselves, CO emission is nevertheless indirectly important for LFI because of its prevalence in the HFI channels, and these are in turn critical to model thermal dust emission for LFI.

3.4. Extra-galactic compact sources

In addition to the Galactic foreground emission mechanisms discussed in Sect. 3.3, several extra-galactic effects are also important for CMB frequencies, including cosmic infrared background (CIB) radiation (Hauser & Dwek 2001) and the Sunyaev-Zeldovich effect (Sunyaev & Zeldovich 1972), and there are also contributions arising from within the Solar system, namely zodiacal light emission (Kelsall et al. 1998; Planck Collaboration X 2016; San et al. 2022). However, the *Planck* LFI data have a very low signal-to-noise ratio to any of these effects, and we therefore ignore these in the following.

For LFI frequencies, the most important class of extragalactic components are compact radio sources. All the emission mechanisms listed above operate in external galaxies, but the radio source population is dominated by active galactic nuclei (AGN). Radio emission from AGN is largely synchrotron, and comes from either the galactic nucleus itself or from jets and lobes associated with the nucleus. While the morphology of individual sources may be complicated, few are resolved by most CMB experiments and hence can be treated as "point" sources. Thus, while individual components of an AGN may exhibit polarized microwave emission, the emission from an unresolved source as a whole is rarely strongly polarized; typical polarization fractions are a few percent (Datta et al. 2019).

AGN have a wide range of SEDs, but most AGN spectra at CMB frequencies can be adequately modeled by a simple power law with a spectral index determined primarily by the energy spectrum of the relativistic electrons generating the synchrotron emission. The spectral indices (in brightness) typically range from 0 to -1, and most fall in a narrower range of -0.5 to -0.7. Hence we adopt a simple power law SED as our model for radio sources, and fit for the amplitude and spectral index of the radio

source contribution,

$$s_{\mathrm{RJ}}^{\mathrm{src}}(\nu) \propto \left(\frac{\nu}{\nu_{0,\mathrm{src}}}\right)^{\alpha-2} \tag{26}$$

Here ν_{src} is a fixed reference frequency, and α is the spectral index defined in intensity units; the conversion between intensity and brightness temperature is proportional to ν^2.

As we move to higher CMB frequencies, or to more sensitive experiments, the counts of extra-galactic sources begin to include dusty galaxies. These objects emit modified blackbody radiation, like Galactic dust, but typically at higher temperatures. Emission from this class of objects is included in the cosmic infrared background discussed below.

Unlike the dusty galaxies, which tend to be clustered, synchrotron-dominated radio sources are quite randomly distributed on the sky, and hence have a flat angular power spectrum. On the other hand, the emission of synchrotron dominated sources is frequently variable, on time scales ranging from days to years. Time variability is not accounted for in the current model, and variable sources are therefore likely to leave residuals in the final maps. For this reason, we will apply a dedicated point source mask during the final CMB parameter estimation, to minimize contamination in the final cosmological parameters.

3.5. Default sky model

Based on the above survey, and unless specified otherwise, the default BeyondPlanck astrophysical sky model (in brightness temperature units) reads as follows,

$$s_{\mathrm{RJ}} = \left(a_{\mathrm{CMB}} + a_{\mathrm{quad}}(\nu)\right)\frac{x^2 e^x}{(e^x-1)^2}+ \tag{27}$$

$$+ a_{\mathrm{s}}\left(\frac{\nu}{\nu_{0,\mathrm{s}}}\right)^{\beta_{\mathrm{s}}}+ \tag{28}$$

$$+ a_{\mathrm{ff}}\left(\frac{\nu_{0,\mathrm{ff}}}{\nu}\right)^2\frac{g_{\mathrm{ff}}(\nu;T_{\mathrm{e}})}{g_{\mathrm{ff}}(\nu_{0,\mathrm{ff}};T_{\mathrm{e}})}+ \tag{29}$$

$$+ a_{\mathrm{ame}}\left(\frac{\nu_{0,\mathrm{ame}}}{\nu}\right)^2\frac{f_{\mathrm{ame}}\left(\nu\cdot\frac{30.0\,\mathrm{GHz}}{\nu_{\mathrm{p}}}\right)}{f_{\mathrm{ame}}\left(\nu_{0,\mathrm{ame}}\cdot\frac{30.0\,\mathrm{GHz}}{\nu_{\mathrm{p}}}\right)}+ \tag{30}$$

$$+ a_{\mathrm{d}}\left(\frac{\nu}{\nu_{0,\mathrm{d}}}\right)^{\beta_{\mathrm{d}}+1}\frac{e^{h\nu_{0,\mathrm{d}}/k_{\mathrm{B}}T_{\mathrm{d}}}-1}{e^{h\nu/k_{\mathrm{B}}T_{\mathrm{d}}}-1}+ \tag{31}$$

$$+ U_{\mathrm{mJy}}\sum_{j=1}^{N_{\mathrm{src}}} a_{j,\mathrm{src}}\left(\frac{\nu}{\nu_{0,\mathrm{src}}}\right)^{\alpha_{j,\mathrm{src}}-2}, \tag{32}$$

where $x = h\nu/kT_0$ and $\nu_{0,i}$ is the reference frequency for component i; U_{mJy} is a conversion factor between flux density in millijansky and brightness temperature. Thus, a_i is the amplitude of component i in units of $\mu\mathrm{K}_{\mathrm{RJ}}$, as observed at a monochromatic frequency $\nu_{0,i}$. The sum in line 32 runs over all sources brighter than some flux threshold as defined by an external source catalog, and both the amplitude and spectral index are fitted individually per source. We adopt the same catalog as Planck Collaboration IV (2018), which is hybrid of the AT20G (Murphy et al. 2010), GB6 (Gregory et al. 1996), NVSS (Condon et al. 1998) and PCCS2 (Planck Collaboration XXVI 2016) catalogs comprising a total of 12 192 individual sources.

Only $\{s_{\mathrm{RJ}}, a_{\mathrm{CMB}}, a_{\mathrm{s}}, a_{\mathrm{d}}\}$ are assumed to be polarized in this model, and these comprise 3-component vectors including

Stokes T, Q, and U parameters. The remaining amplitudes parameters, $\{a_{\mathrm{ff}}, a_{\mathrm{AME}}, a_{\mathrm{src}}^j\}$, are assumed unpolarized, and have vanishing Stokes Q and U parameters.

For algorithmic reasons, we distinguish between linear and nonlinear parameters. The former group includes $\{a_{\mathrm{CMB}}, a_{\mathrm{s}}, a_{\mathrm{ff}}, a_{\mathrm{AME}}, a_{\mathrm{d}}, a_{\mathrm{src}}\}$, collectively denoted a; as described in Sect. 8.3.6, this set of parameters may be estimated jointly and efficiently through a multivariate Gaussian Monte Carlo sampler. In contrast, the nonlinear parameters include $\{\beta_{\mathrm{s}}, T_e, \nu_p, \beta_{\mathrm{d}}, T_{\mathrm{d}}, \beta_{\mathrm{src}}\}$, and these must be estimated independently and with computationally far more expensive algorithms; see Sect. 8.3.5 for specific details. In practice, we fit individual compact source amplitudes jointly with the corresponding spectral indices using a general sampling algorithm, since these are much more correlated with these than with any of the diffuse component parameters.

4. Instrument characterization

We now turn to the second half of the parametric model employed in the BeyondPlanck analysis, which describes the instrument used to collect the measurements. So that the Beyond-Planck analysis may freely be used by others, we aim to keep the presentation and notation as general as possible, and only introduce BeyondPlanck and LFI-specific notation where strictly necessary. We start our discussion by first defining an ideal detector response model, and then increase the level of realism step-by-step, until we reach the final instrument model.

4.1. Ideal instrument model

Let us first consider an ideal monochromatic detector observing at frequency ν a stationary sky signal with local Stokes parameters $\{T, Q, U\}$ at Galactic coordinates (l, b) and polarization angle ψ. We also initially assume infinite angular resolution. In this ideal case, the signal recorded by the detector as a function of time t may be written as

$$d(t) = g(t) \left[T + Q \cos 2\psi + U \sin 2\psi \right] + n(t), \tag{33}$$

where g is a multiplicative factor called the gain, which converts between physical signal units (which in our case will be $\mathrm{K_{CMB}}$) and digitized instrumental detector units (which in our case will be V), and n denotes instrumental noise.

In order to obtain data that may be processed on a computer, it is necessary to discretize the measurements by averaging over some (short) time period, Δt. For most CMB experiments, typical samples rates are between 10 and 200 Hz. A single recorded datum, d_t, thus corresponds to the detector output averaged over a period typically between 0.005 and 0.1 s.

For an ideal detector, the noise may be approximated as Gaussian and uncorrelated in time, and, as such, its variance decreases proportionally to $1/\Delta t$. We define the standard deviation of a single time sample to be σ_0.

A CMB experiment scans the sky according to some scanning strategy, $p(t) = [l(t), b(t), \psi(t)]$, while continuously recording the signal d_t. To describe this behaviour in a convenient notation, we first discretize the sky as described in Sect. 3.1, $s = s_p$, and then re-write Eq. (33) in vector form as follows,

$$d = \mathsf{G}\,\mathsf{P}s + n, \tag{34}$$

where $d = [d_1, d_2, \ldots, d_{n_{\mathrm{TOD}}}]^t$ and $n = [n_1, n_2, \ldots, n_{n_{\mathrm{TOD}}}]^t$ are time-domain vectors of length N_{TOD}, and G is a diagonal $N_{\mathrm{TOD}} \times$

N_{TOD} matrix with g_t on the diagonal. The scanning strategy is encoded in an $N_{\mathrm{TOD}} \times 3N_{\mathrm{pix}}$ matrix that contains $(1, \cos 2\psi, \sin 2\psi)$ in the columns that correspond to pixel p that happens to be observed at time t, and zero elsewhere, i.e.,

$$\mathsf{P} = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 & \cos 2\psi_1 & 0 & \ldots & 0 & \sin 2\psi_1 & 0 \\ 1 & 0 & 0 & \ldots & \cos 2\psi_2 & 0 & 0 & \ldots & \sin 2\psi_2 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & \ldots & 0 & 0 & \cos 2\psi_1 & \ldots & 0 & 0 & \sin 2\psi_1 \end{bmatrix}. \tag{35}$$

This matrix is called the pointing matrix.[6] Correspondingly, the sky vector consists of the three pixelized Stokes parameter maps stacked into a single vector, $s = [T_1, \ldots, T_{N_{\mathrm{pix}}}, Q_1, \ldots, Q_{N_{\mathrm{pix}}}, U_1, \ldots, U_{N_{\mathrm{pix}}}]^t$.

Equation (34) describes an ideal instrument that cannot be realized in actual hardware. The remainder of this section is therefore dedicated to generalizing this equation to the point that it actually does serve as a useful model for real-world CMB experiments.

4.2. Spectral response, bandpass averaging, and unit conversion

The first generalization we will consider is the assumption of monochromaticity. No real detector can measure a single frequency signal, but it is instead sensitive to a range of frequencies. This sensitivity is described by a so-called *bandpass profile* or *spectral transmission*, $\tau(\nu)$, which quantifies how much of the radiation at a given frequency is actually recorded by the detector. We define τ to be normalized to unity when integrated across all frequencies. Adopting brightness temperature units for all quantities (i.e., τ, d, and the monochromatic sky signal, $s(\nu)$), the data model in Eq. (34) generalizes to

$$d = \mathsf{G}\mathsf{P} \int s(\nu)\tau(\nu)\,\mathrm{d}\nu + n, \tag{36}$$

after taking into account the bandpass effect.[7]

However, most data sets are not provided in terms of brightness temperature units, but more often in either thermodynamic temperature or intensity units. As described in detail in Planck Collaboration IX (2014), in order to convert from unit convention X_i to unit convention X_j, one must multiply with a unit conversion factor that is given by

$$U_{ij} = \frac{\int \tau(\nu) \frac{dI_\nu}{dX_i}\,\mathrm{d}\nu}{\int \tau(\nu) \frac{dI_\nu}{dX_j}\,\mathrm{d}\nu}, \tag{37}$$

where dI_ν/dX_i is the intensity derivative expressed in unit convention X_i. In particular, the conversion factors from brightness temperature to thermodynamic temperature and intensity units are given by

$$U_{\mathrm{K_{RJ}} \to \mathrm{K_{CMB}}} = \frac{\int \tau(\nu) \frac{2k\nu^2}{c^2}\,\mathrm{d}\nu}{\int \tau(\nu)\, b'_\nu\,\mathrm{d}\nu} \tag{38}$$

$$U_{\mathrm{K_{RJ}} \to \mathrm{MJy\,sr^{-1}}} = \frac{\int \tau(\nu) \frac{2k\nu^2}{c^2}\,\mathrm{d}\nu}{\int \tau(\nu) \frac{\nu_c}{\nu}\,\mathrm{d}\nu}, \tag{39}$$

---

[6] Only the nonzero entries need to be stored in the pointing matrix, and memory requirements are therefore manageable.

[7] Note that many experiments, including *Planck* HFI, defines the bandpass profile in intensity units rather than brightness temperature units, and in this case an additional factor of $2h\nu^2/c^2$ must be included in the integral, as given by Eq. (9); see Planck Collaboration IX (2014) for details.

where

$$b'_\nu = \left.\frac{\partial B(T,\nu)}{\partial T}\right|_{T=T_0} \tag{40}$$

is the derivative of the blackbody function with respect to temperature, evaluated at the CMB temperature T_0, and ν_c is an arbitrary selected reference frequency for the channel in question. For other conversions, including to $\mathrm{K\,km\,s^{-1}}$ and the SZ y-parameter, we refer the interested reader to Planck Collaboration IX (2014). Taking into account both bandpass integration and unit conversion, the instrument model reads

$$\boldsymbol{d} = U\mathsf{G}\mathsf{P}\int \boldsymbol{s}(\nu)\tau(\nu)\,\mathrm{d}\nu + \boldsymbol{n}. \tag{41}$$

We aim to constrain $\boldsymbol{s}$ given $\boldsymbol{d}$. It is therefore important to be able to quickly evaluate the integral and unit conversion factors in Eq. (41). With this in mind, we consider signal component i as defined by the sky model in Sect. 3.5, and write it in the general form $\boldsymbol{s}_i(\nu) = \boldsymbol{a}_i f_i(\nu; \nu_0, \beta)$, where $\boldsymbol{a}_i$ is the linear amplitude relative to some reference frequency, $\nu_{0,i}$, and $f_i(\nu; \beta)$ is the frequency scaling from that reference frequency to an arbitrary frequency ν, which depends on some set of spectral parameters β. The total signal measured by detector j may then be written as

$$\boldsymbol{s}^j = \sum_{i=1}^{N_{\mathrm{comp}}} \boldsymbol{a}_i \left[U_j \int f_i(\nu; \beta)\,\tau_j(\nu)\,\mathrm{d}\nu \right] \equiv \sum_{i=1}^{N_{\mathrm{comp}}} \mathsf{M}_i^j\, \boldsymbol{a}_i = \mathsf{M}^j\, \boldsymbol{a}, \tag{42}$$

where M_i^j is called the *mixing matrix*. In order to take into account bandpass integration and unit conversion, the idealized data model in Eq. (34) must be generalized as follows,

$$\boldsymbol{d} = \mathsf{G}\mathsf{P}\mathsf{M}\boldsymbol{a} + \boldsymbol{n}. \tag{43}$$

It is evident that M depends only on the spectral parameters β and the bandpass τ, but not the amplitudes. Since most signal components are parameterized with limited number of spectral parameters (see Sect. 3), and these parameters are typically also fairly uniform on the sky, it is possible to pre-compute accurate lookup tables for M for each component and detector. In our current code, we adopt (bi-)cubic splines with regular grids for these lookup tables, and the computational cost of performing a full bandpass integral is thus equal to that of a simple polynomial evaluation.

4.2.1. Bandpass uncertainties and corrections

While the bandpass integral described by Eq. (36) may look simple enough at first glance, it does lead to a wide variety of important complications in practice. The most important among these is the fact that the exact shape of the bandpass profile itself is unknown. In particular, it is highly nontrivial to measure τ accurately in a laboratory for a combined multi-component instrument, and it is obviously impossible to do so after commissioning for satellite missions.

As a concrete real-world illustration of this, Fig. 2 shows the laboratory-determined (normalized) bandpass profiles after averaging over all radiometers for a given LFI channel. First, we see that the profiles for both 44 and 70 GHz are truncated, and therefore significant response is likely present outside the measured range. Second, for all three channels we see notable small scale ripples, which are due to standing waves. These may be due to real standing waves within the optical assembly of the
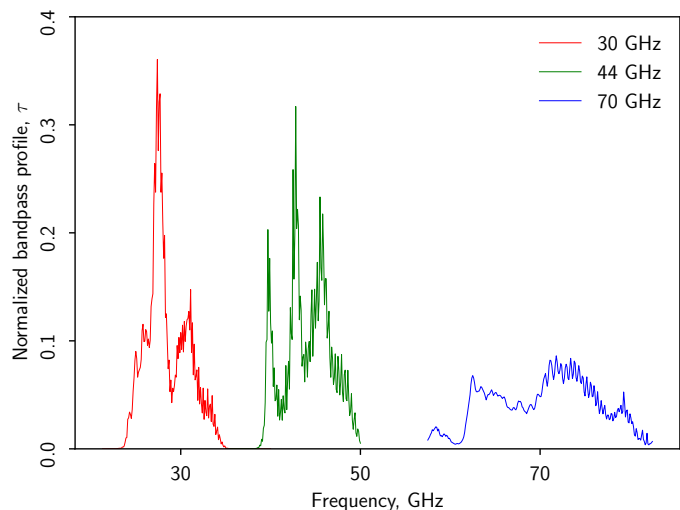

Fig. 2. Detector averaged bandpass profiles, τ, for the three *Planck* LFI frequency channels.

LFI instrument itself; but some part of them may also be due to standing waves in the test equipment used to make the measurements. These artefacts were already noted by Zonca et al. (2009), but they were never actually corrected during the course of the official processing; this is now done as part of the current reprocessing, as described by Svalheim et al. (2022a). In addition to these visually obvious effects, there may also be systematic errors in the actual shape, for instance in the form of a smooth slope across the bands, or in the precise position of the peaks within the band.

As described in Sect. 8.3.1, the CMB dipole serves as our primary calibrator for BEYONDPLANCK, following both *WMAP* and the official *Planck* pipelines. Because the CMB SED very closely follows a blackbody spectrum, which translates into a frequency independent scaling in thermodynamic units, the precise shape of the bandpass is irrelevant for the CMB component. Instead, errors in the bandpass shape effectively translate into incorrectly estimated foreground components, and introduce inaccuracies in the relative foreground SEDs between different frequency channels. In turn, foreground errors can affect the CMB reconstruction.

To account for the uncertainties noted above, we introduce one or more free parameters that can modify the bandpass shape, and allow the data to inform us about, and hence mitigate, potential inaccuracies in the laboratory bandpass measurements. The simplest and most common model we adopt is a simple linear shift, Δ_{bp}, in frequency space,

$$\tau(\nu) = \tau_0(\nu + \Delta_{\mathrm{bp}}), \tag{44}$$

where τ_0 is the default laboratory measurement. One value of Δ_{bp}^i is allowed per radiometer i, but (in most cases) either with the prior that $\sum_i \Delta_{\mathrm{bp}}^i = 0$, or that one particular channel is held fixed. Without any priors, the bandpass parameters are fully degenerate with the spectral parameters β of the foreground model, and no stable solution can be found. Various choices of both bandpass models and priors are considered by Svalheim et al. (2022a). In general, we note that the impact of Δ_{bp} is essentially to scale the amplitude of foregrounds, while leaving the CMB unchanged. At CMB dominated frequency channels, the bandpass shift is therefore non-degenerate with respect to the gain, while at foreground-dominated channels, it is virtually impossible to distinguish between a bandpass error and a gain error.
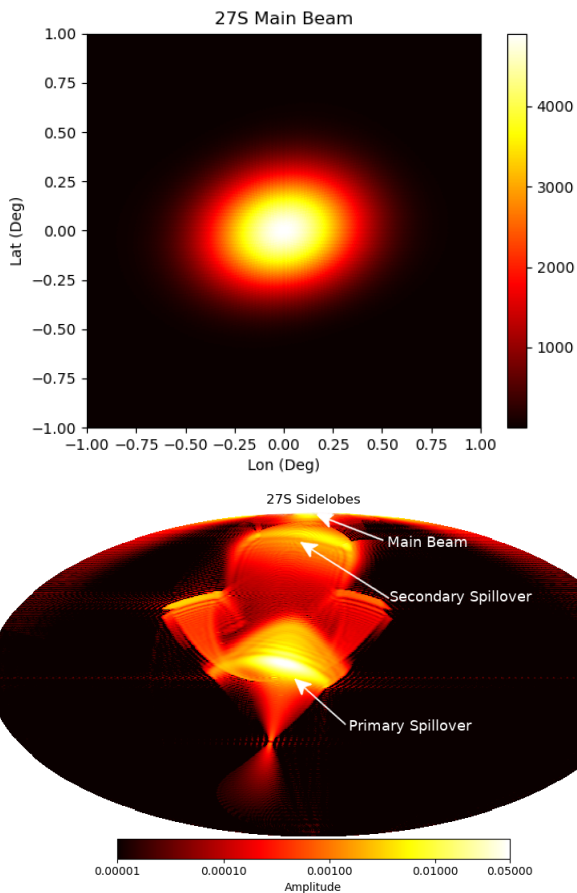
Fig. 3. (*Top*:) Scanning beam (or main beam) of the 30 GHz LFI 27S radiometer in local telescope coordinates, i.e., the instantaneous spatial sensitivity to a point source centered at the beam maximum. *Bottom*: Corresponding 4π beam map, oriented such that the main beam is located on the north pole. The main *Planck* far sidelobes are caused by spillover from (i.e., diffraction around) the primary and secondary mirrors. The beams are normalized such that their combined integral over the full sky equals unity.

In addition to this fundamental uncertainty in the bandpass profile for each detector, we note, first, that different detectors within the same frequency band observe different sky signals, and if not properly accounted for, this can create so-called bandpass mismatch errors in co-added frequency maps (see Sect. 7.2.2). Second, as discussed in the next section, the instrumental beam is also intrinsically frequency dependent, with an angular resolution of the main beam that is inversely proportional to the frequency for diffraction-limited observations, as is the case for LFI. In addition, far sidelobes can vary rapidly with frequency through complicated diffraction patterns. Unless properly accounted for, all these effects can potentially compromise final estimates. In BeyondPlanck we account for sidelobes as modelled by the *Planck* team (Planck Collaboration IV 2016), but we do not explore uncertainties in the beam model itself.

4.3. Beam and pixel window convolution

In the same way that no real detector can measure the signal from only a single monochromatic frequency, no real detector can measure the signal from a single point on the sky. Rather, each detector is associated with a so-called "point spread func-
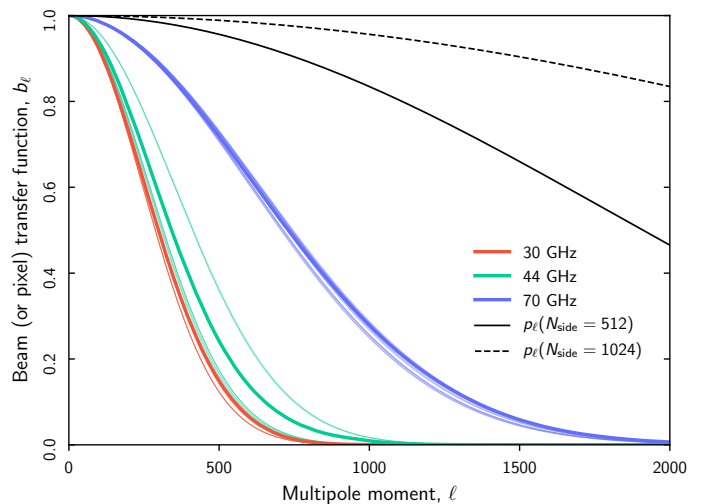
Fig. 4. Azimuthally symmetric and normalized beam transfer functions, b_ℓ for each co-added *Planck* LFI frequency channel (thick colored lines) and for each radiometer (thin colored lines). The former include the effects of non-Gaussian tails, while the latter are strictly Gaussian approximations. Black curves show the `HEALPix` pixel window functions for $N_{\rm side} = 512$ and 1024, respectively.

tion" (PSF) or "beam", $\boldsymbol{b}(\hat{n})$, that characterizes its instantaneous spatial sensitivity. Conventionally, we define $\boldsymbol{b}(\hat{n})$ to be normalized by setting the full-sky integral equal to unity, and to be oriented such that its maximum value is centered on the north pole.

The recorded value of the sky signal, $s_t^{\rm beam}$, as seen through the beam at time t is then given as the convolution of $\boldsymbol{b}$ and $\boldsymbol{s}$,

$$s_t^{\rm beam} = \int_{4\pi} \boldsymbol{b}_t(\hat{n})\boldsymbol{s}(\hat{n}) \, {\rm d}\Omega, \tag{45}$$

where $\boldsymbol{b}_t(\hat{n}) = {\sf R}_t(\hat{n}, \hat{n}')\boldsymbol{b}(\hat{n}')$, and ${\sf R}_t$ is a time-dependent rotation matrix that rotates the beam as given by the scanning strategy of the instrument. Since convolution is a linear operation, we may define a matrix operator, ${\sf B}$, such that $\boldsymbol{s}^{\rm beam} = {\sf B}\boldsymbol{s}$, and the data model in Eq. (43) may therefore be generalized further into its final form,

$$\boldsymbol{d} = {\sf GPBM}\boldsymbol{a} + \boldsymbol{n}, \tag{46}$$

where the position of the operator is defined by noting that the beam only acts on the true sky signal, and not on instrumental effects such as gain or noise.

Noting that modern CMB maps typically comprise up to several hundred million pixels, Eq. (45) is prohibitively expensive to evaluate directly in pixel space. Instead, we take advantage of the convolution theorem, which states that any convolution in pixel space may be expressed as a multiplication in harmonic space, and vice versa. As first demonstrated by Wandelt & Górski (2001), and later optimized by Prézeau & Reinecke (2010), Eq. (45) may be computed efficiently through reduced Wigner matrices, reducing the cost by a factor of $O(\sqrt{N_{\rm pix}})$ per evaluation for a general $\boldsymbol{b}$. A particularly computationally convenient formulation of this algorithm is presented by Galloway et al. (2022b) as part of the current BeyondPlanck work.

Another substantial saving can be made if we additionally assume that $\boldsymbol{b}$ is azimuthally symmetric. In that case, the spherical harmonics expansion of $\boldsymbol{b}$ is independent of m, and may be expressed in terms of its Legendre transform, b_ℓ. The full convolution may (by the convolution theorem) in this case be written

as[8]

$$s^{\text{beam}}(\hat{n}) = \sum_{\ell=0}^{\ell_{\text{max}}} \sum_{m=-\ell}^{\ell} b_\ell s_{\ell m} Y_{\ell m}(\hat{n}), \qquad (47)$$

where $s_{\ell m}$ are the spherical harmonics coefficients of s. Often, b_ℓ is referred to as the beam transfer function.

Note that the bandlimit, ℓ_{max}, in Eq. (47) should be selected sufficiently large that $b_\ell \approx 0$ as compared to the noise level of the instrument. Conversely, if a too low value of ℓ_{max} is adopted for analysis, the most notable artifacts arising from the convolution is ringing around bright point sources, resulting from premature harmonics truncation.

Note also that $s^{\text{beam}}(\hat{n})$ in Eq. (47) is written as a function of position rather than time in the above expression, which is only possible in the case of an azimuthally symmetric beam. To obtain the time-dependent signal, one simply reads off the value of $s^{\text{beam}}(\hat{n})$ given by the beam center position at time t. In this approximation, a full real-space convolution may be carried out at the cost of only two spherical harmonics transforms.

As discussed in Sect. 3.1, all CMB sky maps are pixelized in order to allow for efficient analysis on a computer. Such pixelization corresponds to an additional smoothing operation of the true sky signal that can be approximated with a top-hat convolution kernel of a given pixel size. For `HEALPix`, the effect of this kernel in harmonic space is described in terms of a pixel window function, p_ℓ, that is provided with the library. Implementationally, it is often convenient to redefine $b_\ell \rightarrow b_\ell p_\ell$ internally in computer codes, as the beam and pixel window affect the signal in the same way, and accounting for the pixel window can therefore usually be done with no additional computational cost compared to beam convolution.

In Euclidean space, the Nyquist theorem assures that any bandwidth limited signal may be reconstructed exactly with at least two samples per bandwidth. No corresponding exact theorem exists on the sphere. Instead, a rough rule of thumb for smooth spherical fields is to allow for at least two or three pixels per beam width. Likewise, no exact multipole bandlimit exists for given a `HEALPix` pixelization; however, numerical experiments suggest that multipoles above $\ell \gtrsim 2.5 N_{\text{side}}$ are poorly resolved on the `HEALPix` grid. Combined, these rules of thumb provide useful relationships between a given beam width and the corresponding appropriate values of N_{side} and ℓ_{max}.

Figure 3 shows the beam of the *Planck* 27S radiometer (Planck Collaboration IV 2014). The bottom panel shows the full 4π beam, while the top panel shows a zoom-in on the north pole. Clearly, this beam pattern is not azimuthally symmetric. However, in this respect it is useful to distinguish between the *main beam*, which is highlighted in the top panel, and the *sidelobes*, which are highlighted in the bottom panel. Furthermore, since convolution is a linear operation, contributions from the main beam and sidelobes may be computed separately.

The sidelobes are caused by optical imperfections, typically by diffraction around the main optical elements. In the case of *Planck*, these are the primary and secondary mirrors (see Fig. 3). As such, the resulting beam structures tend to be highly frequency dependent, and also cover large angular scales. While they clearly cannot be described as azimuthally symmetric in any meaningful way, they are associated with relatively modest bandlimits, ℓ_{max}, and this leads to acceptable computational costs for treating this component.

---

[8] This expression applies to temperature convolution; polarization convolution is notationally slightly more involved, but mathematically fully analogous.

The main beam, on the other hand, can often be described reasonably well as azimuthally symmetric, when centered on the north (or south) pole. Of course, the LFI 27S beam shown in the top panel of Fig. 3 exhibits a substantial ellipticity of $\epsilon \approx 1.3$, but this instantaneous beam profile is at least partially symmetrized by averaging due to the scanning strategy. The remaining effects of beam asymmetries may be accounted for, at least in terms of power spectrum bias, by adjusting the transfer function b_ℓ through simulations, as described by, e.g., Mitra et al. (2011).

For simplicity or because of low signal-to-noise, the beam profile is also sometimes approximated in terms of a two-dimensional Gaussian with some full-width-half-maximum (FWHM), or σ_{FWHM}, in the following expressed in radians. In the Gaussian case, one can derive an explicit expression for the beam transfer function in the form

$$b_\ell = \text{e}^{-\frac{1}{2}\ell(\ell+1)\frac{\sigma_{\text{FWHM}}^2}{8\ln 2}}, \qquad (48)$$

where the factor $8\ln 2$ simply accounts for the conversion between the square of the FWHM and the variance for a Gaussian.

Figure 4 compares the azimuthally symmetric beam transfer functions of the three *Planck* LFI channels, co-added over all radiometers, as well as the Gaussian approximations to the individual radiometer beam transfer functions. For reference, we also show the `HEALPix` window transfer functions for $N_{\text{side}} = 512$ and 1024, which are the typical pixelizations used for LFI and *WMAP* analysis.

We see that the general azimuthal approximations tend to have slightly heavier tails than the Gaussian approximations, and this is important to account for when estimating the CMB power spectrum, C_ℓ. At the same time, we also see that for applications for which only percent-level accuracy is required, the Gaussian approximations may very well be sufficient. In the following analyses, we will adopt the general azimuthally symmetric approximations for co-added frequency maps, which will be used for component separation and CMB estimation purposes, but Gaussian approximations for radiometer-specific signal modelling during time-domain processing, where the signal-to-noise ratio per sample is low, and sub-percent precision is irrelevant. The reason for the latter approximation is simply that the *Planck* collaboration only provides FWHM estimates for individual radiometers, not full transfer functions.

In the current work, we assume that the transfer functions provided by the *Planck* collaboration are exact, and do not assign dedicated stochastic parameters to them. This is neither a realistic description, nor a testament to the accuracy of the provided products, but only a statement of currently limited human resources; a high-priority task for future work is to implement full support for dynamic beam modelling and error propagation. As presented in this work, however, beam convolution is assumed to be a fully deterministic operation, dependent on officially available beam characterizations alone.

4.4. Gain and analog-to-digital conversion

While the instrument model in Eq. (46) is structurally complete in terms of components, we still need to introduce a few generalizations before we can apply it to our data. The first regards the gain g, simply by reemphasizing that this should be interpreted as a truly time-dependent object, g_t.

To understand why this is the case, it is useful to consider its origin and physical interpretation, and to focus the discussion we will consider the special case of a perfect total-power receiver.

The output voltage of such a device is given by

$$P = GkT_{\text{sys}}\Delta\nu, \tag{49}$$

where G is a unit-less gain factor, and $\Delta\nu$ is the width of the bandpass. The system temperature is defined as $T_{\text{sys}} = T_{\text{ant}} + T_{\text{recv}}$, where $T_{\text{ant}} = T_{\text{CMB}} + T_{\text{fg}}$ is the antenna temperature, and T_{recv} is the receiver temperature; the latter essentially defines the intrinsic noise level of the receiver.

For a *Planck* LFI 30 GHz radiometer, the bandwidth is 6 GHz, and the receiver temperature is typically 10 K. The antenna temperature is dominated by the CMB temperature, $T_{\text{CMB}} = 2.7$ K, as other sky components typically only make up a few mK at most. Assuming, therefore, a system temperature of about 13 K, Eq. (49) predicts that the power measured by this device is $P = 1.1$ pW or $P = -90$ dBm,[9] assuming no amplification ($G = 1$). However, current microwave detectors are typically only able to reliably record power levels larger than $P \gtrsim -30$ dBm. For this reason, the signal level must be actively amplified by a factor of 60 dB or more between the optical assembly and the detector. For *Planck* LFI, such amplification is achieved through the use of high-electron-mobility transistors (HEMTs).

HEMTs provide high gain factors, while adding only very low levels of additional noise to the data. However, they are not perfectly stable in time. Rather, their effective gains exhibit time-dependent drifts with typical overall variations at the $O(10^{-6})$ level, and correlations in time that are often well described by a so-called $1/f$ spectrum (see Sect. 4.5). Unless explicitly accounted for in the model, these time-dependent gain fluctuations can and will bias the derived sky model.

The gain defined by our original instrument model in Eq. (46), denoted G, is in principle the same gain as in Eq. (49), but with two important differences. First, while G is defined as a pure power amplification, and therefore unit-less, G takes into account the end-to-end conversion from a raw sky signal to final recorded data values. As such, G has units of V K^{-1}, in order to be dimensionally correct.

As far as measuring the actual gain from a given experiment at high precision is concerned, an important practical aspect is suppressing measurement noise through various gain smoothing algorithms. The official *Planck* processing employed a simple boxcar average algorithm for this purpose (Planck Collaboration II 2016; Planck Collaboration VI 2020). As shown by Gjerløw et al. (2022), this method can lead to striping in the final map by averaging over real variations. The current BEYONDPLANCK processing therefore replaces the boxcar algorithm with a tuned Wiener filter which allows variations on much finer time-scales, and this greatly suppresses gain-induced striping. In fact, this modification represents the single most important algorithmic improvement in the BEYONDPLANCK analysis in terms of improving data quality, and it was a key step in understanding long-standing issues regarding the LFI 44 GHz channel (Planck Collaboration II 2020; Planck Collaboration V 2020).

Second, G additionally takes into account the digitization process that converts analog signals to digital bits stored on a computer. This process takes place in a so-called analog-to-digital converter (ADC). An ideal ADC is perfectly linear. Unfortunately, many real-world ADCs exhibit important imperfections, for instance in the form of smooth nonlinear conversion within given signal ranges, or, as for LFI, sharp jumps at specific signal or bit values.

Overall, ADC errors are indistinguishable from gain fluctuations in terms of their direct impact on the recorded data. However, there is one critical difference between the two effects: While gain fluctuations are stochastic and random in time, and do not correlate with the sky signal, ADC errors are perfectly reproducible, and depend directly on the sky signal. Consequently, while the archetypical signature of unmitigated gain fluctuations are coherent stripes or large-scale features in the final sky maps, the corresponding unique signature of unmitigated ADC errors is an asymmetry in the amplitude of the CMB dipole along its positive and negative directions. This effect can be used to characterize and mitigate ADC non-linearity, as done both for *Planck* LFI and HFI (Planck Collaboration II 2020; Planck Collaboration III 2020; Planck Collaboration Int. LVII 2020). For a discussion of ADC corrections within the BEYONDPLANCK framework, see Herman et al. (2022a).

4.5. Instrumental noise

We complete our review of the instrument model by considering the properties of the instrumental noise, $\boldsymbol{n}$. This component may be decomposed into two main contributions, called correlated and white noise,

$$\boldsymbol{n} = \boldsymbol{n}_{\text{corr}} + \boldsymbol{n}_{\text{wn}}. \tag{50}$$

Both terms may be approximated as Gaussian, but they have different covariances.

The dominant physical source of white noise is Johnson (or thermal) noise, typically excited by thermal electron motions within the electric radiometer circuits. This noise is temperature dependent, and cryogenic cooling is usually required to achieve sufficient sensitivity. The dominant source of the correlated noise term are rapid gain fluctuations modulating the system temperature, T_{sys}, as discussed in Sect. 8.3.1.

Based on this decomposition, the standard deviation of the total instrumental noise term for a sample of duration Δt (i.e., σ_0 in Eq. (34)) may be estimated through the so-called radiometer equation,

$$\sigma_0 = T_{\text{sys}} \sqrt{\frac{1}{\Delta\nu\Delta t} + \left(\frac{\Delta g}{g}\right)^2}. \tag{51}$$

Here, Δg is the root-mean-square gain variation over Δt, and $\Delta\nu$ is as usual the receiver bandwidth. Intuitively speaking, this equation summarizes the following facts. First, the noise level is proportional to the system temperature, in recognition of the fact that Johnson noise scales with temperature. Second, the white noise term is inversely proportional to the square root of both bandwidth and integration time; this is simply by virtue of collecting more photons, and noting that Gaussian errors add in quadrature. Third and finally, the correlated noise component is proportional to the overall gain fluctuation level. Typical values of σ_0 for the LFI radiometers range between 600 and 1700 μK per sample in temperature units, or between 50 and 200 μV in detector units. If N_{obs} independent observations are made of the same sky pixel p, then the effective noise of the corresponding pixel integrates down roughly as $\sigma_p = \sigma_0/\sqrt{N_{\text{obs}}}$.

The different correlation structures of the white and correlated noise terms are most conveniently described in frequency domain through the noise power spectrum density (PSD), $P_n(f) = \langle |n_f|^2 \rangle$, where n_f are the Fourier coefficients of $\boldsymbol{n}_t$. This PSD is often modelled in terms of a so-called $1/f$

---

[9] The unit dBm measures power ratios, x, in decibel relative to 1 mW, i.e., $x = 10 \log_{10} \frac{P}{1\,\text{mW}}$.
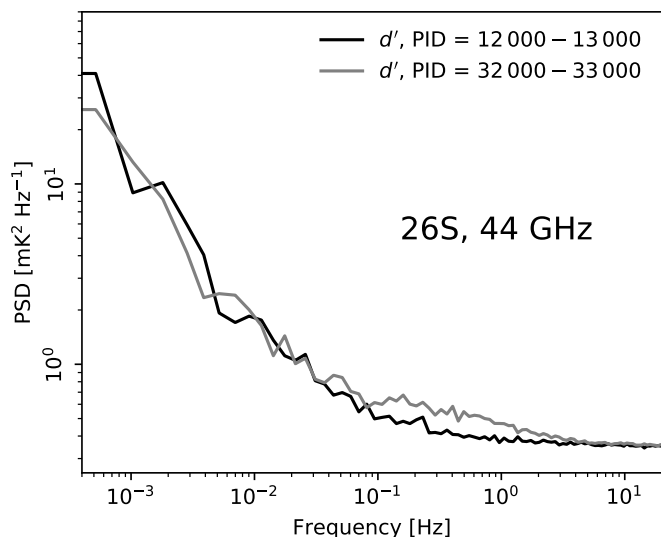
Fig. 5. PSD of signal subtracted data from radiometer 26S, averaged over ten PIDs (at intervals of 100 PIDs) in the ranges 12 000–13 000 (black) and 32 000–33 000 (grey). We see that there is significantly more power in the frequency range 0.1–10 Hz in the later period. For further discussion, see Ihle et al. (2022).

profile, which takes the form

$$P_n(f) = \sigma_0^2 \left[1 + \left(\frac{f}{f_k} \right)^\alpha \right]. \qquad (52)$$

Here, f_k is the knee frequency at which the variance of the correlated noise equals that of the white noise, and α is the slope of the spectrum at low frequencies. Typical best-fit values for LFI radiometers are $f_k \approx 10$ mHz and $\alpha \approx -1$. However, this model is obviously only approximate; if for no other reasons, the real spectrum has to flatten at low frequencies by energy considerations, whereas the power predicted by this model would approach infinity at low frequencies.

An important novel result that emerged from the current BEYONDPLANCK reprocessing is that the correlated noise for the LFI 30 and 44 GHz in fact cannot be adequately described by this simple $1/f$ model (Ihle et al. 2022). Rather, for most detectors we find a significant noise excess for frequencies between 0.01 and 1 Hz that may be described, at least phenomenologically, through the addition of a log-normal power term of the form,

$$P(f) = A_p \exp \left[-\frac{1}{2} \left(\frac{\log_{10} f - \log_{10} f_p}{\sigma_{\mathrm{dex}}} \right)^2 \right], \qquad (53)$$

where A_p is a freely fitted amplitude, and f_p represents the peak frequency of the log-normal term. An example of this power excess for a 44 GHz radiometer is shown in Fig. 5; the black line shows the measured noise power spectrum for a period of time in which the power excess is negligible, while the gray line shows the same for a period in which it is large. Properly modelling this term yields a significantly better χ^2, in particular on angular scales that are important for large-scale polarization measurements; for further discussion, see Ihle et al. (2022).

5. Data

The instrument discussion has until this point for the most part been kept general and applicable to a wide range of different data

sets. In this section, we specialize our discussion to *Planck* LFI. As discussed in Sect. 2, only this data set will be considered in the time-domain, while external data sets will be considered in the form of processed pixelized maps.

We note that the minimal sky model summarized in Sect. 3.5 includes seven distinct astrophysical components, three polarized and four unpolarized. Considering that there are only three LFI frequency channels, we immediately recognize that the LFI data must be augmented with at least four external frequency channels, just in order to make the model minimally constrained. In the default analysis configuration, we therefore include select observations also from *Planck* HFI (Planck Collaboration Int. LVII 2020) and *WMAP* (Bennett et al. 2013), as well as from some ground-based surveys. In this section, we provide a brief overview of these data sets, and refer the interested reader to the respective papers for full details.

The precise combination of data sets used in any particular BEYONDPLANCK analysis will depend on the goal of the respective application. For instance, the main scientific goal of the current paper is to introduce the concept of Bayesian end-to-end CMB analysis, and provide a first demonstration of this framework as applied to the LFI observations. Consequently, we here only include a minimal set of external observations, allowing LFI to play the dominant role, in particular with respect to CMB constraints. Specifically, in this paper we include only

– *Planck* 857 GHz to constrain thermal dust emission in intensity;
– *Planck* 353 GHz in polarization to constrain polarized thermal dust emission;
– *WMAP* 33, 41, and 61 GHz (called *Ka*, *Q* and *V*-bands, respectively) in intensity at full angular resolution to constrain free-free emission and AME;
– the same *WMAP* channels in polarization to increase the signal-to-noise ratio of polarized synchrotron emission, but only at low angular resolution, where a full noise covariance matrix is available; and
– Haslam 408 MHz (Haslam et al. 1982) to constrain synchrotron emission in intensity.

That is, we include neither intermediate HFI channels nor the *WMAP K*-band (23 GHz) channel, because of their higher signal-to-noise ratio relative to the LFI channels. The *WMAP W*-band is excluded because of known systematics effects (Bennett et al. 2013), and it does not have particularly unique features with respect to the signal model that are not already covered by other data sets.

We also note that Andersen et al. (2022), Svalheim et al. (2022b), and Herman et al. (2022b) focus on general foreground constraints, and these papers therefore also consider additional channels. The ultimate long-term goal of the global Bayesian CMB analysis program in general is of course to integrate as many data sets as possible into a single coherent sky model, and thereby produce the strongest possible constraints on the true astrophysical sky. One leading example of such an effort is the COSMOGLOBE[10] project, which specifically aims to combine many state-of-the-art experiments with the ones listed above, including CHIPASS (Calabretta et al. 2014), *COBE*-DIRBE (Hauser et al. 1998) and FIRAS (Mather et al. 1994), PASIPHAE (Tassis et al. 2018), SPIDER (Gualtieri et al. 2018), and many more. The BEYONDPLANCK methodology presented here represents an ideal statistical framework for performing such global data integration.
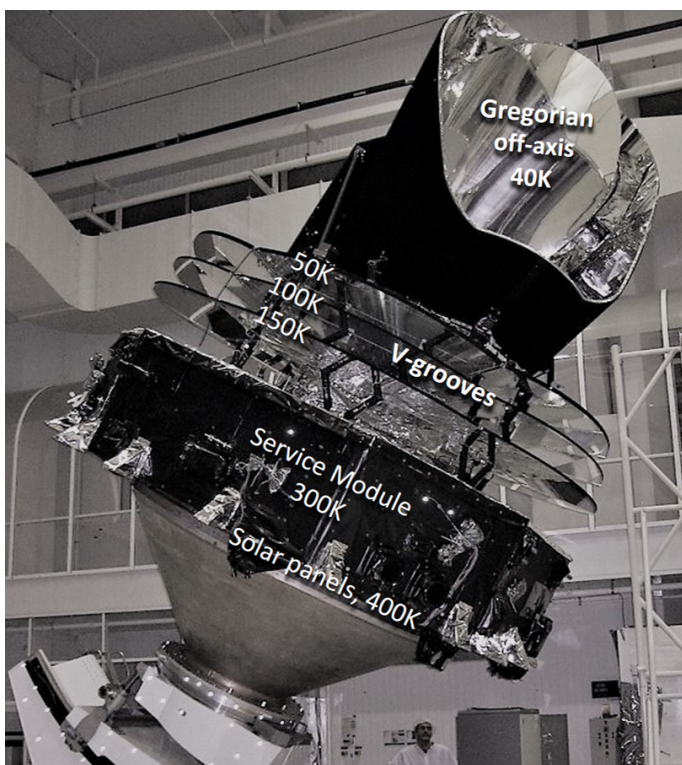
---

[10] http://cosmoglobe.uio.no

5.1. LFI instrument overview

5.1.1. Instrument configuration

We now provide a synthetic description of the LFI instrument configuration, which directly impacts the structure of the LFI data and the potential systematic effects addressed in the BEYONDPLANCK analysis. For more details on the LFI instrument, its ground calibration and in-flight performance, see Bersanelli et al. (2010), Mennella et al. (2011), and references therein; the overall LFI programme is described by Mandolesi et al. (2010).

The heart of the LFI instrument is an array of 22 differential receivers based on high-electron-mobility transistor (HEMT) low noise amplifiers. The instrument operates in three frequency bands, nominally centred at 30, 44 and 70 GHz, with angular resolutions of about 32′, 28′, and 13′ FWHM, respectively. The front end of the receivers is cooled to 20 K, which dramatically reduces the noise temperature of the HEMT amplifiers and of the overall system. In each receiver, the signal coming from different directions of the sky, intercepted by the telescope as the satellite spins, is compared to a stable internal blackbody reference load at 4 K. It is this differential scheme that allows the LFI to achieve its excellent stability.

Radiation from the sky is coupled to 11 corrugated feed horns, shown in Fig. 7. Each horn is followed by an orthomode transducer (OMT), which splits the incoming radiation into two perpendicular linear polarizations that propagate through two independent differential radiometers; see Fig. 8. The OMT provides exquisite polarization purity, with typical isolation of < -30 dB. Each radiometer pair has a front-end module (FEM), cooled to 20 K, and a back-end module (BEM), operated at
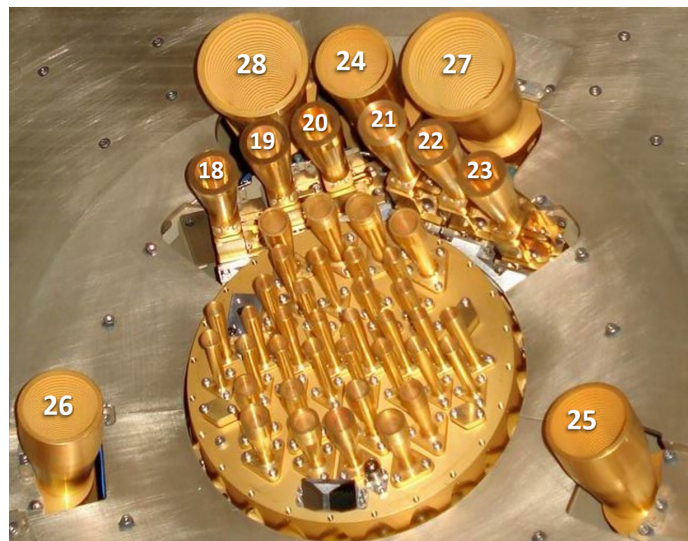
Fig. 7. Top view of the *Planck* focal plane. The central array contains the HFI feed-horns, cooled to 4 K, feeding bolometric detectors cooled to 0.1 K. The LFI horns in the outer part of the array are labelled with numbers; they are cooled to 20 K. The LFI horn numbers 18–23, 24–26, and 27–28 correspond to the 70 GHz, 44 GHz, and 30 GHz channels, respectively.

300 K. The FEM is connected to the BEM by four composite wave-guides (two for each radiometer), thermally coupled to the three *Planck* V-groove radiators to minimize parasitic heat transfer to the cold focal plane (see Fig. 6). The cryogenically cooled front-end modules include the first stage HEMT amplifiers and the differencing system, while the back-end modules provide further radio frequency amplification. Detection is made via two square-law detector diodes for each radiometer.

After detection, an analog circuit in the data acquisition electronics is used to adjust the offset to obtain a nearly null DC output voltage, and a programmable gain is applied on-board to match the signal level to the analog-to-digital converter (ADC) input range. After the ADC, data are digitally down-sampled, re-sampled to match beam resolution ($>$ 3 samples per beam), compressed, and assembled into telemetry packets, which are then downlinked to the ground station.

5.1.2. Stabilization

Cryogenic HEMT amplifiers exhibit excellent low-noise performance, but are affected by significant instability in terms of gain and noise-temperature fluctuations, typically modelled in terms of a $1/f$ spectrum as discussed in Sect. 4.5. The LFI system is designed to efficiently reject such fluctuations in the radiometer response. The main differential process responsible for radiometer stabilization takes place in the front-end modules. The signals from the sky and 4 K reference load are injected into a hybrid coupler, which splits the two signals, and redirects them to both of its output ports (see inset of Fig. 8). Then the two mixed signals are amplified by $\sim$30 dB by the two amplifier chains. Thus, any fluctuation in the FEM amplifiers affects both the sky and the reference load components in exactly the same way. After amplification, a second hybrid coupler reconstructs the sky and reference components, which now contain the same fluctuations. Then the signals are transported by the wave-guides in the warm back-end modules, where they are further amplified and detected by the diodes. Finally, when taking the difference
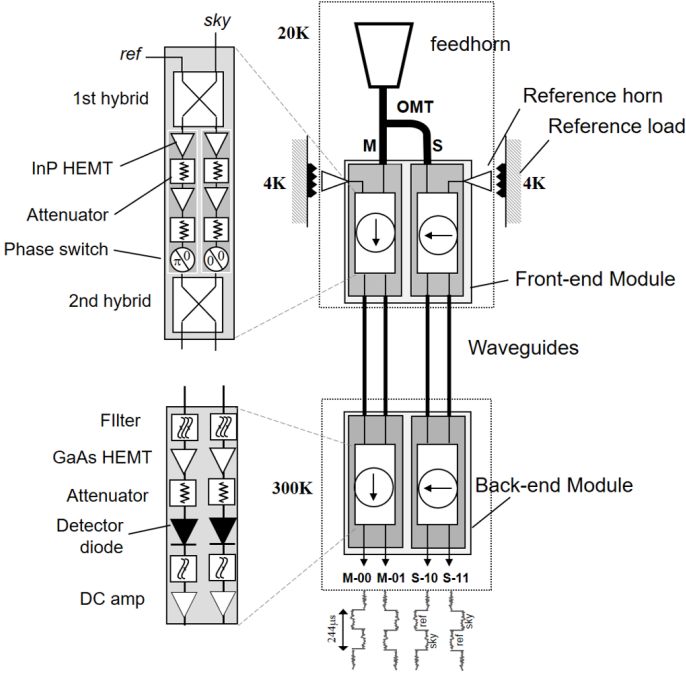
Fig. 8. Schematic of an LFI radiometer chain, consisting of a feedhorn and OMT, and the two associated radiometers, each feeding two diode detectors. The insets show details of a front-end module (top) and of a back end module (bottom).

between the two diodes, the FEM fluctuations cancel out. This "pseudo-correlation" scheme reduces front-end fluctuations by a factor of $\mathcal{O}(10^3)$.

However, instabilities downstream of the FEMs, particularly those originating in the back-end amplifiers and in the detector diodes, would still affect the measurements. For this reason, a further level of stabilization is built into the LFI design. A phase shifter, alternating between $0°$ and $180°$ at a frequency of 4096 Hz, is applied in one of the two amplification chains within the front-end modules, as shown in Fig. 8. In this way, the DC output from each diode rapidly alternates the sky and reference signals, with opposite phase in the two detectors. By taking the difference between time-averaged values of sky and reference, any residual fluctuations on time scales longer than $\sim (1/4096)$ s $= 0.244$ ms are removed.

Of course, any non-ideality in the receiver components will introduce some level of residual fluctuations. Further strategies to suppress remaining instabilities and potential systematics introduced by the receiver are described below.

5.1.3. LFI signal model

Based on the above description, the differential power output for each of the four diodes associated with a feedhorn can be written as the following special case of Eq. (49),

$$P_{\text{out,0}}^{\text{diode}} = a\, G_{\text{tot}}\, k\, \Delta \nu \left[T_{\text{sky}} + T_{\text{noise}} - r(T_{\text{ref}} + T_{\text{noise}}) \right], \tag{54}$$

where G_{tot} is the total gain, k is the Boltzmann constant, $\Delta \nu$ the receiver bandwidth, and a is the diode constant. T_{sky} and T_{ref} are the sky and reference load antenna temperatures at the inputs of the first hybrid and T_{noise} is the receiver noise temperature, averaged over an appropriate integration time. The gain modulation factor, r, is a factor of order unity ($0.8 < r < 1.0$ depending on channel) used to balance in software the temperature

offset between the sky and reference load signals. This has the important effect of minimising the residual $1/f$ noise and other non-idealities in the differential data stream. In the DPC analysis (Planck Collaboration II 2020) r was determined using the approximation

$$r = \frac{T_{\text{sky}} + T_{\text{noise}}}{T_{\text{ref}} + T_{\text{noise}}} \approx \frac{V_{\text{sky}}}{V_{\text{ref}}}, \tag{55}$$

and we adopt the same procedure without modification in BEYONDPLANCK for now. However, we do note that the NPIPE analysis pipeline implements an alternative approach in which V_{ref} is low-pass filtered prior to differencing, and this reduces the amount of high frequency noise in the final maps. Future BEYONDPLANCK versions can and should implement a similar solution.

Although somewhat counter-intuitive, the sensitivity of the LFI radiometers does not depend significantly on the absolute temperature of the reference load. In fact, to first order, the white noise spectral density at the output of each diode is given by

$$\Delta T_0^{\text{diode}} = \frac{2\,(T_{\text{sky}} + T_{\text{noise}})}{\sqrt{\Delta \nu}}. \tag{56}$$

However, a large imbalance between T_{sky} and T_{ref} would have the effect of amplifying residual fluctuations in the differential signal. For this reason the LFI reference loads are cooled to about 4 K, exploiting the HFI pre-cooling stage.

The above description holds for the ideal case where all front-end amplifiers and phase switches have perfectly balanced properties. In presence of some level of mismatch, the separation of the sky and reference load signals after the second hybrid is not perfect and the outputs are slightly mixed. If the front-end imbalance is small, Eq. (56) may be written as

$$\left(\Delta T^{\text{diode}} \right)^2 \approx \left(\Delta T_0^{\text{diode}} \right)^2 \left(1 \pm \frac{\epsilon_{A_1} - \epsilon_{A_2}}{2} + \alpha \epsilon_{T_{\text{n}}} \right), \tag{57}$$

where $\epsilon_{T_{\text{n}}}$ is the imbalance in front end noise temperature between the two radiometer arms, and ϵ_{A_1} and ϵ_{A_2} are the imbalance in signal attenuation in the two states of the phase switch. Eq. (57) shows that the output is identical for the two diodes apart from the sign of the term $(\epsilon_{A_1} - \epsilon_{A_2})/2$, representing the phase switch amplitude imbalance. For this reason, the LFI scientific data streams are obtained by averaging the voltage outputs from the two diodes in each radiometer,

$$V_{\text{out}}^{\text{rad}} = w_1 V_{\text{out}}^{\text{diode 1}} + w_2 V_{\text{out}}^{\text{diode 2}}, \tag{58}$$

where w_1 and w_2 are inverse-variance weights calculated from the data. Thus, the diode-diode anti-correlation is cancelled, and the radiometer white noise becomes

$$\Delta T^{\text{rad}} \approx \frac{\Delta T_0^{\text{diode}}}{\sqrt{2}} \left(1 + \alpha \epsilon_{T_{\text{n}}} \right)^{1/2}. \tag{59}$$

In Eqs. (57) and (59), $\epsilon \ll 1$, while α is a term of order unity defined by a combination of the input signals and noise temperature of the radiometer; for details, see Eq. (8) in Mennella et al. (2011).

In the current BEYONDPLANCK processing, we follow the LFI DPC procedure for all these steps. Future versions of the framework may also account for these pre-processing steps, and jointly estimate r, α, ϵ_i, and w_i, but this is left for future work, simply due to the strong time limitations of the current project (see Sect. 1.6).

5.1.4. Naming convention

As described in the previous section, LFI has 11 horns and associated OMTs, FEMs and BEMs; 22 radiometers (two for each horn); and a total of 44 detectors (two for each radiometer). For historical reasons, the 11 horns are labelled by numbers from 18 to 28 as shown in Fig. 7.

The radiometers associated with each horn are labelled as "M" or "S" depending on the arm of the OMT they are connected to ("Main" or "Side", as shown in Fig. 8). Each radiometer has two output diodes that are labelled with binary codes "00", "01" (radiometer M) and "10", "11" (radiometer S), so that the four outputs of each radiometer pair can be named with the following sequence; M-00, M-01, S-10, S-11.

As the telescope scans, the observed region of the sky sweeps across the focal plane in the horizontal direction as appearing in Fig. 7. Since the reconstruction of the polarization information requires at least two horns, every pair of horns aligned in the scan direction are oriented such that their linear polarizations are rotated by 45° from each other (with the exception of LFI-24, which is an unpaired 44 GHz horn). Thus, LFI can produce independent polarization measurements from the "horn pairs" 18–23, 19–22, 20–21 (at 70 GHz); 25–26 (at 44 GHz); and 27–28 (at 30 GHz).

5.2. Implementation details

Since the BeyondPlanck project aims to establish an open-source, reproducible and externally extendable analysis framework, it is no longer possible to rely on direct access to the existing LFI-DPC database, which both employs proprietary software and runs on one specific computer. To circumvent this issue, we convert the LFI TOD into a convenient HDF5 format (Galloway et al. 2022a) that may be accessed using publicly available tools. This, however, does lead to some adjustments in the scientific pre-processing pipeline, which now uses this new interface. At the same time, we have converted the scientific pipeline to C++11, and a number of optimizations are applied at the same time, exploiting the new possibilities given by that language.

5.2.1. Unprocessed Level-1 data

The extraction of time-ordered Level-1 data from the LFI-DPC database and the conversion to HDF5 format only need to be performed once within the LFI-DPC environment. We create one file for each LFI horn for each Operational Day, i.e., the time between two consecutive daily telecommunication periods. The extracted file contains sky, reference load and quality flags for each of the diodes of the horn and timing information, including On-Board Time, Spacecraft Event Time (SCET) and Modified Julian Date (MJD). It also contains attitude information that is critical for the analysis; Pointing Period ID (PID); start and end time of each Pointing Period; end time of the maneuver of each Pointing Period; and number of data samples.

To optimize the computational time of Level-2 processing, various deterministic operations are implemented during extraction. For instance, missing data are added back into the time streams and flagged as bad data; this ensures that all the timelines for each frequency are of the same length. Also, planet transits are flagged, and instrumental flags are added to the extracted data.

5.2.2. Level-2 data pre-processing

In the DPC pipeline, the main pre-processing of the LFI data occurs at the Level-2 stage (see Planck Collaboration II 2020 and references therein). The same is true in the BeyondPlanck framework, although this is now implemented as an integrated component in the full algorithm. The first step in this process is to correct the low-level diode observations for ADC non-linearities.

The analog signal from each detector is processed by an analog-to-digital converter (ADC), which ideally provides a digitized output exactly proportional to the applied voltage. If the voltage step sizes between successive binary outputs of the ADC are not constant, then the ADC introduces a nonlinear response that leads to calibration errors. In differential measurements such as those of LFI, small localized distortions due to ADC non-linearity can have a significant impact, since the calibration reconstruction depends on the gradient of the ADC response curve at the point at which the differential measurements are made.

A non-linearity of the ADC produces a variation in the white noise level of a detector which does not correspond to a variation in the input voltage level, as one would expect if the effect were due to a gain shift. This subtle effect was observed in some of the LFI radiometer data for the first time in flight, where drops of a few percent were observed in the voltage white noise but not in the output level over periods of few weeks (Planck Collaboration III 2014). Because of their lower detector voltages, the 44 GHz channels showed the strongest effect, reaching levels of 3 to 5 %. The typical amplitude of the region where the non-linearity occurs is on the order of 1 mV, corresponding to about three bits in the ADC.

The ADC non-linearity effect has been characterised from flight data and removed from the data streams. The correct response curves is reconstructed by tracking how the noise amplitude varies with the apparent detector voltage in the TOD. Under the assumption that the radiometers are stable, the intrinsic white noise is taken to be constant, so any voltage variations are taken to be due to a combination of gain drift and ADC effects. A mathematical model of the effect and the details of the correction method are described in Appendix A of Planck Collaboration III (2014). In BeyondPlanck, we adopt a very similar algorithm, but with one notable improvement: While the DPC algorithm allow for one correction per narrow voltage bin across the full observed range, the BeyondPlanck ADC model instead introduces a low-dimensional parametric model for each ADC glitch. As such, the BeyondPlanck model has many fewer degrees of freedom, and there is correspondingly less risk of over-fitting non-ADC-related artefacts; for further details, see Herman et al. (2022a).

The next low-level step is to compute a gain modulation factor, r, from the data streams, and apply this to minimize $1/f$ noise as given by Eq. (55). The outputs from the two detector diodes of each radiometer are then combined with appropriate noise weights, to remove the effect of phase switch mismatch, as given by Eq. (58). For now, we adopt the same weights as in the official processing. However, we note that future work should aim to implement a frequency-dependent weighting scheme as introduced by *Planck* PR4, which leads to about 8 % lower white noise (Planck Collaboration LVII 2020).

In the current BeyondPlanck implementation, neither the ADC correction nor diode weighting algorithms depend on the assumed sky model, and there is therefore no feedback from higher-level analysis steps to these steps. Accordingly, these low-level operations are performed as one-time pre-processing steps, saving both CPU time and RAM. Only the co-added and

ADC-corrected TOD are actually stored in memory during the main processing.

5.2.3. 1 Hz spike correction

The output signal of the LFI receivers exhibits a set of narrow spikes at 1 Hz and harmonics with different amplitude and shape for each detector. These subtle artifacts are due to a common-mode additive effect caused by interference between scientific and housekeeping data in the analog circuits of the LFI data acquisition electronics. The spikes are present at some level in the output from all detectors, but affect the 44 GHz data most strongly because of the low voltage and high post-detection gain values in that channel. The spikes are nearly identical in sky and reference load samples, and therefore almost completely removed by the LFI differencing scheme. However, a residual effect remains in the differenced data, which needs to be carefully considered in the data processing.

These features are synchronous with the On-Board time, with no measurable change in phase over the entire survey, allowing construction of a piecewise-continuous template by stacking the data for a given detector onto a one second interval, and fitting a single overall amplitude. In the DPC analysis the spikes were deemed to produce negligible effects in the 30 and 70 GHz channels, and were removed only from the 44 GHz time-ordered data via template fitting, while the *Planck* PR4 processing corrected all three channels. One notable difference between the two algorithms is that while the DPC analysis created the template shape by binning the TOD in time-domain, the PR4 analysis instead fitted a series of Fourier modes.

The BeyondPlanck 1 Hz correction algorithm combines aspects from both the DPC and PR4 analysis, and also introduces one new feature. First, and similar to the DPC analysis, we build the template shapes in time-domain, and fit one single overall amplitude across the full mission. Second, and in contrast to the DPC method, we do correct all three channels, similar to PR4. However, and unlike both the DPC and PR4, we perform the 1 Hz correction on the *co-added* time streams. The main advantage of this is that the precise binning used to construct the shape template does not need to be aligned between diodes to avoid edge effects, while the main disadvantage is that the ADC correction makes the 1 Hz effect in principle time-dependent. However, this latter issue a second-order correction, and, as shown by Basyrov et al. (2022), the full magnitude of the 1 Hz correction is anyway several orders of magnitude lower than any other systematic effect, and typically amounts to only a $\sim 0.3 \mu$Kcorrection in map domain; a second-order ADC correction with respect to this is entirely negligible.

5.3. Pixel-domain data

In addition to time-domain LFI data, we consider several external data sets in the pixel domain, as described in the introduction to this section, simply in order to be able to constrain the full astrophysical sky model as defined in Sect. 3.5.

5.3.1. *Planck* HFI data

The first external data set we consider is *Planck* HFI, primarily in order to constrain thermal dust emission in the LFI frequencies. The HFI measurements were taken during the first 29 months of *Planck* observations, from August 2009 until January 2011, at which time the helium coolant was depleted. The HFI instrument includes a total of six frequency bands, centered on 100, 143, 217, 353, 545, and 857 GHz, respectively. The first four channels are polarized, while the latter two are (at least nominally) only sensitive to intensity.

While LFI employs coherent radiometers and HEMTs for signal detection, HFI employs bolometers. One important difference between these two detector types is that while the former records both the phase and the amplitude of the incoming electric field, the latter is sensitive only to the amplitude. In practice, this difference translates into different sensitivity as a function of frequency, as well as different instrumental systematics. Generally speaking, bolometers have lower noise levels than coherent radiometers over relevant CMB frequencies, but they also tend to be more susceptible to various systematic errors. For instance, for the LFI 70 GHz radiometers the noise equivalent temperature[11] is $152 \mu K_{\mathrm{CMB}} \mathrm{s}^{-1/2}$ (Planck Collaboration II 2016), while it for the HFI 143 GHz bolometers is $57.5 \mu K_{\mathrm{CMB}} \mathrm{s}^{-1/2}$ (Planck Collaboration VII 2016). At the same time, the size of CMB detectors typically scales with wavelength, and it is therefore possible to fit a larger number of high frequency detectors than low-frequency detectors into the same focal plane area. In sum, HFI nominally has more than six times higher sensitivity than LFI with respect to CMB fluctuations, as measured in terms of white noise alone. However, a non-negligible fraction of this sensitivity advantage is lost because of higher sensitivity to cosmic rays, ADC non-linearities, and long-duration bolometer time constants (Planck Collaboration III 2020).

Several different HFI analysis pipelines were developed within the nominal *Planck* collaboration period, as detailed by Planck Collaboration VI (2014), Planck Collaboration VII (2016), and Planck Collaboration III (2020). The two most recent and advanced efforts are summarized in terms of the SROLL2 (Delouis et al. 2019) and NPIPE (Planck Collaboration Int. LVII 2020) pipelines. For BeyondPlanck, we adopt by default the NPIPE processing as our HFI data set, which is the most recent among the various available options. However, we note that most analyses here will only consider the highest frequency channels (857 GHz in temperature and 353 GHz in polarization), in order to constrain thermal dust emission, and the precise details of the HFI processing are largely irrelevant for these purposes.

The HFI data are pre-processed as follows before integration into the BeyondPlanck pipeline. First, we note that the HFI frequency channels have angular resolutions ranging between 9.7 arcmin at 100 GHz and 4.4 arcmin at 857 GHz. The natural HEALPix pixel resolution for HFI is thus either $N_{\mathrm{side}} = 2048$ or 4096. While our computational codes do support full resolution analysis, such high resolution is computationally wasteful for the purposes of LFI analysis. We therefore smooth the HFI maps to a common angular resolution of 10′ FWHM (which is still smaller than the 14′ beam of the 70 GHz channel), and we re-pixelize each map at $N_{\mathrm{side}} = 1024$. Overall, this reduces both CPU and memory requirements for the component separation phase of the algorithm by about one order of magnitude. Second, we subtract estimates of both zodiacal light and the kinematic CMB quadrupole from each sky map prior to analysis, following Planck Collaboration Int. LVII (2020), but exclude the quadrupole from the astrophysical sky model for these channels.

---

[11] The noise equivalent temperature (NET) represents the noise standard deviation, σ_0, expressed in thermodynamic units of μK_{CMB} with an integration time of $\Delta t = 1$ s.

5.3.2. *Wilkinson Microwave Anisotropy Probe*

Second, we consider observations from the *Wilkinson Microwave Anisotropy Probe* (*WMAP*; Bennett et al. 2013), primarily in order to constrain synchrotron, free-free, and anomalous microwave emission. *WMAP* was funded by the National Aeronautics and Space Administration (NASA), and operated for 9 years between 2001 and 2010. *WMAP* observed the microwave sky in five frequency bands, centered on 23, 33, 41, 61, and 94 GHz, with an angular resolution varying from 53′ at 23 GHz to 13′ at 94 GHz, and with sensitivities that range between 0.8 and 1.6 mK s$^{-1/2}$.

Like LFI, the *WMAP* detectors are based on coherent HEMT technology. However, there are (at least) two critical differences between the practical implementation of the two experiments. First, while the LFI detectors measure the difference between the sky signal in a single direction and that from an internal 4 K reference load, the *WMAP* detection chain is intrinsically differential. That is, each radiometer is coupled to two independent feedhorns that are separated by an angle of 141° on the sky, and each TOD sample is given by the difference between the signals recorded by those two horns. For this reason, each *WMAP* channel is often referred to as a "differencing assembly" (DA), rather than a radiometer. Second, while the basic *Planck* scanning strategy is fixed by its single reaction wheel, supporting smooth rotation only around a single axis, the *WMAP* satellite carried three orthogonal reaction wheels that allow for much more tightly interconnected scanning strategies. In sum, these differences lead to independent instrumental systematics between the two instruments and consequently to different strategies to minimise their impact. The two data sets are thus complementary, and can be used to break each other's internal degeneracies.

As discussed above, we will in this paper only use enough external data to break parameter degeneracies that cannot be resolved by *Planck* LFI alone, thereby leaving enough room to allow this data set to provide the main CMB constraints. Therefore, we include in the following only the *WMAP* channels between 33 and 61 GHz. In intensity, we use the *WMAP* 9-year full-resolution maps with a diagonal noise covariance matrix, while in polarization we use the low-resolution maps with full noise covariance. No pre-processing is applied to any *WMAP* data before integration into the BEYONDPLANCK pipeline.

5.3.3. Low-frequency surveys

As discussed by Planck Collaboration X (2016), because of the roughly similar shapes of the synchrotron, free-free and AME SEDs between 20 and 70 GHz, *Planck* and *WMAP* are not able to resolve these components on their own. Rather, it is critically important to complement these data with at least one low-frequency survey in order to establish a statistically non-degenerate model.

In BEYONDPLANCK, we follow Planck Collaboration X (2016), and include the celebrated 408 MHz survey by Haslam et al. (1982). Although this is widely believed to suffer more from instrumental systematic errors than comparable recent surveys, such as S-PASS (Carretti et al. 2019) or C-BASS (King et al. 2014), it also has the distinct advantages of both being publicly available and covering the full sky. This full-sky coverage was achieved by combining observations taken by the Jodrell Bank MkI 76 m telescope, the Bonn 100 m telescope, and the Parkes 64 m telescope during the 1960's and 1970's. A second advantage is its very low frequency, which allows for a very clean separation of synchrotron emission, with only a minor additional contribution from free-free emission.

We adopt the reprocessed version of the Haslam map that was presented by Remazeilles et al. (2015) for our analyses, and, following Planck Collaboration X (2016), we model the uncertainty of this map with a uniform standard deviation of 0.8 K per pixel, added in quadrature to 1 % of the amplitude in that pixel. Finally, we adopt the monopole and dipole corrections presented by Wehus et al. (2017) to fix the largest angular scales.

6. Bayesian analysis and MCMC sampling

We have now defined an effective parametric model of the astrophysical sky in Sect. 3.5, and an effective instrument model in Eq. (46), and we seek to constrain these models using the data summarized in Sect. 5. Let us for convenience denote the combined set of all free parameters by ω, such that $\omega \equiv \{g, \Delta_{bp}, \boldsymbol{n}_{corr}, \boldsymbol{a}_i, \beta_i, C_\ell, \ldots\}$. In BEYONDPLANCK, we choose to work within the well-established Bayesian framework, and as such, our main goal is to estimate the posterior distribution, $P(\omega \mid \boldsymbol{d})$, where $\boldsymbol{d}$ denotes all available data, both in the form of time-ordered LFI observations and pre-pixelized external sky maps.

Clearly, this distribution involves billions of non-Gaussian and highly correlated parameters. Figure 9 is an informal attempt to visualize some of the main degeneracies of this distribution. Thick arrows indicate particularly strong correlations, while thin arrows indicate weaker ones. This chart is just intended to be a rough illustration, based on our practical experience, rather than a formal posterior exploration, and so it is obviously incomplete. Still, it may serve as useful reminder for new readers about how individual parameters affect other parts of the system. To consider one specific example, the gain has a direct and strong impact on both the CMB and foreground maps by virtue of multiplying the TOD, and this impact goes both ways; if the current CMB or foreground parameters are biased, then the estimated gains will also be biased. The same observations also hold with respect to the correlated noise and bandpasses, although at a lower level. On the other hand, the gains are only weakly dependent on the monopoles or sidelobes. The sidelobes do affect the CMB dipole, however, which is a critically important component for the gain estimation, and so there is a second-order dependency. Similar observations hold for most other parameters; the distribution is tightly integrated, and each parameter affects a wide range of the full model, either directly or indirectly. This integrated nature of the full posterior distribution emphasizes the importance of global end-to-end analysis with full propagation of uncertainties, as implemented in the following.

To start our formal exploration of this full posterior distribution, we write down Bayes' theorem,

$$P(\omega \mid \boldsymbol{d}) = \frac{P(\boldsymbol{d} \mid \omega)P(\omega)}{P(\boldsymbol{d})} \propto \mathcal{L}(\omega)P(\omega), \tag{60}$$

where $P(\boldsymbol{d} \mid \omega) \equiv \mathcal{L}(\omega)$ is called the likelihood; $P(\omega)$ is called the prior; and $P(\boldsymbol{d})$ is a normalization factor usually referred to as the "evidence". By virtue of being independent of ω, the evidence is irrelevant for parameter estimation purposes, and we ignore it in the current work, although we note that it is important for model selection applications.

For a one-, two-, or three-dimensional parametric model, the simplest way to numerically evaluate the posterior distribution is often to compute the right-hand side of Eq. (60) over some grid in ω. However, this approach quickly becomes computationally expensive in higher-dimensional parameter spaces, since the number of grid points grows exponentially with the number of parameters. For models with more than three parameters,
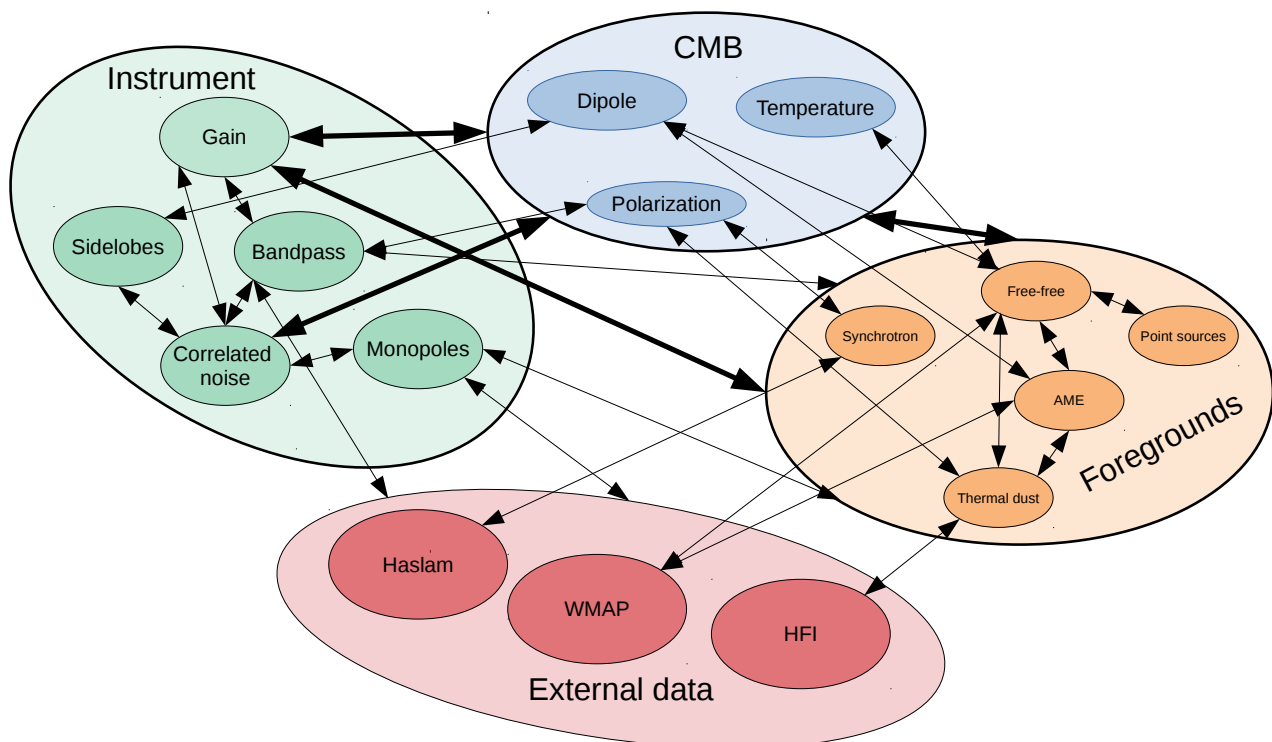
Fig. 9. Schematic overview of the primary parameters and external data sets considered in the current BEYONDPLANCK analysis and their inter-dependencies. This chart is intended to visualize the deeply integrated nature of a modern CMB analysis problem; changing any one of these parameter can lead to significant changes in a wide range of other parameters, and tracing these joint uncertainties is critically important for high-precision experiments.

it is common practice to resort to Markov Chain Monte Carlo (MCMC) sampling techniques rather than grid techniques. The main advantage of these techniques is that computing resources are mostly spent on exploring the peak of the posterior, which is the region in parameter space that actually matters for final parameter estimates. In contrast, gridding techniques spend most of their time evaluating probability densities that are statistically equivalent to zero. In this section, we will briefly review three particularly important examples of such MCMC sampling techniques, as they play a fundamental role in the BEYONDPLANCK pipeline.

6.1. Metropolis sampling

By far the most commonly applied, and widely known, MCMC algorithm is the Metropolis sampler (Metropolis et al. 1953). Let ω_i denote the ith sample in a Markov chain,[12] and $T(\omega_{i+1} \mid \omega_i)$ be a stochastic transition probability density for ω_{i+1} that depends on ω_i, but not on earlier states. Assume further that T is symmetric, such that $T(\omega_{i+1} \mid \omega_i) = T(\omega_i \mid \omega_{i+1})$. The most typical example of such a transition rule is a Gaussian distribution with mean equal to ω_i and with some predefined standard deviation (or "step size"), σ.

With these definitions, the Metropolis sampling algorithm can be summarized in terms of the following steps:

1. Initialize the chain at some arbitrary parameter set, ω_0.

2. Draw a random proposal[13] for the next sample based on the transition rule, i.e., $\omega_{i+1} \leftarrow T(\omega_{i+1} \mid \omega_i)$.

3. Compute the acceptance probability, a, defined by

$$a = \min\left(1, \frac{P(\omega_{i+1} \mid \boldsymbol{d})}{P(\omega_i \mid \boldsymbol{d})}\right) \tag{61}$$

4. Draw a random number, η, from a uniform distribution, $U[0, 1]$. Accept the proposal if $\eta < a$; otherwise, set $\omega_{i+1} = \omega_i$.

5. Repeat steps 2–4 until convergence.

The critical component in this algorithm is the acceptance rule in Eq. (61). On the one hand, this rule ensures that the chain is systematically pushed toward the posterior maximum by always accepting proposals that are more likely than the previous step. In this sense, the Metropolis sampler can be considered a nonlinear optimization algorithm that performs a random walk in the multidimensional parameter space. However, unlike most standard optimization algorithms, the method also does allow samples with lower probability density than the previous state. In particular, by accepting samples with a probability given by the relative posterior ratio of the two samples, one can show that the time spent at a given differential parameter volume is proportional to the underlying distribution density at that state. Thus, the multidimensional histogram of MC samples produced with this algorithm converges to $P(\omega \mid \boldsymbol{d})$ in the limit of an infinite number of samples.

---

[12] A Markov chain is a stochastic sequence of parameter states, $\{\omega_i\}$, in which ω_i only depends on ω_{i-1}, but not earlier states.

[13] The symbol "←" indicates setting the symbol on the left-hand side equal to a sample drawn from the distribution on the right-hand side.

6.2. Metropolis-Hastings sampling

We note that there is no reference to the proposal distribution T in the Metropolis acceptance probability as defined by Eq. (61). This is because we have explicitly assumed that T is symmetric. If we were to choose an asymmetric transition distribution, this equation would no longer hold, as proposals within the heavier tail would be systematically proposed more often than proposals within the lighter tail, and this would overall bias the chain.

For asymmetric transition distributions, we need to replace Eq. (61) with

$$a = \min\left(1, \frac{P(\omega_{i+1} \mid \boldsymbol{d})}{P(\omega_i \mid \boldsymbol{d})} \frac{T(\omega_i \mid \omega_{i+1})}{T(\omega_{i+1} \mid \omega_i)}\right), \qquad (62)$$

as shown by Hastings (1970). Without further changes, the algorithm in Sect. 6.1 is then valid for arbitrary distributions T, and the algorithm is in this case called Metropolis-Hastings sampling.

6.3. Gibbs sampling

While the Metropolis and Metropolis-Hastings samplers are prevalent in modern Bayesian analysis applications, they do require a well-tuned proposal distribution T in order to be computationally efficient. If the step size is too small, it takes a prohibitive number of proposals to move from one tail of the distribution to another, whereas if the step size is too large, then all proposals are in effect rejected by the acceptance rate. The latter issue is particularly critical in high-dimensional spaces, and for this reason Metropolis-type samplers are usually only applied to moderately high-dimensional parameter spaces, for instance 20 or 50 dimensions. For millions of dimensions, traditional non-guided Metropolis sampling becomes entirely intractable.

In order to achieve acceptable efficiencies in such cases, one must typically exploit additional information within the transition probability. For instance, the Hamiltonian sampler exploits the derivative of the posterior distribution to establish proposals (e.g, Liu 2008), while the Langevin Monte Carlo algorithm can also incorporate second-order derivatives (Girolami & Calderhead 2011).

Another effective way of improving computing efficiency is to decompose complicated high-dimensional joint distributions into its various conditional distributions, a process that is called Gibbs sampling (Geman & Geman 1984). In this case, one exploits the shape of the posterior distribution itself to make proposals, but only in the form of conditionals. To illustrate the process, let us for the sake of notational simplicity consider a two-dimensional distribution $P(\alpha, \beta)$. In that case, the Gibbs sampling transition probability takes the form

$$T_{\text{Gibbs}}(\alpha_{i+1}, \beta_{i+1} \mid \alpha_i, \beta_i) = P(\alpha_{i+1} \mid \beta_i) \, \delta(\beta_{i+1} - \beta_i), \qquad (63)$$

where $\delta(x)$ denotes the Dirac delta function, which vanishes for $x \neq 0$, but has a unit integral. The δ function in Eq. (63) ensures that $\beta_{i+1} = \beta_i$, i.e., that β is kept fixed.

This is an asymmetric proposal distribution, and the corresponding acceptance probability is therefore given by inserting

Eq. (63) into the Metropolis-Hastings rule in Eq. (62):

$$a = \frac{P(\alpha_{i+1}, \beta_{i+1})}{P(\alpha_i, \beta_i)} \frac{T_{\text{Gibbs}}(\omega_i \mid \omega_{i+1})}{T_{\text{Gibbs}}(\omega_{i+1} \mid \omega_i)} \qquad (64)$$

$$= \frac{P(\alpha_{i+1}, \beta_{i+1})}{P(\alpha_i, \beta_i)} \frac{P(\alpha_i \mid \beta_{i+1}) \, \delta(\beta_i - \beta_{i+1})}{P(\alpha_{i+1} \mid \beta_i) \, \delta(\beta_{i+1} - \beta_i)} \qquad (65)$$

$$= \frac{P(\alpha_{i+1}, \beta_i)}{P(\alpha_i, \beta_i)} \frac{P(\alpha_i \mid \beta_i)}{P(\alpha_{i+1} \mid \beta_i)} \qquad \beta_{i+1} = \beta_i \qquad (66)$$

$$= \frac{P(\alpha_{i+1} \mid \beta_i) \, P(\beta_i)}{P(\alpha_i \mid \beta_i) \, P(\beta_i)} \frac{P(\alpha_i \mid \beta_i)}{P(\alpha_{i+1} \mid \beta_i)} \qquad P(\alpha, \beta) = P(\alpha \mid \beta) P(\beta) \qquad (67)$$

$$= 1, \qquad (68)$$

where we have used the definitions of both conditional[14] and marginal[15] distributions; the equations marked in gray indicate which relation is used in a given step. From this calculation, we see that when proposing samples from a conditional distribution within a larger global joint distribution, the Metropolis-Hastings acceptance rate is always unity. Consequently, there is no need to even compute it, and this can save large amounts of computing time for complex distributions. However, one does of course have to propose from the proper conditional distribution for this result to hold.

It is also important to note that only a sub-space of the full distribution is explored within a single Markov step with this algorithm. To explore the full distribution, it is therefore necessary to iterate through all possible conditionals, and allow changes in all dimensions. Note, however, that there are no restrictions in terms of order in which the conditionals are explored. Any combination of sampling steps is valid, as long as all dimensions are explored sufficiently to reach convergence.

The Gibbs sampling algorithm forms the main computational framework of the BeyondPlanck analysis pipeline. However, within this larger framework a large variety of different samplers are employed in order to explore the various conditionals. For convenience, Appendix A provides a summary of the most important samplers, while specific implementation details are deferred to the individual companion papers.

We conclude this section by noting that Gibbs sampling only works well for uncorrelated and weakly degenerate distributions. For strongly degenerate distributions, the number of Gibbs iterations required to explore the full distribution becomes prohibitive, as the algorithm only allows parameter moves parallel to coordinate axes. In such cases, it is usually necessary either to reparametrize the model in terms of less degenerate parameters; or, if possible, sample the degenerate parameters jointly. A commonly used solution in that respect is to exploit the identity $P(\alpha, \beta) = P(\alpha \mid \beta) P(\beta)$, which tells us that a joint sample may be established by first sampling β from its *marginal* distribution, and then α from the corresponding conditional distribution as before. The marginal sampling step ensures the Markov chain correlation length becomes unity. This method is used in several places in the BeyondPlanck Gibbs chain, for instance for the combination of instrumental gain and correlated noise (Gjerløw et al. 2022), and for the combination of astrophysical component amplitudes and spectral parameters in intensity (Andersen et al. 2022), both of which are internally strongly correlated.

7. Global model specification

The previous section provides a very general overview of our analysis strategy. In this section, we provide a detailed specifica-

---

[14] Definition of a conditional distribution: $P(\alpha \mid \beta) \equiv P(\alpha, \beta)/P(\beta)$
[15] Definition of a marginal distribution: $P(\beta) \equiv \int P(\alpha, \beta) \, d\alpha$

tion of the parametric BeyondPlanck model that is appropriate for actual implementation and processing.

7.1. Global parametric model

Following the general model introduced in Sects. 3–4, we adopt the following time-ordered data model,

$$
d_{j,t} = g_{j,t} \mathsf{P}_{tp,j} \left[\mathsf{B}^{\mathrm{symm}}_{pp',j} \sum_c \mathsf{M}_{cj}(\beta_{p'}, \Delta^j_{\mathrm{bp}}) a^c_{p'} + \mathsf{B}^{4\pi}_{j,t} s^{\mathrm{orb}}_j + \mathsf{B}^{\mathrm{asymm}}_{j,t} s^{\mathrm{fsl}}_t \right] +
$$
$$
+ a_{1\mathrm{Hz}} s^{1\mathrm{Hz}}_j + n^{\mathrm{corr}}_{j,t} + n^{\mathrm{w}}_{j,t}. \tag{69}
$$

Here j represents a radiometer label, t indicates a single time sample, p denotes a single pixel on the sky, and c represents one single astrophysical signal component. Further,

- $d_{j,t}$ denotes the measured data value in units of V;
- $g_{j,t}$ denotes the instrumental gain in units of $\mathrm{V\,K^{-1}_{cmb}}$;
- $\mathsf{P}_{tp,j}$ is the $N_{\mathrm{TOD}} \times 3N_{\mathrm{pix}}$ pointing matrix defined in Eq. (35), where ψ is the polarization angle of the respective detector with respect to the local meridian;
- B_j denotes the beam convolution in Eq. (45) in the form of a matrix operator, either for the (symmetric) main beam, the (asymmetric) far sidelobes, or the full 4π beam; note that for computational efficiency reasons we only take into account beam asymmetries for the sidelobes and orbital dipole in this paper;
- $\mathsf{M}_{cj}(\beta_p, \Delta_{\mathrm{bp}})$ denotes element (c, j) of an $N_{\mathrm{comp}} \times N_{\mathrm{comp}}$ mixing matrix defined in Eq. (42), describing the amplitude of component c as seen by radiometer j relative to some reference frequency j_0 when assuming some set of bandpass correction parameters Δ_{bp};
- a^c_p is the amplitude of component c in pixel p, measured at the same reference frequency as the mixing matrix M, and expressed in brightness temperature units;
- s^{orb}_j is the orbital CMB dipole signal in units of $\mathrm{K_{cmb}}$, including relativistic quadrupole corrections;
- $s^{1\mathrm{Hz}}_j$ represents the electronic 1 Hz spike correction;
- s^{fsl}_j denotes the contribution from far sidelobes, also in units of $\mathrm{K_{cmb}}$;
- $n^{\mathrm{corr}}_{j,t}$ denotes correlated instrumental noise, as defined by Eqs. (50) and (52); and
- $n^{\mathrm{w}}_{j,t}$ is uncorrelated (white) instrumental noise.

For notational convenience, we also define

$$
s^{\mathrm{sky}}_j = \sum_c \mathsf{M}_{cj}(\beta, \Delta^j_{\mathrm{bp}}) a^c \tag{70}
$$

to be the sky model for detector j without beam convolution, but integrated over the bandpass.

For external data sets, which are defined in terms of pre-pixelized maps, this model simplifies to

$$
d_{j,p} = g_j \mathsf{B}^{\mathrm{symm}}_{pp',j} \sum_c \mathsf{M}_{cj}(\beta_{p'}, \Delta^j_{\mathrm{bp}}) a^c_{p'} + n^{\mathrm{w}}_{j,p}, \tag{71}
$$

which is identical to the `Commander2` data model considered by Seljebotn et al. (2019).

The free parameters in Eq. (69) are $\{g, \Delta_{\mathrm{bp}}, n_{\mathrm{corr}}, a, \beta\}$. All other quantities are either provided as intrinsic parts of the original data sets (e.g., the pointing matrix, the beam profile, and the orbital dipole), or given as a deterministic function of already

available parameters (e.g., the mixing matrix and the far side-lobe component). The only exception to this is the white noise component, which is neither fitted explicitly nor given by prior knowledge, but is simply left as a stochastic uncertainty in the model.

In addition to the parameters defined by Eq. (69), our model includes a handful of parameters that describe the statistical properties of the stochastic random fields included in the model. Specifically, we associate each of the astrophysical component maps a^c with a covariance matrix S^c, which in most cases is assumed to be statistically isotropic. Expanding $a^c_p = \sum_{\ell m} a^c_{\ell m} Y_\ell(p)$ into spherical harmonics, this matrix may then be written as

$$
S^c_{\ell m, \ell' m'} \equiv \left\langle a^c_{\ell m} a^{c,*}_{\ell' m'} \right\rangle = C^c_\ell \delta_{\ell \ell'} \delta_{mm'}, \tag{72}
$$

where C^c_ℓ denotes the angular power spectrum of component c. (Here we have for notational simplicity assumed that the component in question is unpolarized; the appropriate generalization to polarization is straightforward, and will be discussed in Sect. 8.3.8.) This power spectrum is a stochastic parameter on the same footing as a or β, and may as such be included in the model fitted to the data. Alternatively, the power spectrum may be modelled in terms some smaller set of parameters, ξ, through some deterministic function $C_\ell(\xi)$, in which case ξ is the set of stochastic parameters included in the model. For notational simplicity, we will only include the power spectrum in the various posterior distributions below, but we note that C_ℓ may be replaced with ξ without loss of generality.

Finally, similar considerations hold for the two noise components. First, the white noise component is assumed to be piece-wise stationary and Gaussian distributed with vanishing mean and a covariance matrix equal to $N^{\mathrm{w}}_{tt'} = \sigma_0^2 \delta_{tt'}$. In the following, we will assume the stationary period to be given by PIDs, and σ_0 will be fitted independently for each period. Second, the correlated noise component is also assumed to be piece-wise stationary and Gaussian distributed with zero mean, but with a nontrivial covariance structure in time. Traditionally, this has been described by the $1/f$ model in Eq. (52) for *Planck* LFI, but as discussed by Ihle et al. (2022), the BeyondPlanck processing has identified an additional component at intermediate frequencies in the 30 and 44 GHz channels between 0.01 and 1 Hz. In this current analysis, this is modelled in terms of a log-normal contribution, and the full noise power spectral density (PSD) reads

$$
P(f) = \sigma_0^2 \left[1 + \left(\frac{f}{f_{\mathrm{knee}}} \right)^\alpha \right] + A_{\mathrm{p}} \exp\left[-\frac{1}{2} \left(\frac{\log_{10} f - \log_{10} f_{\mathrm{p}}}{\sigma_{\mathrm{dex}}} \right)^2 \right], \tag{73}
$$

where A_{p} is the peak amplitude of the additional term, f_{p} is the peak frequency, and σ_{dex} is the width; for now, the latter two are fixed at fiducial values, but this may change in future analyses. With this approximation, the total noise PSD for the 30 and 44 GHz channels is modelled in terms of a total of four free parameters, namely the white noise level σ_0, a knee frequency f_{knee}, a low-frequency spectral slope α, and the amplitude of the log-normal contribution. We denote the spectral noise parameters collectively as ξ_n. For the 70 GHz channel, the log-normal term is omitted.

So far, the discussion has been kept general, aiming to fit all necessary parameters into one succinct and computationally convenient framework. However, at this point it is useful to remind ourselves that one of the astrophysical component carries

particular importance in this work, namely the CMB. This component is accommodated in Eq. (69) in the form of $\boldsymbol{a} = \boldsymbol{a}^{\mathrm{cmb}}$ and $\mathsf{M}^{\mathrm{cmb}} = 1$ in thermodynamic temperature units, with an angular CMB power spectrum defined as $C_\ell = \langle |a^{\mathrm{cmb}}|^2 \rangle$. Computing $P(C_\ell \mid \boldsymbol{d})$ (or $P(\xi \mid \boldsymbol{d})$, where ξ represents a set of cosmological parameters) properly marginalized over all relevant astrophysical and instrumental parameters, is the single most important scientific goal of the current algorithm.

In summary, the total set of free stochastic parameters adopted in this work is $\omega \equiv \{\boldsymbol{g}, \Delta_{\mathrm{bp}}, \boldsymbol{n}_{\mathrm{corr}}, \xi_n, \boldsymbol{a}, \beta, C_\ell\}$, where each symbol collectively represents a larger set of individual parameters, typically depending on radiometer, time, pixel, or component. For notational convenience, we will usually suppress individual indices, unless explicitly required for context. Likewise, we also note that in most cases, each of the parameters and quantities discussed above is associated with its own technicalities, which have been omitted in the above discussion. Such details will be provided in dedicated companion papers, with appropriate references given where appropriate. Finally, a full specification of the astrophysical component model considered in this analysis is provided in Sect. 3.5.

7.2. Deterministic quantities

Before considering the posterior distribution $P(\omega \mid \boldsymbol{d})$, it is useful to introduce some extra notation regarding various quantities that may either be derived deterministically from ancillary information or from other parameters in our model. These quantities are not stochastic variables in their own right within our model, and are as such not associated with independent degrees of freedom, but they are simply computationally convenient auxiliary variables.

7.2.1. Frequency maps and leakage corrections

Our first derived quantity are frequency maps, which we will denote $\boldsymbol{m}_\nu$. In our framework, frequency maps are not stochastic parameters, but instead they represent a deterministic compression of the full data set from time-ordered data into sky pixels, conditioning on any parameter or quantity that is not stationary, such as the gain, correlated noise, and the orbital dipole.

In order to construct frequency sky maps, we start by computing the following *residual calibrated TOD* for each detector,

$$r_{j,t}^{(0)} = \frac{d_{j,t} - n_{j,t}^{\mathrm{corr}} - s_{j,t}^{\mathrm{1Hz}}}{g_{t,j}} - \left(s_{j,t}^{\mathrm{orb}} + s_{j,t}^{\mathrm{fsl}} \right). \tag{74}$$

According to Eq. (69), $r_{j,t}$ now contains only stationary sky signal and white noise, given the current estimates of all other parameters.

In principle, $r_{j,t}^{(0)}$ could be individually binned into a pixelized map for each radiometer j given the pointing information in P_{tp}^j. Unfortunately, due to the poor cross-linking properties of the *Planck* scanning strategy, it is very difficult to solve for three independent Stokes parameters per pixel based on only information from a single radiometer. In practice, four radiometers are required in order to obtain well-conditioned maps with robust statistical properties. In the following we will mostly consider full-frequency maps, combining all four, six and twelve LFI radiometers into respective 30, 44 and 70 GHz maps.

Unfortunately, combining multiple radiometers into a single pixelized map carries its own complications. Since each radiometer has its own separate bandpass and beam profile, the observed sky will appear slightly different for each radiometer. However, when creating a single joint frequency map, only one single value per pixel is allowed. Any deviation from this mean value will be interpreted within the data model as either correlated or white noise, and consequently be filtered according to ξ_n or down-weighted according to σ_0 during processing, or be split among the various other free parameters, including the CMB map. This typically gives rise to artifacts proportional to the total signal amplitude, but modulated by the scanning strategy of the instrument. These effects are often referred to as bandpass or beam mismatch contamination, respectively. Informally speaking, this is also often referred to as "temperature-to-polarization leakage," in recognition of the fact that the temperature signal is orders of magnitude brighter than the polarization signal, and therefore even a small bandpass or beam difference can induce a large spurious polarization signal.

Fortunately, with the model described above, which includes a full and explicit model of the astrophysical sky signal as part of its parameter space, it is possible to correct for such leakages. As described by Svalheim et al. (2022a), we adopt a very straightforward approach by simply subtracting a correction from each detector TOD, prior to map binning, of the form

$$\delta s_{j,t}^{\mathrm{leak}} = \mathsf{P}_{tp}^j \mathsf{B}_{pp'}^j \left(s_{jp'}^{\mathrm{sky}} - \left\langle s_{jp'}^{\mathrm{sky}} \right\rangle \right), \tag{75}$$

where s_j^{sky} denotes the sky model as seen by detector j, accounting for separate bandpass profiles, and angle brackets indicate an average over all radiometers included in the map. For computational efficiency reasons, the beam is here approximated as azimuthally symmetric, which allows the average over detector indicated by brackets in the equation to be performed pixel-by-pixel. However, since δs^{leak} is already a difference between two very similar sky models with slightly different bandpasses, the error due to asymmetric beams is a second-order effect, and completely negligible compared to instrumental noise.

In order to correct for bandpass and beam leakage effects, we modify Eq. (74) accordingly,

$$r_{j,t} = \frac{d_{j,t} - n_{j,t}^{\mathrm{corr}} - s_{j,t}^{\mathrm{1Hz}}}{g_{t,j}} - \left(s_{j,t}^{\mathrm{orb}} + s_{j,t}^{\mathrm{fsl}} + \delta s_{j,t}^{\mathrm{leak}} \right). \tag{76}$$

After applying this correction, all detector TODs exhibit the same net sky signal, up to the accuracy of the instrument model, which itself is sampled over within the Markov chain. At the same time, the mean signal is not affected by this correction, independent of the accuracy of the instrument model, as $\left\langle \delta s^{\mathrm{leak}} \right\rangle = 0$ when averaged over all detectors.

With calibrated and cleaned TOD ready at hand which contain exclusively equalized signal and white noise for each detector, optimal mapmaking is performed simply by solving the corresponding normal equations pixel-by-pixel (see, e.g., Appendix A.2 or Ashdown et al. 2007a),

$$\left(\sum_{j \in \nu} \mathsf{P}_j^t (\mathsf{N}_j^{\mathrm{w}})^{-1} \mathsf{P}_j \right) \boldsymbol{m}_\nu = \sum_j \mathsf{P}_j^t (\mathsf{N}_j^{\mathrm{w}})^{-1} \boldsymbol{d}_j. \tag{77}$$

For our pointing matrix definition and white noise covariance matrix, this equation may for a single pixel be written explicitly as

$$\begin{bmatrix} \sum \frac{1}{\sigma_{0,j}^2} & \sum \frac{\cos 2\psi_{j,t}}{\sigma_{0,j}^2} & \sum \frac{\sin 2\psi_{j,t}}{\sigma_{0,j}^2} \\ \sum \frac{\cos 2\psi_{j,t}}{\sigma_{0,j}^2} & \sum \frac{\cos^2 2\psi_{j,t}}{\sigma_{0,j}^2} & \sum \frac{\cos 2\psi_{j,t} \sin 2\psi_{j,t}}{\sigma_{0,j}^2} \\ \sum \frac{\sin 2\psi_{j,t}}{\sigma_{0,j}^2} & \sum \frac{\sin 2\psi_{j,t} \cos 2\psi_{j,t}}{\sigma_{0,j}^2} & \sum \frac{\sin^2 2\psi_{j,t}}{\sigma_{0,j}^2} \end{bmatrix} \begin{bmatrix} T \\ Q \\ U \end{bmatrix} = \begin{bmatrix} \sum \frac{d_j}{\sigma_{0,j}^2} \\ \sum \frac{d_j \cos 2\psi_{j,t}}{\sigma_{0,j}^2} \\ \sum \frac{d_j \sin 2\psi_{j,t}}{\sigma_{0,j}^2} \end{bmatrix}, \tag{78}$$

where the sums run over both detector j and all time samples t that point toward pixel p. The associated inverse white noise pixel-pixel covariance matrix, $N^{-1}_{pp'}$, is given simply by the inverse of the matrix on the left-hand side of Eq. (77).

It is important to note that the frequency maps defined by Eq. (77) have a slightly different statistical interpretation than those delivered by earlier CMB analysis pipelines, for instance from the *Planck* DPCs or *WMAP* science team. With our definition, $\boldsymbol{m}_\nu$ represents one possible realization of the frequency sky map *assuming perfect knowledge about the correlated noise, gain, bandpass, leakage effects*, among others; the only unmitigated stochastic quantity is instrumental white noise. The uncertainties due to all those other effects are instead accounted for by the fact that we produce an entire ensemble, $\boldsymbol{m}^i_\nu$, each with different combinations of systematic effects. For full error propagation, it is thus important to analyze the full set of available frequency maps, not just one single realization. In contrast, traditional frequency maps represent an approximation to the overall maximum likelihood solution, and error propagation can only be achieved through analysis of end-to-end simulations.

We conclude this section by emphasizing that $\boldsymbol{s}^{\mathrm{leak}}$ as defined above is not a separate stochastic parameter within our model. It neither increases the total uncertainty in the system, nor does it induce new parameter degeneracies; it is a simple deterministic correction that removes a known bias in co-added frequency maps.

7.2.2. Spurious leakage maps

The correction for spurious leakages from bandpass and beam mismatch defined in Eq. (75) is only exact to the extent that the assumed bandpass and beam profiles are accurate. In order to monitor the efficiency of the leakage correction, it is therefore useful to establish a dedicated goodness-of-fit statistic for this correction. For this purpose, we adopt the "spurious map" approach pioneered by Page et al. (2007), and later adapted within various pipelines, including Planck Collaboration II (2020) and Planck Collaboration Int. LVII (2020).

The central idea underlying this approach is to modify the pointing matrix to allow for a set of additional temperature maps, each corresponding to the difference between the temperature sky as seen by radiometer j and the temperature sky as seen by the mean of the detectors at that frequency. However, to prevent the linear mapmaking equation from becoming degenerate, one can at most include $N_{\mathrm{det}} - 1$ such spurious maps for a configuration involving N_{det} detectors. Thus, we generalize the pointing model for a single observation in terms of the Stokes parameters and spurious maps as follows,

$$s_j = T + Q \cos 2\psi_j + U \sin 2\psi_j + \sum_{i=1}^{N_{\mathrm{det}}-1} S_i \, \delta_{ij}. \tag{79}$$

Given this definition, the mapmaking equation in Eq. (77) generalizes straightforwardly, and for the special case of three detectors, the contribution of a single sample from detector j takes the schematic form

$$\begin{bmatrix} 1 & \cos 2\psi & \sin 2\psi & \delta_{1j} & \delta_{2,j} \\ \cos 2\psi & \cos^2 2\psi & \cos 2\psi \sin 2\psi & \cos 2\psi \, \delta_{1j} & \cos 2\psi \, \delta_{2,j} \\ \sin 2\psi & \sin 2\psi \cos 2\psi & \sin^2 2\psi & \sin 2\psi \, \delta_{1j} & \cos 2\psi \sin 2\psi \, \delta_{2,j} \\ \delta_{1j} & \cos 2\psi \, \delta_{1i} & \sin 2\psi \, \delta_{1j} & \delta_{1j} & 0 \\ \delta_{2j} & \cos 2\psi \, \delta_{2i} & \sin 2\psi \, \delta_{2j} & 0 & \delta_{2j} \end{bmatrix} \begin{bmatrix} T \\ Q \\ U \\ S_1 \\ S_2 \end{bmatrix} = \begin{bmatrix} d \\ d \cos 2\psi \\ d \sin 2\psi \\ d \, \delta_{1j} \\ d \, \delta_{2j} \end{bmatrix}. \tag{80}$$

For *WMAP*, it is in fact possible to solve this equation pixel-by-pixel, due to the highly interconnected *WMAP* scanning strategy (Page et al. 2007). The resulting Stokes parameter maps solved jointly with S were therefore released as primary mission

products (Bennett et al. 2013). Unfortunately, the same is not possible for *Planck* without inducing an unacceptable increase in the overall noise level, as the coupling matrix in Eq. (80) is poorly conditioned over most of the sky. However, the resulting S maps are still very useful for monitoring purposes, and we will in fact use these maps to optimize a small number of bandpass parameters, for which a high level of noise is of no concern; see Svalheim et al. (2022a) and Sects. 4.2 and 8.3.4 for further details.

7.2.3. Orbital dipole

The third derived quantity we will need is the orbital dipole, $s^{\mathrm{orb}}_{j,t}$. Including a relativistic quadrupole correction, this has a closed form as given by

$$s^{\mathrm{orb}}_{j,t} = \frac{T_{\mathrm{CMB}}}{c} \left(\boldsymbol{v}_{\mathrm{sat}} \cdot \hat{\boldsymbol{n}}_{j,t} + q(\boldsymbol{v}_{\mathrm{sat}} \cdot \hat{\boldsymbol{n}}_{j,t})^2 \right), \tag{81}$$

where

$$q = \frac{x(\mathrm{e}^{2x} + 1)}{\mathrm{e}^{2x} - 1}; \quad x = \frac{h\nu}{2kT_{\mathrm{CMB}}} \tag{82}$$

is the frequency dependency of the relativistic quadrupole term. The CMB temperature is in our analysis fixed to $T_{\mathrm{CMB}} = 2.7255$ K, following Fixsen (2009). Finally, c is the speed of light, h is Planck's constant, k is Boltzmann's constant, $\boldsymbol{v}_{\mathrm{sat}}$ is the satellite velocity, and $\hat{\boldsymbol{n}}_{j,t}$ is the pointing vector of detector j at time t. The satellite velocity is known with an absolute precision better than $1 \, \mathrm{cm \, s^{-1}}$ (Planck Collaboration Int. LVII 2020). An efficient convolution algorithm for this component that takes into account the full 4π beam is described by Galloway et al. (2022b).

It is important to note the critical role of this particular signal term. Depending only on the velocity of the satellite (which is known to exceedingly high precision) and the CMB temperature (which is known to a precision of 0.02 %; Fixsen 2009), it provides the best absolute calibration source in microwave astronomy, if not all of astronomy. For BeyondPlanck, as for both *Planck* and *WMAP*, this signal is therefore used to determine the overall absolute calibration of the entire data set.

7.2.4. Far sidelobe corrections

The last derived quantity we will need at this stage is the far sidelobe correction, s^{fsl}, as defined in Sect. 4.3. As shown by Planck Collaboration IV (2016), the *Planck* LFI optics have several significant sidelobes at large angles from the optical axis. The most important is due to spillover around the main reflector, and located about 85° from the main beam. The second most important is due to spillover around the secondary reflector, and located about 20° from the main beam. To account for these, we convolve the parametric sky model with the (near-)4π beam profile, B, of each radiometer (regions closer than 5° from the main beam are excluded),

$$s^{\mathrm{fsl}}_{j,t} = \int_{4\pi} [\mathsf{R}(\Omega_t)\mathsf{B}(\Omega)] \, \boldsymbol{s}^{\mathrm{sky}}_j(\Omega) \, \mathrm{d}\Omega, \tag{83}$$

where $\mathsf{R}(\Omega_t)$ is a rotation matrix that rotates the beam as specified by the satellite pointing at time t. To evaluate this integral, we employ an algorithm that is algebraically equivalent to the `conviqt` approach described by Prézeau & Reinecke (2010), but implemented in terms of spin harmonics, as described by Galloway et al. (2022b).

We stress, however, that uncertainties in the far-sidelobe model are not yet accounted for, and this represents a significant model uncertainty in the current analysis. Generalizing the parametric model in Eq. (69) to allow for new beam-related degrees of freedom is an important goal for future algorithm development.

8. The BeyondPlanck Gibbs sampler

8.1. Global posterior distribution

Given the global parametric model defined in Sect. 7.1, and the ancillary quantities summarized in Sect. 7.2, we are now finally ready to consider the full global BeyondPlanck posterior distribution, $P(\omega \mid d)$, and describe the computational algorithms required to map it out. In practice, this entails writing down explicit expressions for the likelihood and priors in Eq. (60), as well as specifying an explicit Gibbs chain that is able to explore the posterior distribution efficiently.

Starting with the likelihood, $\mathcal{L}(\omega)$, we first note that the data model defined in Eqs. (69)–(71) is given as a linear sum of various components, all of which are specified precisely in terms of our free parameters ω. This applies even to the correlated noise component, n^{corr}, which for the purposes of the likelihood is fully equivalent to any of the other physical components. As such, we may symbolically write $d = s^{\mathrm{tot}}(\omega) + n^{\mathrm{w}}$, where $s^{\mathrm{tot}}(\omega)$ is the sum of all model components in Eq. (69), whether they have a cosmological, astrophysical or instrumental origin. With this notation, we immediately see that

$$P(d \mid \omega) \propto P(n^{\mathrm{w}} \mid \omega) \propto \exp\left(-\frac{1}{2}\left(\frac{d - s^{\mathrm{tot}}(\omega)}{\sigma_0}\right)^2\right), \qquad (84)$$

since $n^{\mathrm{w}} = d - s^{\mathrm{tot}}(\omega)$, $P(n^{\mathrm{w}}) \propto N(0, \sigma^2)$, and s^{tot} is deterministically given by ω.

Next, the prior $P(\omega)$ should encapsulate all our prior knowledge about any of the model parameters. For instance, we may use this term to introduce information regarding the instrumental gain from temperature measurements of the 4 K load onboard the *Planck* satellite during the calibration stage; or we can use it to impose prior knowledge regarding the CIB zero-level amplitude at each frequency during component separation; or we may introduce a prior on the Hubble constant during cosmological parameter estimation; or we may use it to regularize posterior volume effects through the application of a Jeffreys ignorance prior (Jeffreys 1946). A detailed breakdown of the priors used in this particular analysis will be presented in association with the respective steps.

8.2. Overview of Gibbs chain

As already discussed, the posterior distribution defined by Eq. 60 involves millions of tightly correlated and non-Gaussian parameters, and it is clearly unfeasible to optimize or sample from it directly. We therefore resort to the Gibbs sampling algorithm described in Sect. 6.3: We compute a Markov chain of correlated samples by initializing on some arbitrary parameter combination, ω_0, and then iteratively sample from each conditional distribution from the full distribution. In practice, most runs are initialized on the outcome of an earlier analysis, in order to save burn-in time.

The BeyondPlanck Gibbs chain may be written schematically as follows,

$$g \leftarrow P(g \quad \mid d, \qquad \xi_n, a^{1\mathrm{Hz}}, \Delta_{\mathrm{bp}}, a, \beta, C_\ell) \qquad (85)$$

$$n_{\mathrm{corr}} \leftarrow P(n_{\mathrm{corr}} \mid d, g, \qquad \xi_n, a^{1\mathrm{Hz}}, \Delta_{\mathrm{bp}}, a, \beta, C_\ell) \qquad (86)$$

$$\xi_n \leftarrow P(\xi_n \quad \mid d, g, n_{\mathrm{corr}}, \qquad a^{1\mathrm{Hz}}, \Delta_{\mathrm{bp}}, a, \beta, C_\ell) \qquad (87)$$

$$a^{1\mathrm{Hz}} \leftarrow P(a^{1\mathrm{Hz}} \mid d, g, n_{\mathrm{corr}}, \xi_n, \qquad \Delta_{\mathrm{bp}}, a, \beta, C_\ell) \qquad (88)$$

$$\Delta_{\mathrm{bp}} \leftarrow P(\Delta_{\mathrm{bp}} \mid d, g, n_{\mathrm{corr}}, \xi_n, a^{1\mathrm{Hz}}, \qquad a, \beta, C_\ell) \qquad (89)$$

$$\beta \leftarrow P(\beta \quad \mid d, g, n_{\mathrm{corr}}, \xi_n, a^{1\mathrm{Hz}}, \Delta_{\mathrm{bp}}, \qquad C_\ell) \qquad (90)$$

$$a \leftarrow P(a \quad \mid d, g, n_{\mathrm{corr}}, \xi_n, a^{1\mathrm{Hz}}, \Delta_{\mathrm{bp}}, \quad \beta, C_\ell) \qquad (91)$$

$$C_\ell \leftarrow P(C_\ell \quad \mid d, g, n_{\mathrm{corr}}, \xi_n, a^{1\mathrm{Hz}}, \Delta_{\mathrm{bp}}, a, \beta \quad), \qquad (92)$$

where the conditional variables have been vertically aligned for clarity only. As usual, the symbol $\leftarrow$ means setting the variable on the left-hand side equal to a sample from the distribution on the right-hand side. For convenience, in the following we also define the notation "$\omega \setminus \xi$" to imply the set of parameters in ω except ξ.

Note that the first conditional in this Gibbs chain, $P(g \mid d, \ldots)$ represents a marginal distribution with respect to n_{corr}. As such, g and n_{corr} are in effect sampled jointly in the BeyondPlanck Gibbs chain (Gjerløw et al. 2022; Ihle et al. 2022), using the properties discussed in Sect. 6.3. The reason for this choice is that these two parameters are particularly strongly degenerate, and joint sampling therefore leads to a much shorter overall correlation length than strict Gibbs sampling. This outweighs by far the somewhat higher computational cost per iteration that is required for sampling the gain from its marginal distribution.

The same method is applied when sampling astrophysical component parameters, a and β in the case of intensity maps. In this case, we first sample β marginalized over a, and then a conditionally on β (Andersen et al. 2022). Since a is a set of linear parameters, the integral over a may be computed analytically, as first exploited for CMB component separation purposes in the `Miramare` code (Stompor et al. 2009; Stivoli et al. 2010). For polarization, we still sample β conditionally on a, as described by Svalheim et al. (2022b), because the low-resolution *WMAP* data with full covariance matrix prohibits smoothing to a common angular resolution, as needed for the marginal sampling approach.

We will now describe each of these distributions in turn, with the main goal being to build intuition regarding each distribution. For specific implementational details we refer the interested reader to companion papers.

At this point, we note that if a joint maximum likelihood estimate is required as opposed to a sample set, the same methodology applies as described below, with the exception that one should then maximize each conditional, rather than sample from it. The algorithm then becomes equivalent to a (slow but computationally convenient) steepest descent nonlinear optimizer. In our codes, we have implemented support for both modes of operation.

8.3. Specification of conditional sampling steps

8.3.1. Gain and calibration sampling

We start our review of the various Gibbs sampling steps with the gain, g_t. In this paper, we only summarize the main central steps, and we refer the interested reader to Gjerløw et al. (2022) for full algorithmic details.
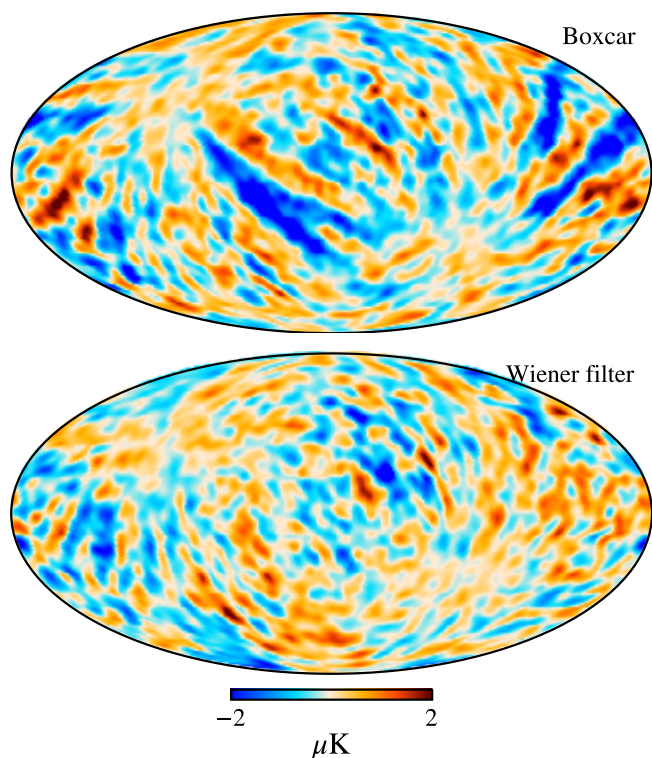
Fig. 10. Random correlated noise Stokes Q Gibbs sample, $\boldsymbol{n}^{\mathrm{corr}}$, for the 44 GHz frequency channel, smoothed to an effective angular resolution of 5° FWHM. The top figure shows the map resulting from a boxcar smoothed gain solution, whereas the bottom figure is the map which results from smoothing the gain solution with a Wiener filter. Reproduced from Gjerløw et al. (2022)

The gain is among the most critical parameters in our model in terms of the overall resulting data quality, and even relative errors at the $O(10^{-4})$ level are highly significant. At the same time, it is also one of the parameters we have the least prior information about, as it is entirely specific for each individual instrument. To estimate the gain robustly, we therefore exploit the following observations: First, we note that the orbital CMB dipole (see Sect. 7.2.3) depends only the satellite velocity, which is known to a precision of 10^{-6} (Godard et al. 2009), and the CMB monopole value, which is known to a precision of 0.02 % (Fixsen 2009). The orbital dipole therefore by far provides the most robust constraints on the mean calibration.[16] However, since the Earth's orbital velocity is $30\,\mathrm{km\,s^{-1}}$ and the CMB monopole is 2.7255 K, the absolute amplitude of the orbital dipole is only 270 μK, which is small compared to typical signal and noise variations. As a result, the orbital dipole is not strong enough to directly determine the gain alone on short time scales.

In contrast, the amplitude of the solar CMB dipole is 3 mK, about ten times brighter than the orbital dipole. Of course, the true solar CMB dipole parameters are unknown, and must be estimated jointly with everything else; but we do know that all detectors observe the *same* solar dipole. We also know that its frequency spectrum is given by a perfect blackbody with temperature T_{CMB}. Together, these two facts provide a strong handle on relative gain variations, both between detectors and in time.

---

[16] The term "calibration" refers in this paper to the time average of the gain.
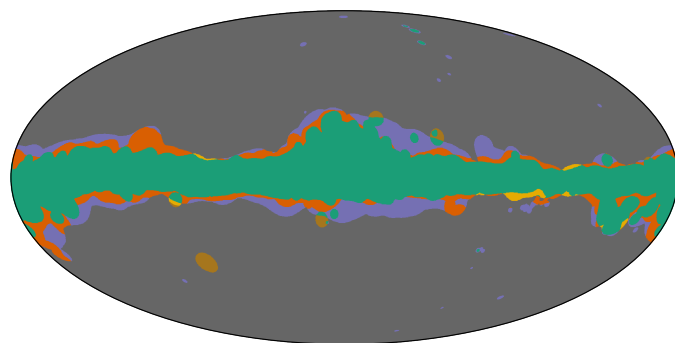

Fig. 11. Processing masks used in gain and correlated noise estimation for each frequency channel. The allowed 30 GHz sky fraction (blue) is 0.73, the 44 GHz sky fraction (green) is 0.81, and the 70 GHz sky fraction (red) is 0.77. Reproduced from Gjerløw et al. (2022).

First, we note that the fundamental data model in Eq. (69) may for each detector be written in the form

$$d_t = g_t s_t^{\mathrm{tot}} + s_t^{\mathrm{1Hz}} + n_t^{\mathrm{corr}} + n_t^{\mathrm{w}} \tag{93}$$

where

$$s_t^{\mathrm{tot}} = \mathsf{P}_{tp} \left[\mathsf{B}_{pp'}^{\mathrm{symm}} s_{p'}^{\mathrm{sky}} + \mathsf{B}_{pp'}^{\mathrm{asymm}} \left(s_t^{\mathrm{orb}} + s_t^{\mathrm{fsl}} \right) \right] \tag{94}$$

is the total sum of all true astrophysical signals entering through the optical assembly of the instrument. Noting that both the gain and correlated noise *a priori* are unknown quantities with no and weak priors, respectively, it is clear from the structure of Eq. (93) that these two parameters are highly degenerate: significant variations in $\boldsymbol{g}$ can be accounted for by adjusting $\boldsymbol{n}_{\mathrm{corr}}$ with only a small penalty in terms of total goodness-of-fit through the noise power spectrum, ξ.

Recognizing the importance of this degeneracy, and the resulting sensitivity to potential modelling errors, we decompose the full time-dependent gain function into three components, and sample each of these with a special-purpose sampler. Specifically, we write the full gain for detector i in the form $g_t^i = g_0 + \Delta g^i + \delta g_t^i$, where g_0 is the gain averaged both over all detectors within a given frequency map and over time; Δg^i is the mean gain differential for detector i averaged over time, with the additional requirement that $\sum_i \Delta g^i = 0$; and δg_t^i represents the time dependence in gain, with the additional requirement that $\sum_t g_t^i = 0$ for each i. In addition, when sampling the these gain parameters, we marginalize over the correlated noise component, as discussed in Sect. 8.3.1, in order to minimize the Gibbs chain correlation length. In total, the data model used for gain sampling therefore reads

$$d_t^i = (g_0 + \Delta g^i + \delta g_t^i) s_t^{\mathrm{tot}} + s_t^{\mathrm{1Hz}} + n_t^{\mathrm{tot}}, \tag{95}$$

where $n_t^{\mathrm{tot}} = n_t^{\mathrm{corr}} + n_t^{\mathrm{wn}}$ is the total noise contribution with a full covariance matrix given by ξ^n.

Formally speaking, the statistically optimal sampling algorithm for any of the three gain parameters is given by correlating the full sky signal with the observed data. In effect, this direct approach was adopted by the LFI DPC pipeline (Planck Collaboration II 2020). A significant disadvantage of this direct approach, however, is sensitivity to foreground and bandpass mismatch errors. Instead, we adopt the following three-step approach, which is structurally similar to the `NPIPE` algorithm (Planck Collaboration Int. LVII 2020).

First, we sample g_0 with the orbital CMB dipole alone as calibration source, based on the residual

$$r_t^i \equiv d_t^i - \hat{g}(s_t^{\mathrm{tot}} - s^{\mathrm{orb}}) - (\Delta \hat{g}^i + \delta \hat{g}_t^i) s_t^{\mathrm{orb}} - s_t^{\mathrm{1Hz}} = g_0 s_t^{\mathrm{orb}} + n_t^{\mathrm{tot}}, \tag{96}$$

where the symbol "^" denotes the respective estimate from the previous Gibbs iteration. Noting that n^{tot} is a Gaussian field with covariance N and s^{orb} is a deterministic quantity, the sampling algorithm for g_0 is given by that of a univariate Gaussian distribution as described in Appendix A.1,

$$g_0 = \frac{\sum_i s_i^{\mathrm{orb},t} \mathsf{N}_i^{-1} r_i^{\mathrm{orb}}}{\sum_i s_i^{\mathrm{orb},t} \mathsf{N}_i^{-1} s_i^{\mathrm{orb}}} + \frac{\eta}{\sqrt{\sum_i s_i^{\mathrm{orb},t} \mathsf{N}_i^{-1} s_i^{\mathrm{orb}}}}, \quad (97)$$

where the sum runs over all detectors in the frequency map, and $\eta \sim N(0,1)$. We note that this expression represents a formal violation of the Gibbs chain, since g_0 in reality affects both s^{tot} and s^{orb}. By fitting only to s^{orb} we effectively achieve robustness with respect to modelling errors at the cost of increased statistical uncertainty.

Note that the noise covariance matrices in Eq. (97) include both correlated and white noise contributions; this is algebraically equivalent to marginalizing over n_{corr} as described above. In constrast, in a classic Gibbs sampling algorithm we would subtract n_{corr} from r^i in Eq. (98), and then only include the white noise component in N_i. This, however, would lead to a much longer Markov correlation length, since the joint uncertainties between g_0 and n_{corr} then would have to be explored by sampling, rather than algebraically.

Second, we sample the detector dependent calibration factors, Δg_i, based on the residual

$$r_t^i \equiv d_t^i - (g_0 + \delta \hat{g}_t^i) s_t^{\mathrm{tot}} - s_t^{\mathrm{1Hz}} = \Delta g_i s_t^{\mathrm{tot}} + n_t^{\mathrm{tot}}, \quad (98)$$

for each detector, which now includes contributions from both the solar CMB dipole and astrophysical foregrounds, and therefore supports a significantly higher signal-to-noise ratio than the orbital dipole alone. At the same time, we impose the additional linear constraint that

$$\sum \Delta g_i = 0, \quad (99)$$

such that possible contamination in this step does not affect the absolute mean calibration of the full frequency channel. The total system may be solved using the Lagrange multiplier technique (e.g., Bertsekas 1996) with a Lagrangian of the form

$$\mathcal{L}(\Delta g_i, \lambda) = \sum_i \left(r^i - \Delta g_i s^{\mathrm{tot}} \right)^t \mathsf{N}_i^{-1} \left(r^i - \Delta g_i s^{\mathrm{tot}} \right) + \lambda \sum_i \Delta g_i, \quad (100)$$

where λ is a Lagrange multiplier. The maximum posterior solution is found by solving the linear equations resulting from setting $\partial \mathcal{L}/\partial x_i = 0$ with $x_i = \{\Delta g_i, \lambda\}$.

Third and finally, the time-dependent gain fluctuations are sampled starting from the residual

$$r_t^i \equiv d_t^i - (g_0 + \Delta g_i) s_t^{\mathrm{tot}} - s_t^{\mathrm{1Hz}} = \delta g_i s_t^{\mathrm{tot}} + n_t^{\mathrm{tot}}, \quad (101)$$

where δg_t^i is assumed to be constant within each PID period, but free to vary between consecutive PIDs. Unfortuately, the noise on the estimated gain per PID is very large, and in practice, these estimates must be smoothed. For a fully stationary system, the mathematically optimal approach to do so would be to sample δg_t^i from a multivariate Gaussian distribution, taking into account their known $1/f$ power dependency and associated temporal correlations. This would correspond to an optimal Wiener filter, and the optimal sampling algorithm takes the following form (Gjerløw et al. 2022),

$$(\mathsf{G}^{-1} + \mathsf{N}^{-1})\delta \hat{g} = \mathsf{N}_i^{-1} r_i + \mathsf{N}_i^{-1/2} \eta_1 + \mathsf{G}^{-1/2} \eta_2. \quad (102)$$

In this expression, G is the covariance matrix of the Gaussian prior on δg_i, N_i is the noise covariance of the individual gain measurements, and η_1 and η_2 are two independent vectors drawn from a normal distribution with unity variance.

Motivated by the fact that gain fluctuations originate from the same $1/f$ fluctuations as correlated noise, we adopt a similar functional form in Fourier space for G as the correlated noise component,

$$\mathsf{G}(f) = \sigma_0^2 \left(\frac{f}{f_0} \right)^\alpha, \quad (103)$$

where α and σ_0 are parameters to be determined, while f_0 is just a reference frequency. These two parameters effectively determine the smoothness of the resulting solution; for instance, the lower the value of σ_0, the smoother the final gain estimate will be. In principle, these parameters could be fitted directly from the data, fully analogous to the correlated noise parameters discussed in the next section. However, the time-dependent gain shows clear evidence of non-thermal variations, for instance in the form of sharp jumps due to abrupt changes in the thermal or electronic environment. To avoid over-smoothing these, we instead choose values that impose less temporal smoothing than dictated by the stationary $1/f$ behaviour; for further details, see Gjerløw et al. (2022).

For comparison, the official *Planck* LFI DPC processing adopted a boxcar average algorithm for smoothing the gain (Planck Collaboration II 2016), effectively averaging over a gliding window in time. The main difference between these two algorithms is illustrated in Fig. 10, which compares the correlated noise components that results from the two methods: The boxcar averaging introduces significant striping by over-smoothing real gain fluctuations, which in turn biases the estimated CMB Solar dipole and Galactic foregrounds for the relevant time periods. The novel Wiener-filter smoothing algorithm presented by Gjerløw et al. (2022) represents one of the most important individual algorithmic improvements in BEYONDPLANCK with respect to the official *Planck* processing, and it was key to understanding the long-standing issues with the 44 GHz frequency channel (Planck Collaboration II 2020).

To prevent foreground modelling errors from affecting the various gain estimates, we apply the processing masks indicated in gray in Fig. 11 in each of the above equations. Any sample that falls within the masked region is omitted from the corresponding inner product, and does not contribute to the overall estimate. The same applies to any sample that may be flagged by the instrument database. Removing individual samples, however, does introduce a slight computational complication because of the $\mathsf{N}_{\mathrm{tot}} = \mathsf{N}_{\mathrm{corr}} + \mathsf{N}_{\mathrm{wn}}$ operator, which denotes a dense noise covariance matrix that includes both correlated and white noise. Application of this operator at full temporal TOD resolution is computationally expensive. However, we note that since the gain is defined only by a single value per PID, small-scale fluctuations can be averaged down with minimal loss of information in all the above equations. We therefore down-sample each time-ordered data object to 1 Hz before evaluating the above equations, and this reduces the overall computational cost for gain sampling by almost two orders of magnitude; see Gjerløw et al. (2022) and Ihle et al. (2022) for further details.

8.3.2. Correlated noise sampling

Since the gain is sampled from a marginal distribution with respect to correlated noise, not a conditional distribution, it is essential to sample the correlated noise immediately following the

gain; otherwise the Gibbs chain would end up in an internally inconsistent state. However, as far as the actual sampling algorithm for the correlated noise is concerned, this is a normal conditional with respect to the gain, akin to any other standard Gibbs step, and was first described in a CMB setting by Wehus et al. (2012). The same algorithm has now also, for the first time, been used to solve the CMB mapmaking problem by Keihänen et al. (2022).

To derive the appropriate sampling equation for $\boldsymbol{n}^{\mathrm{corr}}$, we return to the full data model in Eq. (69), and note that it may be written on the form

$$r_t \equiv d_t - g_t s_t^{\mathrm{tot}} - s_t^{\mathrm{1Hz}} = n_t^{\mathrm{corr}} + n_t^{\mathrm{w}}, \tag{104}$$

where s_t^{tot} is defined in Eq. (94). As discussed in Sect. 7.1, $\boldsymbol{n}^{\mathrm{corr}}$ is assumed to be Gaussian distributed with zero mean and covariance $\mathsf{N}^{\mathrm{corr}}$, while the white noise term is uncorrelated Gaussian with variance σ_0^2. Eq. (104) therefore also describes a correlated Gaussian distribution, and the sampling equation is in this case given by Eq. (A.11) with a template matrix $\mathsf{T} = 1$, a signal covariance matrix $\mathsf{S} = \mathsf{N}^{\mathrm{corr}}$, a noise covariance matrix $\mathsf{N} = \mathsf{N}^{\mathrm{w}}$, and data $\boldsymbol{d} = \boldsymbol{r}$.

Let us first consider the ideal case of a single PID with no missing data due to either instrument flags or processing mask. In that case, Eq. (A.11) can be solved very conveniently in the Fourier domain, and the appropriate sampling equation for the kth Fourier mode reads

$$n_k = \frac{r_k + \eta_1 \sigma_0^2 / \sqrt{P_k^{\mathrm{corr}}} + \eta_2 \sigma_0}{1 + \sigma_0^2 / P_k^{\mathrm{corr}}}. \tag{105}$$

For this case, we note that the computational cost is equivalent to two Fourier transforms of the full time-ordered data.

As usual, the first term in the numerator of Eq. (105) is simply a Wiener filtered version of the residual, $\boldsymbol{r}$. As such, it represents a biased estimate of $\boldsymbol{n}^{\mathrm{corr}}$, with a noise suppression factor given by the relative inverse signal-to-noise ratio, $P_k^{\mathrm{corr}}/\sigma_0^2$. The two last terms are stochastic random fluctuations that ensure that the resulting sample has the appropriate covariance structure.

Equation (105) only applies to data with no missing time samples, as the Fourier transforms require continuous inputs. In practice, however, we drop all samples that are removed by either the instrument flags or by the processing mask shown in Fig. 11. In this case, the optimal solution is therefore given by Eq. (A.11), where rows and columns corresponding to masked samples are set to zero in N^{-1}. The resulting equation is therefore solved efficiently by a Conjugate Gradient (CG) technique, as described by Keihänen et al. (2022). As reported by Galloway et al. (2022a), and summarized in Table 2, this particular step accounts for about 40 % of the total computational cost of the BEYONDPLANCK Gibbs sampler, and it is as such by far the most expensive single component in the entire analysis.

In the current framework, correlated noise estimation plays the role of the traditional CMB mapmaking problem with correlated noise in a traditional pipeline. In this respect, it is worth noting that the correlated noise sample is constructed based on the signal-subtracted data, $\boldsymbol{r}$, alone. Under the assumption of a perfect signal model, inaccuracies in the correlated noise model can therefore not introduce any signal bias. Using the analogy of traditional destriping codes (e.g., Maino et al. 1999; Keihänen et al. 2004, 2005, 2010), the signal subtraction plays the same role in the Gibbs sampling approach as the projection operator $\mathsf{Z} = \mathsf{I} - \mathsf{P}(\mathsf{P}^t\mathsf{N}^{-1}\mathsf{P})^{-1}\mathsf{P}^t\mathsf{N}^{-1}$ does for destriping, shielding any stationary signal from the noise filter. The main fundamental difference between the two approaches lies in the fact that while the traditional destriper only exploits information from a single frequency channel at any given time, the Gibbs sampling approach simultaneously exploits information from all frequencies to construct a joint signal model, which then is used to shield the signal during correlated noise estimation. The Gibbs sampling approach is thus mathematically equivalent to destriping all frequencies at once. The effect of this global correlated noise estimation will become evident later, in the form of lower correlated noise residuals in the joint approach.

Second, it is important to note that the correlated noise solution resulting from Eq. (105) is moderately robust against model errors, whether they are due to foreground modelling errors or inaccuracies in the bandpass or beam profile. The reason is simply that Eq. (105) is a Wiener filter, and therefore has greatly suppressed power in any frequency mode for which $P_k^{\mathrm{corr}} \ll \sigma_0^2$. Intuitively, this means that any feature that cannot be readily identified in the raw time-ordered data as compared with σ_0, will only be weakly affected by the correlated noise component. Minor errors in the signal model, beam or bandpass profiles are therefore mostly negligible.

There are, however, two important exceptions to this general rule. First, some point sources, such as Tau A or the Galactic center, are sufficiently bright that uncertainties in the beam or foreground model can be large compared to the white noise level. If so, the resulting errors will typically translate into bright stripes passing through the respective source, extending along the scanning path of the satellite. To avoid this, it is critically important to mask all bright sources as part of the processing mask, and replace those regions with a proper constrained realization as described above.

The second important exception is the CMB dipole. This signal is both bright, with a peak-to-peak amplitude of about 3 mK, and concentrated at a very low frequency that corresponds to the satellite spin rate of 1/60 Hz. This is typically comparable to (or lower than) the correlated noise knee frequencies (Planck Collaboration II 2020). Furthermore, the ring-based *Planck* scanning strategy provides minimal modulation of the dipole signal on frequencies beyond the spin frequency. The combination of these facts leads to a strong degeneracy between the CMB dipole parameters, the time-dependent gain, and the correlated noise. Indeed, experience shows that *Planck* is, for all practical purposes, unable to break this degeneracy through statistical power alone. Instead, various strong priors are typically imposed to regularize these degeneracies. For instance, the LFI DPC processing impose the requirement that $\boldsymbol{m}_{\mathrm{D}} \cdot \boldsymbol{m} = 0$, where $\boldsymbol{m}_{\mathrm{D}}$ is a map of the CMB dipole and $\boldsymbol{m}$ is the sky map; this effectively leaves the full instrumental noise component aligned with the CMB dipole in the final sky map (Planck Collaboration V 2016). Additionally, the LFI pipeline makes no explicit corrections for bandpass mismatch during gain calibration. For the HFI 2018 DPC processing, the dominant assumption is that the gain is fully independent of time, and the only source of apparent gain fluctuations are ADC non-linearities (Planck Collaboration III 2020). For NPIPE, two important assumptions are that polarized foregrounds at frequencies between 44 and 217 GHz may be fully modelled in terms of the signal observed by 30 and 353 GHz, and that CMB polarization may be ignored during calibration (Planck Collaboration Int. LVII 2020). Obviously, none of these assumptions are formally correct, and they will necessarily lead to systematic biases at some level.

In BEYONDPLANCK, we adopt a different approach to the problem, by actually exploiting information beyond *Planck*. Specifically, as described in Sect. 5, we will in the following perform a joint analysis of *WMAP* and *Planck* observations, and thereby take advantage of information in one experiment to break de-

generacies in the other. Most notably, the *WMAP* scanning strategy covers 70 % of the sky every hour, as compared to less than 1 % per hour for *Planck*. This strategy is thus obviously better suited for measuring the very largest angular scales on the sky, despite higher white noise. On the other hand, the differential structure of the *WMAP* differencing assemblies leads to particularly large uncertainties for some specific modes, including $E_{\ell=5}$ and $B_{\ell=3}$ (Jarosik et al. 2011). In BEYONDPLANCK we therefore choose to combine *Planck* and *WMAP* data while taking into account the full covariance information of each experiment, and thereby optimally leverage the individual strengths of each experiment. Still, we emphasize the importance of visually inspecting binned sky maps of $\boldsymbol{n}_{\mathrm{corr}}$ for dipole-like residuals, which is the archetypical signature of calibration errors; such residuals may occur if the assumed signal model is inadequate for the data set in question.

8.3.3. Noise PSD sampling

The third conditional distribution, $P(\xi^n \mid \boldsymbol{d}, \omega \setminus \xi^n)$, in the BEYONDPLANCK Gibbs chain describes the noise power spectrum density parameters, P_k and σ_0, collectively denoted ξ^n. In the following, we will make the assumptions that ξ^n is constant within each PID and uncorrelated between PIDs. Being closely connected to the previous sampling step, the following procedure was also first presented for CMB applications by Wehus et al. (2012).

To sample from $P(\xi^n \mid \boldsymbol{d}, \omega \setminus \xi^n)$, we recall that $\boldsymbol{n}_{\mathrm{corr}} \sim N(0, \mathsf{N}_{\mathrm{corr}})$. Therefore,

$$P(\xi \mid \boldsymbol{d}, \omega \setminus \xi^n) \propto P(\xi \mid \boldsymbol{n}_{\mathrm{corr}}) \tag{106}$$

$$\propto \frac{e^{-\frac{1}{2}\boldsymbol{n}_{\mathrm{corr}}^t \mathsf{N}_{\mathrm{corr}}^{-1} \boldsymbol{n}_{\mathrm{corr}}}}{\sqrt{|\mathsf{N}_{\mathrm{corr}}|}}, \tag{107}$$

where $\mathsf{N}_{\mathrm{corr}} = \mathsf{N}_{\mathrm{corr}}(\xi^n)$. To sample from this distribution, we could for instance run a Metropolis sampler over ξ^n, using Eq. (107) to define the acceptance probability. However, at this stage we introduce an approximation to the exact solution, trading a small amount of statistical optimality for increased robustness to modelling errors and minimal parameter degeneracies. Specifically, we decouple the white noise variance from the correlated noise model simply by defining

$$\sigma_0^2 \equiv \frac{\mathrm{Var}(r_t - r_{t-1})}{2}, \tag{108}$$

where we define

$$r_t \equiv d_t - g s_t^{\mathrm{tot}} - n_t^{\mathrm{corr}} - s_t^{1\mathrm{Hz}} \tag{109}$$

to be the residual time stream after subtracting both the current total sky signal and correlated noise estimates. Thus, we take the variance of the difference between any two neighboring residual samples to be our white noise variance. On the one hand, this represents the single most robust estimate of the total white noise level one can form from a finite data set. On the other hand, it is of course only an approximation to the true white noise level, since the correlated noise component may also happen to include a flat and non-negligible power spectrum density at the highest frequency mode. This situation typically arises more often for bolometers (as for instance employed by the *Planck* HFI detectors) than for coherent detectors (as employed by the *Planck* LFI detectors and considered here), but the principle is the same both cases.

Thus, we define the component of the correlated noise at half the Nyquist frequency to be part of the white noise, and the correlated noise is consequently defined as the difference between the total noise and the white noise. For error propagation into other parameters in the model, only the sum of the two components is significant. This split is thus essentially just a computational short-cut that eliminates internal degeneracies between the two noise components, and maximizes the relative contribution of the white noise component. This has two main numerical advantages. First, noting that white noise marginalization is performed algebraically, while correlated noise marginalization is done through sampling, a high relative white noise fraction leads to a shorter overall Markov chain correlation length for all steps in the algorithm. Second, by fixing the white noise level, we break degeneracies within the ξ^n parameters, which otherwise lead a very long correlation length between σ_0, α, f_{knee} and A_p (see Sect. 4.5), making convergence assessment difficult.

Given this definition of the white noise variance, the correlated noise level may now be sampled from Eq. (107) by fixing N^{w}. Specifically, as discussed by Ihle et al. (2022), the conditional posterior may be written in Fourier space as

$$-\ln P(\xi^n) = \sum_{f>0} \left[\frac{|n_f^{\mathrm{corr}}|^2}{P_{\mathrm{corr}}(f)} + \ln P_{\mathrm{corr}}(f) \right], \tag{110}$$

up to an irrelevant constant, where n_f^{corr} are the Fourier coefficients of the correlated noise estimate, $\boldsymbol{n}_{\mathrm{corr}}$, and $P_{\mathrm{corr}}(f)$ represent all non-white contributions to the noise PSD. We sample from this distribution with a simple inversion sampler (see Appendix A.3), iteratively Gibbs sampling over α, f_{knee}, and A_{p}. Masking and in-painting is handled by the $\boldsymbol{n}_{\mathrm{corr}}$ sampling step described in Sect. 8.3.2.

8.3.4. Bandpass sampling

Next, we consider the bandpass correction conditional distribution, $P(\Delta_{\mathrm{bp}} \mid \boldsymbol{d}, \omega \setminus \Delta_{\mathrm{bp}})$, and in the following we will consider the most basic form of bandpass correction, namely a linear shift as defined by Eq. (44); see Svalheim et al. (2022a) for further details.

Similar to the gain case, we find it useful to decompose the full bandpass shift for detector j as follows,

$$\Delta_{\mathrm{bp}}^j = \bar{\Delta}_{\mathrm{bp}} + \delta_{\mathrm{bp}}^j. \tag{111}$$

Here the first term is the average over all radiometers within a given frequency channel and the second term is constrained by $\sum_j \delta_{\mathrm{bp}}^j = 0$. The motivation for this decomposition is that the two terms impact the data in qualitatively different ways. The average bandpass shift, $\bar{\Delta}_{\mathrm{bp}}$, change the overall effective frequency of the full frequency channel, and is as such highly degenerate with the foreground SED parameters; a given bandpass frequency shift may often be counteracted by adjusting the values of the synchrotron or thermal dust spectral indices. This mean bandpass shift does not in general change the polarization properties of the resulting frequency map. The relative bandpass corrections, however, have a strong impact in terms of polarization through temperature-to-polarization leakage, as discussed in Sect. 4.2 and by Svalheim et al. (2022a).

For this reason, we have implemented two different sampling algorithms for these parameters. First, the mean bandpass cor-

rection is sampled with the full time-domain residual on the form

$$r_j = d_j - n_j^{\text{corr}} - \mathsf{G}_j\mathsf{P}_j\mathsf{B}_j^{\text{asymm}}\left(s_j^{\text{orb}} + s_j^{\text{fsl}}\right) - s_j^{\text{1Hz}} \tag{112}$$

$$= \mathsf{G}_j\mathsf{P}_j\mathsf{B}_j^{\text{symm}} \sum_c \mathsf{M}_{cj}(\beta, \Delta_{\text{bp}}^j)a^c + n_j^{\text{w}}. \tag{113}$$

Clearly, this residual is highly nonlinear in Δ_{bp}, and no analytic distribution or sampler exist. We therefore once again resort to the Metropolis sampler described in Sect. 6.1. Specifically, we propose small variations to the current mean bandpass shift (while keeping the relative differences between radiometers fixed); we compute the resulting mixing matrices M and sky maps for the new proposals; and we finally then apply the Metropolis acceptance rule as given by the resulting χ^2. Only samples within the small processing mask in Fig. 11 are included in the χ^2. Since mixing matrix updates are computationally expensive, bandpass corrections are among of the most difficult parameters to explore within the entire model. However, as discussed by Svalheim et al. (2022a), the degeneracies between CMB, free-free, AME and $\bar{\Delta}_{\text{bp}}$ are too strong to support a robust determination of $\bar{\Delta}_{\text{bp}}$ when including only LFI and *WMAP* data. In the final BeyondPlanck production runs, we therefore adopt priors based on the *Planck* 2015 analysis (Planck Collaboration X 2016), which used HFI data to break these degeneracies. In practice, we only apply an overall mean correction of 0.24 GHz to the 30 GHz channel, and no mean corrections to the 44 and 70 GHz channels. In future analyses also including the full HFI set, these priors will obviously be removed.

For the relative bandpass corrections, δ_{bp}, we adopt an alternative approach that is specifically tuned to increase robustness in the final polarization maps. Specifically, after proposing changes to each of the detector-specific bandpasses (under the constraint that their sum vanishes), we compute the resulting *IQUS* map that was defined in Eq. (80) for both the old and new parameter values. Next, we define a special purpose χ^2 of the form

$$\chi^2 = \sum_{j=1}^{N_{\text{det}}-1} \sum_p \left(\frac{S_j(p)}{\sigma_j(p)}\right)^2, \tag{114}$$

where $S_j(p)$ is the spurious map corresponding to radiometer j in pixel p, and $\sigma_j(p)$ is the associated uncertainty resulting from the *IQUS* solution. This χ^2 defines the Metropolis acceptance probability as follows,

$$a = \min\left(1, e^{-\frac{1}{2}(\chi_{\text{prop}}^2 - \chi_{\text{i}-1}^2)}\right), \tag{115}$$

where χ_{prop}^2 and $\chi_{\text{i}-1}^2$ are the χ^2's of the proposed and previous parameter states, respectively.

Overall, this approach builds on the same fundamental ideas as the *IQUS* approach pioneered by *WMAP* (Page et al. 2007), but using vastly fewer free parameters: Rather than fitting one free parameter per pixel, this algorithm introduces only one additional free parameter per radiometer. To achieve acceptable precision, it instead uses the current foreground model to predict the resulting corrections in each pixel. Thus, while the direct *IQUS* method is not applicable for *Planck* due to its poorly interconnected scanning strategy, our approach implements the same basic idea but without excessively increasing the overall white noise level of the final maps. For further discussion of the method, we refer the interested reader to Svalheim et al. (2022a).

8.3.5. Diffuse component spectral parameter sampling

The fifth conditional distribution in the BeyondPlanck Gibbs chain concerns the foreground SED parameters, $P(\beta \mid d, \omega \setminus \beta)$. Noting that the linear amplitudes a and spectral parameters β are in general highly degenerate for high signal-to-noise models, we employ the same computational method for intensity sampling as for the gain and correlated noise, and sample these jointly. In practice, this is achieved by first sampling β from the marginal probability distribution with respect to a, and then a conditionally on β. For specific details regarding the following algorithm, we refer the interested reader to Andersen et al. (2022). For polarization, we employ a standard Metropolis sampler that is conditional on the foreground amplitudes; see Svalheim et al. (2022b) for details.

For CMB component separation applications, the two-step marginal sampling approach was first described by Stompor et al. (2009) and later implemented in the Miramare code by Stivoli et al. (2010). To see how their methodology connects with our notation, as defined by Eq. (69), we can write the relevant residual in the following form,

$$r_j = \left(d_j - n_j^{\text{corr}} - s_j^{\text{1Hz}}\right)/g_j - \left(s_j^{\text{orb}} + s_j^{\text{sl}}\right)$$
$$= g_j\mathsf{P}_j\mathsf{B}_j s_j^{\text{sky}}(\beta) + n_j^{\text{w}}. \tag{116}$$

The left-hand side in this equation is identical to the residual in Eq. (74), which is the input to the binned mapmaker defined by Eq. (77). Under the assumption of azimuthally symmetric beams,[17] B_j, this expression may therefore be rewritten in terms of binned sky maps on the form

$$m_\nu = \mathsf{A}_\nu(\beta)a + n_\nu^{\text{w}}, \tag{117}$$

where $\mathsf{A}(\beta) \equiv \mathsf{B}_\nu\mathsf{M}_\nu(\beta)$ is an effective mixing matrix that accounts for both beam convolution and astrophysical component SEDs. Given this expression, the marginal log-posterior for β then reads (Stompor et al. 2009)

$$-2\ln P(\beta \mid m) = \sum_\nu \left(\mathsf{A}_\nu^t\mathsf{N}_\nu^{-1}m_\nu\right)^t \left(\mathsf{A}_\nu^t\mathsf{N}_\nu^{-1}\mathsf{A}_\nu\right)^{-1} \left(\mathsf{A}_\nu^t\mathsf{N}_\nu^{-1}m_\nu\right). \tag{118}$$

However, the derivation of this expression relies on an assumption of identical beam responses across all frequency channels, and it is therefore necessary to smooth all input maps to a common angular resolution before evaluating this expression. We therefore use this expression only for intensity sampling, coupled to a tuned Metropolis sampler.

For polarization, we employ a likelihood given by the original residual defined by Eq. (117),

$$-2\ln P(\beta \mid m, a) = \sum_\nu \left(\frac{m_\nu - \mathsf{A}_\nu(\beta)a}{\sigma_\nu(p)}\right)^2 \tag{119}$$

where $\sigma_\nu(p)$ is the standard deviation map of channel ν. When estimating the spectral index of synchrotron emission, we partition the sky into four large disjoint regions, and sample one constant value of β_s per region, while still allowing for smooth transitioning between regions. Sky partitioning allows us both to tune the signal-to-noise ratio per fitted parameter, and also to

---

[17] In the current BeyondPlanck implementation, we assume azimuthally symmetric beams for all component separation steps, following all previous CMB analysis pipelines.

reduce the overall computational cost. All other free spectral parameters are fitted using a single constant value across the full sky. For both temperature and polarization, we employ tuned Metropolis samplers to explore the posterior distribution (Andersen et al. 2022; Svalheim et al. 2022b).

Finally, we note that even with low-dimensional spectral parameter models, it is useful to impose additional priors on β to stabilize the fits. Specifically, we consider two types of priors in the following. First, in order to be able to pre-compute efficient mixing matrix lookup tables for each parameter, we impose a hard uniform prior on each parameter as discussed in Sect. 8.3.8. Second, we impose informative Gaussian priors on β, with parameters informed from the literature; see Andersen et al. (2022) and Svalheim et al. (2022b) for further details.

8.3.6. Diffuse component amplitude sampling

Since we sample β from a marginal distribution with respect to $\boldsymbol{a}$ for the intensity case, we must also sample $P(\boldsymbol{a} \mid \boldsymbol{d}, \omega \setminus \boldsymbol{a})$ directly following β. The relevant data model for $\boldsymbol{a}$ is (similar to β) given by Eq. (117), but this time interpreted as a function of $\boldsymbol{a}$ instead of β. As applied to CMB estimation, this model was first introduced into the CMB literature by Jewell et al. (2004), Wandelt et al. (2004), and Eriksen et al. (2004), and later generalized to joint CMB power spectrum estimation and astrophysical component separation by Eriksen et al. (2008). With the uniformized notation defined above, the same formalism applies both to CMB and diffuse astrophysical foregrounds, just with different parametric forms for the mixing matrices, M, signal covariance matrices, S, and optional priors.

Noting that $\boldsymbol{n}_\nu^{\mathrm{w}}$ represents Gaussian white noise and $\sum_\nu \mathsf{B}_\nu \mathsf{M}_\nu$ is a deterministic linear operation given $\omega \setminus \boldsymbol{a}$, the appropriate sampling equation for $\boldsymbol{a}$ is yet again given by the multivariate Gaussian sampler in Eq. (A.11) with a template matrix $\mathsf{T} = \sum_\nu \mathsf{B}_\nu \mathsf{M}_\nu$, i.e.,

$$\left(\mathsf{S}^{-1} + \sum_\nu \mathsf{M}_\nu^t \mathsf{B}_\nu^t \mathsf{N}_\nu^{-1} \mathsf{B}_\nu \mathsf{M}_\nu\right) \boldsymbol{a} =$$
$$\sum_\nu \mathsf{M}_\nu^t \mathsf{B}_\nu^t \mathsf{N}_\nu^{-1} \boldsymbol{m}_\nu + \mathsf{S}^{-1}\mu + \sum_\nu \mathsf{M}_\nu^t \mathsf{B}_\nu^t \mathsf{N}_\nu^{-1/2}\eta_\nu + \mathsf{S}^{-1/2}\eta_0. \quad (120)$$

Here we have included the signal covariance matrix, $\mathsf{S} = \mathsf{S}(C_\ell)$, which is a prior that depends on the angular power spectrum of the respective component. If no spatial prior is desired, S^{-1} may simply be set to zero.

Equation (120) arguably represents the single most challenging step in the entire BEYONDPLANCK analysis pipeline in terms of computational complexity. Fortunately, an efficient iterative solver was recently developed by Seljebotn et al. (2019) for precisely this equation, and this algorithm forms the computational engine of `Commander2` (see Sect. 2.2). The main new idea in that work is the use of a pseudo-inverse preconditioner coupled to a Conjugate Gradient (CG) solver that easily supports multi-resolution observations, as required for Eq. (120). For specific details, we refer the interested reader to Seljebotn et al. (2019).

Computationally speaking, the main complicating factor associated with Eq. (120) is the application of an analysis mask. For CMB likelihood estimation purposes, it is necessary to exclude pixels with particularly bright astrophysical foregrounds by setting $\mathsf{N}_\nu^{-1} = 0$, in order not to contaminate the resulting CMB map. Unfortunately, this makes the coefficient matrix on the left-hand side of Eq. (120) poorly conditioned, and the resulting CG search expensive. At the same time, we are also scientifically interested in the properties of astrophysical foregrounds

inside the Galactic mask, and simply discarding all this useful information is clearly undesirable.

Rather than directly applying a processing mask, we therefore instead choose to solve Eq. (120) twice. First, within the main Gibbs loop (as defined in Sect. 8.2) we solve Eq. (120) imposing neither a spatial prior on the CMB component, nor an analysis mask. In this configuration the CG search converges typically within $O(10^2)$ iterations, which corresponds to a computational cost that is smaller than the TOD processing steps by one order of magnitude (Galloway et al. 2022a). The resulting CMB sky map samples correspond to prior-free, full-sky CMB maps, similar to those produced by classic component separation algorithms; see, e.g., Planck Collaboration IX (2016) and Planck Collaboration IV (2018).

However, in order to produce the clean full-sky CMB map and power spectrum samples that are required for high-resolution CMB likelihood estimation purposes (see Sect. 8.3.8 and Colombo et al. 2022), we additionally solve Eq. (120) with S^{-1} and a mask, but *condition* on all non-CMB parameters. Statistically speaking, this is equivalent to writing the full joint posterior distribution in Eq. (60) in the form

$$P(\boldsymbol{a}^{\mathrm{CMB}}, \omega \setminus \boldsymbol{a}^{\mathrm{CMB}} \mid \boldsymbol{d}) = P(\boldsymbol{a}^{\mathrm{CMB}} \mid \boldsymbol{d}, \omega \setminus \boldsymbol{a}^{\mathrm{CMB}}) P(\omega \setminus \boldsymbol{a}^{\mathrm{CMB}} \mid \boldsymbol{d}), \quad (121)$$

and using the first main Gibbs loop to draw samples from the second factor on the right-hand side, and the second solution of Eq. (120) to sample from the first factor.

Formally speaking, we note that this approach is only approximate, since C_ℓ should in principle also be conditioned upon in the second factor in Eq. (121). The penalty of not doing so is slightly more noise in the non-CMB parameters, since the prior-free CMB sky map sample is less smooth than it is with the prior. However, the practical benefits gained by separating the TOD processing steps from the CMB likelihood estimation step more than outweighs a small increase in statistical uncertainties for several reasons: 1) it greatly reduces overall computational costs for the joint Gibbs chain; 2) it allows CMB estimation from individual frequency channels or channel combinations; and 3) it allows rapid exploration of different analysis masks and/or cosmological models without having to rerun the costly TOD processing steps. Thus, this split plays the same role in the BEYOND-PLANCK pipeline as the split between mapmaking and likelihood estimation does in a traditional CMB analysis pipeline.

We employ a similar method also for low-resolution likelihood analysis, and re-sample CMB multipoles below $\ell \leq 64$, while conditioning on all higher multipole CMB modes and other parameters. In this case, we do not impose the C_ℓ prior term, but rather set $\mathsf{S}^{-1} = 0$ as in the original analysis. This allows us to generate tens of thousands of low-resolution samples at a greatly reduced computational cost, and derive a well-converged brute-force low-ℓ likelihood from a relatively limited number of full-scale samples. For further details, see Sect. 9.4, Colombo et al. (2022), and Paradiso et al. (2022).

For two of the astrophysical foregrounds, namely free-free emission and AME, we use informative priors to stabilize the model (Andersen et al. 2022). For free-free emission, we adopt the *Planck* 2015 model (Planck Collaboration X 2016) as a spatial template for the prior mean, while the AME prior is based on the *Planck* HFI 857 GHz map, but with a free scaling factor, under the assumption that the AME surface brightness correlates strongly with thermal dust emission (Planck Collaboration X 2016). In both cases, the signal covariance matrices are empirically tuned to allow sufficient variations to statistically fit

the data, while at the same time not introducing too many un-constrained degrees-of-freedom.[18]

8.3.7. Compact source sampling

The two previous sections described sampling algorithms for diffuse components (such as CMB, synchrotron or thermal dust emission) in terms of their amplitude and SED parameters. These algorithms are strongly tuned toward global modelling in terms of spherical harmonics expansions through the use of computationally efficient spherical harmonics transforms. However, as discussed in Sect. 3.4, a multitude of compact objects also scatters the sky, and some of these are extremely bright. Formally speaking, these may of course also be described in terms of a spherical harmonics decomposition, since the instrumental beam ensures that they are indeed bandwidth limited in the observed data. However, in practice this would require an extremely high bandwidth limit for the diffuse components, and this is therefore impractical because of the high associated computational costs.

Instead, we follow Planck Collaboration IV (2018), and individually model the brightest compact sources based on a pre-existing catalog of object candidates. Each source candidate is mathematically modelled spatially as a delta function convolved with the instrumental beam evaluated at the source location, and with a power-law SED given by an amplitude, a^{src}, and a spectral index, α. For *Planck* frequencies, we take into account the full asymmetric beam profiles as evaluated with FEBeCoP (Mitra et al. 2011), while for non-*Planck* frequency maps, we adopt azimuthally symmetric beams.

The conditional posterior for the ith compact object is given by subtracting all diffuse components and all other compact objects from the map-based data model in Eq. (117), such that the effective residual at frequency ν reads

$$\boldsymbol{r}_i = \boldsymbol{m}_\nu - \sum_{c \neq i} \mathsf{B}_\nu \mathsf{M}_\nu^c \boldsymbol{a}_c, \tag{122}$$

where c runs both over all diffuse components and all compact objects except the i'th source. The likelihood then takes the form

$$-2 \ln P(a_i, \alpha_i \mid \boldsymbol{m}, \omega \setminus \{a_i, \alpha_i\}) = \sum_\nu \left(\frac{\boldsymbol{m}_\nu - U_\nu a_i \left(\frac{\nu}{\nu_{\mathrm{ptsrc}}} \right)_i^{\alpha-2} \boldsymbol{t}_\nu^i}{\sigma_\nu(p)} \right)^2, \tag{123}$$

where ν_{ptsrc} is the reference frequency adopted for the point source component, $\boldsymbol{t}_\nu^i$ is the spatial (pre-computed) beam template for the current source, and U_ν is the unit conversion factor for frequency ν. (As usual, bandpass integration is suppressed in the notation for readability, but is of course taken into account in the actual calculations, as described in Sect. 4.2.)

In addition, we impose a Gaussian prior on the spectral index of $P(\alpha) = N(-0.1, 0.3^2)$, motivated by Bennett et al. (2013), and a positivity prior on the amplitude, $a_i \geq 0$.

The full conditional posterior is sampled using a Metropolis sampler for (a_i, α_i), running 100 MCMC steps for each source, while completing 3 full scans through the full source set per full

Gibbs cycle. This step represents a relatively minor computational cost, due to extensive pre-computation of both effective beam and bandpass profiles.

8.3.8. C_ℓ and cosmological parameter sampling

The final conditional distribution in the BeyondPlanck Gibbs chain concerns the angular power spectrum, C_ℓ, of each component, possibly as parameterized in terms of a smaller number of general parameters. In the following, we will actually apply this only to the angular CMB power spectrum, but we note that the formalism applies without changes to any other statistically isotropic component, for instance the CIB.

Before we start the discussion, we remind the reader that, as mentioned in Sect. 8.3.6, we apply three different sampling steps for the CMB amplitude map:

1. full-resolution solution of Eq. (120) with no spatial CMB prior, $\mathsf{S}_{\mathrm{CMB}}^{-1} = 0$; the resulting samples are primarily used for CMB prior-free component separation and deriving unbiased frequency maps, but not directly for cosmological parameter estimation;
2. low-resolution solution of Eq. (120) with no spatial CMB prior,[19] $\mathsf{S}_{\mathrm{CMB}}^{-1} = 0$, but only including multipoles $\ell \leq 64$, and conditioning on all other parameters; typically, 50 low-resolution samples are drawn based on each high-resolution sample. These samples form the basis for the low-ℓ temperature-plus-polarization CMB likelihood described below.
3. full-resolution solution of Eq. (120) with a spatial CMB prior, $\mathsf{S}_{\mathrm{CMB}}^{-1} \neq 0$, where C_ℓ is sampled with an inverse Wishart sampler as summarized below. The resulting samples form the basis for our high-ℓ temperature likelihood.

In practice, the first step is run together with the full Gibbs analysis, including both TOD and component separation steps, while the other two are performed by re-running the code after the main run has been completed. From the point of view of CMB estimation alone, the primary purpose of the main Gibbs run is thus to derive an ensemble of frequency maps and corresponding astrophysical sky models, that later can be re-sampled with respect to CMB parameters.

Low-resolution temperature-plus-polarization likelihood
From step 2 above, we typically have a sample set of $O(10^4)$ CMB-only samples, each corresponding to one possible combination of TOD, foreground and high-ℓ CMB parameters. Clearly, the information contained in this sample set may be combined into an effective CMB likelihood in many different ways, each with its own algorithmic advantages and disadvantages. For instance, they could form the basis of a highly robust cross-spectrum estimator, by analysing two halves of the data set at a time, and cross-correlating the resulting CMB map; for a recent example of such cross-spectrum approach applied to the *Planck* data, see, e.g., Planck Collaboration V (2020).

However, since our main goal of this work is to introduce a statistically well-motivated end-to-end Bayesian approach, we prefer to stay as close as possible to the exact Bayesian solution. And, practically speaking, that corresponds most closely to a Gaussian multivariate distribution on the form,

$$P(C_\ell \mid \hat{\boldsymbol{s}}_{\mathrm{CMB}}) \propto \frac{e^{-\frac{1}{2} \hat{\boldsymbol{s}}_{\mathrm{CMB}}^t (\mathsf{S}(C_\ell) + \mathsf{N})^{-1} \hat{\boldsymbol{s}}_{\mathrm{CMB}}}}{\sqrt{|\mathsf{S}(C_\ell) + \mathsf{N}|}}, \tag{124}$$

---

[18] Note that S plays a fully analogous role in a multivariate Gaussian prior as the usual standard deviation in a univariate Gaussian prior, and can be used to adjust the strength of the prior.

[19] In practice, we do formally apply a prior also in this case, but with a sufficiently large numerical value that $\mathsf{S}_{\mathrm{CMB}}^{-1} \approx 0$.

where $\hat{s}_{\text{CMB}}$ represents a CMB-plus-noise map and N is its corresponding effective noise covariance map.[20] Since we at this point have access to a full ensemble of low-resolution CMB samples that span the full allowed posterior volume, we may estimate these quantities as

$$\hat{s}_{\text{CMB}} = \left\langle s^i_{\text{CMB}} \right\rangle \tag{125}$$

$$\mathsf{N} = \left\langle (s^i_{\text{CMB}} - \hat{s}_{\text{CMB}})(s^i_{\text{CMB}} - \hat{s}_{\text{CMB}})^t \right\rangle, \tag{126}$$

where brackets indicate average over the sample set. In the limit of an infinite number of samples, these quantities will converge to the Gaussian approximation of the full pixel-based CMB posterior. The resulting covariance matrices are shown and discussed by Colombo et al. (2022).

This approach is conceptually very similar to that adopted by both the *Planck* LFI DPC Planck Collaboration V (2020) and the *WMAP* science team Hinshaw et al. (2013) for low-ℓ likelihood estimation, both of which rely on brute-force likelihood estimation according to Eq. (124). However, there is one critically important difference: with our approach, all sources of uncertainty that are sampled over in the Gibbs chain with ω are seamlessly propagated to the CMB likelihood, including gain and bandpass uncertainties; foreground uncertainties; correlated noise etc. For the traditional approaches, typically only correlated noise and overall calibration is accounted for in the covariance matrix.

An important question regarding the practicality of Eq. (124) is how many samples are required for convergence. As discussed by Sellentin & Heavens (2016), an absolute minimum criterion for a sampled $n \times n$ covariance matrix simply to be invertible is that $N_{\text{samp}} > n$. However, this is by no means sufficient to obtain a *robust* estimate, and, more typically, numerical experiments indicate that many times this is required for matrices of moderate size and relatively weak correlations; the precise value, however, is something that must be tested on a case-by-case matrix.

In any case, since we have a relatively limited number of samples available, it is of great interest to compress the relevant information in $\hat{s}_{\text{CMB}}$ into as few spatial modes as possible, while still retaining the lion's share of its full information content. With this in mind, we note that the main scientific target for low-ℓ likelihood estimation for *Planck* is the reionization optical depth, τ. In this case, τ typically only depends on the first 6 or 8 multipoles, because of the limited sensitivity of the instrument (Planck Collaboration V 2020). As such, a first natural compression is to retain only modes with $\ell \leq 8$, which corresponds to a total of $3(\ell_{\text{max}} + 1)^2 \approx 240$ modes. However, many of these modes fall within a typical analysis mask (Colombo et al. 2022), and therefore carry no statistical weight in the final answer.

One particularly convenient method of isolating the actually useful modes is through Karhunen-Loève compression, as discussed by Tegmark et al. (1997) and Gjerløw et al. (2015). This approach essentially corresponds to retaining the eigenvectors of $\mathsf{S} + \mathsf{N}$ with the highest eigenvalues, where S is evaluated for a typical model of interest. Adopting the notation of Gjerløw et al. (2015), we organize the eigenmodes with eigenvalues higher than some user-specified threshold row-by-row into a projection operator, P, and apply this to the CMB samples derived

---

[20] We note that this expression does not correspond to the exact Bayesian solution, strictly speaking, because the true uncertainty of a given pixel may be non-Gaussian due to the presence of both foregrounds and TOD corrections. To account for this, cosmological parameters should ideally be sampled within the full-resolution Gibbs chain, for instance using the algorithms proposed by Racine et al. (2016); this, however, is left for future work, and we adopt a Gaussian approximation for now.

above. The compressed data and covariance matrix then reads

$$\tilde{s}_{\text{CMB}} = \mathsf{P}\hat{s}_{\text{CMB}} \tag{127}$$

$$\tilde{\mathsf{N}} = \mathsf{P}\mathsf{N}\mathsf{P}^t \tag{128}$$

$$\tilde{\mathsf{S}} = \mathsf{P}\mathsf{S}\mathsf{P}^t. \tag{129}$$

Adopting a multipole threshold of $\ell_{\text{max}} = 8$ and a signal-to-noise threshold of 10^{-6} typically leaves around 170 spatial modes in the full data set, for which we that convergence is typically reached with about $100\,000$ fast samples, corresponding to 2000 full samples including all systematic effects; see Paradiso et al. (2022). The computational cost of a single likelihood evaluation is also correspondingly reduced because of this compression, and only takes a few hundredths of a second.

High-resolution Blackwell-Rao estimator The above estimator can only be employed at low angular resolution because of its strong dependence on the size of the covariance matrix. For high angular resolution analysis, we use another well-established solution, namely the Blackwell-Rao estimator (Chu et al. 2005), which works very well for high signal-to-noise data. In practice, we only use this for temperature analysis in the current paper, since the signal-to-noise ratio for high-ℓ polarization is very low with only LFI and *WMAP* data. However, we keep the following presentation general, such that it can be used for both temperature and polarization analysis for other experiments.

To derive the appropriate sampling algorithm for $P(C_\ell \mid \boldsymbol{d}, \omega \setminus C_\ell)$ from first principles, we first note that $P(C_\ell \mid \boldsymbol{d}, \omega \setminus C_\ell) = P(C_\ell \mid \boldsymbol{a}^{\text{CMB}})$; if the true CMB map, $\boldsymbol{s}_{\text{CMB}}$, is perfectly known, then no further knowledge regarding the measured data can possibly provide more useful information about the angular CMB power spectrum, C_ℓ. Second, as discussed in Sect. 3.2, we assume that the CMB fluctuation field is isotropic and Gaussian distributed, and its probability distribution is therefore given by Eq. (12). Noting that individual $a_{\ell m}$'s are statistically independent by the assumption of isotropy, we can write (Wandelt et al. 2004)

$$P(C_\ell \mid \boldsymbol{a}^{\text{CMB}}) \propto P(\boldsymbol{a}^{\text{CMB}} \mid C_\ell)P(C_\ell) \tag{130}$$

$$= \prod_{m=-\ell}^{\ell} \frac{e^{-\frac{1}{2}\boldsymbol{a}_{\ell m}^{\dagger} \mathsf{C}_\ell^{-1} \boldsymbol{a}_{\ell m}}}{\sqrt{|\mathsf{C}_\ell|}} P(\mathsf{C}_\ell) \tag{131}$$

$$= \frac{e^{-\frac{2\ell+1}{2}\text{tr}(\sigma_\ell \mathsf{C}_\ell^{-1})}}{|\mathsf{C}_\ell|^{\frac{2\ell+1}{2}}} P(\mathsf{C}_\ell), \tag{132}$$

where $\boldsymbol{a}_{\ell m} = \{a_{\ell m}^T, a_{\ell m}^E, a_{\ell m}^B\}$ and

$$\mathsf{C}_\ell \equiv \begin{bmatrix} C_\ell^{TT} & C_\ell^{TE} & C_\ell^{TB} \\ C_\ell^{TE} & C_\ell^{EE} & C_\ell^{EB} \\ C_\ell^{TB} & C_\ell^{EB} & C_\ell^{BB} \end{bmatrix}; \tag{133}$$

$$\sigma_\ell \equiv \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} \begin{bmatrix} (a_{\ell m}^T)^* a_{\ell m}^T & (a_{\ell m}^T)^* a_{\ell m}^E & (a_{\ell m}^T)^* a_{\ell m}^B \\ (a_{\ell m}^E)^* a_{\ell m}^T & (a_{\ell m}^E)^* a_{\ell m}^E & (a_{\ell m}^E)^* a_{\ell m}^B \\ (a_{\ell m}^B)^* a_{\ell m}^T & (a_{\ell m}^B)^* a_{\ell m}^E & (a_{\ell m}^B)^* a_{\ell m}^B \end{bmatrix}. \tag{134}$$

We typically adopt uniform priors on C_ℓ (although for a discussion of non-uniform priors, see Larson et al. 2007), and the distribution in Eq. (132) is then known as the inverse Wishart distribution, which has a very simple sampling algorithm:

1. Draw $2\ell - n$ Gaussian random vectors, η_i, from the empirical covariance matrix $(2\ell + 1)\sigma_\ell$, each of length n, where n is the dimension of C_ℓ;

2. Compute the outer product of these vectors, $\rho_\ell = \sum_{i=1}^{2\ell-n} \eta_i \eta_i^t$;
3. Set $\mathsf{C}_\ell = \sigma_\ell / \rho_\ell$.

Note that if C is block-diagonal, as for instance is the case if $C_\ell^{TB} = C_\ell^{EB} = 0$, then this algorithm should be applied separately block-by-block. Also, if binning is desired, for instance to increase the effective signal-to-noise ratio of a given power spectrum coefficient, this is most conveniently done in terms of $D_\ell = C_\ell \ell(\ell+1)/2\pi$; for details, see Larson et al. (2007).

The above algorithm describes a self-consistent Gibbs-based approach to CMB power spectrum sampling, as originally suggested by Wandelt et al. (2004). The product from this procedure is a set of joint samples $(s_{\mathrm{CMB}}, C_\ell)_i$. However, the algorithm does not specify how to constrain cosmological parameters from these samples. Indeed, many different approaches may be adopted for this purpose, each making different assumptions and choices with regard to computational cost and robustness to systematic errors. Some approaches presented in the literature include

– *the Blackwell-Rao estimator* (Chu et al. 2005): Direct averaging over σ_ℓ samples given the analytic smoothing kernel in Eq. (132). Exact, but converges slowly in low signal-to-noise regime. Used by *WMAP* low-ℓ TT likelihood (Hinshaw et al. 2013).
– *the Gaussianized Blackwell-Rao estimator* (Rudjord et al. 2009): Multivariate Gaussian approximation to the above, following a Gaussian change-of-variable defined by univariate marginal distribution. Converges much faster than direct Blackwell-Rao estimator, and is highly accurate for typical masks. Used by *Planck* low-ℓ TT likelihood (e.g., Planck Collaboration V 2020).
– *joint Metropolis-Hastings sampling* of $\{a_{\ell m}, \mathsf{C}_\ell\}$ (Jewell et al. 2009; Racine et al. 2016): Efficient in both low and high signal-to-noise regimes; may be applied to both C_ℓ and cosmological parameter estimation.

The first two of these methods define a CMB power spectrum likelihood function, $\mathcal{L}(\mathsf{C}_\ell)$, which then must be coupled to a cosmological parameter estimation code. We employ the widely employed `CosmoMC` (Lewis & Bridle 2002) code for this purpose, as detailed in Paradiso et al. (2022). In contrast, when applied to cosmological parameter estimation, the third method requires a means to convert between cosmological parameters and angular power spectra, such as `CAMB` (Lewis et al. 2000). In this paper, we adopt the Gaussianized Blackwell-Rao estimator as our default solution, and leave the full integrated MCMC sampling approach for future work.

8.4. Computational requirements and optimization

The end-to-end algorithm summarized in the last few sections represents a significant computational challenge, both in terms of fundamental hardware requirements and in terms of software optimization. In this section we briefly review some critical computational features implemented in the current code, while in-depth presentations are provided by Galloway et al. (2022a) and Gerakakis et al. (2022). In addition, we highly recommend the interested reader to consult the source code.[21] At the same time, we emphasize that these codes are most definitely works in progress,

---

[21] The BeyondPlanck software is available under a GNU Public Library (GPL) open-source license at https://github.com/Cosmoglobe/Commander.

and still undergo rapid development. Nearly every single component and function have room for further improvement and optimization. However, it is our hope and intention that by providing all codes to the general community under an open-source license, new collaborations, efforts and ideas will emerge, and this will leading to more mature, efficient and generally applicable code.

With these caveats in mind, Table 2 summarize the overall computational cost of the current implementation, both in terms of initialization and cost per sample. These benchmarks were obtained by running the pipeline on a single compute node with 128 AMD EPYC 7H12 2.6 GHz cores and 2 TB of RAM. All time related costs are provided in units of wall-time, and must therefore be multiplied with 128 to convert to CPU time.

Overall, the computational complexity of the BeyondPlanck Gibbs sampler is determined by three fundamentally different types of operations. First, the low-level analysis is dominated by TOD memory management. Second, the high-level amplitude sampling step is dominated by spherical harmonic transforms. Third, the spectral index sampling step is dominated by map-based operations, typically either spherical harmonic transforms or χ^2 evaluations. Efficient parallelization of each of these three types of operations is therefore the critical design driver for the current implementation. We now briefly review how the BeyondPlanck pipeline optimizes each of these aspects, and refer the interested reader to Galloway et al. (2022a) for further details.

8.4.1. Low-level optimization

Starting with the low-level TOD-oriented operations, we first note in Table 2 that the full data volume of four years of *Planck* LFI observations is 8 TB. This number includes all science and housekeeping data. A single read of the full data set from spinning disks on a typical intermediate-sized high-performance computing (HPC) cluster therefore requires a few hours of wall time, assuming $O(1\,\mathrm{GB\,s^{-1}})$ read speed. While acceptable as a one-time initialization cost, integrating such expenses into the Gibbs loop clearly leads to impractical run times. A first requirement for efficient end-to-end TOD processing is thus that the entire data set may be stored in RAM. Likewise, noting that the memory bus from the RAM to the CPU is relatively slow compared to CPU operations, a corollary requirement is that the overall memory footprint should be aggressively minimized.

With these observations in mind, we first choose to read only those parts of the data that are strictly required for the analysis in question; all unnecessary housekeeping data are omitted. For each *Planck* LFI radiometer the only retained quantities therefore include 1) differenced detector voltages, d_t (one float per sample); 2) pointing, P_t (three double precision values per sample); and 3) flags, f_t (one integer per sample). Nominally, a total of 32 bytes/sample/radiometer are required to store the TOD information.

However, as detailed by Galloway et al. (2022a), because the pointing and flags are both very smooth functions of time, they lend themselves to highly efficient compression. We exploit this by transforming and discretizing each relevant quantity into integers; taking the difference between consecutive samples to minimize their dynamic range; and finally Huffman compressing (Huffman 1952) the resulting time streams, i.e., we assign bit patterns of variable lengths to each integer according to their relative frequency. The average number of bits per sample is thus reduced by a factor of 5–6. These compressed TOD arrays are then stored in memory PID-by-PID, and only decompressed when needed. The total data volume is in this way re-

Table 2. Computational resources required for end-to-end BEYONDPLANCK processing. All times correspond to CPU hours. All reported times are averaged over more than 100 samples, and vary by $\lesssim 5\,\%$ from sample to sample. Reproduced from Galloway et al. (2022a).

ITEM	30 GHz	44 GHz	70 GHz	SUM	REFERENCE
Data volume					
Uncompressed TOD volume	761 GB	1 633 GB	5 522 GB	7 915 GB	
Compressed TOD volume	86 GB	178 GB	597 GB	861 GB	
Non-TOD-related RAM usage				659 GB	
Total RAM requirements				**1 520 GB**	
Processing time (cost per run)					
TOD initialization/IO time	3.8 h	4.3 h	12.5 h	20.6 h	
Other initialization				43.4 h	
Total initialization				**64.0 h**	
Gibbs sampling steps (cost per sample)					
Huffman decompression	1.1 h	1.8 h	7.1 h	10.0 h	Galloway et al. (2022a)
TOD projection (P operation)	0.3 h	0.7 h	3.1 h	4.1 h	BeyondPlanck (2022)
Sidelobe evaluation ($s_{\rm sl}$)	1.1 h	2.1 h	6.5 h	9.7 h	Galloway et al. (2022b)
Orbital dipole ($s_{\rm orb}$)	0.5 h	1.1 h	4.6 h	6.2 h	Gjerløw et al. (2022)
Gain sampling (g)	0.6 h	0.7 h	4.7 h	6.0 h	Gjerløw et al. (2022)
1 Hz spike sampling ($s_{\rm 1hz}$)	0.2 h	0.3 h	1.9 h	2.4 h	BeyondPlanck (2022)
Correlated noise sampling ($n_{\rm corr}$)	1.7 h	3.6 h	24.8 h	30.1 h	Ihle et al. (2022)
Correlated noise PSD sampling ($\xi_{\rm n}$)	3.3 h	4.0 h	1.1 h	8.4 h	Ihle et al. (2022)
TOD binning (P^t operation)	0.2 h	0.5 h	4.1 h	4.8 h	Basyrov et al. (2022)
Sum of other TOD processing	1.3 h	2.5 h	10.9 h	14.7 h	Galloway et al. (2022a)
TOD processing cost per sample	**10.4 h**	**17.4 h**	**69.1 h**	**96.9 h**	
Amplitude sampling, $P(a \mid d, \omega \setminus a)$				23.9 h	Andersen et al. (2022)
Spectral index sampling, $P(\beta \mid d, \omega \setminus \beta)$				40.3 h	Svalheim et al. (2022b)
Other steps				0.6 h	BeyondPlanck (2022)
Total cost per sample				**163.9 h**	

duced from 8 TB to 861 GB, which fits into the RAM of a single modern compute node. The decompression cost accounts for about 5 % of the total analysis wall time, which we consider well worth the memory savings. However, as discussed by Galloway et al. (2022a), this compression does have notable implications in terms of the overall Gibbs sampling structure, as the full decompressed TOD set can never be stored in memory at once, nor is it possible to store multiple copies of the TOD. Accordingly, careful relative ordering of the various Gibbs sampling steps is necessary. In practice, four full scans are made through the entire TOD within each Gibbs iteration, where each scan corresponds to sampling one global TOD-related parameter, namely three gain components (see Sect. 8.3.1) and the bandpass correction parameter; none of these can be sampled simultaneously without breaking the Gibbs chain.

Next, the low-level parallelization scheme for TOD processing is organized according to PIDs, such that each computing core processes a distinct subset of PIDs. Load balancing is achieved by first measuring the effective computing time for each PID, and then distributing them according to cost in a round-robin manner among the various computing cores.

Inspecting the costs of individual steps in Table 2, we see that the dominant TOD operation is associated with sampling $n_{\rm corr}$, which makes intuitive sense: While most operations scale linearly in the number of samples, $O(N_{\rm tod})$, the correlated noise step requires two Fourier transforms, and therefore scales as $O(N_{\rm tod} \log N_{\rm tod})$. To optimize this step, we first of all employ the FFTW library (Frigo & Johnson 2005) for all FFT operations. Second, we note that the speed required for a single FFT transform depends sensitively and highly non-linearly on $N_{\rm tod}$. Values of $N_{\rm tod}$ that happen to factorize into particularly favorable combinations of primes may happen to be, say, three to five times faster than neighboring values. We exploit this by first measuring the

time required per FFT for every length between 1 and 10^6, and construct a table of optimal lengths, with at least one value per 100th sample. At read time, we then truncate the length of each PID until it equals the closest lower optimal length. As such, we lose on average one second of data per PID, corresponding to about 0.03 % of the total data volume, while gaining a factor of three or more in overall TOD processing time.

After the FFT-based operations, the dominant TOD operations are the sidelobe and orbital dipole evaluations, as well as the pointing projections operators, P and P^t. Here it is worth noting that the TOD analysis is currently memory-bus limited. That is, the cost is associated simply with transferring data from RAM into the CPU. As such, the specific algorithmic details of each step are largely irrelevant, and the important factor is simply the total data volume. To improve the performance of these steps, the best approach would be to run across multiple nodes, which thereby increase the number of memory buses available. On the other hand, this also leads to lower performance for the CPU dominated operations, and most notably the spherical harmonics transforms. A future optimal solution should implement a better tuned parallelization strategy where SHTs are parallelized within nodes, while TOD operations are parallelized across nodes; this is left for future development.

Next, the two TOD projection operators warrant a few comments. First, we recall that P converts a map into a time stream. This represents a computational challenge in itself, because each core then needs access to all pixels in the map. However, actually storing the full map per core would require substantial amounts of memory. To solve this, we exploit a MPI-3 shared memory feature, and only store one copy of the map per compute node, rather than one per core. However, we do observe that the memory access speed associated with these shared-memory arrays is
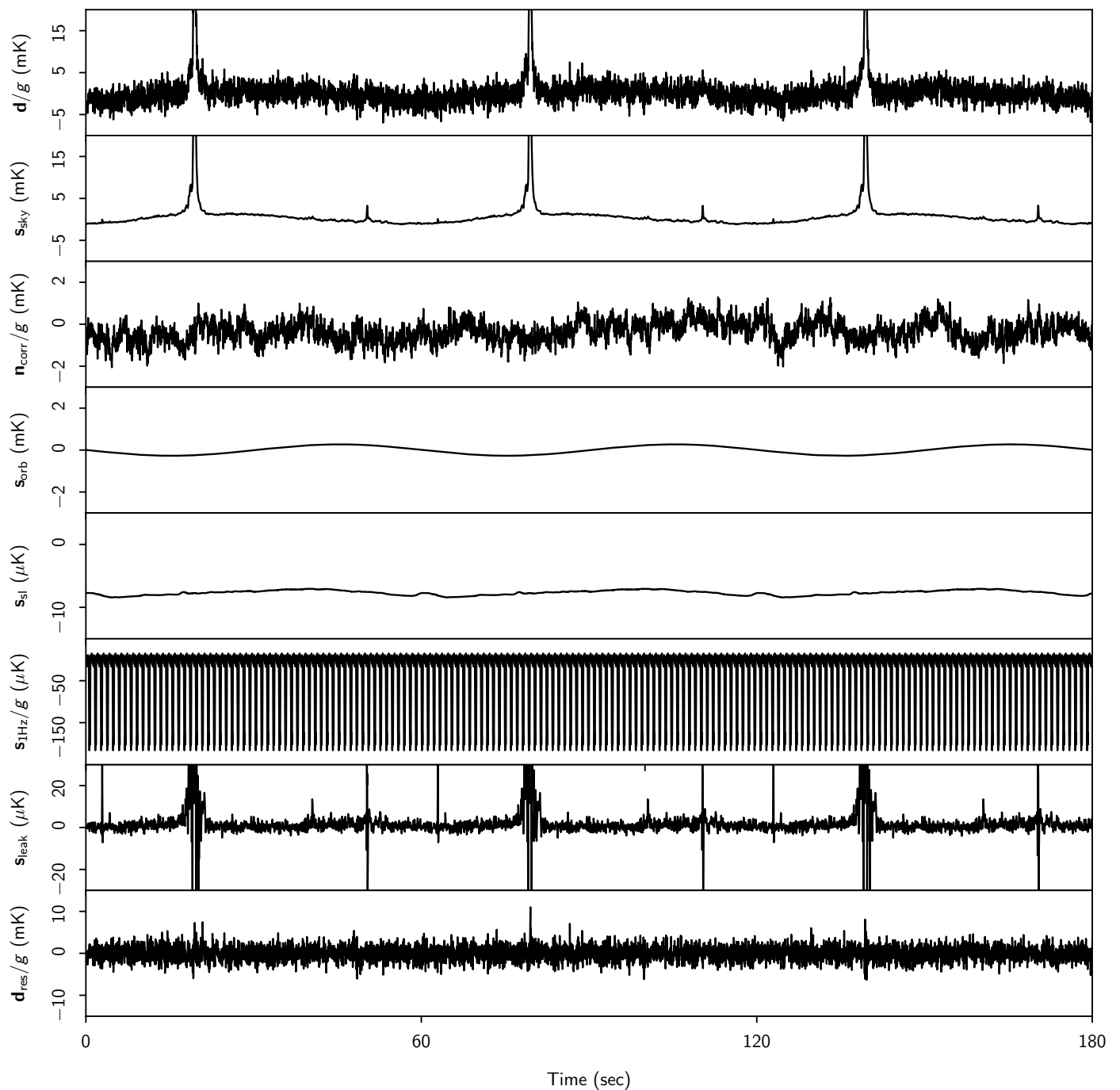
Fig. 12. Time-ordered data segment for the 30 GHz LFI 27M radiometer. From top to bottom, the panels show 1) raw calibrated TOD, $\boldsymbol{d}/g$; 2) sky signal, $\boldsymbol{s}_{\mathrm{sky}}$; 3) calibrated correlated noise, $\boldsymbol{n}_{\mathrm{corr}}/g$; 4) orbital CMB dipole signal, $\boldsymbol{s}_{\mathrm{orb}}$; 5) sidelobe correction, $\boldsymbol{s}_{\mathrm{sl}}$; 6) electronic 1 Hz spike correction, $\boldsymbol{s}_{\mathrm{1Hz}}$; 7) leakage mismatch correction, $\boldsymbol{s}_{\mathrm{leak}}$; and 8) residual TOD, $\boldsymbol{d}_{\mathrm{res}} = (\boldsymbol{d} - \boldsymbol{n}_{\mathrm{corr}} - \boldsymbol{s}_{\mathrm{1Hz}})/g - \boldsymbol{s}_{\mathrm{sky}} - \boldsymbol{s}_{\mathrm{orb}} - \boldsymbol{s}_{\mathrm{leak}} - \boldsymbol{s}_{\mathrm{sl}}$. Note that the vertical range vary significantly from panel to panel. Reproduced from Basyrov et al. (2022).

typically five times slower than for local arrays, and further optimizations are therefore possible.

In contrast, the P^t operation co-adds samples in a timestream into a map. In terms of practical code, this is a more complex operation than P, since all cores need to update the values stored in each sky map pixel, not only read them. This can easily lead to race conditions in which different cores simultaneously write to the same parts of memory, resulting in corrupt data, and a direct shared array approach is therefore impractical. At the same time, allocating a full sky map per core is not an option due to the same memory constraints discussed above. As a compromise, we instead first scan the full pointing stored by each core, and accumulate a list of all locally observed pixels. Due to the sparse *Planck* scanning strategy, this typically amounts to only 5–10 % of all pixels for each core. Allocating and maintaining a sub-map of this limited size is acceptable in terms of total memory footprint. Co-addition over cores is then achieved using a combination of shared arrays within each computing node, and a single `MPI_ALLREDUCE` operation between nodes. Clearly, further optimization is very likely possible also with respect to this operation.
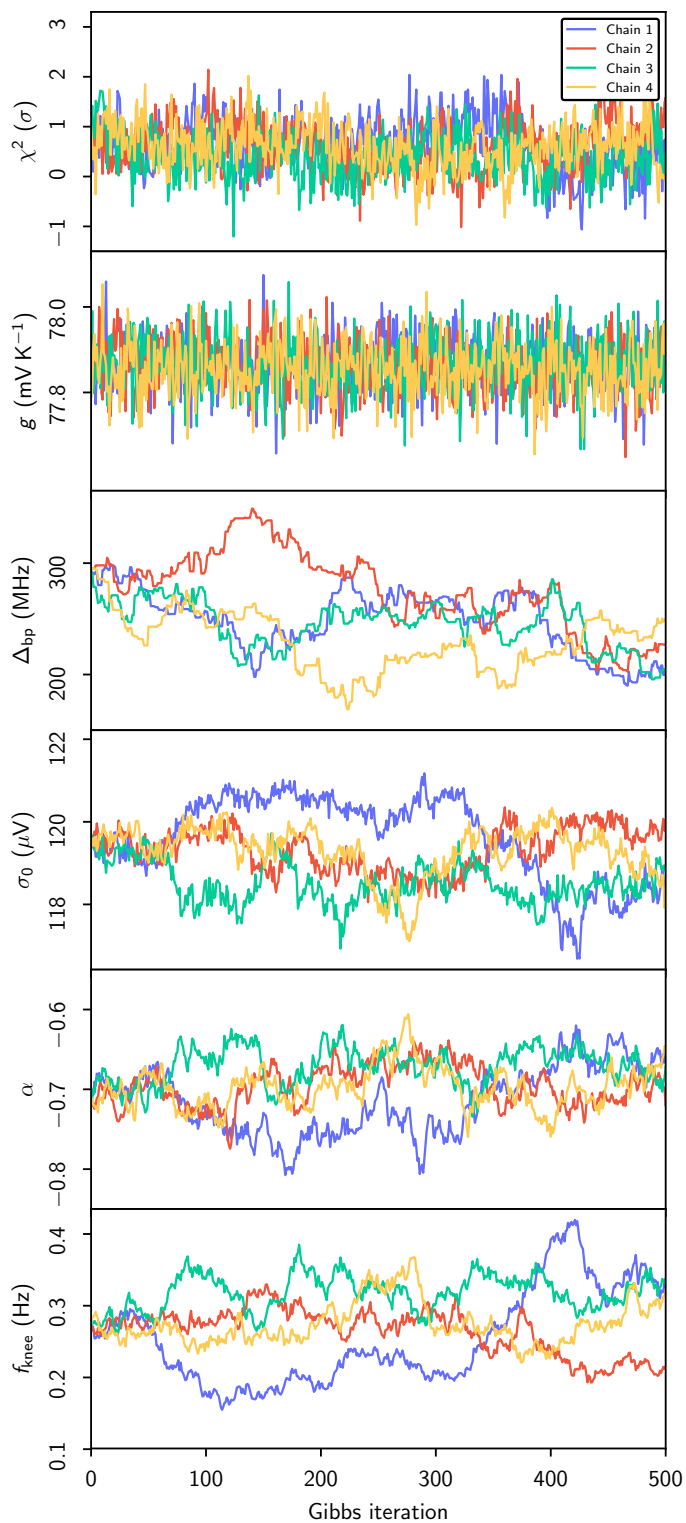
Fig. 13. Example of TOD parameters Gibbs chain for the 30 GHz LFI 27M radiometer. From top to bottom, the panels show normalized reduced χ^2 for a single Pointing Period; gain for the same PID in units of $mV\,K^{-1}$; bandpass correction in MHz; white noise level, σ_0; correlated noise slope, α; and correlated noise knee frequency, f_{knee}. The four different colored curves correspond to independent Gibbs chains.

8.4.2. High-level parallelization and optimization

Next, we consider optimization of the high-level routines, and in particular of the amplitude and spectral index sampling steps.
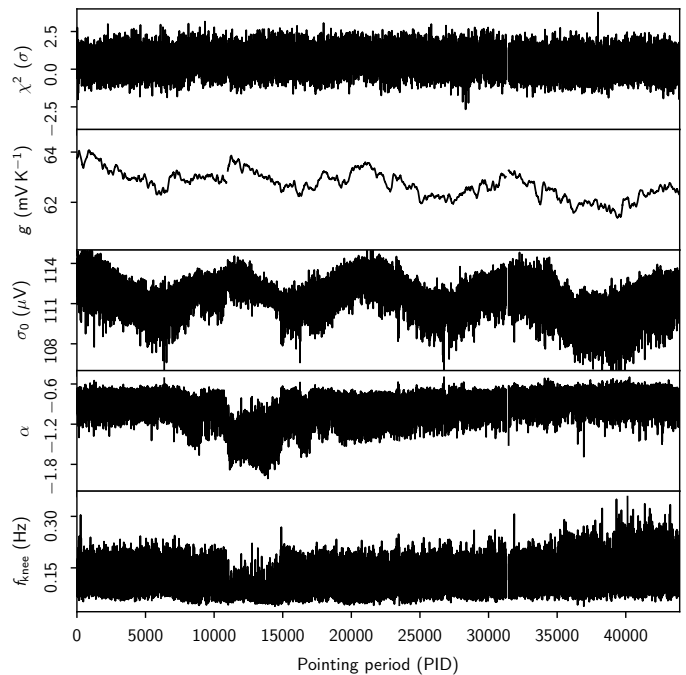
Fig. 14. A single Gibbs sample for the 30 GHz LFI 28M radiometer as a function of PID. From top to bottom the panels show 1) normalized reduced χ^2; 2) gain, g; 3) white noise level, σ_0; 4) correlated noise slope, α; and 5) correlated noise knee frequency, f_{knee}.

These are largely overlapping in terms of essential low-level routines, and so we will also discuss them jointly.

The single most important computational routine involved in these operations is the spherical harmonics transform, needed both for solving the Wiener filter defined by Eq. (120) and for smoothing maps to a common angular resolution as required for Eq. (118). Indeed, the importance of this operation is so critically important that we base our entire map parallelization strategy of our codes around it. With this in mind, we adopt the `libsharp2` (Reinecke & Seljebotn 2013) spherical harmonics library for all harmonic transforms, which is the most efficient library for this purpose available today. This library is based on a deep parallelization level in both pixel and harmonic space, distributing both constant-latitude rings and constant-m harmonics across different cores. We adopt these parallelization conventions without modification.

The second most important operation involved in these operations is multiplication with the mixing matrix, $M(\beta; \Delta_{bp})$. As described in Sect. 4.2, this expression involves integration of an ideal parametric SED with the bandpass of each instrumental detector. It also varies from pixel-to-pixel, depending on the local properties of the spectral parameters, β. For this reason, we precompute the full mixing matrix prior to each full amplitude sampling step, pixel-by-pixel. Taking advantage of the `libsharp` parallelization scheme, which distributes rings across all available cores, the memory requirements for this is fairly limited. Furthermore, employing the spline-based library discussed in Sect. 4.2, the actual evaluation of this matrix only carries a cost equal to a polynomial evaluation per pixel. However, it is important to note that actually changing the bandpass correction parameters, Δ_{bp}, requires a full re-evaluation of the underlying splines, as well as all higher-level mixing matrices, and this particular operation is therefore very computationally intensive. As a result, it is done as infrequently as possible.
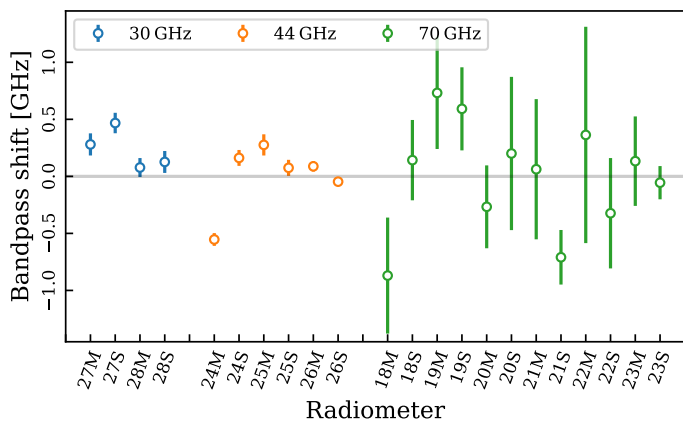
Fig. 15. Estimated bandpass corrections for each LFI radiometer. Error bars indicate $\pm 3\sigma$ uncertainties. Note these parameters are constrained to have vanishing mean within each frequency, and are as such strongly correlated. Reproduced from Svalheim et al. (2022a).

Finally, as described above, many of the various sampling steps are carried out with a standard Metropolis sampler. Although conceptionally and implementationally straightforward, this sampler does have the drawback of requiring specific tuning of the step size to be efficient. For most of these samplers, we therefore typically run a short tuning chain during the first iteration, if the computational cost of the sampler is limited (which, for instance, is the case for the point source sampler), or insert a pre-calculated proposal matrix into the run through a parameter file (which, for instance, is the case for the bandpass correction sampler). Such tuning is essential to achieve acceptable mixing for the overall chain.

9. Results

We are now finally ready to present the main results resulting from applying the algorithms summarized in Sects. 7–8 to the data combination described in Sect. 5. For the analysis shown here, we have produced a total of four independent Monte Carlo Markov chains of samples drawn from the posterior distribution $P(\omega \mid d)$, as described in Sect. 8.2. Each chain has 1000 samples, and we conservatively discard the first 200 samples for burn-in. We thus retain a total of 3200 accepted samples for final processing. With a computational cost of 1.3 wall hours/sample (Galloway et al. 2022a), this set took about three months of continuous run time to produce on two nodes, for a total computational cost of 670 000 CPU hours. Although not directly comparable, it is still interesting to note that the production of the *Planck* FFP8 simulation set required a total of 25 million CPU hours, and the cost of constructing only a single component of a single Monte Carlo realization of the 70 GHz channel cost 9360 CPU-hours (Planck Collaboration XII 2016). The full analysis shown in the following thus carries a total computational cost that is equivalent to about 70 *Planck* FFP8 70 GHz simulations. This clearly demonstrates the computational feasibility of the Bayesian end-to-end approach, and the algorithms shown here do not require the use of a massive super-computer center to be useful. At the same time, it is also clear that future work should concentrate on increasing the concurrency of the current implementation through better parallelization schemes, such that the wall time can be reduced to hours or days, as opposed to months, when more resources are available.

9.1. Instrumental parameters

We start our review by inspecting the data and model at the lowest level, and for this purpose we select the 30 GHz channel as a reference test case, for which the foreground contamination is the largest, and therefore the calibration challenge the hardest. The top panel of Fig. 12 (reproduced from Basyrov et al. 2022) shows a 3 minute chunk of the 30 GHz LFI 27M TOD, in which the only pre-processing steps are differencing with the 4 K load signal and ADC corrections (see Sect. 5 for details).

The top panel shows the raw measurements, which are visually dominated by the CMB dipole, as seen by the slow sinusoidal oscillations; the Galactic plane, as seen by the sharp spikes; and instrumental noise. The second panel shows the estimated sky signal for one random sample; here we see small-amplitude perturbations in addition to the large dipole and galactic contributions, and these are dominated by the CMB temperature fluctuations, which are the main scientific target of the entire analysis. Rows 3–7 show various corrections arising from correlated noise, the orbital CMB dipole, sidelobes, the 1 Hz spike signal, and bandpass and beam leakage. The last panel shows the residual after subtracting all the above terms from the raw data, and this highlights anything that is not explicitly captured by the parametric model; overall, this is largely consistent with white noise, except for a few spikes near the Galactic plane crossings, which are masked both in low-level and high-level processing. For further discussion of this plot, see Basyrov et al. (2022).

Figure 12 represents one single Gibbs sample in the full chain. In contrast, Fig. 13 shows samples from all four Gibbs chains for the instrumental parameters for one PID, but this time plotted as a function of Gibbs iteration. For perfect Markov chain mixing, these should all scatter around a well-defined mean value with a short correlation length.

The top panel shows the normalized reduced χ^2 as defined by

$$\chi^2 \equiv \frac{\sum_{t=1}^{N_{\text{tod}}} \left(\frac{d_t - s_t^{\text{tot}}}{\sigma_0} \right)^2 - N_{\text{tod}}}{\sqrt{2 N_{\text{tod}}}}. \tag{135}$$

Recalling that the χ^2 distribution with n degrees of freedom converges towards a Gaussian with mean equal to n and variance equal to $2n$, this quantity should be approximately distributed as $N(0, 1)$ for ideal data, with deviations measured in units of σ. We adopt this χ^2 as a convenient goodness-of-fit measure. We see that the mean value is $\chi^2 \approx 0.5\sigma$, which indicates good a good fit overall.

The second panel shows the gain g for the same PID. In this case, the Markov correlation length appears very short. Considering that we have 3200 full Gibbs samples available, this implies that the number of independent gain samples per PID is quite high, and their histogram provides a useful estimate of the true underlying distribution; in this case the marginal posterior may be summarized as $g = 77.85 \pm 0.02 \, \text{mV} \, \text{K}^{-1}$. For further discussion of the gain posteriors, see Gjerløw et al. (2022).

The third panel shows the bandpass shift, Δ_{bp}, for the 30 GHz LFI 27M radiometer. As already noted in Sect. 4.2, this parameter is the single most difficult quantity to estimate in the entire framework, because of the highly non-linear, non-Gaussian and global nature of its impact; virtually *all* stochastic variables in the entire model depend on the instrumental bandpass in one form or another, and changes in this parameter therefore take a substantial amount of time to propagate throughout the model. Furthermore, the sampling algorithm used for this parameter is a basic Metropolis sampler, simply because of a lack of better
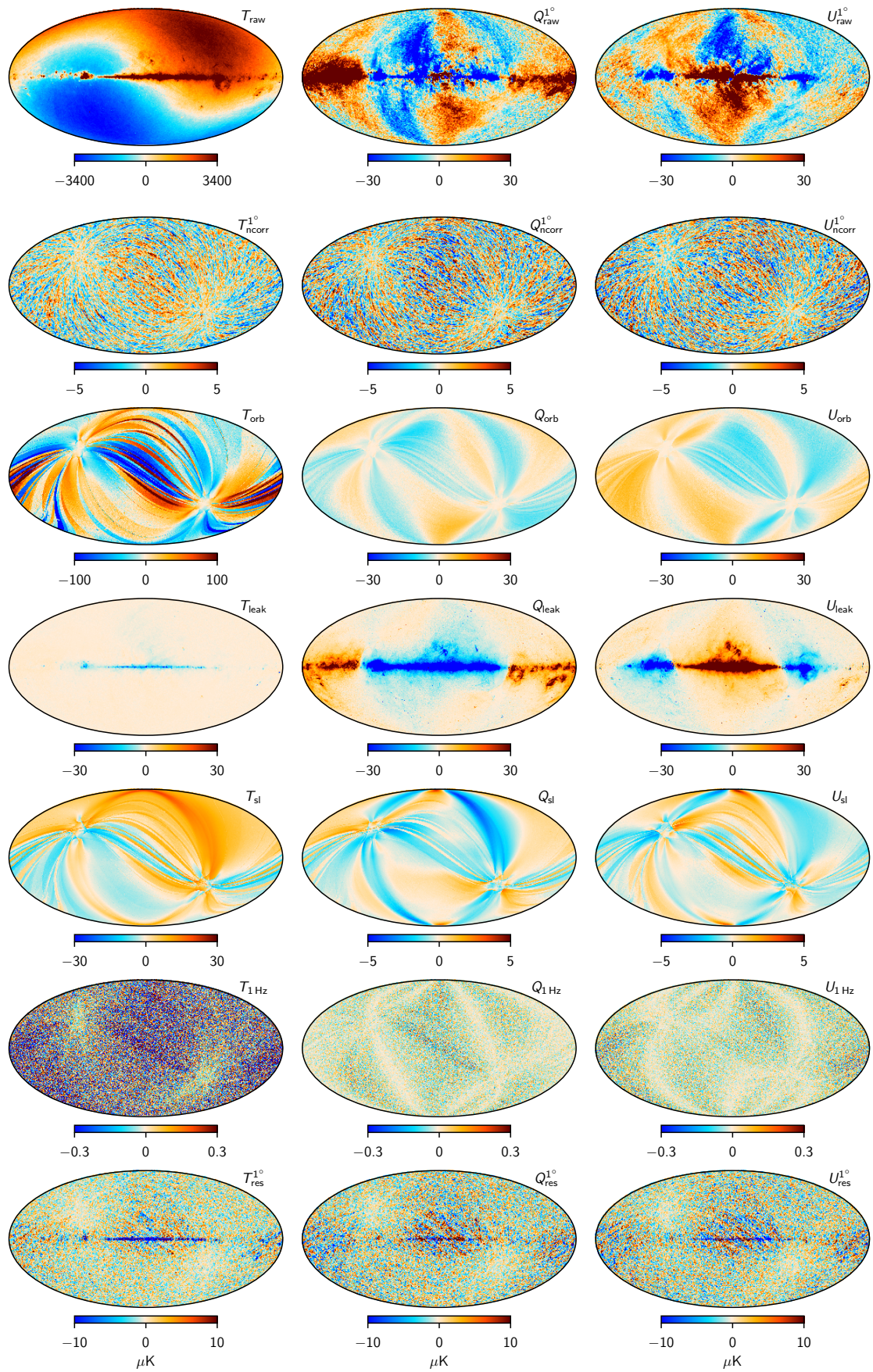
Fig. 16. Comparison between TOD corrections for the 30 GHz channel for a single Gibbs sample, projected into sky maps. Columns show Stokes T, Q, and U parameters. Rows show, from top to bottom, 1) raw TOD; 2) correlated noise; 3) the orbital dipole; 4) bandpass and beam mismatch leakage; 5) sidelobe corrections; and 6) 1 Hz electronic spike correction. The bottom row shows the shows the residual obtained when binning the sky and systematics subtracted TOD into a sky map. Note that some components have been smoothed to an angular resolution of 1° FWHM.
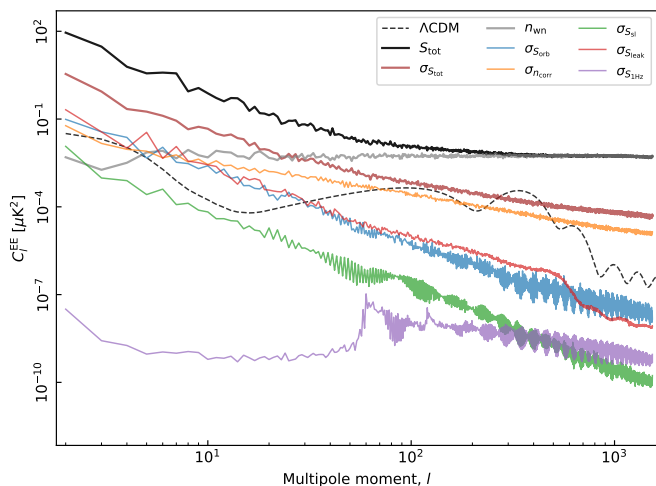
Fig. 17. *EE* pseudo-spectrum standard deviation for each instrumental systematic correction at 30 GHz shown (*thin colored lines*). For comparison, the thick black line shows the spectrum for the full co-added frequency map; the thick red line shows the standard deviation of the same (i.e., the full systematic uncertainty); the gray line shows white noise; and the dashed black line shows the best-fit *Planck* 2018 ΛCDM power spectrum convolved with the instrument beam. All spectra have been derived outside the CMB confidence mask (Andersen et al. 2022) using the HEALPix `anafast` utility, correcting only for sky fraction and not for mask mode coupling. Reproduced from (Basyrov et al. 2022).

alternatives. The result is a long correlation length of about 100 samples, resulting in perhaps as few as 32 uncorrelated samples in the full sample set. Still, even with this crude sampler, we do see that the four chains mix reasonably well, and it is possible to establish a useful estimate for the marginal posterior, which in this case may be summarized as $\Delta_{\rm bp} = 240 \pm 30$ MHz. However, in this case the sampling uncertainty accounts for a large fraction of the error bar. For further discussion of the bandpass posteriors, see Svalheim et al. (2022a).

The three last panels show the three noise PSD parameters, σ_0, α and $f_{\rm knee}$, for the same radiometer. These also show long correlation lengths. However, as discussed by Brilenkov et al. (2022); Ihle et al. (2022), this long correlation length is due to internal degeneracies among the correlated noise parameters, ξ_n, and since all other parameters, such as the gain or CMB component, only care about the total noise power spectral density, $P_n(f)$, and not the individual ξ_n parameters, this poor mixing does not represent a significant limitation for the analysis; see Fig. 15 in Brilenkov et al. (2022). For further discussion of the noise posteriors, see Ihle et al. (2022).

Figure 14 shows corresponding values as a function of PID for one single Gibbs sample, this time for the 30 GHz LFI 28M radiometer. The top panel shows the normalized reduced χ^2, as defined by Eq. (135). As discussed above, this function should ideally be independent between PIDs, and distributed according to $N(0, 1)$. This plot is therefore a powerful monitor for identifying unmitigated and non-stationary instrumental systematic effects in a given radiometer. In this case, the distribution scatters around zero with roughly the expected variations, although there is a slight shift towards positive values of about 0.5–1 σ; overall, the model appears to perform wellx.

The second panel shows the gain for the same 28M radiometer as a function of PID. Here we see clear evidence of a systematic oscillation with a period of one year, and a maximum variation of about 1–2 % throughout the mission. The oscillatory

behaviour is primarily due to variations in the incoming solar radiation during the year, effectively changing the heating of the instrument depending on its precise orientation with respect to the Sun.

The three bottom panels of Fig. 14 show corresponding plots of the three $1/f$ noise parameters as a function of PID. Similar features as observed in the gain are seen also here, although with lower signal-to-noise ratio. Overall, it is visually obvious that the noise properties of this channel are not stationary throughout the mission, but rather vary significantly in time. In particular, the white noise level varies by 3–4 % throughout the mission, and mirrors the gain variations seen above. For the slope, α, the most noteworthy feature are overall steeper values between PIDs 11 000 and 15 000; as shown by Ihle et al. (2022), these can be traced to changes in the thermal environment of the satellite using house-keeping data.

Next, Fig. 15 shows the marginal posterior mean and standard deviation for the bandpass correction, $\Delta_{\rm bp}$, of each radiometer (Svalheim et al. 2022a). Recalling that we fix the absolute corrections for the 44 and 70 GHz at zero and only fit the 30 GHz offset, we find an overall 30 GHz frequency shift of 0.24 ± 0.03 GHz, in agreement with Planck Collaboration X (2016). Regarding the relative corrections, we note that these are depend sensitively on the foreground amplitude at a given frequency, and as a result, the relative bandpass uncertainties are small at 30 GHz, where the foregrounds are bright, while they are large at 70 GHz, where the foregrounds are weak.

Next, we consider the spatial structure of each of the various TOD model terms in pixel space, and Fig. 16 shows each of the TOD objects binned into a 3-component Stokes *IQU* sky map for the 30 GHz channel and one arbitrarily selected Gibbs sample. This plot corresponds essentially to a binned version of Fig. 12, and shows, from top to bottom, 1) the raw data; 2) correlated noise; 3) the orbital dipole; 4) bandpass and beam leakage corrections; 5) far sidelobe corrections; 6) 1 Hz spike corrections; and 7) the total unmodelled residual. Note that the various terms are plotted with very different color ranges, and, for instance, the range used for the 1 Hz correction is only 0.3 μK, and this correction is therefore for most practical purposes negligible. On the other hand, the range of the polarized bandpass leakage correction is 30 μK, and this represents as such the single most important large-scale polarization correction.

Figure 17 summarizes the uncertainty of each systematic effect for the 30 GHz channel and for the *EE* angular power spectrum. The black line shows the power spectrum of the full amplitude map, while the thick red line shows the corresponding posterior standard deviation, which essentially summarizes the total systematic uncertainty. The thin colored lines break down this into individual contributions. Here we see that for *EE* polarization at the *Planck* 30 GHz channel, the three dominant uncertainty contributions are correlated noise (orange), gain fluctuations (through the orbital dipole; blue), and bandpass corrections (thin red). All these are roughly of the same order of magnitude at $\ell \lesssim 10$, and jointly accounting for all three is therefore essential in order to properly capture the full uncertainties. For reference, we note that only the correlated noise contribution was fully accounted in the official LFI low-resolution covariance matrices. A complete survey of similar maps and power spectra for all three LFI frequencies is provided by Basyrov et al. (2022).

9.2. Frequency maps

We now turn our attention to co-added frequency maps, as solved for deterministically through Eq. (77). For many users, these rep-
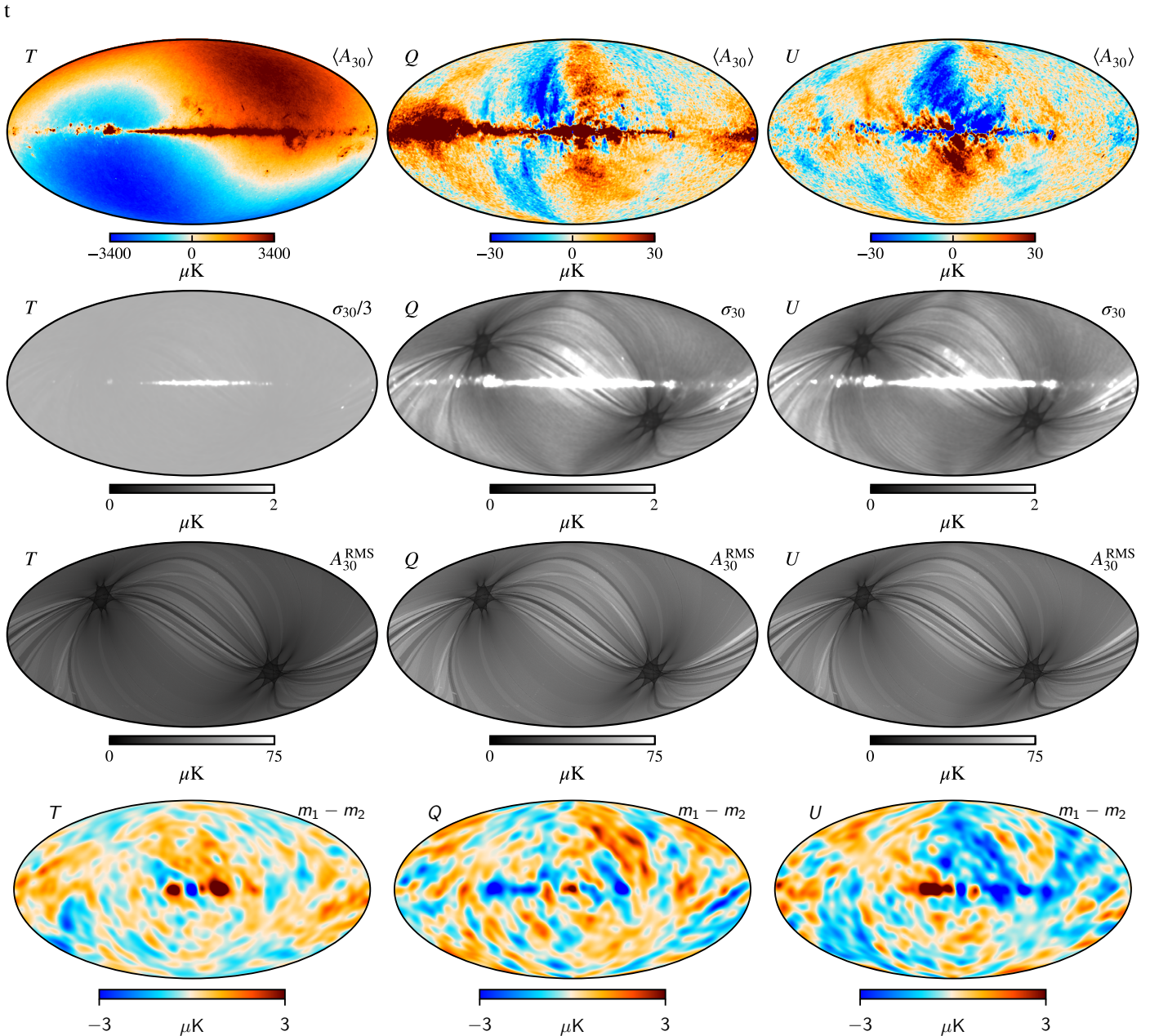
t

Fig. 18. Posterior summary maps for the LFI 30 GHz channel. Columns show the temperature and Stokes Q and U parameters, while rows show, from top to bottom, the posterior mean, the posterior standard deviation, the white noise rms per pixel, and the difference between two random individual Gibbs samples. The polarization maps in the two two rows have been smoothed to an angular resolution of $1°$ FWHM to visually reduce the noise level, while the bottom row is smoothed to $7°$ FWHM.

resent the most convenient form of the BEYONDPLANCK products, and we provide these maps both in the form of individual samples, each corresponding to one possible realization of all modelled systematic effects, and as more traditional posterior mean and standard deviation maps,

$$\hat{m}_\nu = \left\langle m_\nu^i \right\rangle \qquad (136)$$

$$\sigma_\nu(p) = \sqrt{\left\langle \left(m_\nu^i(p) - \hat{m}_\nu(p) \right)^2 \right\rangle}, \qquad (137)$$

where brackets indicate averaging over Monte Carlo samples. Note that σ_ν, as defined here, only accounts for systematic uncertainties per pixel, not white noise uncertainties as defined by the diagonal of the inverse coupling matrix in Eq. (77), $\sigma_\nu^{\text{wn}}(p)$.

To obtain the full uncertainty, these two terms must be added in quadrature,

$$\sigma_\nu^{\text{tot}}(p) = \sqrt{\sigma_\nu(p)^2 + \sigma_\nu^{\text{wn}}(p)^2}. \qquad (138)$$

We stress, however, that analysis of these posterior mean maps is likely to be sub-optimal for most scientific applications, and will not exploit the full power of the BEYONDPLANCK framework. Instead, we highly recommend users to analyze the full ensemble of individual posterior samples; that is by far the most robust and statistically correct method for propagating BEYONDPLANCK uncertainties into any higher-level analysis.

With these caveats in mind, Fig. 18 shows various posterior summary maps for the 30 GHz channel; the full set of maps are shown and discussed by Basyrov et al. (2022). From top to
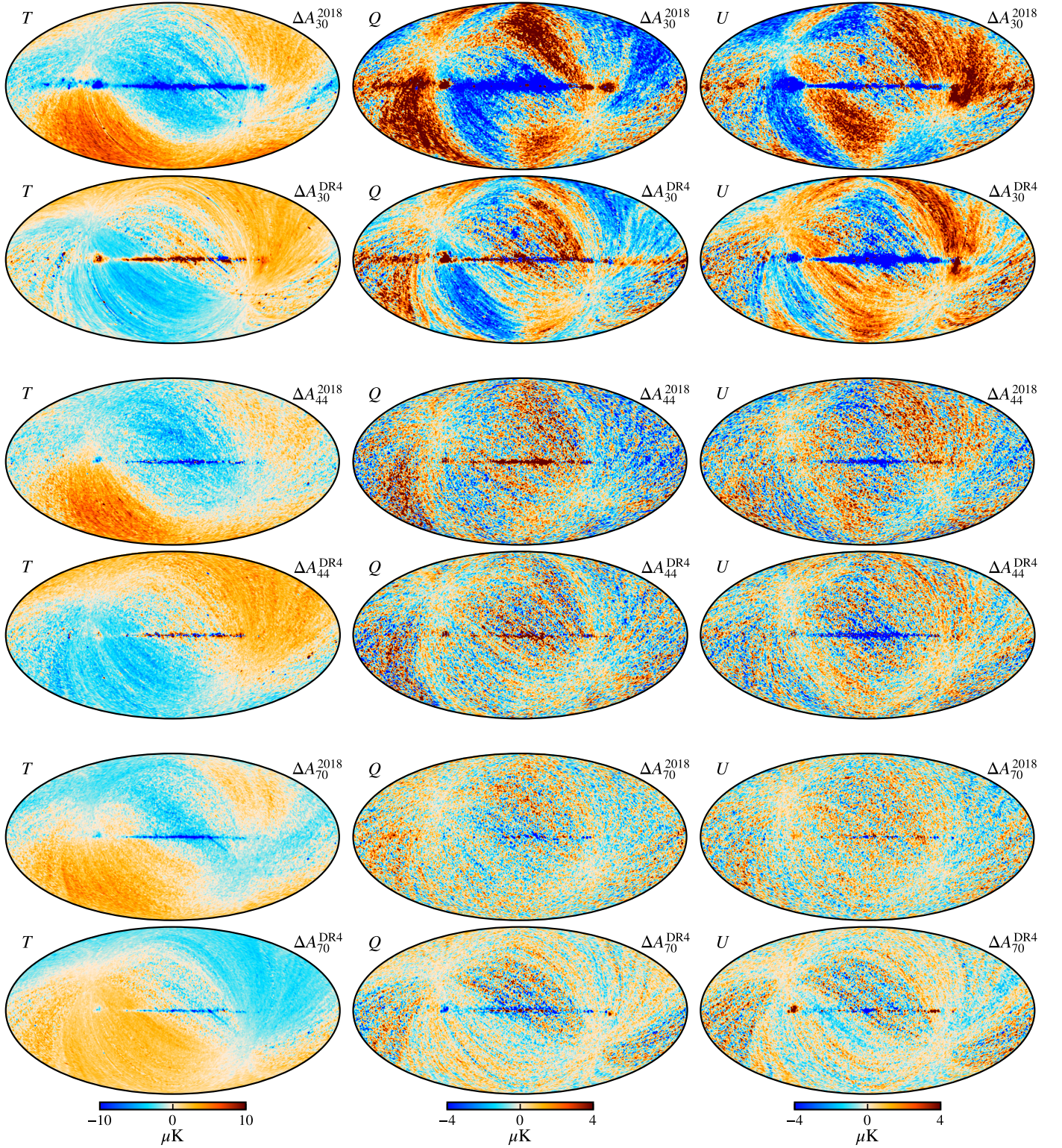
Fig. 19. Differences between BEYONDPLANCK and 2018 or NPIPE frequency maps, smoothed to a common angular resolution of 2° FWHM. Columns show Stokes T, Q and U parameters, respectively, while rows show pair-wise differences with respect to the pipeline indicated in the panel labels. A constant offset has been removed from the temperature maps, while all other modes are retained. The 2018 maps have been scaled by their respective beam normalization prior to subtraction.
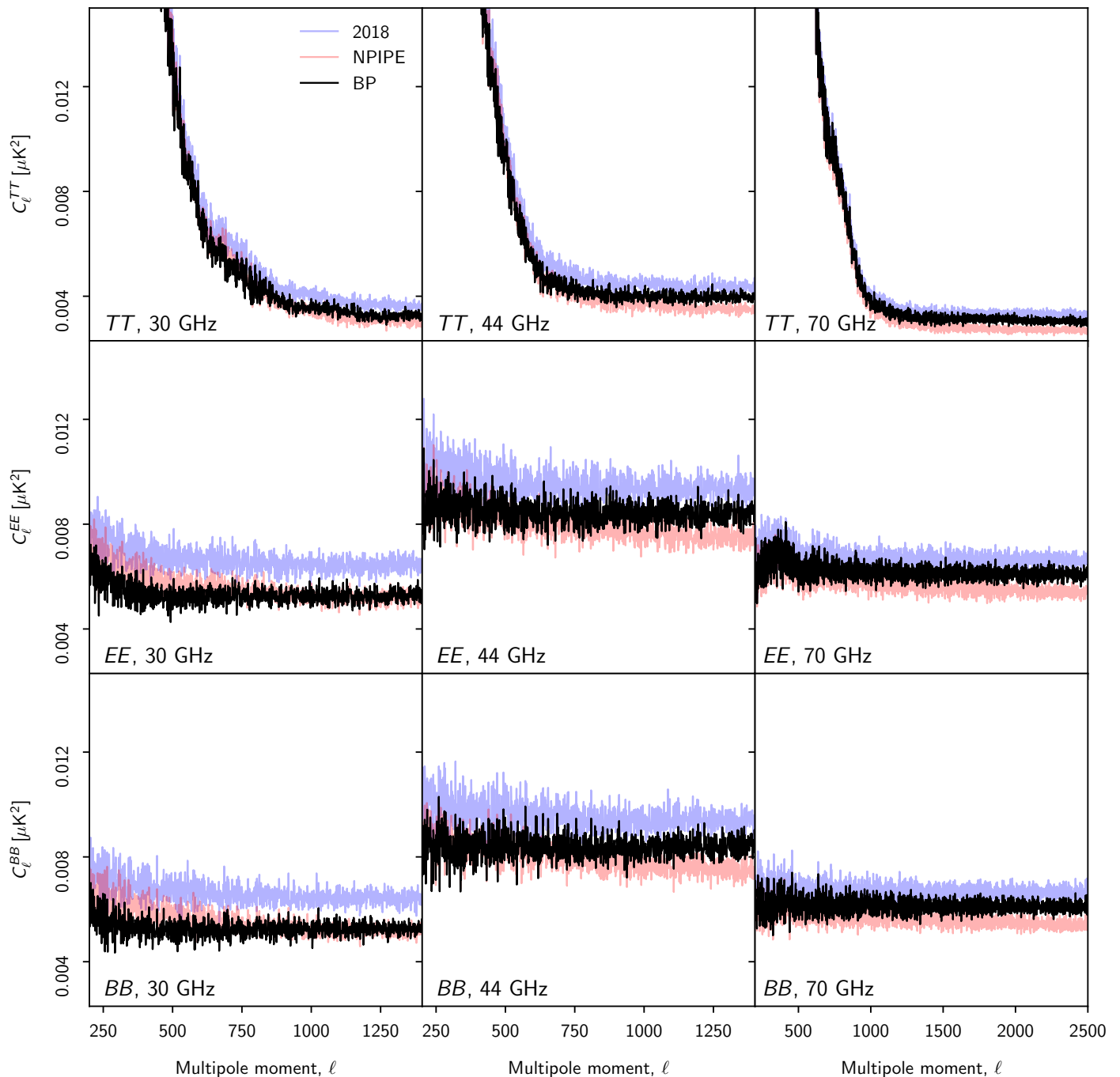
Fig. 20. Comparison between angular auto-spectra computed from the BᴇʏᴏɴᴅPʟᴀɴᴄᴋ (*black*), *Planck* 2018 (*blue*), and NPIPE (*red*) full-frequency maps. Rows show different frequencies, while columns show *TT*, *EE*, and *BB* spectra. All spectra have been estimated with PolSpice using the *Planck* 2018 common component separation confidence mask (Planck Collaboration IV 2018). Reproduced from Basyrov et al. (2022).

bottom, the various rows show the posterior mean, the posterior standard deviation, the white noise rms, and a straight difference map between two independent Gibbs samples. The polarization maps in the top two rows have been smoothed to an angular resolution of 1° FWHM to reduce noise, while the bottom row has been smoothed to 7° FWHM.

Regarding the posterior mean map, perhaps the most striking feature is that the BᴇʏᴏɴᴅPʟᴀɴᴄᴋ temperature map retains the CMB dipole, similar to *Planck* PR4 Planck Collaboration Int. LVII (2020), but contrary to the *Planck* 2018 and *WMAP* frequency maps. Leaving this component in the maps ensures

that the full information content of the data is available for subsequent component separation and calibration applications.

The posterior standard deviation maps in the second row summarize the combined effect of all the various systematic corrections made to the frequency map. The most striking features include:

1. a large monopole variation in the 30 GHz temperature map, resulting in a nearly uniform morphology dipole variations;
2. excess variance for rings aligned with the Galactic plane in polarization, reflecting the higher uncertainties in the time-variable gain resulting from the processing mask;

3. excess variance along the Galactic plane, reflecting the higher uncertainties here due to gain and bandpass variations;

all of which are super-imposed on the general *Planck* scanning pattern, which itself reflects correlated noise variations. We also note that the upper limit of the temperature color scale is only $2\,\mu$K, which indicates that these variations are much lower than the intrinsic variance of the CMB temperature fluctuations, which is about $30\,$muK on these angular scales, and minor details in the systematic model are therefore unlikely to affect final cosmological results. In contrast, the standard deviation of the polarization maps at high Galactic latitudes is typically about $0.5\,\mu$K, which is of the same order of magnitude as the expected polarization imprint from cosmic reionization.

The third row shows the corresponding white noise standard deviation maps. These maps are fully specified by the detector white noise level σ_0, the time-variable gains g_t, and the number of observations per pixel. An important point regarding these white noise rms maps is that their amplitude scales directly with the pixel size, while the posterior standard deviation maps in the second row do not. When smoothed over sufficiently large angular scales, the systematic uncertainties will therefore eventually start to dominate over the white noise level.

The bottom row shows the difference between two frequency map samples, smoothed to a common angular resolution of $7°$ FWHM. Here we clearly see correlated noise stripes along the *Planck* scan direction in all three frequency channels, but significantly more pronounced in the $30\,$GHz channel than in the other two frequencies. We also see fluctuations along the Galactic plane, which are dominated by uncertainties in the bandpass correction parameters, Δ_{bp}. Clearly, modelling such correlated fluctuations in terms of a single standard deviation per pixel is unlikely to be adequate for any high-precision analysis, and the only way to robustly propagate uncertainties is through analysis of the full ensemble of Gibbs samples.

Figure 19 shows differences between the BeyondPlanck frequency maps and those presented in the *Planck* 2018 and NPIPE data releases. To ensure that this comparison is well defined, the 2018 maps have been scaled by the uncorrected beam efficiencies, and the best-fit *Planck* 2018 solar dipole has been added to each map, before computing the differences. Overall, we see that the BeyondPlanck maps agree with the other two pipelines to $\lesssim 10\,\mu$K in temperature, and to $\lesssim 4\,\mu$K in polarization. In temperature, we see that the main difference between NPIPE and BeyondPlanck is an overall dipole, while differences with respect to the 2018 maps show greater morphological differences. The sign of the NPIPE dipole differences changes with frequency. This result is consistent with the original characterization of the NPIPE maps derived through multi-frequency component separation in Planck Collaboration Int. LVII (2020); that paper reports a relative calibration difference between the 44 and 70 GHz channel of $0.31\,\%$, which corresponds to $10\,\mu$K in the map-domain. Overall, in temperature BeyondPlanck is thus morphologically similar to NPIPE, but it improves a previously reported relative calibration uncertainty between the various channels by performing joint analysis.

In polarization, the dominant large-scale structures appear to be dominated by effectively different offset determinations per PID, which may originate from different gain or correlated noise solutions. It is worth noting that the overall morphology of these difference maps is structurally similar between frequencies, and that the apparent amplitude of the differences falls with frequency. This strongly suggests that different foreground mod-

elling plays a crucial role. In this respect, two observations are particularly noteworthy: First, while both the *Planck* 2018 and NPIPE pipelines incorporate component separation as an external input as defined by the *Planck* 2015 data release (Planck Collaboration X 2016), BeyondPlanck performs a joint fit of both astrophysical foregrounds and instrumental parameters. Second, both the LFI DPC and the NPIPE pipeline consider only *Planck* observations alone, while BeyondPlanck also exploits *WMAP* information to establish the sky model, which is particularly important to break scanning-induced degeneracies in polarization.

In Fig. 20 we show the high-ℓ auto-correlation spectra for each of the three generations of *Planck* LFI maps (2018, PR4, and BeyondPlanck) as computed with PolSpice (Chon et al. 2004). In all cases, we apply the *Planck* 2018 common component separation confidence mask (Planck Collaboration IV 2018), which accepts a sky fraction of $80\,\%$. The most notable features here are, first, that the overall noise levels of the BeyondPlanck maps are slightly lower than in the *Planck* 2018 maps, but also somewhat higher than PR4. Secondly, we also note that the BeyondPlanck spectra are notably flatter than the other two pipelines, and in particular than NPIPE, which shows a clearly decreasing trend toward high multipoles.

As discussed by Planck Collaboration Int. LVII (2020) and Basyrov et al. (2022), *Planck* PR4 achieves lower noise than *Planck* 2018 primarily through three modifications, namely use of the *Planck* repointing periods (which accounts for $8\,\%$ of the total data volume); use of the so-called "ninth survey" (which accounts for $3\,\%$ of the total data volume); and by smoothing the LFI reference load data prior to diode differencing. Among these, BeyondPlanck only implements the repointing period extension, for which we find no measurable issues. For the ninth survey data, on the other hand, we find that the TOD χ^2 statistics vary more strongly from PID to PID, which suggests poorer instrumental stability. We therefore exclude these data, following *Planck* 2018. Finally, we note that the decreased white noise floor that results from load smoothing also comes at a cost of increased colored (or correlated) noise at high multipoles, and this is both computationally expensive to model properly. For further discussion regarding these frequency maps and power spectra, see Basyrov et al. (2022).

9.3. Astrophysical component posteriors

We now turn our attention to the astrophysical component posteriors. However, before presenting the results, we recall that a main design feature of the current analysis was to let the LFI data play the main role in the CMB reconstruction. In practice, this means that neither the CMB-dominated HFI frequencies, nor the *WMAP K*-band observations, are included in the analysis. As a result, we note that the derived foreground posterior constraints shown here are significantly weaker than those presented by the *Planck* team in Planck Collaboration X (2016), Planck Collaboration IV (2018), and Planck Collaboration Int. LVII (2020). Full joint analysis of all data sets is left for future work.

With that caveat in mind, Fig. 21 shows the posterior mean maps for each of the four modelled temperature foregrounds, namely synchrotron, free-free, AME, and thermal dust emission. As discussed by Andersen et al. (2022), these are consistent with earlier results of the same type (Planck Collaboration X 2016), but with notably higher uncertainties, because of the more limited data set employed here.

Similarly, Fig. 22 shows the posterior mean amplitude for polarized synchrotron emission, and Fig. 23 summarizes the posterior mean (left panel) and standard deviation (right panel) for
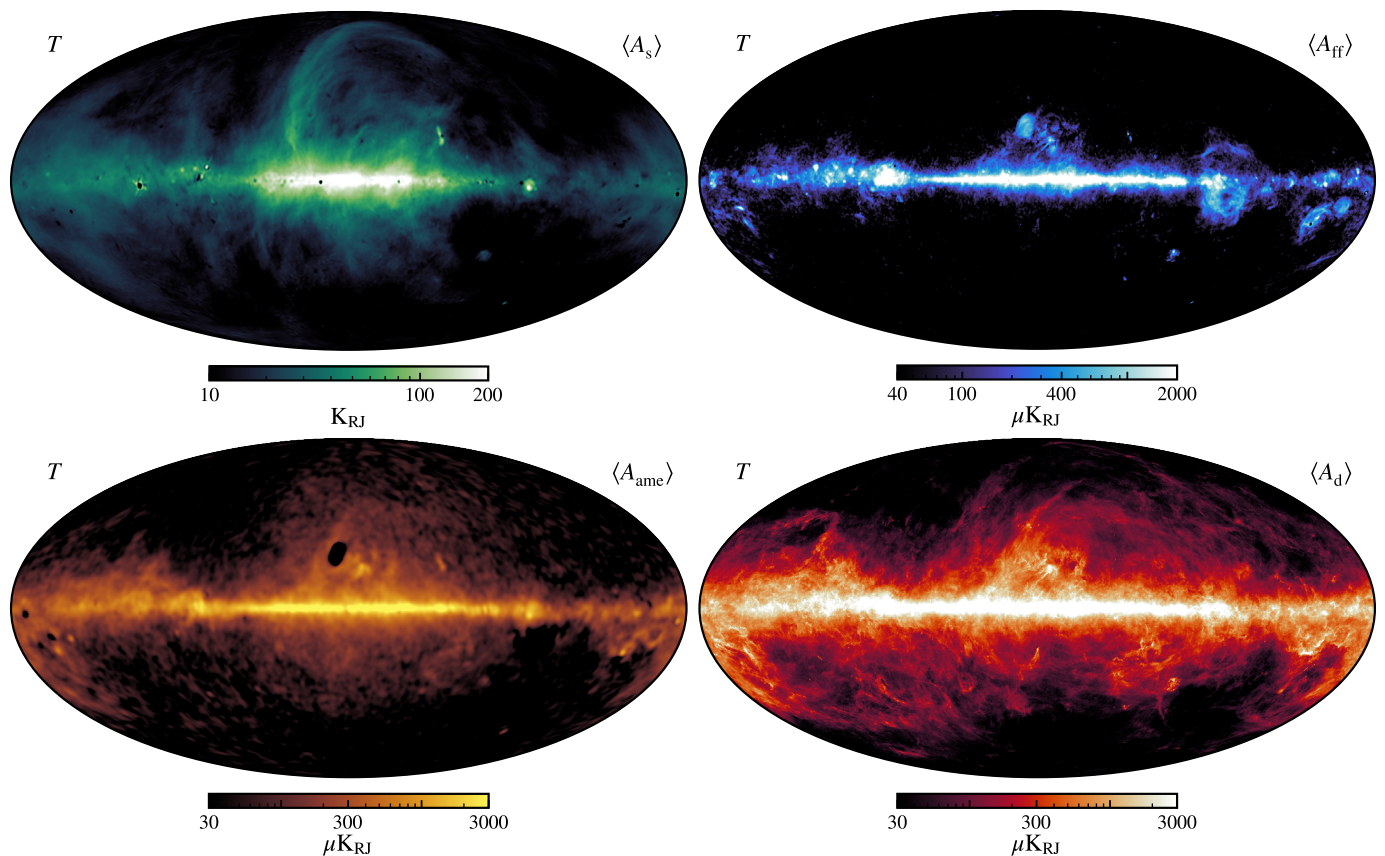
Fig. 21. Posterior mean maps of the amplitude of each of the four intensity foreground components included in the BeyondPlanck analysis. (*Top left:*) Synchrotron amplitude, evaluated at 30 GHz and smoothed to 2° FWHM resolution. (*Top right:*) Free-free amplitude, evaluated at 40 GHz and smoothed to 30′ FWHM resolution. (*Bottom left:*) AME amplitude, evaluated at 22 GHz and smoothed to 2° FWHM resolution. (*Bottom right:*) Thermal dust amplitude, evaluated at 545 GHz and smoothed to 10′ FWHM resolution. Note that the color bars vary between panels. See Andersen et al. (2022) for further discussion of these maps.
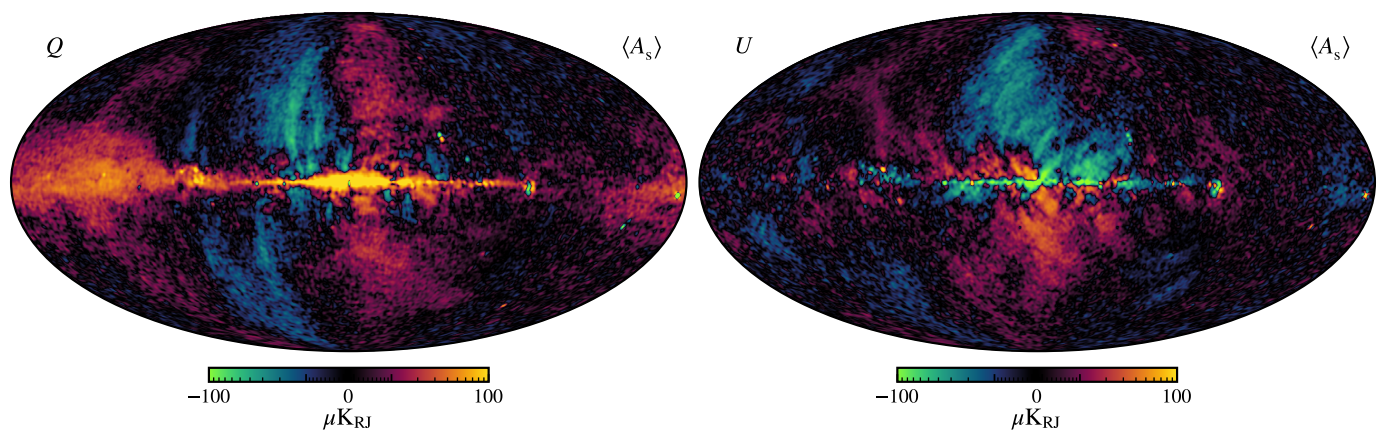
Fig. 22. Posterior mean maps of polarized synchrotron amplitude derived from BeyondPlanck, evaluated at 30 GHz and smoothed to an angular resolution of 1° FWHM. The two columns show Stokes Q and U parameters, respectively; see Svalheim et al. (2022b) for further discussion of these maps.

the power-law index of polarized synchrotron emission. In this case, it is worth pointing out that the *Planck* team never published a joint polarized synchrotron solution that included both *Planck* and *WMAP* observations, for the simple reason that these data sets could never made to agree statistically to a satisfactory degree when analyzed separately; when attempting to fit a single synchrotron spectral index across both data sets, the resulting constraints were clearly nonphysical, and led to large χ^2 excesses.

Thus, the BeyondPlanck analysis represents the first reduction of the *Planck* LFI data set for which a joint foreground polarization analysis with *WMAP* yields statistically meaningful results. However, as shown by Svalheim et al. (2022b), even the combination of the two data sets does not constrain the spectral index very strongly, and for this reason we choose to fit only a small number of independent spectral indices across the sky. Specifically, we partition the sky into four disjoint regions, corresponding to the Galactic Center (GC), the Galactic Plane
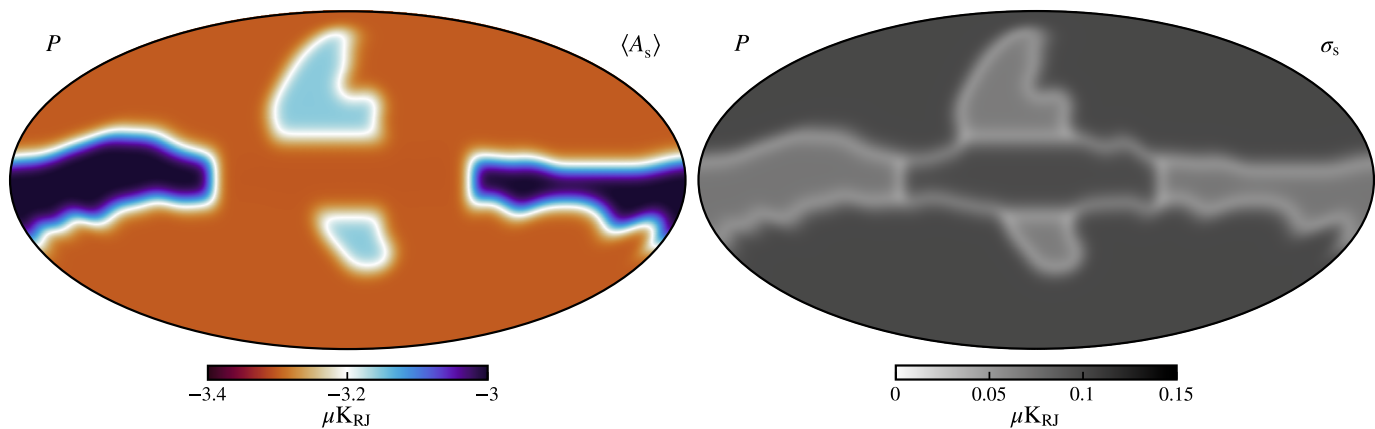
Fig. 23. Posterior mean and standard deviation maps for the spectral index of polarized synchrotron emission, β_s. Note that β_s is fitted in terms of four disjoint regions, each with a constant value but smoothed with a 10° FWHM Gaussian beam to avoid edge effects. The effect of this smoothing is seen in both the mean and standard deviation maps. Reproduced from Svalheim et al. (2022b).
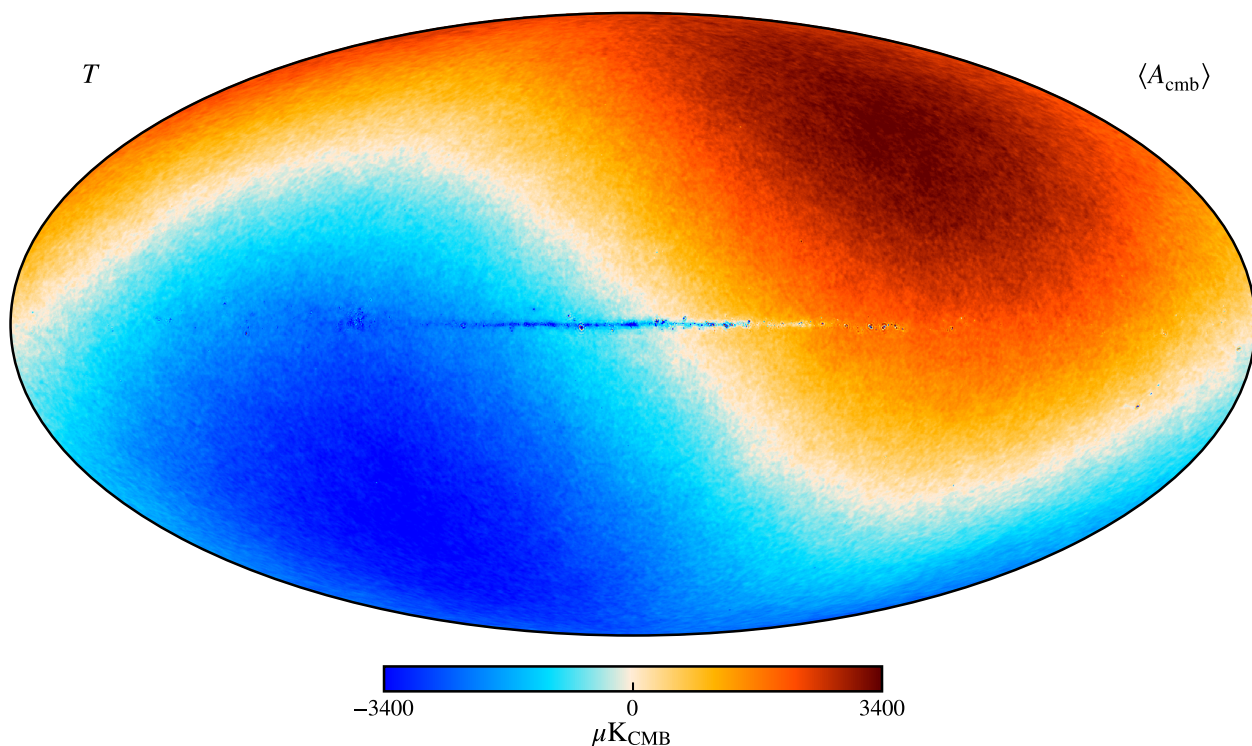

Fig. 24. Posterior mean CMB BeyondPlanck temperature map, smoothed to an angular resolution of 14′ FWHM. Reproduced from Colombo et al. (2022).

(GP), the North Galactic Spur (NGS), and High Galactic Latitudes (HGL), and treat each region separately. Adopting Planck Collaboration X (2016) as a reference, we enforce a Gaussian prior of $\beta_s \sim N(-3.1, 0.1^2)$. Finally, each spectral index sample is smoothed with a Gaussian beam of 10° FWHM to avoid edge effects.

For the GP and NGS regions, which both have significant signal-to-noise ratio with respect to polarized synchrotron emission and low systematic effects, we fit β_s using the full posterior distribution as described in Sect. 8.3.5. However, for the HGL region, in which the effective synchrotron signal-to-noise ratio is very low, we simply marginalize over the prior, and exclude the likelihood term. The reason for this is simply that unconstrained degeneracies with other parameters, such as the gain, tend to bias

β_s toward high values ($\beta_s^{\mathrm{HGL}} \approx -2.5$; see Svalheim et al. 2022b) when fitted freely.

We also do the same for the GC region, for which temperature-to-polarization leakage and bandpass effects are particularly important, and the synchrotron signal may also be biased by Faraday rotation. When fitting this region freely, we find an effective spectral index of $\beta_s^{\mathrm{GC}} \approx -4$, which is also clearly unphysical. Rather than letting these unmodelled systematic effects feed into the other components, we marginalize over the physically motivated prior.

This leaves us with two main regions usable for scientific interpretation, and these may be seen as blue regions in the standard deviation map in Fig. 23. Specifically, we find $\beta_s^{\mathrm{GP}} = -3.03 \pm 0.07$ and $\beta_s^{\mathrm{NGS}} = -3.17 \pm 0.06$, respectively (Svalheim et al. 2022b). We note that these values are broadly
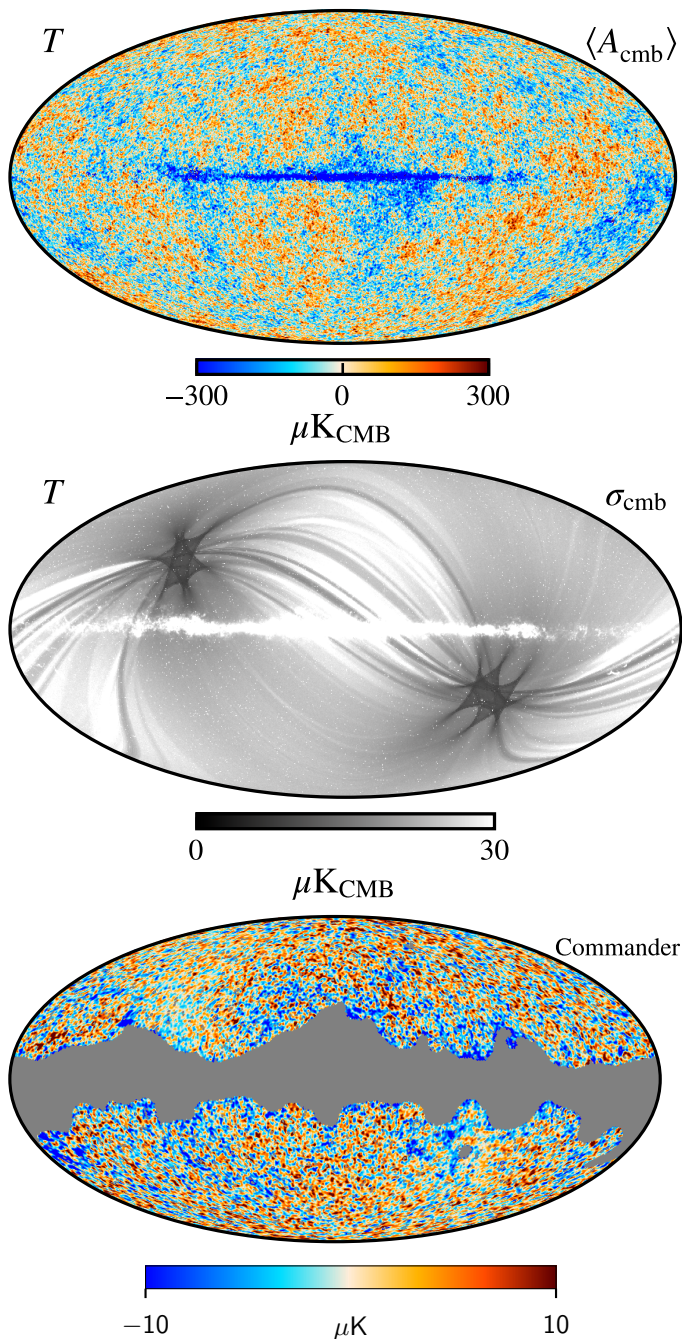
Fig. 25. Difference map between the BᴇʏᴏɴᴅPʟᴀɴᴄᴋ and the *Planck* 2018 Commander CMB temperature map (Planck Collaboration IV 2018), smoothed to a common angular resolution of 1° FWHM.

Fig. 26. CMB dipole amplitude as a function of sky fraction, reproduced from Colombo et al. (2022). The gray band indicates the 68 % posterior confidence region.

9.4. CMB temperature posteriors

Finally, we arrive at the main scientific target application of the paper, the CMB posteriors. For a full discussion of the following results, we refer the interested reader to Colombo et al. (2022); Paradiso et al. (2022), and here we only provide a brief overview of the main points.

First, Fig. 24 shows the marginal CMB temperature fluctuation posterior mean map as derived in BᴇʏᴏɴᴅPʟᴀɴᴄᴋ, given the data, model and priors described above. This map is massively dominated by the CMB solar dipole, with only a small imprint of the Galactic plane being visible in the very center. At high latitudes, CMB temperature fluctuations may be seen as tiny ripples superimposed on the dipole.

Figure 25 shows, from top to bottom, the BᴇʏᴏɴᴅPʟᴀɴᴄᴋ posterior mean temperature map (after subtracting the CMB solar dipole), the posterior standard deviation, and the difference with respect to the *Planck* 2018 Commander CMB map. Both the posterior mean and standard deviation maps show clear evidence for foreground residuals near the Galactic plane, but these are well covered by the confidence mask. The differences seen in the bottom panel range between $\pm 10\,\mu\text{K}$, which is the same order of magnitude as observed between the various component separation algorithms used internally in the *Planck* collaboration (Planck Collaboration IV 2018).

Next, Fig. 26 shows the CMB Solar dipole amplitude as derived from masks with different sky fractions, ranging between 20 and 95 %, and the value corresponding to the fiducial analysis mask of a 68 % sky fraction is tabulated in Table 3. Overall, we see that the BᴇʏᴏɴᴅPʟᴀɴᴄᴋ estimates agree well with previous estimates, both in terms of amplitude and direction, although there is a $1.5\,\sigma$ shift in the . The main difference between the BᴇʏᴏɴᴅPʟᴀɴᴄᴋ and previous results lies thus not in central values, but rather in their uncertainties. In particular, it is important to note that the CMB dipole is within the BᴇʏᴏɴᴅPʟᴀɴᴄᴋ framework estimated on completely the same footing as any other mode in the CMB sky, and is represented in terms of three spherical harmonic coefficients in s_{CMB}. No special-purpose component separation algorithms are applied to derive the CMB dipole, nor are there any *ad-hoc* instrumental uncertainty corrections involved in the estimation of the error bars; both the posterior mean

consistent with previous temperature-only constraints, such as those reported by Planck Collaboration X (2016), who found $\beta_s = -3.1$. They are also consistent with the observation that the spectral index is flatter towards the Galactic plane than in the North Galactic Spur (e.g., Kogut 2012; Fuskeland et al. 2014, 2021), although the statistical significance of this observation is marginal. In this respect, it is worth noting that the low Galactic latitudes are particularly sensitive to both systematic and astrophysical modelling errors, both in temperature and polarization. For a full discussion of these results, we refer the interested reader to Svalheim et al. (2022b).
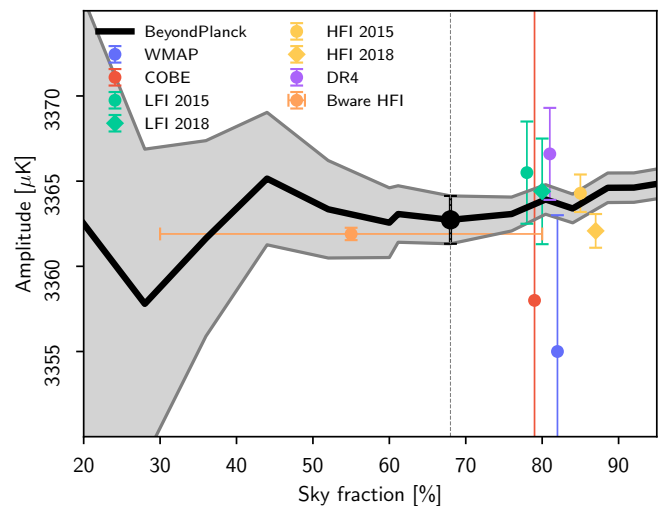
Table 3. Comparison of Solar dipole measurements from *COBE*, *WMAP*, and *Planck*.

| | | GALACTIC COORDINATES | | |
EXPERIMENT	AMPLITUDE $[\mu K_{CMB}]$	l [deg]	b [deg]	REFERENCE
COBE[a,b]	3358 ± 23	264.31 ± 0.16	48.05 ± 0.09	Lineweaver et al. (1996)
WMAP [c]	3355 ± 8	263.99 ± 0.14	48.26 ± 0.03	Hinshaw et al. (2009)
LFI 2015 [b]	3365.5 ± 3.0	264.01 ± 0.05	48.26 ± 0.02	Planck Collaboration II (2016)
HFI 2015 [d]	3364.29 ± 1.1	263.914 ± 0.013	48.265 ± 0.002	Planck Collaboration VIII (2016)
LFI 2018 [b]	3364.4 ± 3.1	263.998 ± 0.051	48.265 ± 0.015	Planck Collaboration II (2020)
HFI 2018 [d]	3362.08 ± 0.99	264.021 ± 0.011	48.253 ± 0.005	Planck Collaboration III (2020)
Bware	3361.90 ± 0.40	263.959 ± 0.019	48.260 ± 0.008	Delouis et al. (2021)
Planck PR4 [a,c] . . .	3366.6 ± 2.6	263.986 ± 0.035	48.247 ± 0.023	Planck Collaboration Int. LVII (2020)
BEYONDPLANCK [e] . .	**3362.7 ± 1.4**	**264.11 ± 0.07**	**48.279 ± 0.026**	Colombo et al. (2022)

[a] Statistical and systematic uncertainty estimates are added in quadrature.

[b] Computed with a naive dipole estimator that does not account for higher-order CMB fluctuations.

[c] Computed with a Wiener-filter estimator that estimates, and marginalizes over, higher-order CMB fluctuations jointly with the dipole.

[d] Higher-order fluctuations as estimated by subtracting a dipole-adjusted CMB-fluctuation map from frequency maps prior to dipole evaluation.

[e] Estimated with a sky fraction of 68 %. Error bars include only statistical uncertainties, as defined by the global BEYONDPLANCK posterior framework, and they thus account for instrumental noise, gain fluctuations, parametric foreground variations etc.
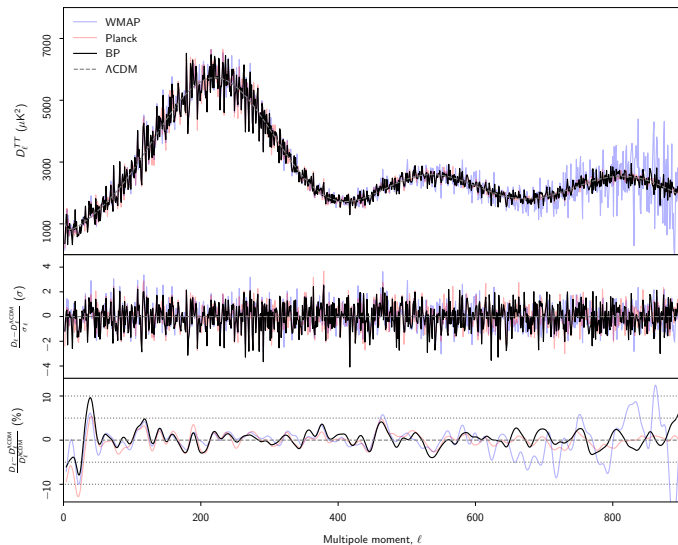

Fig. 27. (*Top*:) Angular CMB temperature power spectrum, D_ℓ^{TT}, as derived by BEYONDPLANCK (black), *Planck* (red), and *WMAP* (blue). The best-fit *Planck* 2018 ΛCDM power spectrum is shown in dashed gray. (*Middle*:) Residual power spectrum relative to ΛCDM, measured relative to full quoted error bars, $r_\ell \equiv (D_\ell - D_\ell^{\Lambda CDM})/\sigma_\ell$. For pipelines that report asymmetric error bars, σ_ℓ is taken to be the average of the upper and lower error bar. (*Bottom*:) Fractional difference with respect to the *Planck* ΛCDM spectrum. In this panel, each curve has been boxcar averaged with a window of $\Delta\ell = 100$ to suppress random fluctuations.

and standard deviation are evaluated directly from the Gibbs samples.

Figure 27 shows the angular temperature power spectrum derived from the sample set with a Blackwell-Rao estimator (Chu et al. 2005; Colombo et al. 2022), and compared with the similar spectra presented by *Planck* (Planck Collaboration V 2020) and *WMAP* (Hinshaw et al. 2013). For reference, the gray dashed line shows the best-fit *Planck* 2018 ΛCDM spectrum. The middle panel shows the difference of each measured spectrum with respect to the model spectrum in units of each pipeline's respective

error bars, while the bottom panel shows the corresponding fractional difference with respect to the best-fit *Planck* 2018 ΛCDM spectrum in units of percent. At $\ell \lesssim 500$, where these data sets are all signal-dominated, the three spectra follow each other almost ℓ-by-ℓ, while at higher multipoles, where *WMAP* becomes noise-dominated, larger variations are seen within multipoles. Overall, the agreement between the three estimates is very good, both as measured by fractional differences and in units of σ. Correspondingly, standard ΛCDM parameter constraints from the BEYONDPLANCK analysis are fully compatible with those derived from the standard *Planck* processing when considering the same multipole ranges (Paradiso et al. 2022).

9.5. CMB polarization posteriors

Turning to low-resolution polarization estimation, which we stated served as the main scientific target of the entire BEYONDPLANCK processing, we start by showing a slice through the CMB Stokes Q covariance matrix in the bottom panel Fig. 28; for comparison, the top panel shows the corresponding DPC 70 GHz covariance matrix that takes into account correlated noise and foreground template subtraction. These matrices are sliced through the pixel marked in gray in the upper right quadrant. As discussed by Colombo et al. (2022), the dominant features in the BEYONDPLANCK matrix are 1) correlated noise at high latitudes, tracking the same scanning path as seen in the DPC matrix; 2) bandpass and foreground uncertainties along the Galactic plane; and 3) gain and calibration uncertainties at high latitudes. For the LFI 30 GHz, the latter effect is in fact slightly stronger than the correlated noise contribution, as already pointed out in Fig. 17, and neglecting this contribution significantly underestimates the total uncertainties.

Figure 29 compares the posterior distribution for the optical depth of reionization, τ, with similar estimates derived using different data sets and methods, while Table 4 compares numerically τ, r, and the reduced χ^2 goodness-of-fit statistics for the same analyses. First and foremost, we see in Fig. 29 that the BEYONDPLANCK estimates of τ agree very well with previous analyses in terms of the posterior mean value for τ, with a central value

Table 4. Summary of cosmological parameters dominated by large-scale polarization and goodness-of-fit statistics. Columns list, from left to right, 1) analysis name; 2) basic data sets included in the analysis; 3) effective accepted sky fraction; 4) posterior mean estimate of the optical depth of reionization with 68 % error bars; 5) upper limit on tensor-to-scalar ratio at 95,% confidence; 6) χ^2 goodness-of-fit statistic as measured in terms of probability-to-exceed; and 7) primary reference. Reproduced from Paradiso et al. (2022).

| ANALYSIS NAME	DATA SETS	f_{sky}^{pol}	τ	$r_{95\%}^{BB}$	χ^2 PTE	REFERENCE
WMAP 9-yr	*WMAP Ka–V*	0.76	0.089 ± 0.014			Hinshaw et al. (2013)
Natale et al.	LFI 70, *WMAP Ka–V*	0.54	0.069 ± 0.011	< 0.79		Natale et al. (2020)
Planck 2018	HFI 100×143	0.50	0.050 ± 0.009	< 0.41		Planck Collaboration V (2020)
SROLL2	HFI 100×143	0.50	0.059 ± 0.006			Pagano et al. (2020)
NPIPE (Commander CMB)	LFI+HFI	0.50	0.058 ± 0.006	< 0.16		Tristram et al. (2021)
BEYONDPLANCK, $\ell = 2\text{–}8$	**LFI, *WMAP Ka–V***	**0.68**	$\mathbf{0.066 \pm 0.013}$	**< 0.84**	**0.32**	**Paradiso et al. (2022)**
BEYONDPLANCK, $\ell = 3\text{–}8$	LFI, *WMAP Ka–V*	0.68	0.066 ± 0.014	< 1.0	0.32	Paradiso et al. (2022)
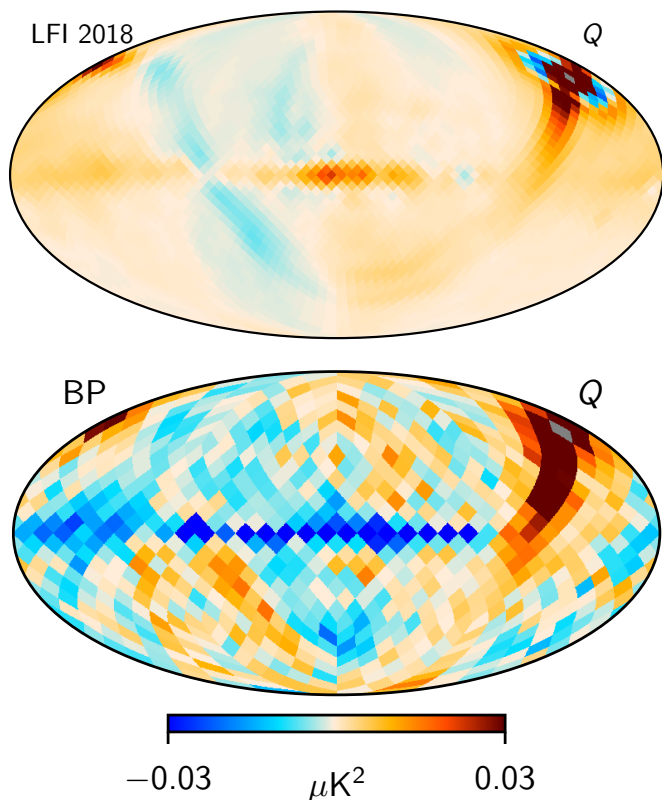
Fig. 28. Single column of the low-resolution CMB noise covariance matrix, as estimated by the LFI DPC (*top row*) and BEYONDPLANCK (*bottom row*). The column corresponds to the Stokes Q pixel marked in gray, which is located in the top right quadrant near the 'Q' label. Note that the DPC covariance matrix is constructed at $N_{side} = 16$ and includes a cosine apodization filter, while the BEYONDPLANCK matrix is constructed at $N_{side} = 8$ with no additional filter.
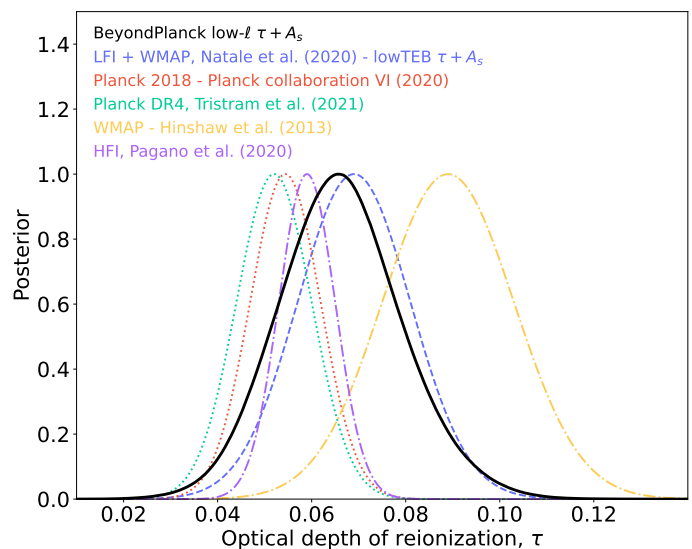

Fig. 29. Comparison of (unnormalized) marginal posterior distributions of the reionization optical depth from BEYONDPLANCK (solid black line) and various previously published estimates (colored dotted/dashed curves). Reproduced from Paradiso et al. (2022).

of $\tau = 0.066 \pm 0.013$. We also see that the tensor-to-scalar ratio is consistent with zero, and the χ^2 statistic has a probability-to-exceed of 32 %, both of which suggests that the adopted model performs well.

In particular, the agreement in terms of τ is very close with respect to Natale et al. (2020), who reported a value of $\tau = 0.069 \pm 0.011$ using an almost identical data selection. However, as discussed by Paradiso et al. (2022), it is important to note that the BEYONDPLANCK uncertainty is in fact significantly larger than that reported by Natale et al. (2020). Not only is the actually reported values 18 % larger in itself, but the BEYONDPLANCK

analysis mask is also more permissive, allowing 68 % of the sky to be used, which is to be compared with a sky fraction of 54 % in the previous analysis. When accounting for all effects, the BEYONDPLANCK uncertainty is about 30 % larger than the Natale et al. result, and we interpret this as being caused by marginalizing over a more complete set of uncertainties.

This effect is explicitly demonstrated in Fig. 30 by analyzing different models within the BEYONDPLANCK framework alone: The blue curve shows the posterior distribution when only accounting for white noise; the red curve shows the same when additionally marginalizing over foreground uncertainties; and, finally, the black curve shows the posterior distribution with marginalizing over all sources of uncertainty; white noise, foregrounds, and instrumental parameters. Neglecting the instrumental effects, such as gain and bandpass uncertainties, significantly under-estimates the true uncertainty, and properly accounting for all these effects is a main goal of the BEYONDPLANCK framework.

10. Reproducibility and Open Science

As discussed in Sect. 1, the main long-term scientific goal and motivation of the BEYONDPLANCK program is to establish an end-
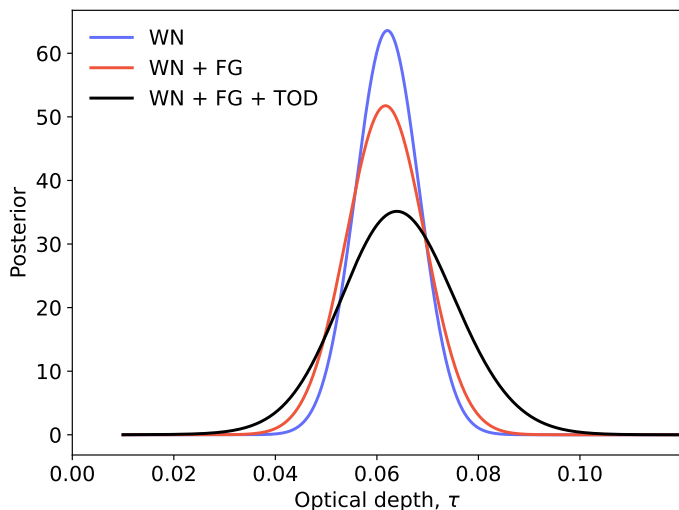
Fig. 30. Estimates of τ under different uncertainty assumptions. The blue curve shows marginalization over white noise only; the red curve shows marginalization over white noise and astrophysical uncertainties; and, finally, the black curve shows marginalization over all contributions, including low-level instrumental uncertainties, as in the final BeyondPlanck analysis.

to-end analysis framework for CMB observations that, we hope, will be useful for the general community. This framework is designed to be sufficiently flexible to allow analysis of different and complementary experiments, and thereby exploit the strengths of one instrument to break degeneracies in another. Organizing and promoting this work is a key goal for the Cosmoglobe project, which aims at performing a similar analysis as BeyondPlanck, but for a much wider selection of datasets, and eventually build a single statistically consistent and multi-experiment model of the radio, microwave, and sub-millimeter sky. An early example of this work has already been demonstrated in the current paper suite by Watts et al. (2022), who have extended the current framework to re-analyze the *WMAP Q*-band data with very promising results.

For this project to succeed, substantial efforts have been spent within the BeyondPlanck program on the issue of *reproducibility*. These efforts are summarized by Gerakakis et al. (2022), both in terms of the internal process itself and some lessons learned, and also in terms of the final practical solutions that have been implemented. Here we provide a brief summary of the main points.

10.1. Reproducibility

For the BeyondPlanck and Cosmoglobe framework to be useful for other experiments it must be *reproducible*: Researchers outside of the current collaboration must be able to repeat our analysis, before improving and extending it. To support this, we have focused on four main items:

1. *Documented open-source code* – the full `Commander`[22] source code, as well as various pre- and post-processing tools,[23] are made publicly available in a GitHub repository under a GPL license, and may be freely downloaded and extended within the general restriction of that license. Pre-

liminary documentation is provided,[24] although it is under continuous development, as is the source code itself.
2. *Cmake compilation* – easy compilation is supported through the Cmake environment; required external libraries are automatically downloaded and compiled.
3. *Data downloader* – a Python-based tool is provided that automatically downloads all BeyondPlanck input data to a user-specified directory, together with the parameter files that are needed to run the code.

In addition, all main results (both full chain files and selected post-processed posterior mean and standard deviation maps) are available from the Cosmoglobe homepage,[25] and eventually through the *Planck* Legacy Archive.[26] For further details regarding the reproducibility aspects of the work, we refer the interested reader to Gerakakis et al. (2022).

10.2. Software

A second requirement for the BeyondPlanck framework to be useful for other users is that the software is computationally efficient so that it can be run on generally available hardware, and also that the source code is extendable without expert knowledge. Regarding the former point, we note that great emphasis has been put on minimizing the required computational resources throughout the implementation. This appears to be at least partially successful, as summarized in Sect. 8.4 and by Galloway et al. (2022a): The full BeyondPlanck analysis, as presented here, has a computational cost of 670 000 CPU hours, which is roughly equivalent to the cost of producing 70 end-to-end *Planck* FFP8 70 GHz realizations using the traditional pipeline (Planck Collaboration XII 2016). Furthermore, by compressing the TOD inputs the memory footprint of the LFI data set has been reduced by about an order of magnitude (see Table 2 and Galloway et al. 2022a), and now requires only about 1.5 TB of RAM to run. Computers with this amount of memory and clock cycles are now available outside super-computing centers, and a full *Planck* LFI analysis therefore no longer requires the use of expensive super-computers – although they will of course be beneficial when available.

Regarding the software itself, the current main code base is written in object-oriented Fortran 2003. Clearly, this may represent a significant hurdle for many users, as most astrophysics students today are typically more exposed to languages like Python or C than Fortran. This choice of language is primarily historical, and due to the fact that a large part of the legacy code base was originally written in Fortran, most notably `HEALPix` (Górski et al. 2005) and `Commander` (Eriksen et al. 2004, 2008). However, a second important motivation for adopting Fortran is that it remains one of the fastest languages even today in terms of computational speed and memory management. As far as readability and extendability goes, the code has been designed with a strong focus on object-orientation, and we believe that adding support for several types of new sub-classes is relatively straight-forward. This includes classes for new signal components; noise or beam representations; or TOD models. On the other hand, modifying the underlying memory management, component separation infrastructure, or parallelization paradigm, is likely to be difficult without expert knowledge. A guide to the current software is provided by Galloway et al. (2022a). As a real-world demonstration of the extendability of

---

the framework, we present a preliminary application to *WMAP* in (Watts et al. 2022).

As useful as we hope the current version will be, we do believe that developing a massively parallel version of `Commander` in Python would be a useful, interesting and intellectually challenging task, and we would encourage (and support!) work in this direction. For reference, the current `Commander` Fortran source code spans 45 000 lines,[27] which can likely be reduced by a significant factor if written in a less verbose language; porting this to Python would obviously be a major undertaking, but certainly feasible for even just a small team of talented and motivated researchers.

11. Conclusions, summary and outlook

The *Planck* project represents a landmark achievement in international cosmology, mapping out the primary temperature fluctuations in the CMB to a precision determined by astrophysical constraints. This achievement was made possible by the dedication and long-term contributions from ESA and NASA; from tens of national funding agencies; and many hundreds of scientists and engineers working for more than two decades. At the end of the mission, a massive amount of knowledge and expertise regarding optimal analysis of CMB experiments had been generated within the collaboration, as clearly demonstrated through more than 150 refereed scientific publications.

A central goal of the BEYONDPLANCK project was to translate a significant part of this aggregated experience into a practical computer code that can analyse *Planck* data from end-to-end, and to make this code available to the community in general. Main advantages of an integrated end-to-end Bayesian approach include:

1. *Faithful error propagation*: BEYONDPLANCK implements global end-to-end Bayesian CMB analysis framework. The single most important advantage of this is faithful propagation of uncertainties from raw TOD to final cosmological parameters. Instrumental and astrophysical systematic errors are propagated to the final CMB likelihood on the same footing as any other nuisance parameter. While already important for *Planck*, this issue will become absolutely critical for future planned high-precision *B*-mode experiments, such as *LiteBIRD* or *PICO*.

2. *Breaking degeneracies and saving costs by exploiting synergistic observations*: Combining data from complementary sources is essential to break fundamental degeneracies within a given experiment. For instance, both *Planck* and *WMAP* have degenerate polarization modes that they cannot measure well on their own, due to peculiarities in their respective scanning strategies—but there are no degenerate modes in the combined data set. In general, however, the usefulness of joint analysis with external data is often limited by systematic errors. The BEYONDPLANCK framework addresses this by providing a common platform for performing joint low-level analysis of different experiments. Also noting that the lion's share of the analysis cost of any real-world CMB experiment is associated with understanding degeneracies and systematic errors, we believe that a global approach will lead to better and cheaper science for each experiment.

3. *Fewer human errors*: Tight analysis integration also leads to many important practical advantages, including less room for human errors or miscommunication; greater transparency of both explicit and implicit priors; better optimization of computing resources; and significantly reduced end-to-end wall-clock time by eliminating intermediate human interaction.

4. *"Faster, better and cheaper" through open-source science*: True inter-experiment global analysis will clearly not succeed without active contributions and support from a large part of the general community. For this reason, we make our source codes publicly available under a GPL open-source license to ensure long-term stability of the currently released software. It also means that future improvements must be released under a similarly generous license, in recognition of the fact that this project is intended to be collaborative, open, and inclusive. The use of stakeholder knowledge is critically important—and we hope that many stakeholders will indeed be interested in actively contributing to the program, ultimately leading to "faster, better, and cheaper" science for everyone.

As discussed above, the BEYONDPLANCK program has focused on the *Planck* LFI data. The reasons for doing so were three-fold. First and foremost, many BEYONDPLANCK collaborators have been working with the LFI data for one or two decades, and the aggregated experience with this data set within the collaboration implied a low start-up cost; results could be produced quickly. Second, the full LFI data volume is fairly limited in size, comprising less than 1 TB after compression, which is good for fast debugging and testing. Third, the LFI instrument is based on HEMT radiometers, which generally both have a relatively high noise contribution and low systematic errors per sample. The combination of these three points made LFI a natural starting point for the work.

However, now that the computational framework already exists, it will require substantially less effort to generalize it to other and complementary data sets. Currently on-going generalizations by various teams involve *WMAP*, *LiteBIRD*, CHIPASS, and COBE-DIRBE, and we warmly welcome initiatives targeting any other experiment as well. In this respect, it may be useful to distinguish between four types of experiments, each with their own set of algorithmic complexities.

First, many radio, microwave and sub-millimeter experiments may be modelled within nearly the same sky and instrument model as BEYONDPLANCK. Examples include C-BASS, QUIET and QUIJOTE, all of which simply provide additional signal-to-noise and/or frequency coverage, as far as the underlying algorithms are concerned. For these, analysis within the BEYONDPLANCK framework may turn out to amount simply to writing one or more TOD processing modules (for instance using the current LFI module as a template) to take into account the various instrument-specific systematic effects of the experiment in question. These experiments should be, relatively speaking, the easiest to integrate into the current framework.

Other experiments may build on the same sky model and component separation procedures as BEYONDPLANCK, but require a different mapmaking algorithm. One prominent example of this is *WMAP*, which is differential in nature, and therefore requires a different Conjugate Gradient mapmaking algorithm to translate cleaned TOD into pixelized maps; this work is already on-going (Watts et al. 2022). Experiments of this type should also be relatively straightforward to integrate.

A third class of experiments are those that can use the same type of sky models, but requires a significantly different instrumental model. The most prominent example of such are TES bolometer-based instruments. These often have both higher

---

[27] Interestingly, only about 6000 lines are directly associated with TOD processing, while 14 000 lines are directly associated with component separation; the rest is spent on general data infrastructure and tools.
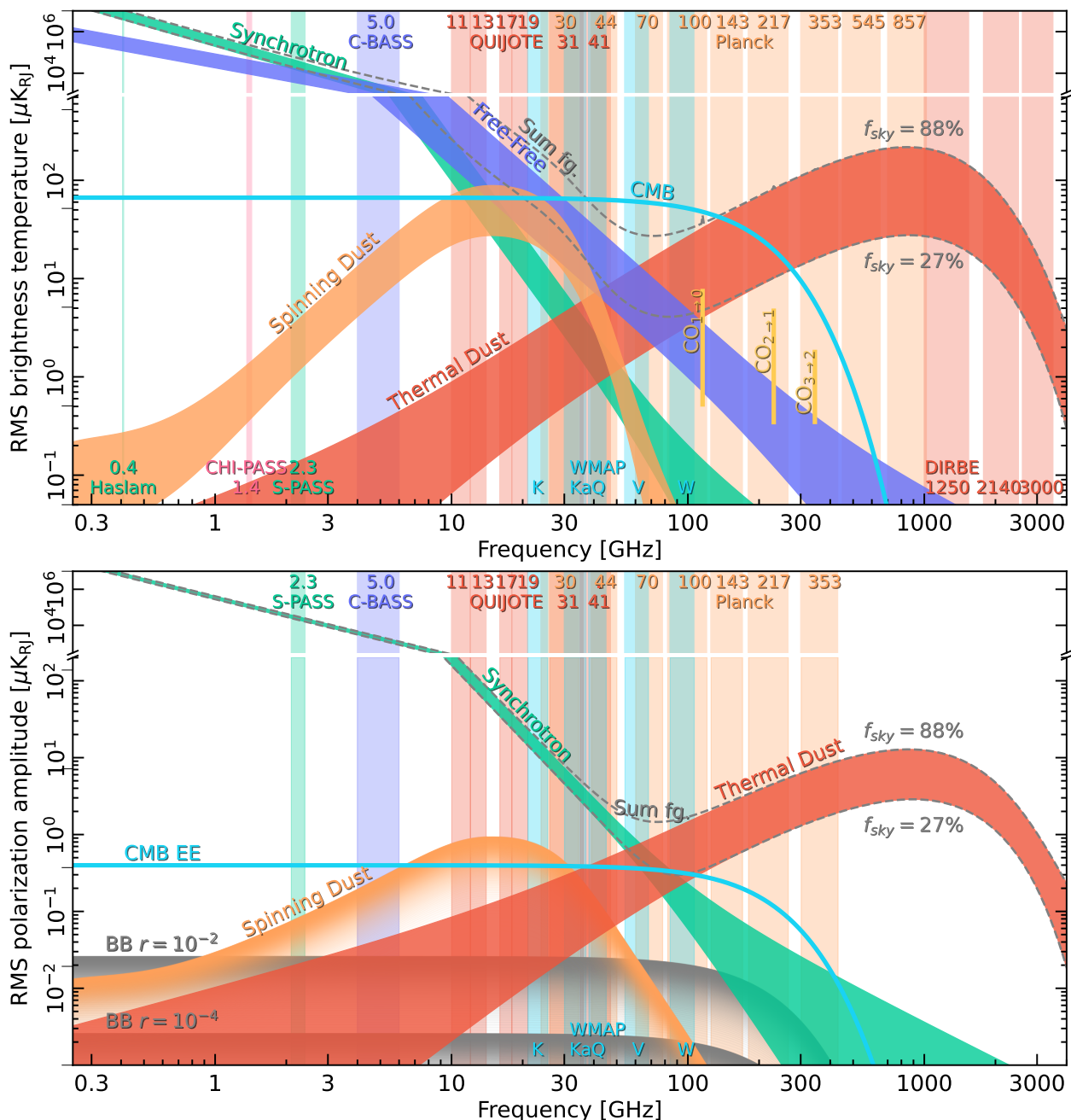
Fig. 31. Brightness temperature (*top panel*) and polarization amplitude (*bottom panel*) RMS as a function of frequency and astrophysical component, and as derived from the BᴇʏᴏɴᴅPʟᴀɴᴄᴋ and *Planck* sky models. Vertical bands indicate the frequency ranges of various experiment. All components have been evaluated at a common angular resolution of 1° FWHM. The widths of each confidence region correspond to the spread when evaluating the RMS over two different masks with sky fractions of 88 and 27 %, respectively. The cyan curve shows the level of CMB fluctuations evaluated for the best-fit *Planck* ΛCDM spectrum. In the current BᴇʏᴏɴᴅPʟᴀɴᴄᴋ analysis, only the three LFI channels are modelled in the time-domain. For polarization, the spinning dust component (orange curve) indicates an upper limit as presented by Herman et al. (2022b), not a detection. A long-term future goal is to include all publicly available and relevant data (for instance *WMAP* and *Planck* HFI) into this model; preferably in the form of time-ordered data, but if this is not technically or financially possible, then at least in the form of pre-processed sky maps. This work will be organized within the Cᴏsᴍᴏɢʟᴏʙᴇ project.

signal-to-noise ratios and systematic errors per sample, and therefore require a richer set of systematics corrections. They also typically have a significant multiplicative transfer function, which means that unbiased maps cannot be produced simply by introducing additive TOD corrections, as is done in the current implementation. Instead, they will also require a dedicated Conjugate Gradient mapmaker to take into account the multiplicative effects. Examples of potentially relevant experiments include for

instance BICEP2, CLASS, SPIDER, and *LiteBIRD*. Integrating these will thus be more challenging than HEMT-based experiments like LFI or *WMAP*, but it should certainly be feasible, and the scientific rewards will be massive.

The fourth and final group of experiments are those that either produce massive amounts of time-ordered data, or very high-resolution data. Important examples are ACT , SPT, Simons Observatory, and CMB-S4. These will all require a fundamental

redesign of the existing code base, simply to handle the massive amounts of memory and network communication efficiently. Additionally, experiments that observe only a fraction of the sky, but at high angular resolution, cannot employ the spherical harmonics basis that we currently use for component separation without introducing large degeneracies and singular modes; all spatial modes need to be constrained by at least one experiment for the current implementation to work properly. Developing a new version of the Bayesian framework that can handle higher levels of parallelization, and also use more general basis sets, is thus an important goal for future work.

Returning to the specific scientific results derived by the BeyondPlanck project, we note that cosmological constraints derived from LFI and *WMAP* alone will never be competitive in terms of overall uncertainties as compared to an HFI-based analysis. Nevertheless, many interesting results have been established during the course of the project. Some of the most noteworthy among these are the following:

1. We have succeeded in integrating the LFI 44 GHz channel into a statistically viable low-ℓ CMB likelihood. Two key algorithmic steps that were required for this were, firstly, to implement a more flexible gain smoothing algorithm stiff than a hard boxcar average, and, secondly, to model a previously unknown additional correlated noise contribution on 0.01–1 sec time scales. Less important, but still notable, improvements include more efficient ADC corrections and removal of known bandpass artefacts. After making all these corrections, we are not currently aware of any significant outstanding systematic effects in the *Planck* LFI data.

2. We have for the first time constructed full dense low-resolution CMB and frequency map covariance matrices that account for marginalization over a wide range of important systematic time-ordered effects, including gain, bandpass, and foreground corrections, and correlated noise. This results in a low-ℓ polarization likelihood that yields results consistent with the latest HFI analyses, and a best-fit value of the reionization optical depth of $\tau = 0.066 \pm 0.013$. The associated χ^2 goodness-of-fit statistics are statistically acceptable, although there might be weak hints of excess power, possibly due to the break-down of the $1/f$ noise model.

3. We have produced a statistically consistent and joint estimate of the CMB dipole using both *Planck* and *WMAP* data. The best-fit dipole amplitude of $3362.7 \pm 1.4 \,\mu$K is consistent with all published results, including the latest HFI-based measurements, and the quoted error estimate is derived strictly within the well-defined Bayesian statistical framework.

4. We are for the first time able to fit a physically meaningful spectral index of polarized synchrotron emission using both *WMAP* and *Planck*. This is the direct result of performing a truly joint analysis with LFI and *WMAP* as described above, using information from one experiment to break degeneracies within the other.

While the BeyondPlanck project itself contractually ended on November 30th, 2020, the work will in general continue with various alternative funding sources, and, we hope, also with the help of a continuously growing community of supporting collaborators and experiments. Figure 31 shows a compilation of the current BeyondPlanck sky model and data sets in both temperature (top panel) and polarization (bottom panel) together with selected external products. The long-term goal of this work is to populate this plot with all available experimental data, and thereby gradually refine the sky model. The ERC-funded CosmoGlobe project aims to coordinate these efforts, and will serve

as a stable platform for all parties interested in global Bayesian CMB analysis. CosmoGlobe will also serve as the long-term home for all BeyondPlanck material and products, long after the current BeyondPlanck web portal vanishes.

Finally, we end with an important *caveat emptor*, and emphasize that `Commander` is very much a work-in-progress—and it will remain so for all foreseeable future. Essentially every single step in the pipeline can and will be replaced by smarter and more capable sampling algorithms; the user-interface could most certainly be made more intuitive; and so on. This is an unavoidable side-effect of being at the cutting edge of algorithmic research, where new ideas are continuously being explored, implemented and tested. However, at the same time, it is also our belief that the current platform is now finally sufficiently mature to allow external users and developers to use it productively for their own analyses, and to extend it as they see fit. In other words, we believe that now is the right time for Bayesian end-to-end CMB analysis to go OpenSource, and we invite all interested parties to participate in this work.

References

Ade, P. A. R., Ahmed, Z., Amiri, M., et al. 2021, Phys. Rev. Lett., 127, 151301
Ali-Haïmoud, Y. 2010, Astrophysics Source Code Library [`ascl:1010.016`]
Ali-Haïmoud, Y., Hirata, C. M., & Dickinson, C. 2009, MNRAS, 395, 1055
Alvarez, M. A., Komatsu, E., Doré, O., & Shapiro, P. R. 2006, ApJ, 647, 840
Andersen et al. 2022, A&A, in press [`arXiv:2201.08188`]
Andersson, B. G., Lazarian, A., & Vaillancourt, J. E. 2015, ARA&A, 53, 501
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
Basyrov et al. 2022, A&A, submitted [`arXiv:2208.14293`]
Beck, R., Anderson, J., Heald, G., et al. 2013, Astronomische Nachrichten, 334, 548
Bennett, C. L., Larson, D., Weiland, J. L., et al. 2013, ApJS, 208, 20
Bersanelli, M., Mandolesi, N., Butler, R. C., et al. 2010, A&A, 520, A4
Bertsekas, D. P. 1996, Constrained Optimization and Lagrange Multiplier Methods (Optimization and Neural Computation Series), 1st edn. (Athena Scientific)
BeyondPlanck. 2022, A&A, submitted [`arXiv:2011.05609`]
Brilenkov et al. 2022, A&A, submitted [`arXiv:2209.04437`]
Calabretta, M. R., Staveley-Smith, L., & Barnes, D. G. 2014, PASA, 31, e007
Carretti, E., Haverkorn, M., Staveley-Smith, L., et al. 2019, MNRAS, 489, 2330
Chon, G., Challinor, A., Prunet, S., Hivon, E., & Szapudi, I. 2004, MNRAS, 350, 914
Chu, M., Eriksen, H. K., Knox, L., et al. 2005, Phys. Rev. D, 71, 103002
Colombo et al. 2022, A&A, submitted [`arXiv:2208.14276`]
Condon, J. J., Cotton, W. D., Greisen, E. W., et al. 1998, AJ, 115, 1693
Datta, R., Aiola, S., Choi, S. K., et al. 2019, MNRAS, 486, 5239
de Bernardis, P., Ade, P. A. R., Bock, J. J., et al. 2000, Nature, 404, 955
Delouis, J. M., Pagano, L., Mottet, S., Puget, J. L., & Vibert, L. 2019, A&A, 629, A38
Delouis, J. M., Puget, J. L., & Vibert, L. 2021, A&A, 650, A82
Dickinson, C., Davies, R. D., & Davis, R. J. 2003, MNRAS, 341, 369
Dolginov, A. Z. & Mitrofanov, I. G. 1976, Ap&SS, 43, 291
Draine, B. T. 2011, Physics of the Interstellar and Intergalactic Medium (Princeton University Press)
Draine, B. T. & Hensley, B. S. 2016, The Astrophysical Journal, 831, 59
Draine, B. T. & Lazarian, A. 1998, ApJ, 494, L19
Erickson, W. C. 1957, ApJ, 126, 480
Eriksen, H. K., Jewell, J. B., Dickinson, C., et al. 2008, ApJ, 676, 10
Eriksen, H. K., O'Dwyer, I. J., Jewell, J. B., et al. 2004, ApJS, 155, 227

Fixsen, D. J. 2009, ApJ, 707, 916

Frigo, M. & Johnson, S. G. 2005, Proceedings of the IEEE, 93, 216, special issue on "Program Generation, Optimization, and Platform Adaptation"

Fuskeland, U., Andersen, K. J., Aurlien, R., et al. 2021, A&A, 646, A69

Fuskeland, U., Wehus, I. K., Eriksen, H. K., & Næss, S. K. 2014, ApJ, 790, 104

Galloway et al. 2022a, A&A, in press [arXiv:2201.03509]

Galloway et al. 2022b, A&A, in press [arXiv:2201.03478]

Geman, S. & Geman, D. 1984, IEEE Trans. Pattern Anal. Mach. Intell., 6, 721

Génova-Santos, R., Rubiño-Martín, J. A., Peláez-Santos, A., et al. 2017, MN-RAS, 464, 4107

Génova-Santos, R., Rubiño-Martín, J. A., Rebolo, R., et al. 2015, MNRAS, 452, 4169

Gerakakis et al. 2022, arxiv [arXiv:2205.11262]

Girolami, M. & Calderhead, B. 2011, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73, 123

Gjerløw, E., Colombo, L. P. L., Eriksen, H. K., et al. 2015, ApJS, 221, 5

Gjerløw et al. 2022, A&A, submitted [arXiv:2011.08082]

Godard, B., Croon, M., Budnik, F., & Morley, T. 2009, in Proceedings of the 21st International Symposium on Space Flight Dynamics (ISSFD), Toulouse

Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759

Greaves, J. S., Holland, W. S., Friberg, P., & Dent, W. R. F. 1999, The Astrophysical Journal, 512, L139

Gregory, P. C., Scott, W. K., Douglas, K., & Condon, J. J. 1996, ApJS, 103, 427

Gualtieri, R., Filippini, J. P., Ade, P. A. R., et al. 2018, Journal of Low Temperature Physics, 193, 1112

Guillet, V., Fanciullo, L., Verstraete, L., et al. 2018, A&A, 610, A16

Gunn, J. E. & Peterson, B. A. 1965, ApJ, 142, 1633

Haslam, C. G. T., Salter, C. J., Stoffel, H., & Wilson, W. E. 1982, A&AS, 47, 1

Hastings, W. K. 1970, Biometrika, 57, 97

Hauser, M. G., Arendt, R. G., Kelsall, T., et al. 1998, ApJ, 508, 25

Hauser, M. G. & Dwek, E. 2001, ARA&A, 39, 249

Hensley, B. S. & Draine, B. T. 2020, ApJ, 895, 38

Hensley, B. S. & Draine, B. T. 2022, arXiv e-prints, arXiv:2208.12365

Herman et al. 2022a, A&A, in press [arXiv:2203.13635]

Herman et al. 2022b, A&A, in press [arXiv:2201.03530]

Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, ApJS, 208, 19

Hinshaw, G., Weiland, J. L., Hill, R. S., et al. 2009, ApJS, 180, 225

Hoang, T., Lazarian, A., & Martin, P. G. 2013, ApJ, 779, 152

Hu, W. & Dodelson, S. 2002, ARA&A, 40, 171

Huffman, D. A. 1952, Proceedings of the IRE, 40, 1098

Ihle et al. 2022, A&A, in press [arXiv:2011.06650]

Jarosik, N., Barnes, C., Greason, M. R., et al. 2007, ApJS, 170, 263

Jarosik, N., Bennett, C. L., Dunkley, J., et al. 2011, ApJS, 192, 14

Jeffreys, H. 1946, Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 186, 453

Jewell, J., Levin, S., & Anderson, C. H. 2004, ApJ, 609, 1

Jewell, J. B., Eriksen, H. K., Wandelt, B. D., et al. 2009, ApJ, 697, 258

Kamionkowski, M. & Kovetz, E. D. 2016, ARA&A, 54, 227

Keating, B., Timbie, P., Polnarev, A., & Steinberger, J. 1998, ApJ, 495, 580

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

Keihänen, E., Kurki-Suonio, H., Poutanen, T., Maino, D., & Burigana, C. 2004, A&A, 428, 287

Keihänen et al. 2022, A&A, in press [arXiv:2011.06024]

Kelsall, T., Weiland, J. L., Franz, B. A., et al. 1998, ApJ, 508, 44

King, O. G., Jones, M. E., Blackhurst, E. J., et al. 2014, MNRAS, 438, 2426

Kogut, A. 2012, ApJ, 753, 110

Kogut, A., Spergel, D. N., Barnes, C., et al. 2003, ApJS, 148, 161

Komatsu, E., Dunkley, J., Nolta, M. R., et al. 2009, ApJS, 180, 330

Kovac, J. M., Leitch, E. M., Pryke, C., et al. 2002, Nature, 420, 772

Larson, D., Dunkley, J., Hinshaw, G., et al. 2011, ApJS, 192, 16

Larson, D. L., Eriksen, H. K., Wandelt, B. D., et al. 2007, ApJ, 656, 653

Leach, S. M., Cardoso, J., Baccigalupi, C., et al. 2008, A&A, 491, 597

Leitch, E. M., Readhead, A. C. S., Pearson, T. J., & Myers, S. T. 1997, ApJ, 486, L23

Lewis, A. & Bridle, S. 2002, Phys. Rev. D, 66, 103511

Lewis, A., Challinor, A., & Lasenby, A. 2000, ApJ, 538, 473

Lineweaver, C. H., Tenorio, L., Smoot, G. F., et al. 1996, ApJ, 470, 38

Liu, J. S. 2008, Monte Carlo Strategies in Scientific Computing (Springer Publishing Company, Incorporated)

Loeb, A. & Barkana, R. 2001, Annual Review of Astronomy and Astrophysics, 39, 19

Macellari, N., Pierpaoli, E., Dickinson, C., & Vaillancourt, J. E. 2011, MNRAS, 418, 888

Macià Escatllar, A. & Bromley, S. T. 2020, A&A, 634, A77

Maino, D., Burigana, C., Maltoni, M., et al. 1999, A&AS, 140, 383

Mandolesi, N., Bersanelli, M., Butler, R. C., et al. 2010, A&A, 520, A3

Mather, J. C., Cheng, E. S., Cottingham, D. A., et al. 1994, ApJ, 420, 439

Mennella, A., Butler, R. C., Curto, A., et al. 2011, A&A, 536, A3

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. 1953, J. Chem. Phys., 21, 1087

Mitra, S., Rocha, G., Górski, K. M., et al. 2011, ApJS, 193, 5

Murphy, T., Sadler, E. M., Ekers, R. D., et al. 2010, MNRAS, 402, 2403

Natale, U., Pagano, L., Lattanzi, M., et al. 2020, A&A, 644, A32

Notari, A. & Quartin, M. 2015, Journal of Cosmology and Astroparticle Physics, 2015, 047–047

Pagano, L., Delouis, J. M., Mottet, S., Puget, J. L., & Vibert, L. 2020, A&A, 635, A99

Page, L., Hinshaw, G., Komatsu, E., et al. 2007, ApJS, 170, 335

Paradiso et al. 2022, A&A, submitted [arXiv:2205.10104]

Penzias, A. A. & Wilson, R. W. 1965, ApJ, 142, 419

Planck Collaboration. 2005, ESA publication ESA-SCI(2005)/01 [arXiv:astro-ph/0604069]

Planck Collaboration II. 2011, A&A, 536, A2

Planck Collaboration I. 2014, A&A, 571, A1

Planck Collaboration III. 2014, A&A, 571, A3

Planck Collaboration IV. 2014, A&A, 571, A4

Planck Collaboration VI. 2014, A&A, 571, A6

Planck Collaboration IX. 2014, A&A, 571, A9

Planck Collaboration XI. 2014, A&A, 571, A11

Planck Collaboration XV. 2014, A&A, 571, A15

Planck Collaboration XVI. 2014, A&A, 571, A16

Planck Collaboration XVII. 2014, A&A, 571, A17

Planck Collaboration I. 2016, A&A, 594, A1

Planck Collaboration II. 2016, A&A, 594, A2

Planck Collaboration IV. 2016, A&A, 594, A4

Planck Collaboration V. 2016, A&A, 594, A5

Planck Collaboration VII. 2016, A&A, 594, A7

Planck Collaboration VIII. 2016, A&A, 594, A8

Planck Collaboration IX. 2016, A&A, 594, A9

Planck Collaboration X. 2016, A&A, 594, A10

Planck Collaboration XI. 2016, A&A, 594, A11

Planck Collaboration XII. 2016, A&A, 594, A12

Planck Collaboration XXV. 2016, A&A, 594, A25

Planck Collaboration XXVI. 2016, A&A, 594, A26

Planck Collaboration I. 2020, A&A, 641, A1

Planck Collaboration II. 2020, A&A, 641, A2

Planck Collaboration III. 2020, A&A, 641, A3

Planck Collaboration IV. 2018, A&A, 641, A4

Planck Collaboration V. 2020, A&A, 641, A5

Planck Collaboration VI. 2020, A&A, 641, A6

Planck Collaboration VII. 2020, A&A, 641, A7

Planck Collaboration IX. 2020, A&A, 641, A9

Planck Collaboration XI. 2020, A&A, 641, A11

Planck Collaboration Int. XLVI. 2016, A&A, 596, A107

Planck Collaboration Int. XLVII. 2016, A&A, 596, A108

Planck Collaboration Int. XLVIII. 2016, A&A, 596, A109

Planck Collaboration Int. LVII. 2020, A&A, 643, A42

Planck Collaboration LVII. 2020, A&A, 643, A42

Prézeau, G. & Reinecke, M. 2010, ApJS, 190, 267

Racine, B., Jewell, J. B., Eriksen, H. K., & Wehus, I. K. 2016, ApJ, 820, 31

Reinecke, M. & Seljebotn, D. S. 2013, A&A, 554, A112

Remazeilles, M., Dickinson, C., Banday, A. J., Bigot-Sazy, M.-A., & Ghosh, T. 2015, MNRAS, 451, 4311

Rudjord, Ø., Groeneboom, N. E., Eriksen, H. K., et al. 2009, ApJ, 692, 1669

San, M., Herman, D., Erikstad, G. B., Galloway, M., & Watts, D. 2022, arXiv e-prints, arXiv:2205.12962

Seljak, U. & Zaldarriaga, M. 1996, ApJ, 469, 437

Seljebotn, D. S., Bærland, T., Eriksen, H. K., Mardal, K. A., & Wehus, I. K. 2019, A&A, 627, A98

Sellentin, E. & Heavens, A. F. 2016, MNRAS, 456, L132

Shewchuk, J. R. 1994, An Introduction to the Conjugate Gradient Method Without the Agonizing Pain, Edition $1\frac{1}{4}$, http://www.cs.cmu.edu/~quake-papers/painless-conjugate-gradient.pdf

Silsbee, K., Ali-Haïmoud, Y., & Hirata, C. M. 2011, MNRAS, 411, 2750

Smoot, G. F., Bennett, C. L., Kogut, A., et al. 1992, ApJ, 396, L1

Stivoli, F., Grain, J., Leach, S. M., et al. 2010, MNRAS, 408, 2319

Stompor, R., Leach, S., Stivoli, F., & Baccigalupi, C. 2009, MNRAS, 392, 216

Sunyaev, R. A. & Zeldovich, Y. B. 1972, Comments on Astrophysics and Space Physics, 4, 173

Svalheim et al. 2022a, A&A, in press [arXiv:2201.03417]

Svalheim et al. 2022b, A&A, in press [arXiv:2011.08503]

Szűcs, L., Glover, S. C. O., & Klessen, R. S. 2014, MNRAS, 445, 4055

Tassis, K., Ramaprakash, A. N., Readhead, A. C. S., et al. 2018, arXiv e-prints, arXiv:1810.05652

Tegmark, M., Taylor, A. N., & Heavens, A. F. 1997, ApJ, 480, 22

Tristram, M., Banday, A. J., Górski, K. M., et al. 2022, Phys. Rev. D, 105, 083524

Tristram, M., Banday, A. J., Górski, K. M., et al. 2021, A&A, 647, A128

Wandelt, B. D. & Górski, K. M. 2001, Phys. Rev. D, 63, 123002

Wandelt, B. D., Larson, D. L., & Lakshminarayanan, A. 2004, Phys. Rev. D, 70, 083511

Wang, X., Tegmark, M., Jain, B., & Zaldarriaga, M. 2003, Phys. Rev. D, 68, 123001

Watts et al. 2022, A&A, in press [arXiv:2202.11979]

Wehus, I. K., Fuskeland, U., Eriksen, H. K., et al. 2017, A&A, 597, A131

Wehus, I. K., Næss, S. K., & Eriksen, H. K. 2012, ApJS, 199, 15

Zaldarriaga, M. & Seljak, U. 1997, Phys. Rev. D, 55, 1830

Zonca, A., Franceschet, C., Battaglia, P., et al. 2009, Journal of Instrumentation, 4, 2010

Appendix A: Review of frequently used textbook sampling algorithms

As described in Sect. 6.3, the BeyondPlanck pipeline is designed in the form of a Gibbs sampler in which each parameter is sampled conditionally on all other parameters. Each parameter must therefore be associated with a specific sampling algorithm that samples from the correct distribution. In this appendix, we therefore review some of the most common sampling techniques that are used in the BeyondPlanck framework, while noting that all of this is textbook material; this is just provided for reference purposes. In all cases below, we assume that samplers for both the uniform distribution, $U[0, 1]$, and the standard univariate normal (Gaussian) distribution, $N(0, 1)$, are already available through some numerical library; we use routines provided in HEALPix.

Appendix A.1: Univariate and low-dimensional Gaussian sampling

Perhaps the single most common distribution in any Bayesian pipeline is the univariate Gaussian distribution $N(\mu, \sigma^2)$ with mean μ and standard deviation σ. A sample x from this distribution can be trivially generated by

$$x = \mu + \sigma \eta, \tag{A.1}$$

where $\eta \sim N(0, 1)$ is a standard normal variate. Note that $\langle x \rangle = \mu$, because $\langle \eta \rangle = 0$ and $\langle (x - \mu)^2 \rangle = \sigma^2$ because $\langle \eta^2 \rangle = 1$.

A sample from a multivariate normal distribution $N(\boldsymbol{\mu}, \mathsf{C})$ with mean vector $\boldsymbol{\mu}$ and covariance matrix C may be produced in a fully analogous manner,

$$\boldsymbol{x} = \boldsymbol{\mu} + \mathsf{C}^{\frac{1}{2}} \boldsymbol{\eta}, \tag{A.2}$$

where now $\boldsymbol{\eta}$ is a vector of independent $N(0, 1)$ variates, and $\mathsf{C}^{\frac{1}{2}}$ denotes some matrix for which $\mathsf{C} = \mathsf{C}^{\frac{1}{2}} (\mathsf{C}^{\frac{1}{2}})^t$. The two most typical examples are the Cholesky decomposition ($\mathsf{C} = \mathsf{LL}^t$ for positive definite matrices, where $\mathsf{C}^{\frac{1}{2}} = \mathsf{L}$) and singular-value decomposition ($\mathsf{C} = \mathsf{V\Sigma V}^t$ for singular matrices, where $\mathsf{C}^{\frac{1}{2}} = \mathsf{V\Sigma}^{\frac{1}{2}} \mathsf{V}^t$). A notable advantage regarding the latter is that it is symmetric, and therefore less bug-prone than the Cholesky factor; on the other hand, it is slightly more computationally expensive.

Appendix A.2: High-dimensional Gaussian sampling

It is important to note that evaluating a "square root" of a matrix, whether it is through Cholesky or eigenvector decomposition, is an $O(n^3)$ operation, where n is the dimension of the matrix. As such, the direct approach is only computationally practical for relatively low-dimensional distributions, and just with a few thousand elements or less. For distributions with millions of correlated variables, the above prescription is entirely impractical. In the following, we therefore describe a widely used method to sample from high-dimensional distributions, effectively by inverting the covariance matrix iteratively by Conjugate Gradients.

Again, let $\boldsymbol{x}$ be a random Gaussian field of n elements with an $n \times n$ covariance matrix S, i.e., $\boldsymbol{x} \sim N(\boldsymbol{0}, \mathsf{S})$. Further, to put the notation into a familiar context, we assume we have some observations $\boldsymbol{d}$ that can be modeled as

$$\boldsymbol{d} = \mathsf{T}\boldsymbol{x} + \boldsymbol{n}, \tag{A.3}$$

where $\boldsymbol{n}$ is a stochastic noise vector of size n_d (which in general is different from n) which is drawn from a Gaussian distribution

with zero mean and covariance N, and T is a matrix of size $n_d \times n$, which effectively translates $\boldsymbol{x}$ into the vector space of $\boldsymbol{d}$. In other words, we assume that the data may be modelled as a linear combination of $\boldsymbol{x}$ plus a well-defined noise contribution.

Note that this assumption about $\boldsymbol{d}$ does not preclude the cases where we have observations that can be written as $\boldsymbol{d} = \mathsf{T}\boldsymbol{x} + \boldsymbol{n} + \boldsymbol{b}$, where $\boldsymbol{b}$ is known and independent of $\boldsymbol{x}$ - in this case, we are free to redefine $\boldsymbol{d}$: $\boldsymbol{d}' \to \boldsymbol{d} - \boldsymbol{b}$, in which case our assumption in Eq. (A.3) would be met for $\boldsymbol{d}'$.

In general, T will not depend on $\boldsymbol{x}$. In the context of the Gibbs framework of this paper, however, T typically *will* depend on other quantities that we do sample, but which we assume to be known with respect to the current conditional of the Gibbs chain.

Our goal is then to draw a sample from $P(\boldsymbol{x} \mid \boldsymbol{d}, \mathsf{T}, \mathsf{S}, \mathsf{N})$, the posterior of $\boldsymbol{x}$, given $\boldsymbol{d}$ and the other quantities, denoted $P(\boldsymbol{x} \mid \boldsymbol{d})$ as a shorthand. Using Bayes' theorem, we can write this as

$$P(\boldsymbol{x} \mid \boldsymbol{d}) \propto P(\boldsymbol{d} \mid \boldsymbol{x}) P(\boldsymbol{x}). \tag{A.4}$$

Here $P(\boldsymbol{x})$ is a prior for x, which we assume takes the form $\mathcal{N}(\boldsymbol{0}, \mathsf{S})$, whereas the likelihood term, $P(\boldsymbol{d} \mid \boldsymbol{x})$, is simply given by a Gaussian distribution with covariance N and mean $\mathsf{T}\boldsymbol{x}$. This gives (neglecting the pre-factors of the exponentials, as they are independent of $\boldsymbol{x}$ and end up as normalization constants)

$$\begin{aligned} -2 \ln P(\boldsymbol{x} \mid \boldsymbol{d}) &= \boldsymbol{x}^t \mathsf{S}^{-1} \boldsymbol{x} + (\boldsymbol{d} - \mathsf{T}\boldsymbol{x})^t \mathsf{N}^{-1} (\boldsymbol{d} - \mathsf{T}\boldsymbol{x}) \\ &= \boldsymbol{x}^t \mathsf{S}^{-1} \boldsymbol{x} + \boldsymbol{d}^t \mathsf{N}^{-1} \boldsymbol{d} + \boldsymbol{x}^t \mathsf{T}^t \mathsf{N}^{-1} \mathsf{T}\boldsymbol{x} - \\ &\quad \boldsymbol{d}^t \mathsf{N}^{-1} \mathsf{T}\boldsymbol{x} - \boldsymbol{x}^t \mathsf{T}^t \mathsf{N}^{-1} \boldsymbol{d} \\ &= \boldsymbol{x}^t (\mathsf{S}^{-1} + \mathsf{T}^t \mathsf{N}^{-1} \mathsf{T}) \boldsymbol{x} - 2 \boldsymbol{x}^t \mathsf{T}^t \mathsf{N}^{-1} \boldsymbol{d}, \end{aligned} \tag{A.5}$$

where, in the last transition, we neglect also the terms that do not include $\boldsymbol{x}$. We also use the identity $\boldsymbol{a}^t \mathsf{C} \boldsymbol{b} = \boldsymbol{b}^t \mathsf{C} \boldsymbol{a}$, which is valid for a symmetric matrix C, in order to gather the terms that are linear in $\boldsymbol{x}$.

This expression for $P(\boldsymbol{x} \mid \boldsymbol{d})$ can be written as a Gaussian distribution by "completing the square": We are looking for a matrix F and a vector $\boldsymbol{c}$ such that

$$\begin{aligned} P(\boldsymbol{x} \mid \boldsymbol{d}) &= \exp\left[-\frac{1}{2} (\boldsymbol{x} - \boldsymbol{c})^t \mathsf{F}^{-1} (\boldsymbol{x} - \boldsymbol{c}) \right] \\ &\propto \exp\left[-\frac{1}{2} \left(\boldsymbol{x}^t \mathsf{F}^{-1} \boldsymbol{x} - 2 \boldsymbol{x}^t \mathsf{F}^{-1} \boldsymbol{c} \right) \right]. \end{aligned} \tag{A.6}$$

Comparing terms in Eqs. (A.5) and (A.6), we find that the terms that are quadratic in $\boldsymbol{x}$ enforce

$$\mathsf{F}^{-1} = \mathsf{S}^{-1} + \mathsf{T}^t \mathsf{N}^{-1} \mathsf{T}. \tag{A.7}$$

Inserting this into the terms that are linear in $\boldsymbol{x}$, we find

$$\boldsymbol{c} = (\mathsf{S}^{-1} + \mathsf{T}^t \mathsf{N}^{-1} \mathsf{T})^{-1} \mathsf{T}^t \mathsf{N}^{-1} \boldsymbol{d}. \tag{A.8}$$

Thus, the posterior of $\boldsymbol{x}$ is a Gaussian distribution with covariance given by Eq. (A.7) and mean (and mode) given by Eq. (A.8).

In order to draw a sample, $\tilde{\boldsymbol{x}}$, from this distribution, we can in principle use the standard prescription for sampling from multivariate Gaussian distributions,

$$\boldsymbol{x} = (\mathsf{S}^{-1} + \mathsf{T}^t \mathsf{N}^{-1} \mathsf{T})^{-1} \mathsf{T}^t \mathsf{N}^{-1} \boldsymbol{d} + (\mathsf{S}^{-1} + \mathsf{T}^t \mathsf{N}^{-1} \mathsf{T})^{-1/2} \boldsymbol{\eta}, \tag{A.9}$$

as summarized in the previous section. However, inverting the covariance matrix, $\mathsf{S}^{-1} + \mathsf{T}^t \mathsf{N}^{-1} \mathsf{T}$, is once again an $O(n^3)$ operation. To circumvent this problem, we instead consider the same equation in the form

$$(\mathsf{S}^{-1} + \mathsf{T}^t \mathsf{N}^{-1} \mathsf{T}) \boldsymbol{x} = \mathsf{T}^t \mathsf{N}^{-1} \boldsymbol{d} + (\mathsf{S}^{-1} + \mathsf{T}^t \mathsf{N}^{-1} \mathsf{T})^{1/2} \boldsymbol{\eta}. \tag{A.10}$$

Since the matrix on the left-hand side is both symmetric and semi-positive definite, this equation can be solved iteratively by Conjugate Gradients; for a brilliant review of this algorithm, see Shewchuk (1994). Additionally, to obtain the correct covariance structure, one can instead add one random zero-mean covariance term for each element in the covariance matrix to the right-hand side of the equation,

$$(\mathsf{S}^{-1} + \mathsf{T}^t\mathsf{N}^{-1}\mathsf{T})\boldsymbol{x} = \mathsf{T}^t\mathsf{N}^{-1}\boldsymbol{d} + \mathsf{T}^t\mathsf{N}^{-1/2}\boldsymbol{\eta}_1 + \mathsf{S}^{-1/2}\boldsymbol{\eta}_2. \qquad (\text{A.11})$$

With this definition, $\langle\boldsymbol{x}\rangle = \boldsymbol{c}$, and $\langle\boldsymbol{x}\boldsymbol{x}^t\rangle = (\mathsf{S}^{-1} + \mathsf{T}^t\mathsf{N}^{-1}\mathsf{T})^{-1} = \mathsf{F}$, as desired.

A fully analogous calculation may be done also with a non-zero prior mean, $\boldsymbol{m}$, in which case an additional term is introduced on the right-hand side of Eq. (A.10),

$$(\mathsf{S}^{-1}+\mathsf{T}^t\mathsf{N}^{-1}\mathsf{T})\boldsymbol{x} = \mathsf{T}^t\mathsf{N}^{-1}\boldsymbol{d}+\mathsf{S}^{-1}\boldsymbol{m}+\mathsf{T}^t\mathsf{N}^{-1/2}\boldsymbol{\eta}_1+\mathsf{S}^{-1/2}\boldsymbol{\eta}_2. \quad (\text{A.12})$$

The relative strength of the data and prior terms is thus effectively determined by the overall signal-to-noise ratio of the data as measured by S and N, and in the limit of vanishing signal-to-noise (i.e., $\mathsf{N}^{-1} \to 0$), $\langle\boldsymbol{x}\rangle = \boldsymbol{m}$, as desired. Note, also, that S quantifies the covariance of the *fluctuations around the mean*, not the co-variance of the entire field $\boldsymbol{x}$ itself. In the limit of $\mathsf{S} \to 0$ (or, equivalently, $\mathsf{S}^{-1} \to \infty$), we therefore also have $\langle\boldsymbol{x}\rangle = \boldsymbol{m}$. Thus, the magnitude of S represents a direct handle for adjusting the strength of the prior.

Appendix A.3: Inversion sampling

The samplers discussed in the two previous sections only concerns Gaussian distributions. In contrast, the so-called *inversion* sampler is a completely general sampler that works for all univariate distributions.

Let $P(x)$ be a general probability distribution for some random variable x. The inversion sampler is then defined as follows:

1. Compute $P(x)$ over a grid in x, making sure to probe the tails to sufficient accuracy.
2. Compute the cumulative probability distribution, $F(x) = \int_{-\infty}^{x} P(x')\,\mathrm{d}x'$.
3. Draw a random uniform variate, $\eta \sim U[0, 1]$.
4. Solve the nonlinear equation $\eta = F(x)$ for x.

Clearly, this is a computationally very expensive algorithm, noting that it actually requires the user to map the full distribution, $P(x)$, in the first step. This typically requires a preliminary bisection search to first identify a sufficiently wide region in x to cover all significant parts of P. Then another 50–100 evaluations are required to grid the (log-)probability distribution.

However, the facts that this sampler requires no manual tuning, and that it produces independent samples, make it an attractive component in many Gibbs samplers; typically, the overall computational cost of the entire Gibbs chain is dominated by completely different operations.