Astronomy & Astrophysics manuscript no. ms
November 12, 2020

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

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

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

November 12, 2020

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 $P(\omega \mid \boldsymbol{d})$, 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, paying special attention to the LFI 44 GHz channel. We find evidence of several significant residual systematic effects that are still not accounted for in the current processing, but must be addressed in future work. These include a break-down of the $1/f$ correlated noise model at 30 and 44 GHz, and scan-aligned stripes in the Southern Galactic hemisphere at 44 GHz. On the Northern hemisphere, however, we find that all results are consistent with the ΛCDM model, and we constrain the reionization optical depth to $\tau = 0.067 \pm 0.016$, with a low-resolution χ^2 probability-to-exceed of 16 %. The marginal CMB dipole amplitude is $3359.5 \pm 1.9\,\mu$K. This analysis framework can play a central role in the analysis of several current and future high-sensitivity CMB experiments, including *LiteBIRD*. 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

⋆ Corresponding author: H. K. Eriksen; `h.k.k.eriksen@astro.uio.no`

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 *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 $O(30\ \mu\mathrm{K})$ 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.

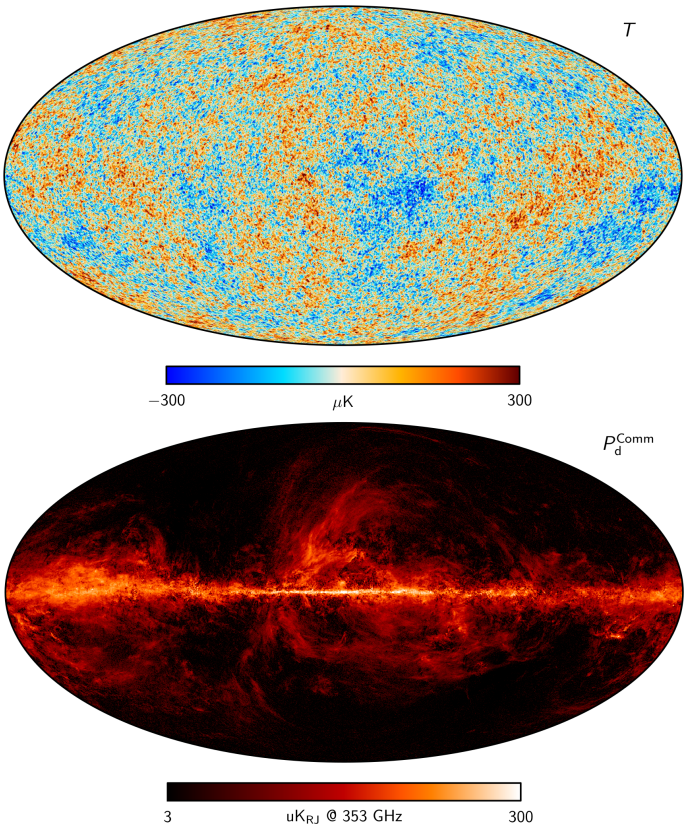

Fig. 1. *Planck* 2018 `Commander` CMB temperature (*top panel*) and thermal dust polarization amplitude (*bottom panel*) maps. Reproductions from Planck Collaboration IV (2020).

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 currently 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 $O(10^{-5})$ temperature fluctuations 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 2020 is defined by *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 some 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 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. 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 at the level of $O(0.5\ \mu\mathrm{K})$ 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 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 $t_{\mathrm{r}} = 180^{+220}_{-80}$ 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 Collabo-

ration 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 and final 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 2020), 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 the so-called NPIPE pipeline (Planck Collaboration Int. LVII 2020). 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). One unique feature of this pipeline is 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 has led to further reduction of systematic errors in both data sets, as reported in Planck Collaboration Int. LVII (2020). The resulting best-fit estimate of the reionization optical depth from NPIPE reads $\tau = 0.058 \pm 0.006$ (Tristram et al. 2020).

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, in that BeyondPlanck primarily 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, two examples of this are already presented within the current project, with preliminary applications to *WMAP* (Bennett et al. 2013; Watts et al. 2020) and *LiteBIRD* (Sugai et al. 2020; Aurlien et al. 2020).

Because instrumental systematics and residual foreground contamination have such a dramatic impact on the large-scale CMB polarization estimates, we will in this paper 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 avoid the use of cross-spectrum techniques, which generally reduce the sensitivity of the final products to instrumental systematics. 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

Historically speaking, to understand the background and motivation for the BeyondPlanck program, it is useful to revisit the "Lessons learned from *Planck*,"[1] 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 integrated 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.*

[1] https://www.cosmos.esa.int/web/planck/lessons-learned

(. . .) 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 limit on the tensor-to-scalar ratio is $r < 0.044$ at 95 % confidence (Tristram et al. 2020), 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 the main long-term scientific motivation for the BeyondPlanck program; current experiments such as *Planck* and *WMAP* provide real-world test-beds that ensure that the BeyondPlanck approach is both realistic and practical.

1.6. The BeyondPlanck program

In this context, 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* scientific results;
2. to provide a computationally convenient interface to the raw *Planck* data that can be accessed and extended by external users;
3. to develop a framework that allows joint analysis of *Planck* 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 on November 30th 2020; the total duration of the program is 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. With an effective run-time of six to eight weeks to achieve convergence on our current computer systems, we have been able to complete two full end-to-end data reprocessings since that time. While two full iterations certainly are a technical achievement, they are insufficient for detailed fine-tuning and thus the results are uncharacteristically unpolished relative to a typical data release. They however demonstrate the power of the analysis process itself. Further, the current BeyondPlanck release is not intended to be a static and final analysis solution for one specific data set, i.e., *Planck* LFI, but

rather a common community-wide platform that will allow scientists to explore different data sets both individually and jointly. As such, we expect numerous updates to emerge in the coming months and years, both from BeyondPlanck members and from external researchers, that will gradually refine the current products in a collaborative effort.

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.

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 (2020), 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`.

Finally, `Commander3` 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) OpenSource license.[2]

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, as specific details are instead 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) analysis of external data. 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

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

Table 1. Overview of BeyondPlanck papers.

Reference	Title
Pipeline	
BeyondPlanck Collaboration (2020)	I. Global Bayesian analysis of the *Planck* Low Frequency Instrument data
Keihänen et al. (2020)	II. CMB mapmaking through Gibbs sampling
Galloway et al. (2020a)	III. Computational infrastructure and `Commander3`
Brilenkov et al. (2020)	IV. Time-ordered data simulations
Gerakakis et al. (2020)	V. Open Science and reproducibility
Instrument characterization	
Ihle et al. (2020)	VI. Noise characterization and modelling
Gjerløw et al. (2020)	VII. Calibration
Galloway et al. (2020b)	VIII. Sidelobe modelling
Svalheim et al. (2020a)	IX. Bandpass and beam leakage modelling
Cosmological and astrophysical results	
Suur-Uski et al. (2020)	X. LFI frequency map posteriors and sample-based error propagation
Colombo et al. (2020)	XI. CMB analysis with end-to-end error propagation: Temperature anisotropies
Paradiso et al. (2020)	XII. CMB analysis with end-to-end error propagation: Likelihood and cosmological parameters
Andersen et al. (2020)	XIII. Intensity foregrounds, degeneracies and priors
Svalheim et al. (2020b)	XIV. Polarized foreground emission between 30 and 70 GHz
Herman et al. (2020)	XV. Limits on polarized anomalous microwave emission
External analysis	
Aurlien et al. (2020)	XVI. Application to simulated *LiteBIRD* observations
Watts et al. (2020)	XVII. Application to *WMAP*
Galeotta et al. (2020)	XVIII. End-to-end validation of BeyondPlanck

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.

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. For now, the main emphasis is put on *WMAP* (Watts et al. 2020) and *LiteBIRD* (Aurlien et al. 2020).

We note that, in the spirit of reproducibility and accessibility, 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[3] (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 $N_{\rm side}$, and the total number of pixels on the sky is $N_{\rm pix} = 12N_{\rm side}^2$. We organize the Stokes parameters into vectors of length $3N_{\rm pix}$, simply by stacking $\{T, Q, U\}$ into a map vector $\boldsymbol{s}(\hat{n})$, 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

[3] `http://healpix.jpl.nasa.gov`

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,

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

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

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

$$a^E_{\ell m} = -\frac{1}{2}\left({}_2a_{\ell m} + {}_{-2}a_{\ell m}\right) \tag{4}$$

$$a^B_{\ell m} = \frac{i}{2}\left({}_2a_{\ell m} - {}_{-2}a_{\ell m}\right), \tag{5}$$

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

$$\sigma^{XY}_\ell = \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} \left(a^X_\ell\right)^* a^Y_{\ell m}, \tag{6}$$

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

$$C^{XY}_\ell \equiv \left\langle \left(a^X_\ell\right)^* a^Y_{\ell m} \right\rangle = \left\langle \sigma^{XY}_\ell \right\rangle, \tag{7}$$

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 $\boldsymbol{s}$ 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 $\mathrm{MJy\,sr^{-1}} \equiv 10^{-20}\,\mathrm{W\,m^{-2}\,Hz^{-1}\,sr^{-1}}$, 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,

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

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 $\mathrm{K_{CMB}}$ or simply K.

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

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

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 $\mathrm{K_{CMB}}$, they are usually limited to a few orders of magnitude when expressed in either $\mathrm{MJy\,sr^{-1}}$ or $\mathrm{K_{RJ}}$. To avoid numerical problems, all astrophysical components are therefore expressed in units of $\mathrm{K_{RJ}}$ internally in `Commander`, and only converted to the respective natural unit before outputting results to disk. Monochromatic conversion between $\mathrm{K_{RJ}}$ and $\mathrm{MJy\,sr^{-1}}$ is performed through Eq. (9), while monochromatic conversion between $\mathrm{K_{RJ}}$ and $\mathrm{K_{CMB}}$ is given by

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

where $x = h\nu/kT_0$, and $T_0 = 2.7255\,\mathrm{K}$ 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 tempeerature 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 $\boldsymbol{s}^{\rm CMB}$ to be a $3N_{\rm pix}$ 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),

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

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.)

For component separation purposes, this is 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 flucutations are both Gaussian distributed and statistically isotropic.

The assumption of Gaussianity determines the conditional probability distribution for the CMB signal,

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

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,

$$\mathsf{S}^{XY}_{\ell m,\ell' m'} \equiv \left\langle \left(a^X_\ell\right)^* a^Y_{\ell' m'} \right\rangle = C^{XY}_{\ell m} \delta_{\ell\ell'} \delta_{mm'}. \tag{13}$$

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 eventually accounts for the entire process from raw time-ordered data to final cosmological parameters; in its current form, the final product of the BeyondPlanck pipeline will be the so-called CMB power spectrum likelihood, $\mathcal{L}(C_\ell) \equiv P(\boldsymbol{d} \mid C_\ell)$, and the final parameter estimation process will still be performed using `CosmoMC`.

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

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

where $\nu_{0,\mathrm{s}}$ 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.

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 birefringence, 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, but reaches hundreds of degrees at 2.3 GHz (Carretti et al. 2019; Fuskeland et al. 2019). Therefore, where relevant, we account for Faraday rotation when comparing our results with low-frequency surveys such as S-PASS (2.3 GHz; Carretti et al. 2019), but we neglect it for higher-frequency surveys such as *Planck* and *WMAP*.

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_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

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

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

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

where

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

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

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

and we correspondingly quantify the free-free amplitude in terms of the observed signal at a given reference frequency in $\mu{\rm K}_{\rm RJ}$, 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 this paper, 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 equilibriate 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 Draine & Hensley (2020).

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

$$I^{\rm d}_\nu \propto \tau \nu^{\beta_{\rm d}} B_\nu(T_{\rm d}). \tag{20}$$

This function is simply a blackbody spectrum with temperature $T_{\rm d}$, modulated by a power-law having index $\beta_{\rm d}$. In physical terms, this corresponds to dust having an opacity that scales as $\nu^\beta_{\rm d}$, 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 $\tau \sim 10^{-6}$, $\beta_{\rm d} \sim 1.6$, and $T_{\rm d} \sim 20$ K. Intuitively speaking, $\beta_{\rm d}$ determines the slope (or first derivative in log-log space) of the SED below 200 GHz, while $T_{\rm d}$ 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

$$s^{\rm d}_{\rm RJ}(\nu) \propto \frac{\nu^{\beta_{\rm d}+1}}{{\rm e}^{h\nu/kT_{\rm d}} - 1}. \tag{21}$$

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 $T_{\rm d}$ for intensity and polarization, while $\beta_{\rm d}$ 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, $s_0^{\rm sd}(\nu)$, computed with the `SpDust2` code (Ali-Haïmoud et al. 2009; Ali-Haimoud 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, $\nu_{\rm p}$, that shifts the spectrum rigidly in $\log\nu$–$\log s$ space,

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

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 a companion BeyondPlanck paper, Herman et al. (2020), we perform a non-parametric analysis of the observed diffuse AME spectrum in polarization. However, for all main analyses presented in this work, we adopt the spinning dust model in Eq. (22) for the AME component.

Despite sensitive searches in individual objects (Génova-Santos et al. 2015; Génova-Santos et al. 2016) 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. (2020), 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 isotopologe 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 $m_{\rm C}$ and $m_{\rm O}$, respectively, and the corresponding atomic distances from their center of mass by $r_{\rm C}$ and $r_{\rm O}$. We also define $r_{\rm CO} = r_{\rm C} + r_{\rm O}$ to be the effective atom size and $m_{\rm CO} = m_{\rm C}m_{\rm O}/(m_{\rm C} + m_{\rm O})$ its reduced mass.

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

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

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

$$\nu_0 = \frac{\Delta E_{\rm rot}}{h} = \frac{\hbar J}{2\pi I} = \frac{\hbar J}{2\pi m r_{\rm CO}^2}, \quad J = 1, 2, \ldots \tag{24}$$

For the ^{12}CO J=1←0 transition, one finds $\nu_0 = 115.27$ GHz, while for the ^{13}CO J=1←0 transition, it is $\nu_0 = 110.20$ GHz. Higher-order transitions, such as J=2←1, 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,

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

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 foreground emission

In addition to the Galactic foreground emission mechanisms discussed in Sect. 3.3, several extra-galactic effects are also important for CMB frequencies.

3.4.1. Extra-galactic compact sources

For LFI frequencies, the most important class of extra-galactic 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,

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

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.

3.4.2. Sunyaev-Zeldovich effect

Some CMB photons happen to pass through one or more clusters of galaxies on their way through the universe. Such clusters are very hot, some reaching temperatures as high as 10^8 K. At such high temperatures, the inter-cluster medium is highly ionized.

CMB photons have a non-negligible probability of scattering on these free electrons, and when they do, they gain energy from the free electrons. As a result, their spectrum is shifted to slightly higher frequencies compared to the standard blackbody form. This is called the thermal Sunyaev-Zeldovich (tSZ) effect (Sunyaev & Zeldovich 1972), and it is an effective probe of the intergalactic medium in high-redshift clusters.

In this work, we will mostly ignore the tSZ effect, as it has a relatively modest impact on the LFI measurements, due to both their limited sensitivity and angular resolution. In the cases where we do consider it, we adopt the non-relativistic model of the effect, which in brightness temperature units takes the form

$$s_{\mathrm{RJ}}^{\mathrm{SZ}}(\nu) \propto \frac{x^2 \mathrm{e}^x}{(\mathrm{e}^x-1)^2}\left(\frac{x\mathrm{e}^x+1}{\mathrm{e}^x-1}-4\right), \tag{27}$$

where $x = h\nu/kT_0$.

In additional to the thermal SZ effect, non-zero cluster velocities give rise to an additional contribution called the kinetic SZ effect. This does not affect the SED shape of the underlying photons, but simply changes the apparent temperature fluctuation at a given position. For typical cluster velocities of $\lesssim 10^3$ km s^{-1}, these modifications are however small, at the level of a few μK, and we therefore neglect this effect in the following. Likewise, we also neglect small polarization effects in the thermal SZ case, which are expected to be well below 1 μK in amplitude.

3.4.3. Cosmic infrared background

The last extra-galactic component that will be encountered in this analysis is the cosmic infrared background (CIB). This is the thermal dust emission emitted by many distant galaxies. The CIB may be spatially approximated as a continuous field, similar to the CMB, but with an SED that is defined as an average of a large number of independent thermal dust SEDs, each redshifted according to the distance of the emitting galaxy.

The CIB affects CMB observations in two different ways. First, uncertainties in the CIB monopole translate into uncertainties in the zero-level of each frequency channel. In particular, current models predict a CIB monopole of about 400 μK at 353 GHz, but with a model uncertainty of about 20 %. If left unmitigated, such large uncertainties would translate into massive uncertainties in the thermal dust spectral parameters, β_{d} and T_{d}. In practice, the HFI monopoles are currently determined through cross-correlation with H I (Planck Collaboration III 2020); however, this approach is of course associated with its own uncertainties.

Second, CIB fluctuations dominate over Galactic thermal dust fluctuations near the Galactic poles, where local thermal dust emission is low. In the foreground model employed in the current work, we do not account separately for CIB fluctuations, as we do not have a high enough signal-to-noise ratio to robustly separate them from Galactic emission. The thermal dust emission estimates presented in the following therefore correspond to the sum of Galactic thermal dust emission and CIB fluctuations. Since CIB and thermal dust emission have very similar SEDs, this has only a small effect on other components, most importantly on the CMB. However, it does complicate the physical interpretation of the resulting thermal dust parameter maps and dust parameters derived from them.

3.5. Zodiacal light emission

The last emission component we will consider is zodiacal light emission (ZLE). Similar to both CO and CIB emission, this component is far more important for HFI than LFI frequencies, and its mean amplitude is only about 0.5 μK at 70 GHz (Planck Collaboration X 2016).

The ZLE is emitted by dust particles located within the Solar system, primarily in the asteroid belt between Mars and Jupiter. These grains are heated by solar radiation to a temperature of about 150 K, and this energy is then thermally re-emitted with a thermal dust-like SED with the corresponding temperature. As such, its frequency spectrum is similar to both Galactic thermal dust and CIB fluctuations across CMB bands.

However, unlike Galactic dust and CIB, the ZLE is not stationary on the sky throughout a given survey. As *Planck* moves with the Earth around the Sun throughout a year, the ZLE is observed through different lines of sight. This both allows for, and indeed requires, dynamic modelling of the effect, taking into account the precise location of the satellite as a function of time. For this purpose, we adopt the *COBE*-DIRBE ZLE model (Kelsall et al. 1998), as integrated and re-implemented natively into `Commander` for efficiency purposes.

The *COBE* model treats the ZLE in terms of six distinct constituents (a smooth cloud, three asteroidal dust bands, a circumsolar ring, and a trailing blob), each specified in terms of its own geometry and optical properties. In total, the model has 43 free parameters, and in this work we adopt the parameter set described by Planck Collaboration XIV (2014), which takes into account information from both *COBE*-DIRBE and *Planck*. Future work will aim to estimate these parameters internally in our analysis framework, but this will require time-domain *Planck* HFI and *COBE*-DIRBE observations, and is outside the scope of the present work. In the current work, we adopt the same cor-

rections as described in Planck Collaboration Int. LVII (2020) for the HFI channels, and neglect ZLE for the LFI channels.

3.6. 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,

$$\boldsymbol{s}_{\mathrm{RJ}} = \boldsymbol{a}_{\mathrm{CMB}} \frac{x^2 \mathrm{e}^x}{(\mathrm{e}^x-1)^2} \frac{(\mathrm{e}^{x_0}-1)^2}{x_0^2 \mathrm{e}^{x_0}} + \tag{28}$$

$$+ \boldsymbol{a}_{\mathrm{s}} \left(\frac{\nu}{\nu_{0,\mathrm{s}}}\right)^{\beta_{\mathrm{s}}} + \tag{29}$$

$$+ \boldsymbol{a}_{\mathrm{ff}} \frac{g_{\mathrm{ff}}(\nu; T_e)}{g_{\mathrm{ff}}(\nu_{0,\mathrm{ff}}; T_e)} \left(\frac{\nu_{0,\mathrm{ff}}}{\nu}\right)^2 + \tag{30}$$

$$+ \boldsymbol{a}_{\mathrm{AME}} \left(\frac{\nu_{0,\mathrm{sd}}}{\nu}\right)^2 \frac{s_0^{\mathrm{sd}}\left(\nu \cdot \frac{\nu_p}{30.0\,\mathrm{GHz}}\right)}{s_0^{\mathrm{sd}}\left(\nu_{0,\mathrm{sd}} \cdot \frac{\nu_p}{30.0\,\mathrm{GHz}}\right)} + \tag{31}$$

$$+ \boldsymbol{a}_{\mathrm{d}} \left(\frac{\nu}{\nu_{0,\mathrm{d}}}\right)^{\beta_{\mathrm{d}}+1} \frac{\mathrm{e}^{h\nu_{0,\mathrm{d}}/kT_{\mathrm{d}}}-1}{\mathrm{e}^{h\nu/kT_{\mathrm{d}}}-1} + \tag{32}$$

$$+ \sum_{j=1}^{N_{\mathrm{src}}} \boldsymbol{a}_{\mathrm{src}}^j \left(\frac{\nu}{\nu_{0,\mathrm{src}}}\right)^{\alpha_{j,\mathrm{src}}-2}, \tag{33}$$

where $x = h\nu/kT_0$ and $\nu_{0,i}$ is the reference frequency for component i. Thus, $\boldsymbol{a}_i$ 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 33 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 (2020), 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 $\{\boldsymbol{s}_{\mathrm{RJ}}, \boldsymbol{a}_{\mathrm{CMB}}, \boldsymbol{a}_{\mathrm{s}}, \boldsymbol{a}_{\mathrm{d}}\}$ 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, $\{\boldsymbol{a}_{\mathrm{ff}}, \boldsymbol{a}_{\mathrm{AME}}, \boldsymbol{a}_{\mathrm{src}}^j\}$, 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 $\{\boldsymbol{a}_{\mathrm{CMB}}, \boldsymbol{a}_{\mathrm{s}}, \boldsymbol{a}_{\mathrm{ff}}, \boldsymbol{a}_{\mathrm{AME}}, \boldsymbol{a}_{\mathrm{d}}, \boldsymbol{a}_{\mathrm{src}}\}$, collectively denoted $\boldsymbol{a}$; 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 BeyondPlanck 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

$$d(t) = g(t)\,[T + Q\,\cos 2\psi + U\,\sin 2\psi] + n(t), \tag{34}$$

where g is a multiplicative factor called the gain, which converts between physical signal units (which in our case will be $\mathrm{K}_{\mathrm{CMB}}$) 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, $\boldsymbol{s} = \boldsymbol{s}_p$, and then re-write Eq. (34) in vector form as follows,

$$\boldsymbol{d} = \mathsf{G}\,\mathsf{P}\boldsymbol{s} + \boldsymbol{n}, \tag{35}$$

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

$$\mathsf{P} = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 & \cos 2\psi_1 & 0 & \ldots & 0 & \sin 2\psi_1 & 0 \\ 1 & 0 & 0 & \ldots & \cos 2\psi_2 & 0 & 0 & \ldots & \sin 2\psi_2 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & \ldots & 0 & 0 & \cos 2\psi_1 & \ldots & 0 & 0 & \sin 2\psi_1 \end{bmatrix}. \tag{36}$$

This matrix is called the pointing matrix.[4] Correspondingly, the sky vector consists of the three pixelized Stokes parameter maps stacked into a single vector, $\boldsymbol{s} = [T_1, \ldots, T_{N_{\mathrm{pix}}}, Q_1, \ldots, Q_{N_{\mathrm{pix}}}, U_1, \ldots, U_{N_{\mathrm{pix}}}]^t$.

Equation (35) 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] Only the nonzero entries need to be stored in the pointing matrix, and memory requirements are therefore manageable.

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, $\boldsymbol{d}$, and the monochromatic sky signal, $\boldsymbol{s}(\nu)$), the data model in Eq. (35) generalizes to

$$\boldsymbol{d} = \mathsf{GP} \int \boldsymbol{s}(\nu)\tau(\nu)\,\mathrm{d}\nu + \boldsymbol{n}, \tag{37}$$

after taking into account the bandpass effect.[5]

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

$$U_{ij} = \frac{\int \tau(\nu) \frac{dI_\nu}{dX_i}\,\mathrm{d}\nu}{\int \tau(\nu) \frac{dI_\nu}{dX_j}\,\mathrm{d}\nu}, \tag{38}$$

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

$$U_{\mathrm{K_{RJ}}\rightarrow\mathrm{K_{CMB}}} = \frac{\int \tau(\nu)\,\frac{2h\nu^2}{c^2}\,\mathrm{d}\nu}{\int \tau(\nu)\, b'_\nu\,\mathrm{d}\nu} \tag{39}$$

$$U_{\mathrm{K_{RJ}}\rightarrow\mathrm{MJy\,sr^{-1}}} = \frac{\int \tau(\nu)\,\frac{2h\nu^2}{c^2}\,\mathrm{d}\nu}{\int \tau(\nu)\,\frac{\nu_c}{\nu}\,\mathrm{d}\nu}, \tag{40}$$

where

$$b'_\nu = \left.\frac{\partial B(T,\nu)}{\partial T}\right|_{T=T_0} \tag{41}$$

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

$$\boldsymbol{d} = U\mathsf{GP} \int \boldsymbol{s}(\nu)\tau(\nu)\,\mathrm{d}\nu + \boldsymbol{n}. \tag{42}$$

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. (42). With this in mind, we consider signal component i as defined by the sky model in Sect. 3.6, 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

$$\boldsymbol{s}^j = \sum_{i=1}^{N_{\mathrm{comp}}} \boldsymbol{a}_i \left[U_j \int f_i(\nu;\beta)\,\tau_j(\nu)\,\mathrm{d}\nu \right] \equiv \sum_{i=1}^{N_{\mathrm{comp}}} \mathsf{M}_i^j\,\boldsymbol{a}_i = \mathsf{M}^j\,\boldsymbol{a}, \tag{43}$$

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. (35) must be generalized as follows,

$$\boldsymbol{d} = \mathsf{GPM}\boldsymbol{a} + \boldsymbol{n}. \tag{44}$$

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.

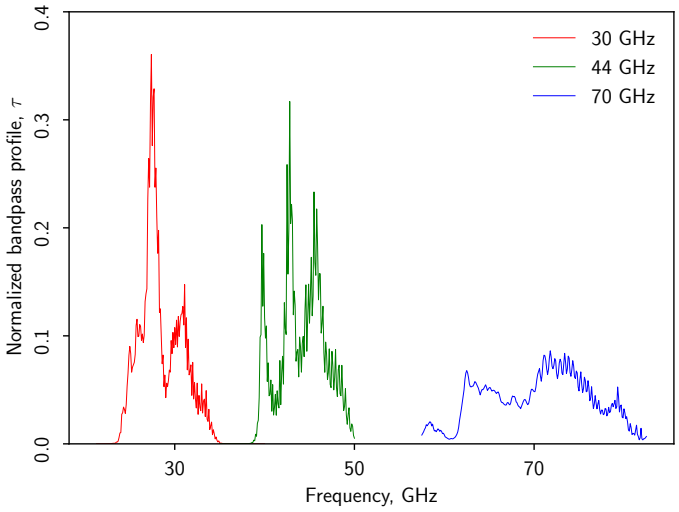

Fig. 2. Detector averaged bandpass profiles, τ, for the three *Planck* LFI frequency channels.

4.2.1. Bandpass uncertainties and corrections

While the bandpass integral described by Eq. (37) 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

[5] 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.

scale ripples, which are due to standing waves. These may be due to real standing waves within the optical assembly of the LFI instrument itself; but some part of them may also be due to standing waves in the test equipment used to make the measurements. In addition to these two 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, $\Delta_{\rm bp}$, in frequency space,

$$\tau(\nu) = \tau_0(\nu + \Delta_{\rm bp}), \tag{45}$$

where τ_0 is the default laboratory measurement. One value of $\Delta^i_{\rm bp}$ is allowed per radiometer i, but (in most cases) either with the prior that $\sum_i \Delta^i_{\rm bp} = 0$, 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. (2020a). In general, we note that the impact of $\Delta_{\rm bp}$ 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.

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 function" (PSF) or "beam", $\boldsymbol{b}(\hat{n})$, that characterizes its instantaneous

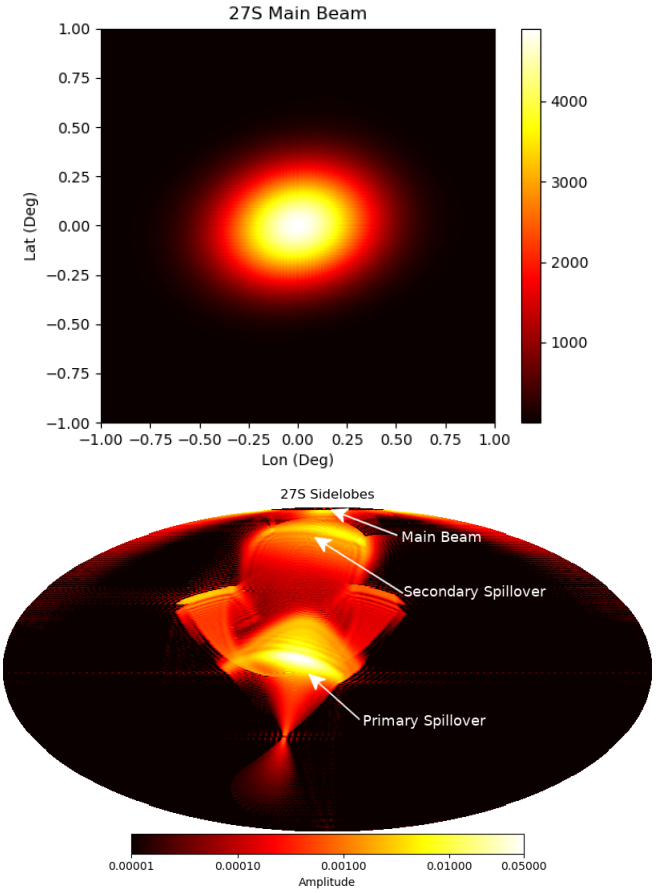

Fig. 3. (*Top*:) Scanning beam (or main beam) of the 30 GHz LFI 27S radiometer in local telescope coordinates, i.e., the instantaneous spatial sensitivity sensitivity to a point source centered at the beam maximum. *Bottom*: Corresponding 4π beam map, oriented such that the main beam is located on the 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.

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 $\boldsymbol{s}$,

$$s_t^{\rm beam} = \int_{4\pi} \boldsymbol{b}_t(\hat{n})\boldsymbol{s}(\hat{n})\,{\rm d}\Omega, \tag{46}$$

where $\boldsymbol{b}_t(\hat{n}) = \mathsf{R}_t(\hat{n},\hat{n}')\boldsymbol{b}(\hat{n}')$, and 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, B, such that $\boldsymbol{s}^{\rm beam} = \mathsf{B}\boldsymbol{s}$, and the data model in Eq. (44) may therefore be generalized further into its final form,

$$\boldsymbol{d} = \mathsf{GPBM}\boldsymbol{a} + \boldsymbol{n}, \tag{47}$$

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. (46) is prohibitively expen-

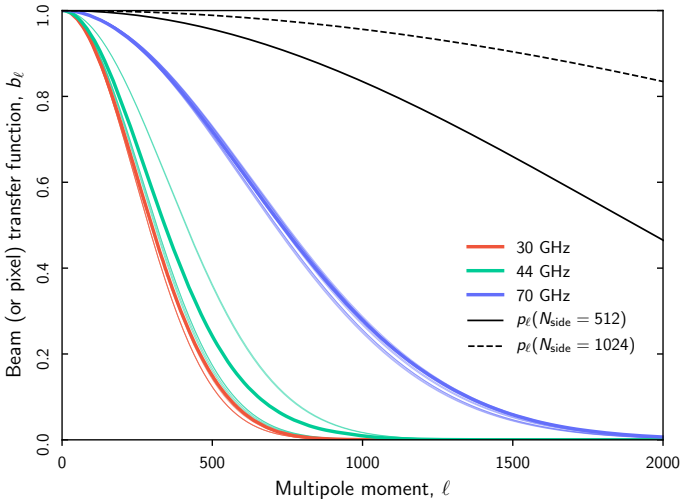

Fig. 4. Azimuthally symmetric and normalized beam transfer functions, b_ℓ for each co-added *Planck* LFI frequency channel (thick colored lines) and for each radiometer (thin colored lines). The former include the effects of non-Gaussian tails, while the latter are strictly Gaussian approximations. Black curves show the `HEALPix` pixel window functions for $N_{\rm side} = 512$ and 1024, respectively.

sive 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. (46) 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 $\boldsymbol{b}$.

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[6]

$$s^{\rm beam}(\hat{n}) = \sum_{\ell=0}^{\ell_{\rm max}} \sum_{m=-\ell}^{\ell} b_\ell s_{\ell m} Y_{\ell m}(\hat{n}), \tag{48}$$

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

Note that the bandlimit, $\ell_{\rm max}$, in Eq. (48) 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 $\ell_{\rm max}$ 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 $s^{\rm beam}(\hat{n})$ in Eq. (48) 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 $s^{\rm beam}(\hat{n})$ 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 $\ell \gtrsim 2.5 N_{\rm side}$ 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 $N_{\rm side}$ and $\ell_{\rm max}$.

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, $\ell_{\rm max}$, and this leads to acceptable computational costs for treating this component.

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 $\sigma_{\rm FWHM}$, in the following expressed in radians. In the Gaussian case, one can derive an explicit expression for the the beam transfer function in the form

$$b_\ell = \mathrm{e}^{-\frac{1}{2}\ell(\ell+1)\frac{\sigma^2_{\rm FWHM}}{8\ln 2}}, \tag{49}$$

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 $N_{\rm side} = 512$ and 1024, which are the typical pixelizations used for LFI and *WMAP* analysis.

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

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. (47) 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

$$P = GkT_{\rm sys}\Delta\nu, \tag{50}$$

where G is a unit-less gain factor, and $\Delta\nu$ is the width of the bandpass. The system temperature is defined as $T_{\rm sys} = T_{\rm ant} + T_{\rm recv}$, where $T_{\rm ant} = T_{\rm CMB} + T_{\rm fg}$ is the antenna temperature, and $T_{\rm recv}$ 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, $T_{\rm CMB} = 2.7$ K, as other sky components typically only make up a few mK at most. Assuming, therefore, a system temperature of about 13 K, Eq. (50) predicts that the power measured by this device is $P = 1.1$ pW or $P = -90$ dBm,[7] 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. (47), denoted G, is in principle the same gain as in Eq. (50), 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.

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).

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,

$$\boldsymbol{n} = \boldsymbol{n}_{\rm corr} + \boldsymbol{n}_{\rm wn}. \tag{51}$$

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, $T_{\rm sys}$, 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. (35)) may be estimated through the so-called radiometer equation,

$$\sigma_0 = T_{\rm sys}\sqrt{\frac{1}{\Delta\nu\Delta t} + \left(\frac{\Delta g}{g}\right)^2}. \tag{52}$$

Here, Δg is the root-mean-square gain variation over Δt, and $\Delta\nu$ is as usual the receiver bandwidth. Intuitively speaking, this

[7] The unit dBm measures power ratios, x, in decibel relative to 1 mW, i.e., $x = 10\,\log_{10}\frac{P}{1\,\mathrm{mW}}$.

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 $N_{\rm obs}$ independent observations are made of the same sky pixel p, then the effective noise of the corresponding pixel integrates down roughly as $\sigma_p = \sigma_0/\sqrt{N_{\rm obs}}$.

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), $P_n(f) = \left\langle |n_f|^2 \right\rangle$, where n_f are the Fourier coefficients of $\boldsymbol{n}_t$. This PSD is often modelled in terms of a so-called $1/f$ profile, which takes the form

$$P_n(f) = \sigma_0^2 \left[1 + \left(\frac{f}{f_{\rm k}}\right)^{\alpha}\right]. \quad (53)$$

Here, $f_{\rm k}$ 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 $f_{\rm k} \approx 10$ 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.

5. Data

The instrument discussion has until this point been general, and applicable to a wide range of different data sets. In this section, we specialize our discussion to one particular combination of data sets, with *Planck* LFI being the primary target of interest. 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. 3.6 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. (2020), Svalheim et al. (2020b), and Herman et al. (2020) 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[8] project, which specifically aims to combine many state-of-the-art experiments with the ones listed above, including *Planck* HFI (Planck Collaboration III 2020), *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.

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. 6. Each horn is followed by an orthomode transducer (OMT), which splits the incoming radiation into two

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

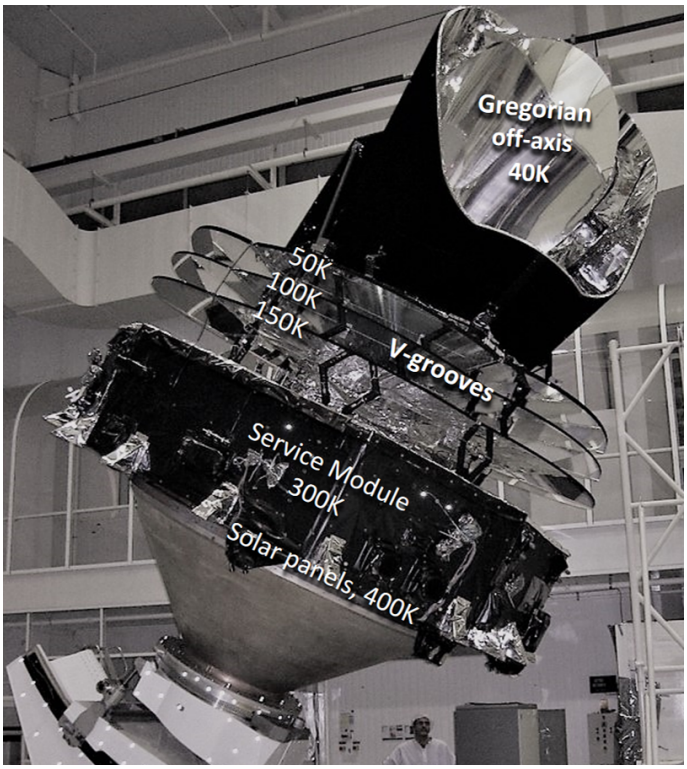

Fig. 5. Flight model of the *Planck* spacecraft. The satellite size is about 4.2×4.2 m, and its mass at launch was 1950 kg. *Planck* was launched on May 14, 2009, and operated for 4.4 years from a Lissajous orbit around the Lagrangian point L2 of the Sun–Earth system. Shown are the approximate temperatures of different critical parts of the satellite during nominal operation in space (see Planck Collaboration II 2011).

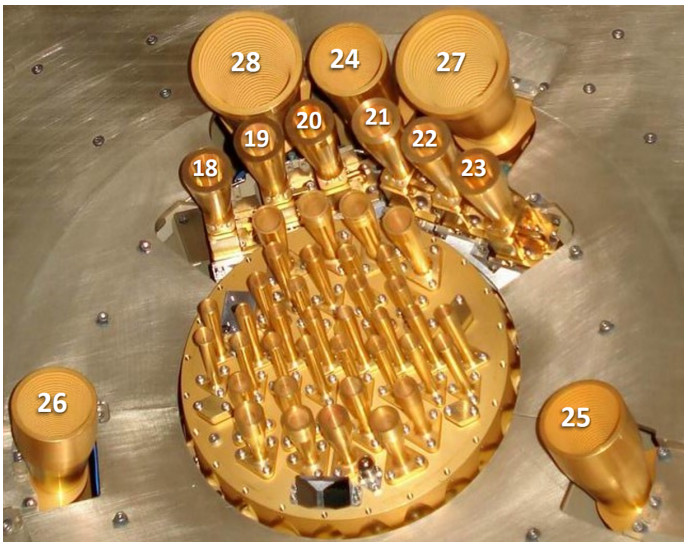

Fig. 6. Top view of the *Planck* focal plane. The central array contains the HFI feed-horns, cooled to 4 K, feeding bolometric detectors cooled to 0.1 K. The LFI horns in the outer part of the array are labelled with numbers; they are cooled to 20 K. The LFI horn numbers 18–23, 24–26, and 27–28 correspond to the 70 GHz, 44 GHz, and 30 GHz channels, respectively.

perpendicular linear polarizations that propagate through two independent differential radiometers; see Fig. 7. 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 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. 5). 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. 7). Then the two mixed signals are amplified by ~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 between the two diodes, the FEM fluctuations cancel out. This "pseudo-correlation" scheme reduces front-end fluctuations by a factor of $O(10^3)$.

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. 7. 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

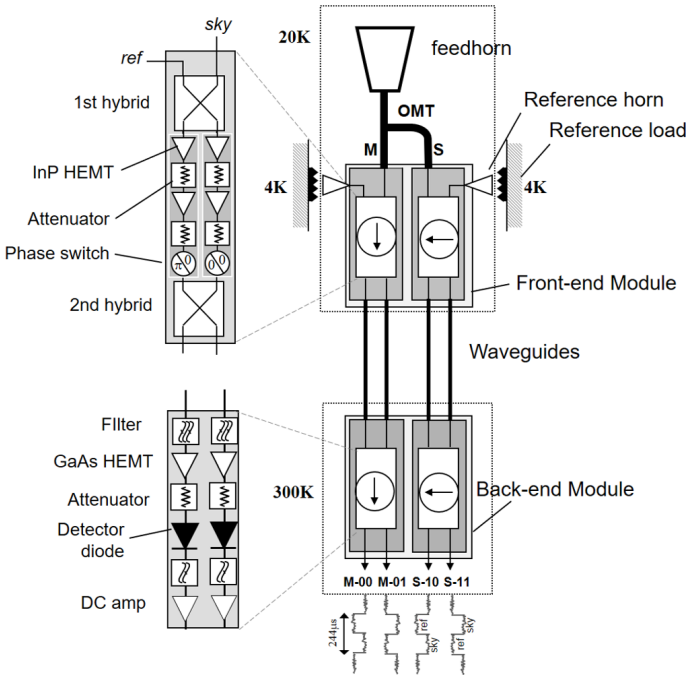

Fig. 7. Schematic of an LFI radiometer chain, consisting of a feedhorn and OMT, and the two associated radiometers, each feeding two diode detectors. The insets show details of a front-end module (top) and of a back end module (bottom).

as the following special case of Eq. (50),

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

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

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

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

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

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

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

where $\epsilon_{T_{\mathrm{n}}}$ 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,

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

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

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

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. 6.

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. 7). 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. 6. 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. 2020a) 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 in 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. Pre-processed Level-2 data

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. First, the data are corrected for the effect of ADC non-linearities and 1 Hz frequency spikes, as outlined in the following sections.

Then the appropriate detector pointing for each sample is computed, based on auxiliary data and beam information. The Level-2 pipeline also calculates the gain modulation factor, r, from the data streams and applies it to minimize $1/f$ noise; see 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).

As for Level-1 data, the output is one HDF5 file for each LFI horn for each Operational Day. Each file contains the differentiated signal, detector pointing (θ, ϕ, ψ) and quality flags for each radiometer, time information (On-Board Time, SCET and MJD) and the same attitude information as in the input data.

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. In the DPC analysis the spikes were found 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. We adopt the same procedure without changes.

5.2.4. Analog-to-digital conversion correction

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). Again, we adopt these corrections without modification, and leave full posterior sampling of ADC corrections to future work.

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. 3.6.

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[9] is $152\,\mu\mathrm{K_{CMB}\,s^{-1/2}}$ (Planck Collaboration II 2016), while it for the HFI 143 GHz bolometers is $57.5\,\mu\mathrm{K_{CMB}\,s^{-1/2}}$ (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).

[9] The noise equivalent temperature (NET) represents the noise standard deviation, σ_0, expressed in thermodynamic units of $\mu\mathrm{K_{CMB}}$ 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 $\mathrm{mK\,s^{-1/2}}$.

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.

Very recently, the S-PASS survey by Carretti et al. (2019) was made publicly available. This provides a detailed map of the southern celestial hemisphere in both intensity and polarization at 2.3 GHz. In principle, this map could play a similar role to the 408 MHz survey for BeyondPlanck. However, its limited sky fraction leads to significant complications. Additionally, the S-PASS polarization map is significantly affected by Faraday rotation (Krachmalnicoff et al. 2018; Fuskeland et al. 2019), which require detailed modelling before inclusion into the main BeyondPlanck pipeline is possible.

6. Bayesian analysis and MCMC sampling

We have now defined an effective parametric model of the astrophysical sky in Sect. 3.6, and an effective instrument model in Eq. (47). We now 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 $\omega \equiv \{g, \Delta_{\rm bp}, \boldsymbol{n}_{\rm corr}, \boldsymbol{a}_i, \beta_i, C_\ell, \ldots\}$. 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 8 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, 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,[10] 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[11] 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

[10] 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.

[11] The symbol "$\leftarrow$" indicates setting the symbol on the left-hand side equal to a sample drawn from the distribution on the right-hand side.

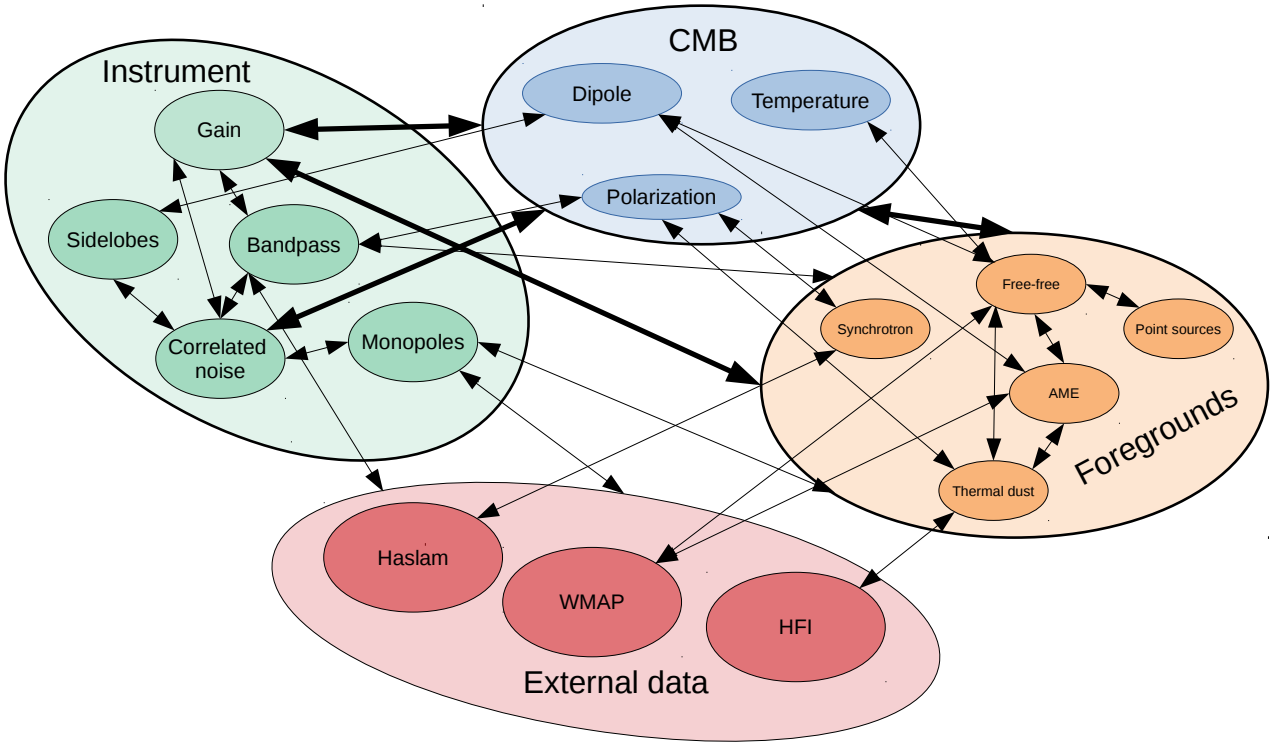

Fig. 8. Schematic overview of the primary parameters and external data sets considered in the current BeyondPlanck analysis and their interdependencies. This chart is intended to visualize the deeply integrated nature of a modern CMB analysis problem; changing any one of these parameter can lead to significant changes in a wide range of other parameters, and tracing these joint uncertainties is critically important for high-precision experiments.

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.

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

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

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

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

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):

$$\begin{aligned}
a &= \frac{P(\alpha_{i+1},\beta_{i+1})}{P(\alpha_i,\beta_i)}\frac{T_{\rm Gibbs}(\omega_i \mid \omega_{i+1})}{T_{\rm Gibbs}(\omega_{i+1} \mid \omega_i)} && (64)\\
&= \frac{P(\alpha_{i+1},\beta_{i+1})}{P(\alpha_i,\beta_i)}\frac{P(\alpha_i \mid \beta_{i+1})\,\delta(\beta_i-\beta_{i+1})}{P(\alpha_{i+1} \mid \beta_i)\,\delta(\beta_{i+1}-\beta_i)} && (65)\\
&= \frac{P(\alpha_{i+1},\beta_i)}{P(\alpha_i,\beta_i)}\frac{P(\alpha_i \mid \beta_i)}{P(\alpha_{i+1} \mid \beta_i)} \qquad \beta_{i+1} = \beta_i && (66)\\
&= \frac{P(\alpha_{i+1} \mid \beta_i)\,P(\beta_i)}{P(\alpha_i \mid \beta_i)\,P(\beta_i)}\frac{P(\alpha_i \mid \beta_i)}{P(\alpha_{i+1} \mid \beta_i)} \qquad P(\alpha,\beta) = P(\alpha \mid \beta)P(\beta) && (67)\\
&= 1, && (68)
\end{aligned}$$

where we have used the definitions of both conditional[12] and marginal[13] 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 trick 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 trick is used in several places in the BeyondPlanck Gibbs chain, for instance for the combination of instrumental gain and correlated noise (Gjerløw et al. 2020), and for the combination of astrophysical component amplitudes and spectral parameters in intensity (Andersen et al. 2020), both of which are internally strongly correlated.

[12] Definition of a conditional distribution: $P(\alpha \mid \beta) \equiv P(\alpha,\beta)/P(\beta)$

[13] Definition of a marginal distribution: $P(\beta) \equiv \int P(\alpha,\beta)\,{\rm d}\alpha$

7. Global model specification

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

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

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 $N_{\rm TOD} \times 3N_{\rm pix}$ pointing matrix defined in Eq. (36), where ψ is the polarization angle of the respective detector with respect to the local meridian;
- B_j denotes the beam convolution in Eq. (46) in the form of a matrix operator; note that for computational efficiency reasons we only take into account beam asymmetries for the sidelobes and orbital dipole in this paper;
- $\mathsf{M}_{cj}(\beta_p,\Delta_{\rm bp})$ denotes element (c,j) of an $N_{\rm comp} \times N_{\rm comp}$ mixing matrix defined in Eq. (43), 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 $\Delta_{\rm bp}$;
- 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;
- $\boldsymbol{s}^{\rm orb}_j$ is the orbital CMB dipole signal in units of $\mathrm{K_{cmb}}$, including relativistic quadrupole corrections;
- $\boldsymbol{s}^{\rm fsl}_j$ denotes the contribution from far sidelobes, also in units of $\mathrm{K_{cmb}}$;
- $n^{\rm corr}_{j,t}$ denotes correlated instrumental noise, as defined by Eqs. (51) and (53); and
- $n^{\rm w}_{j,t}$ is uncorrelated (white) instrumental noise.

For notational convenience, we also define

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

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

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

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

The free parameters in Eq. (69) are $\{\boldsymbol{g}, \Delta_{\rm bp}, \boldsymbol{n}_{\rm corr}, \boldsymbol{a}, \beta\}$. 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 sidelobe 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 $\boldsymbol{a}^c$ 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 $\boldsymbol{a}$ 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 piecewise stationary and Gaussian distributed with vanishing mean and a covariance matrix equal to $\mathsf{N}^{\rm w}_{tt'} = \sigma_0^2 \delta_{tt'}$. 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, for instance as given by the $1/f$ model in Eq. (53). With this approximation, the total noise PSD is modelled in terms of a total of three free parameters, namely the white noise level σ_0, a knee frequency $f_{\rm knee}$, and a low-frequency spectral slope α. We denote the spectral noise parameters collectively as ξ_n.

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 $\boldsymbol{a} = \boldsymbol{a}^{\rm cmb}$ and $\mathsf{M}^{\rm cmb} = 1$ in thermodynamic temperature units, with an angular CMB power spectrum defined as $C_\ell = \left\langle |a^{\rm cmb}|^2 \right\rangle$. 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 $\omega \equiv \{\boldsymbol{g}, \Delta_{\rm bp}, \boldsymbol{n}_{\rm corr}, \xi_n, \boldsymbol{a}, \beta, C_\ell\}$, 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. 3.6.

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

The first derived quantity we consider 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,

$$r^{(0)}_{j,t} = \frac{d_{j,t} - n^{\rm corr}_{j,t}}{g_{t,j}} - \left(s^{\rm orb}_{j,t} + s^{\rm fsl}_{j,t}\right). \tag{73}$$

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, $r^{(0)}_{j,t}$ could be individually binned into a pixelized map for each radiometer j given the pointing information in P^j_{tp}. 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. (2020a), we adopt a very straightforward approach by simply subtracting a correction from each detector TOD, prior to map binning, of the form

$$\delta s^{\rm leak}_{j,t} = \mathsf{P}^j_{tp}\mathsf{B}^j_{pp'}\left(s^{\rm sky}_{jp'} - \left\langle s^{\rm sky}_{jp'}\right\rangle\right), \tag{74}$$

where $s^{\rm sky}_j$ 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 $\delta\boldsymbol{s}^{\rm leak}$ 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. (73) accordingly,

$$r_{j,t} = \frac{d_{j,t} - n^{\rm corr}_{j,t}}{g_{t,j}} - \left(s^{\rm orb}_{j,t} + s^{\rm fsl}_{j,t} + \delta s^{\rm leak}_{j,t}\right). \tag{75}$$

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 $\left\langle \delta s^{\rm leak}\right\rangle = 0$ 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),

$$\left(\sum_{j\in\nu}\mathsf{P}^t_j(\mathsf{N}^{\rm w}_j)^{-1}\mathsf{P}_j\right)\boldsymbol{m}_\nu = \sum_j \mathsf{P}^t_j(\mathsf{N}^{\rm w}_j)^{-1}\boldsymbol{d}_j. \tag{76}$$

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

$$\begin{bmatrix} \sum\frac{1}{\sigma^2_{0,j}} & \sum\frac{\cos 2\psi_{j,t}}{\sigma^2_{0,j}} & \sum\frac{\sin 2\psi_{j,t}}{\sigma^2_{0,j}} \\ \sum\frac{\cos 2\psi_{j,t}}{\sigma^2_{0,j}} & \sum\frac{\cos^2 2\psi_{j,t}}{\sigma^2_{0,j}} & \sum\frac{\cos 2\psi_{j,t}\sin 2\psi_{j,t}}{\sigma^2_{0,j}} \\ \sum\frac{\sin 2\psi_{j,t}}{\sigma^2_{0,j}} & \sum\frac{\sin 2\psi_{j,t}\cos 2\psi_{j,t}}{\sigma^2_{0,j}} & \sum\frac{\sin^2 2\psi_{j,t}}{\sigma^2_{0,j}} \end{bmatrix} \begin{bmatrix} T \\ Q \\ U \end{bmatrix} = \begin{bmatrix} \sum\frac{d_j}{\sigma^2_{0,j}} \\ \sum\frac{d_j\cos 2\psi_{j,t}}{\sigma^2_{0,j}} \\ \sum\frac{d_j\sin 2\psi_{j,t}}{\sigma^2_{0,j}} \end{bmatrix}, \tag{77}$$

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, $\mathsf{N}^{-1}_{pp'}$, is given simply by the inverse of the matrix on the left-hand side of Eq. (76).

It is important to note that the frequency maps defined by Eq. (76) 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 $\boldsymbol{s}^{\rm leak}$ 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. (74) 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 $N_{\rm det} - 1$ such spurious maps for a configuration involving $N_{\rm det}$ detectors. Thus, we generalize the pointing model for a single observation in terms of the Stokes parameters and spurious maps as follows,

$$s_j = T + Q\cos 2\psi_j + U\sin 2\psi_j + \sum_{i=1}^{N_{\rm det}-1} S_i\,\delta_{ij}. \tag{78}$$

Given this definition, the mapmaking equation in Eq. (76) generalizes straightforwardly, and for the special case of three detectors, the contribution of a single sample from detector j takes the schematic form

$$\begin{bmatrix} 1 & \cos 2\psi & \sin 2\psi & \delta_{1j} & \delta_{2,j} \\ \cos 2\psi & \cos^2 2\psi & \cos 2\psi\sin 2\psi & \cos 2\psi\,\delta_{1j} & \cos 2\psi\,\delta_{2,j} \\ \sin 2\psi & \sin 2\psi\cos 2\psi & \sin^2 2\psi & \sin 2\psi\,\delta_{1j} & \cos 2\psi\sin 2\psi\,\delta_{2,j} \\ \delta_{1j} & \cos 2\psi\,\delta_{1i} & \sin 2\psi\,\delta_{1j} & \delta_{1j} & 0 \\ \delta_{2j} & \cos 2\psi\,\delta_{2i} & \sin 2\psi\,\delta_{2j} & 0 & \delta_{2j} \end{bmatrix} \begin{bmatrix} T \\ Q \\ U \\ S_1 \\ S_2 \end{bmatrix} = \begin{bmatrix} d \\ d\cos 2\psi \\ d\sin 2\psi \\ d\,\delta_{1j} \\ d\,\delta_{2j} \end{bmatrix}. \tag{79}$$

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. (79) 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. (2020a) 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, $s^{\rm orb}_{j,t}$. Including a relativistic quadrupole correction, this has a closed form as given by

$$s^{\rm orb}_{j,t} = \frac{T_{\rm CMB}}{c}\left(\boldsymbol{v}_{\rm sat}\cdot\hat{\boldsymbol{n}}_{j,t} + q(\boldsymbol{v}_{\rm sat}\cdot\hat{\boldsymbol{n}}_{j,t})^2\right), \tag{80}$$

where

$$q = \frac{x(\mathrm{e}^{2x}+1)}{\mathrm{e}^{2x}-1}; \quad x = \frac{h\nu}{2kT_{\rm CMB}} \tag{81}$$

is the frequency dependency of the relativistic quadrupole term. The CMB temperature is in our analysis fixed to $T_{\rm CMB} = 2.7255\,\mathrm{K}$, following Fixsen (2009). Finally, c is the speed of light, h is Planck's constant, k is Boltzmann's constant, $\boldsymbol{v}_{\rm sat}$ 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. (2020b).

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, $\boldsymbol{s}^{\rm fsl}$, 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),

$$s^{\rm fsl}_{j,t} = \int_{4\pi} [\mathsf{R}(\Omega_t)\mathsf{B}(\Omega)]\, s^{\rm sky}_j(\Omega)\,\mathrm{d}\Omega, \tag{82}$$

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. (2020b).

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, $P(\omega \mid \boldsymbol{d})$, 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, $\boldsymbol{n}^{\rm corr}$, which for the purposes of the likelihood is fully equivalent to any of the other physical components. As such, we may symbolically write $\boldsymbol{d} = \boldsymbol{s}^{\rm tot}(\omega) + \boldsymbol{n}^{\rm w}$, where $\boldsymbol{s}^{\rm tot}(\omega)$ 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

$$P(\boldsymbol{d}\mid\omega) \propto P(\boldsymbol{n}^{\rm w}\mid\omega) \propto \exp\left(-\frac{1}{2}\left(\frac{\boldsymbol{d}-\boldsymbol{s}^{\rm tot}(\omega)}{\sigma_0}\right)^2\right), \tag{83}$$

since $\boldsymbol{n}^{\rm w} = \boldsymbol{d} - \boldsymbol{s}^{\rm tot}(\omega)$, $P(\boldsymbol{n}^{\rm w}) \propto N(0,\sigma^2)$, and $\boldsymbol{s}^{\rm tot}$ 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,

$$\begin{aligned}
\boldsymbol{g} &\leftarrow P(\boldsymbol{g} \mid \boldsymbol{d}, \xi_n, \Delta_{\rm bp}, \boldsymbol{a}, \beta, C_\ell) && (84)\\
\boldsymbol{n}_{\rm corr} &\leftarrow P(\boldsymbol{n}_{\rm corr} \mid \boldsymbol{d}, \boldsymbol{g}, \xi_n, \Delta_{\rm bp}, \boldsymbol{a}, \beta, C_\ell) && (85)\\
\xi_n &\leftarrow P(\xi_n \mid \boldsymbol{d}, \boldsymbol{g}, \boldsymbol{n}_{\rm corr}, \Delta_{\rm bp}, \boldsymbol{a}, \beta, C_\ell) && (86)\\
\Delta_{\rm bp} &\leftarrow P(\Delta_{\rm bp} \mid \boldsymbol{d}, \boldsymbol{g}, \boldsymbol{n}_{\rm corr}, \xi_n, \boldsymbol{a}, \beta, C_\ell) && (87)\\
\beta &\leftarrow P(\beta \mid \boldsymbol{d}, \boldsymbol{g}, \boldsymbol{n}_{\rm corr}, \xi_n, \Delta_{\rm bp}, C_\ell) && (88)\\
\boldsymbol{a} &\leftarrow P(\boldsymbol{a} \mid \boldsymbol{d}, \boldsymbol{g}, \boldsymbol{n}_{\rm corr}, \xi_n, \Delta_{\rm bp}, \beta, C_\ell) && (89)\\
C_\ell &\leftarrow P(C_\ell \mid \boldsymbol{d}, \boldsymbol{g}, \boldsymbol{n}_{\rm corr}, \xi_n, \Delta_{\rm bp}, \boldsymbol{a}, \beta), && (90)
\end{aligned}$$

where the conditional variables have been vertically aligned for clarity only. As usual, the symbol ← 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, $P(\boldsymbol{g} \mid \boldsymbol{d}, \ldots)$ represents a marginal distribution with respect to $\boldsymbol{n}_{\rm corr}$. As such, $\boldsymbol{g}$ and $\boldsymbol{n}_{\rm corr}$ are in effect sampled jointly in the BeyondPlanck Gibbs chain (Gjerløw et al. 2020; Ihle et al. 2020), 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 trick is applied when sampling astrophysical component parameters, $\boldsymbol{a}$ and β in the case of intensity maps. In this case, we first sample β marginalized over $\boldsymbol{a}$, and then $\boldsymbol{a}$ conditionally on β (Andersen et al. 2020). Since $\boldsymbol{a}$ is a set of linear parameters, the integral over $\boldsymbol{a}$ 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 $\boldsymbol{a}$, as described by Svalheim et al. (2020b), 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. (2020) for full algorithmic details.

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.[14] However, since the Earth's orbital velocity is 30 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.

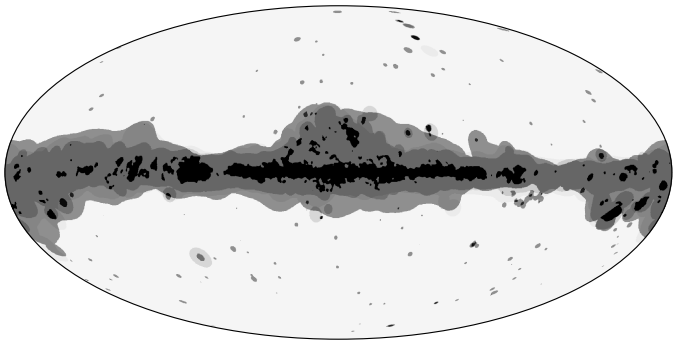

Fig. 9. Processing masks used for low-level TOD processing. The gray regions show the main processing mask used for gain and correlated noise sampling (with 30 GHz shown in lightest gray, and 70 GHz in darkest gray), which remove 27, 19, and 23 % of the sky for 30, 44 and 70 GHz, respectively. The black regions show the bandpass correction sampling mask, which only removes 4.7 % of the sky.

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 $T_{\rm CMB}$. Together, these two facts provide a strong handle on relative gain variations, both between detectors and in time.

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

$$d_t = g_t s_t^{\rm tot} + n_t^{\rm corr} + n_t^{\rm w} \tag{91}$$

where

$$s_t^{\rm tot} = \mathsf{P}_{tp}\left[\mathsf{B}_{pp'}^{\rm symm} s_{p'}^{\rm sky} + \mathsf{B}_{pp'}^{\rm asymm}\left(s_t^{\rm orb} + s_t^{\rm fsl}\right)\right] \tag{92}$$

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. (91) that these two parameters are highly degenerate: significant variations in $\boldsymbol{g}$ can be accounted for by adjusting $\boldsymbol{n}_{\rm corr}$ 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

[14] The term "calibration" refers in this paper to the time average of the 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

$$d_t^i = (g_0 + \Delta g^i + \delta g_t^i) s_t^{\rm tot} + n_t^{\rm tot}, \tag{93}$$

where $n_t^{\rm tot} = n_t^{\rm corr} + n_t^{\rm wn}$ 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

$$r_t^i \equiv d_t^i - \hat{g}(s_t^{\rm tot} - s^{\rm orb}) - (\Delta\hat{g}^i + \delta\hat{g}_t^i) s_t^{\rm orb} = g_0 s_t^{\rm orb} + n_t^{\rm tot}, \tag{94}$$

where the symbol "^" denotes the respective estimate from the previous Gibbs iteration. Noting that $n^{\rm tot}$ is a Gaussian field with covariance N and $\boldsymbol{s}^{\rm orb}$ 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,

$$g_0 = \frac{\sum_i \boldsymbol{s}_i^{{\rm orb},t} \mathsf{N}_i^{-1} \boldsymbol{r}_i^{\rm orb}}{\sum_i \boldsymbol{s}_i^{{\rm orb},t} \mathsf{N}_i^{-1} \boldsymbol{s}_i^{\rm orb}} + \frac{\eta}{\sqrt{\sum_i \boldsymbol{s}_i^{{\rm orb},t} \mathsf{N}_i^{-1} \boldsymbol{s}_i^{\rm orb}}}, \tag{95}$$

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 $\boldsymbol{s}^{\rm tot}$ and $\boldsymbol{s}^{\rm orb}$. By fitting only to $\boldsymbol{s}^{\rm orb}$ we effectively gain robustness with respect to modelling errors at the cost of increased statistical uncertainty.

Note that the noise covariance matrices in Eq. (95) include both correlated and white noise contributions; this is algebraically equivalent to marginalizing over $\boldsymbol{n}_{\rm corr}$ as described above. In constrast, in a classic Gibbs sampling algorithm we would subtract $\boldsymbol{n}_{\rm corr}$ from r^i in Eq. (96), 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 $\boldsymbol{n}_{\rm corr}$ 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

$$r_t^i \equiv d_t^i - (g_0 + \delta\hat{g}_t^i) s_t^{\rm tot} = \Delta g_i s_t^{\rm tot} + n_t^{\rm tot}, \tag{96}$$

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

$$\sum \Delta g_i = 0, \tag{97}$$

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

$$\mathcal{L}(\Delta g_i, \lambda) = \sum_i \left(\boldsymbol{r}^i - \Delta g_i \boldsymbol{s}^{\rm tot}\right)^t \mathsf{N}_i^{-1} \left(\boldsymbol{r}^i - \Delta g_i \boldsymbol{s}^{\rm tot}\right) + \lambda \sum_i \Delta g_i, \tag{98}$$

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

$$r_t^i \equiv d_t^i - (g_0 + \Delta g_i)\, s_t^{\rm tot} = \delta g_i s_t^{\rm tot} + n_t^{\rm tot}, \tag{99}$$

where δg_t^i is assumed to be constant within each PID period, but free to vary between consecutive PIDs. Ideally, δg_t^i should be sampled from a multivariate Gaussian distribution, taking into account their known $1/f$ power dependency. In the current implementation, however, we adopt a similar solution as the LFI DPC pipeline (Planck Collaboration II 2020). That is, we first estimate δg_t^i independently for each time segment, and then smooth the resulting gain estimates with a moving boxcar average. In contrast, NPIPE does not smooth the estimates (Planck Collaboration Int. LVII 2020).

To prevent foreground modelling errors from affecting the various gain estimates, we apply the processing masks indicated in gray in Fig. 9 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 $\mathsf{N}_{\rm tot} = \mathsf{N}_{\rm corr} + \mathsf{N}_{\rm wn}$ 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. (2020); Ihle et al. (2020) for further details.

Finally, we note that several discrete events took place during the *Planck* observations, for instance maneuvers or cooler maintenance, and many of these led to sharp jumps in the LFI gain (Planck Collaboration II 2020). To avoid smoothing across such steps, we boxcar average independently over each continuous PID range between two events. A total of 23 events are included in this manner. However, we do note that this represents a significant unmodelled uncertainty in the current BeyondPlanck analysis: The presence of unknown jumps can contaminate the final resulting polarization maps at significant level. Indeed, as discussed in Sect. 9.5, unmodelled gain jumps are viable candidates for explaining one of the main residual systematic effect in the 44 GHz channel, namely strong stripes visible in the Southern Galactic hemisphere. Automatically identifying jumps as an intrinsic step of the sampling algorithm is an important task for future work.

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. (2020).

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

$$r_t \equiv d_t - g_t s_t^{\rm tot} = n_t^{\rm corr} + n_t^{\rm w}, \tag{100}$$

where $s_t^{\rm tot}$ is defined in Eq. (92). As discussed in Sect. 7.1, $\boldsymbol{n}^{\rm corr}$ is assumed to be Gaussian distributed with zero mean and covariance $\mathsf{N}^{\rm corr}$, while the white noise term is uncorrelated Gaussian with variance σ_0^2. Eq. (100) therefore also describes a correlated Gaussian distribution, and the sampling equation is in this case given by Eq. (A.10) with a template matrix $\mathsf{T} = 1$, a signal covariance matrix $\mathsf{S} = \mathsf{N}^{\rm corr}$, a noise covariance matrix $\mathsf{N} = \mathsf{N}^{\rm w}$, 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.10) can be solved very conveniently in the Fourier domain, and the appropriate sampling equation for the kth Fourier mode reads

$$n_k = \frac{r_k + \eta_1\sigma_0^2/\sqrt{P_k^{\rm corr}} + \eta_2\sigma_0}{1 + \sigma_0^2/P_k^{\rm corr}}. \tag{101}$$

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

Equation (101) 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. 9. In this case, the optimal solution is therefore given by Eq. (A.10), where rows and columns corresponding to masked samples are set to zero in N. The resulting equation is therefore solved efficiently by a Conjugate Gradient (CG) technique, as described by Keihänen et al. (2020). As reported by Galloway et al. (2020a), 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, 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. (101) 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. (101) is a Wiener filter, and therefore has greatly suppressed power in any frequency mode for which $P_k^{\rm corr} \ll \sigma_0^2$. 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 $\boldsymbol{m}_{\rm D} \cdot \boldsymbol{m} = 0$, where $\boldsymbol{m}_{\rm D}$ 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 degeneracies 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 $\boldsymbol{n}_{\rm corr}$ 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 $\boldsymbol{n}_{\rm corr} \sim N(0, \mathsf{N}_{\rm corr})$. Therefore,

$$P(\xi \mid \boldsymbol{d}, \omega \setminus \xi^n) \propto P(\xi \mid \boldsymbol{n}_{\rm corr}) \tag{102}$$

$$\propto \frac{\mathrm{e}^{-\frac{1}{2}\boldsymbol{n}_{\rm corr}^t \mathsf{N}_{\rm corr}^{-1} \boldsymbol{n}_{\rm corr}}}{\sqrt{|\mathsf{N}_{\rm corr}|}}, \tag{103}$$

where $\mathsf{N}_{\rm corr} = \mathsf{N}_{\rm corr}(\xi^n)$. To sample from this distribution, we could for instance run a Metropolis sampler over ξ^n, using Eq. (103) 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

$$\sigma_0^2 \equiv \frac{\mathrm{Var}(r_t - r_{t-1})}{2}, \tag{104}$$

where we define

$$r_t \equiv d_t - s_t^{\rm tot} - n_t^{\rm corr} \tag{105}$$

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 any flat component of the correlated noise 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 trick 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, and $f_{\rm knee}$, making making convergence assessment difficult.

Given this definition of the white noise variance, the correlated noise level may now be sampled from Eq. (103) by fixing $N^{\rm w}$. Specifically, as discussed by Ihle et al. (2020), the conditional posterior may be written in Fourier space as

$$-\ln P(\xi^n) = \sum_{\nu>0}\left[\frac{|n_\nu^{\rm corr}|^2}{\sigma_0^2\left(\frac{\nu}{\nu_{\rm knee}}\right)^\alpha} + \ln\sigma_0^2\left(\frac{\nu}{\nu_{\rm knee}}\right)^\alpha\right], \tag{106}$$

up to an irrelevant constant, where $n_\nu^{\rm corr}$ are the Fourier coefficients of the correlated noise estimate, $\boldsymbol{n}_{\rm corr}$. We sample from this distribution with a simple inversion sampler (see Appendix A.3), iteratively Gibbs sampling over α and $\nu_{\rm knee}$. Masking and inpainting is handled by the $\boldsymbol{n}_{\rm corr}$ sampling step described in Sect. 8.3.2. Additionally, only frequencies below a fixed upper frequency limit (defined by twice the DPC estimate of $f_{\rm knee}$) is included when sampling α and $\nu_{\rm knee}$, to avoid potential signal leakage and modelling errors from contaminating the correlated noise estimates; for further details, see Ihle et al. (2020).

Before concluding this section, we note once again that uncertainties due to modelling errors as such are not meaningfully accounted for in a traditional Bayesian analysis. This point is important for LFI noise estimation, because Ihle et al. (2020) find that the $1/f$ noise model is not an adequate statistical description for the LFI 30 and 44 GHz frequency channels. In particular, they identify significant noise excess at intermediate temporal frequencies, between 0.1 and 1 Hz, which correspond to angular scales between 6 and 60° on the sky given the *Planck* scanning strategy. As such, this excess noise is highly relevant for large-scale polarization estimation with the 30 and 44 GHz channels. Generalizing the $1/f$ noise model to account for this excess noise is a top priority issue for future work. On the other hand, we do find that the $1/f$ model appears to be good description of the noise at 70 GHz, and by far most of the LFI-specific CMB science has been derived from this particular channel (e.g., Planck Collaboration V 2020).

8.3.4. Bandpass sampling

Next, we consider the bandpass correction conditional distribution, $P(\Delta_{\rm bp} \mid \boldsymbol{d}, \omega \setminus \Delta_{\rm bp})$, and in the following we will consider the most basic form of bandpass correction, namely a linear shift as defined by Eq. (45); see Svalheim et al. (2020a) for further details.

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

$$\Delta_{\rm bp}^j = \bar{\Delta}_{\rm bp} + \delta_{\rm bp}^j. \tag{107}$$

Here the first term is the average over all radiometers within a given frequency channel and the second term is constrained by $\sum_j \delta_{\rm bp}^j = 0$. The motivation for this decomposition is that the two

terms impact the data in qualitatively different ways. The average bandpass shift, $\bar{\Delta}_{\rm bp}$, 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. (2020a).

For this reason, we have implemented two different sampling algorithms for these parameters. First, the mean bandpass correction is sampled with the full time-domain residual on the form

$$\boldsymbol{r}_j = \boldsymbol{d}_j - \boldsymbol{n}_j^{\rm corr} - \mathsf{G}_j\mathsf{P}_j\mathsf{B}_j^{\rm asymm}\left(\boldsymbol{s}_j^{\rm orb} + \boldsymbol{s}_j^{\rm fsl}\right) \tag{108}$$

$$= \mathsf{G}_j\mathsf{P}_j\mathsf{B}_j^{\rm symm}\sum_c \mathsf{M}_{cj}(\beta, \Delta_{\rm bp}^j)\boldsymbol{a}^c + \boldsymbol{n}_j^{\rm w}. \tag{109}$$

Clearly, this residual is highly nonlinear in $\Delta_{\rm bp}$, 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. 9 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. (2020a), the degeneracies between CMB, free-free, AME and $\bar{\Delta}_{\rm bp}$ are too strong to support a robust determination of $\bar{\Delta}_{\rm bp}$ 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 0.3 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, $\delta_{\rm bp}$, 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. (79) for both the old and new parameter values. Next, we define a special purpose χ^2 of the form

$$\chi^2 = \sum_{j=1}^{N_{\rm det}-1}\sum_p \left(\frac{S_j(p)}{\sigma_j(p)}\right)^2, \tag{110}$$

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,

$$a = \min\left(1, \mathrm{e}^{-\frac{1}{2}(\chi^2_{\rm prop}-\chi^2_{\rm i-1})}\right), \tag{111}$$

where $\chi^2_{\rm prop}$ and $\chi^2_{\rm i-1}$ 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. (2020a).

8.3.5. Diffuse component spectral parameter sampling

The fifth conditional distribution in the BeyondPlanck Gibbs chain concerns the foreground SED parameters, $P(\beta \mid \boldsymbol{d}, \omega \setminus \beta)$. Noting that the linear amplitudes $\boldsymbol{a}$ and spectral parameters β are in general highly degenerate for high signal-to-noise models, we employ the same computational trick 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 $\boldsymbol{a}$, and then $\boldsymbol{a}$ conditionally on β. For specific details regarding the following algorithm, we refer the interested reader to Andersen et al. (2020). For polarization, we employ a standard Metropolis sampler that is conditional on the foreground amplitudes; see Svalheim et al. (2020b) 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,

$$\boldsymbol{r}_j = \left(\boldsymbol{d}_j - \boldsymbol{n}_j^{\rm corr}\right)/g_j - \left(\boldsymbol{s}_j^{\rm orb} + \boldsymbol{s}_j^{\rm sl} + \boldsymbol{s}_j^{\rm mono}\right)$$

$$= g_j\mathsf{P}_j\mathsf{B}_j\boldsymbol{s}_j^{\rm sky}(\beta) + \boldsymbol{n}_j^{\rm w}. \tag{112}$$

The left-hand side in this equation is identical to the residual in Eq. (73), which is the input to the binned mapmaker defined by Eq. (76). Under the assumption of azimuthally symmetric beams,[15] B_j, this expression may therefore be rewritten in terms of binned sky maps on the form

$$\boldsymbol{m}_\nu = \mathsf{A}_\nu(\beta)\boldsymbol{a} + \boldsymbol{n}_\nu^{\rm w}, \tag{113}$$

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)

$$P(\beta \mid \boldsymbol{m}) \propto \exp\left(\sum_\nu \left(\mathsf{A}_\nu^t\mathsf{N}_\nu^{-1}\boldsymbol{m}_\nu\right)^t\left(\mathsf{A}_\nu^t\mathsf{N}_\nu^{-1}\mathsf{A}_\nu\right)^{-1}\left(\mathsf{A}_\nu^t\mathsf{N}_\nu^{-1}\boldsymbol{m}_\nu\right)\right). \tag{114}$$

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.

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

For polarization, we employ a likelihood given by the original residual defined by Eq. (113),

$$-2\ln P(\beta \mid \boldsymbol{m}, \boldsymbol{a}) = \sum_{\nu} \left(\frac{\boldsymbol{m}_{\nu} - \mathsf{A}_{\nu}(\beta)\boldsymbol{a}}{\sigma_{\nu}(p)} \right)^2 \tag{115}$$

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 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. 2020; Svalheim et al. 2020b).

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. (2020) and Svalheim et al. (2020b) 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 $P(\boldsymbol{a} \mid \boldsymbol{d}, \omega\backslash\boldsymbol{a})$ directly following β. The relevant data model for $\boldsymbol{a}$ is (similar to β) given by Eq. (113), 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); 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.10) with a template matrix $\mathsf{T} = \sum_{\nu} \mathsf{B}_{\nu}\mathsf{M}_{\nu}$, i.e.,

$$\left(\mathsf{S}^{-1} + \sum_{\nu} \mathsf{M}_{\nu}^{t}\mathsf{B}_{\nu}^{t}\mathsf{N}_{\nu}^{-1}\mathsf{B}_{\nu}\mathsf{M}_{\nu}\right)\boldsymbol{a} = \sum_{\nu} \mathsf{M}_{\nu}^{t}\mathsf{B}_{\nu}^{t}\mathsf{N}_{\nu}^{-1}\boldsymbol{m}_{\nu} + \mathsf{S}^{-1}\mu + \sum_{\nu} \mathsf{M}_{\nu}^{t}\mathsf{B}_{\nu}^{t}\mathsf{N}_{\nu}^{-1/2}\eta_{\nu} + \mathsf{S}^{-1/2}\eta_{0}. \tag{116}$$

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 (116) 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. (116). For specific details, we refer the interested reader to Seljebotn et al. (2019).

Computationally speaking, the main complicating factor associated with Eq. (116) 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. (116) 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. (116) twice. First, within the main Gibbs loop (as defined in Sect. 8.2) we solve Eq. (116) 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. 2020a). 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 (2020).

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. 2020), we additionally solve Eq. (116) 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

$$P(\boldsymbol{a}^{\mathrm{CMB}}, \omega \setminus \boldsymbol{a}^{\mathrm{CMB}} \mid \boldsymbol{d}) = P(\boldsymbol{a}^{\mathrm{CMB}} \mid \boldsymbol{d}, \omega \setminus \boldsymbol{a}^{\mathrm{CMB}}) P(\omega \setminus \boldsymbol{a}^{\mathrm{CMB}} \mid \boldsymbol{d}), \tag{117}$$

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. (116) 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. (117). 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 BeyondPlanck pipeline as the split between mapmaking and likelihood estimation does in a traditional CMB analysis pipeline.

We employ a similar trick 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. 9.5, Colombo et al. (2020), and Paradiso et al. (2020).

For two of the astrophysical foregrounds, namely free-free emission and AME, we use informative priors to stabilize the model (Andersen et al. 2020). 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 unconstrained degrees-of-freedom.[16]

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. 3.4.1, 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 (2020), 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, $\boldsymbol{a}^{\rm src}$, 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. (113), such that the effective residual at frequency ν reads

$$\boldsymbol{r}_i = \boldsymbol{m}_\nu - \sum_{c\neq i} \mathsf{B}_\nu \mathsf{M}_\nu^c \boldsymbol{a}_c, \tag{118}$$

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

$$-2\ln P(a_i,\alpha_i \mid \boldsymbol{m},\omega\setminus\{a_i,\alpha_i\}) = \sum_\nu \left(\frac{\boldsymbol{m}_\nu - U_\nu a_i \left(\frac{\nu}{\nu_{\rm ptsrc}}\right)_i^{\alpha-2} \boldsymbol{t}_\nu^i}{\sigma_\nu(p)}\right)^2, \tag{119}$$

where $\nu_{\rm ptsrc}$ 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 different sampling steps for the CMB amplitude map:

1. full-resolution solution of Eq. (116) with no spatial CMB prior, $\mathsf{S}^{-1}_{\rm CMB} = 0$; 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. (116) with no spatial CMB prior,[17] $\mathsf{S}^{-1}_{\rm CMB} = 0$, 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. (116) with a spatial CMB prior, $\mathsf{S}^{-1}_{\rm CMB} \neq 0$, 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

[16] Note that S plays a fully analoguous role in a multi-variate Gaussian prior as the usual standard deviation in a univariate Gaussian prior, and can be used to adjust the strength of the prior.

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

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 paper is to introduce the end-to-end Bayesian approach as such, rather than deriving the most sensitive CMB likelihood possible, we prefer in this paper 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,

$$P(C_\ell \mid \hat{\mathbf{s}}_{\rm CMB}) \propto \frac{\mathrm{e}^{-\frac{1}{2}\hat{\mathbf{s}}^t_{\rm CMB}(\mathsf{S}(C_\ell)+\mathsf{N})^{-1}\hat{\mathbf{s}}_{\rm CMB}}}{\sqrt{|\mathsf{S}(C_\ell)+\mathsf{N}|}}, \tag{120}$$

where $\hat{\mathbf{s}}_{\rm CMB}$ represents a CMB-plus-noise map and N is its corresponding effective noise covariance map.[18] 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

$$\hat{\mathbf{s}}_{\rm CMB} = \left\langle \mathbf{s}^i_{\rm CMB} \right\rangle \tag{121}$$

$$\mathsf{N} = \left\langle (\mathbf{s}^i_{\rm CMB} - \hat{\mathbf{s}}_{\rm CMB})(\mathbf{s}^i_{\rm CMB} - \hat{\mathbf{s}}_{\rm CMB})^t \right\rangle, \tag{122}$$

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.

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. (120). 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. (120) 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 $N_{\rm samp} > n$. 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 $\hat{\mathbf{s}}_{\rm CMB}$ 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 $3(\ell_{\rm max}+1)^2 \approx 240$ modes. However, many of these modes fall within a typical analysis mask (Colombo et al. 2020), 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 eigen-vectors 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 above. The compressed data and covariance matrix then reads

$$\tilde{\mathbf{s}}_{\rm CMB} = \mathsf{P}\hat{\mathbf{s}}_{\rm CMB} \tag{123}$$

$$\tilde{\mathsf{N}} = \mathsf{P}\mathsf{N}\mathsf{P}^t \tag{124}$$

$$\tilde{\mathsf{S}} = \mathsf{P}\mathsf{S}\mathsf{P}^t. \tag{125}$$

Adopting a multipole threshold of $\ell_{\rm max} = 8$ 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 30 000 fast samples, corresponding to 600 full samples including all systematic effects; see Sect. 9.5.3 and Colombo et al. (2020). 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 $P(C_\ell \mid \boldsymbol{d}, \omega \setminus C_\ell) = P(C_\ell \mid \boldsymbol{a}^{\rm CMB})$; if the true CMB map, $\boldsymbol{s}_{\rm CMB}$, 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 therefore write

$$P(C_\ell \mid \boldsymbol{a}^{\rm CMB}) \propto P(\boldsymbol{a}^{\rm CMB} \mid C_\ell) P(C_\ell) \tag{126}$$

$$= \prod_{m=-\ell}^{\ell} \frac{\mathrm{e}^{-\frac{1}{2}\boldsymbol{a}^\dagger_{\ell m}\mathsf{C}_\ell^{-1}\boldsymbol{a}_{\ell m}}}{\sqrt{|\mathsf{C}_\ell|}} P(\mathsf{C}_\ell) \tag{127}$$

$$= \frac{\mathrm{e}^{-\frac{2\ell+1}{2}\mathrm{tr}(\sigma_\ell \mathsf{C}_\ell^{-1})}}{|\mathsf{C}_\ell|^{\frac{2\ell+1}{2}}} P(\mathsf{C}_\ell), \tag{128}$$

[18] 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.

where $\boldsymbol{a}_{\ell m} = \{a^T_{\ell m}, a^E_{\ell m}, a^B_{\ell m}\}$ and

$$\mathsf{C}_\ell \equiv \begin{bmatrix} C^{TT}_\ell & C^{TE}_\ell & C^{TB}_\ell \\ C^{TE}_\ell & C^{EE}_\ell & C^{EB}_\ell \\ C^{TB}_\ell & C^{EB}_\ell & C^{BB}_\ell \end{bmatrix}; \tag{129}$$

$$\sigma_\ell \equiv \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} \begin{bmatrix} (a^T_{\ell m})^* a^T_{\ell m} & (a^T_{\ell m})^* a^E_{\ell m} & (a^T_{\ell m})^* a^B_{\ell m} \\ (a^E_{\ell m})^* a^T_{\ell m} & (a^E_{\ell m})^* a^E_{\ell m} & (a^E_{\ell m})^* a^B_{\ell m} \\ (a^B_{\ell m})^* a^T_{\ell m} & (a^B_{\ell m})^* a^E_{\ell m} & (a^B_{\ell m})^* a^B_{\ell m} \end{bmatrix}. \tag{130}$$

We typically adopt uniform priors on C_ℓ (although for a discussion of non-uniform priors, see Wandelt et al. 2004 and Larson et al. 2007), and the distribution in Eq. (128) is then known as the inverse Wishart distribution, which has a very simple sampling algorithm (Larson et al. 2007):

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 $C^{TB}_\ell = C^{EB}_\ell = 0$, 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).

In the current work, which applies this only to the CMB temperature power spectrum, we only bin multipoles above $\ell \geq 600$, where we apply uniform binning of $\Delta\ell = 50$. To improve the Monte Carlo mixing in the low signal regime, we interleave the straight inverse Wishart sampler with the joint $(a_{\ell m}, C_\ell)$ re-scaling MCMC sampler introduced by Jewell et al. (2009), where the proposal rule is defined as

$$a^{\rm new}_{\ell m} = \sqrt{\frac{C^{\rm new}_\ell}{C^{\rm old}_\ell}}\, a^{\rm old}_{\ell m}. \tag{131}$$

Essentially, this proposal corresponds to a move along the diagonal of the joint $(a_{\ell m}, C_\ell)$ space, and is efficient wherever these two parameters are highly degenerate, which is the case in the low signal-to-noise regime. However, we do note that for final cosmological parameter constraints, we will only use the LFI-based likelihood presented here up to $\ell < 600$, and rather use the HFI-based *Planck* 2018 likelihood at higher multipoles, where appropriate.

The above algorithm describes an efficient 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 $(\boldsymbol{s}_{\rm CMB}, C_\ell)_i$. 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. (128). 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 $\{\boldsymbol{a}_{\ell m}, \mathsf{C}_\ell\}$ (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. (2020). 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. (2020a) and Gerakakis et al. (2020). In addition, we highly recommend the interested reader to consult the source code.[19] At the same time, we emphasize that these codes are most definitely works in progress, 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 72 Intel Xeon E7-8870 2.10 GHz cores and 1.5 TB of RAM. All time related costs are provided in units of wall-time, and must therefore be multiplied with 72 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. (2020a) for further details.

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

Table 2. Computational resources required for end-to-end BeyondPlanck processing. All times correspond to wall-clock duration while running on a single 72-core node. All times are also approximate, and vary by $\lesssim 5\,\%$ from sample to sample; the values reported correspond to one specific but arbitrarily selected sample. We also note that the TOD processing stage is memory-bus limited, and the reported run-times are therefore significantly correlated between various operations.

Item	30 GHz	44 GHz	70 GHz	Sum
Data volume				
Uncompressed data volume	761 GB	1 633 GB	5 522 GB	7 915 GB
Compressed data volume/RAM requirements	86 GB	178 GB	597 GB	**861 GB**
Processing time (cost per run)				
TOD initialization/IO time	176 sec	288 sec	753 sec	1217 sec
Other initialization				663 sec
Total initialization				**1880 sec**
Gibbs sampling steps (cost per sample)				
Data decompression	36 sec	105 sec	252 sec	393 sec
TOD projection (P operation)	33 sec	49 sec	248 sec	330 sec
Sidelobe evaluation ($\boldsymbol{s}_{\rm sl}$)	58 sec	85 sec	337 sec	480 sec
Orbital dipole ($\boldsymbol{s}_{\rm orb}$)	45 sec	61 sec	343 sec	449 sec
Gain sampling (g)	13 sec	10 sec	71 sec	94 sec
Correlated noise sampling ($\boldsymbol{n}_{\rm corr}$)	355 sec	390 sec	2393 sec	3138 sec
TOD binning (P^t operation)	22 sec	34 sec	442 sec	498 sec
Loss due to poor load-balancing	62 sec	305 sec	135 sec	502 sec
Sum of other TOD steps	32 sec	135 sec	139 sec	306 sec
TOD processing cost per sample	656 sec	1074 sec	4666 sec	**6396 sec**
Amplitude sampling, $P(\boldsymbol{a} \mid \boldsymbol{d}, \omega \setminus \boldsymbol{a})$				527 sec
Spectral index sampling, $P(\beta \mid \boldsymbol{d}, \omega \setminus \beta)$				1080 sec
Other steps				149 sec
Total cost per sample				**8168 sec**

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, $\boldsymbol{d}_t$ (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. (2020a), 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 reduced 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. (2020a), 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 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 that dominant TOD operation is associated with sampling $\boldsymbol{n}_{\rm corr}$, 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 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 time-stream 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 simulatenously 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.

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. 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. (116) and for smoothing maps to a common angular resolution as required for Eq. (114). 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, $\mathsf{M}(\beta;\Delta_{\rm bp})$. 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, $\Delta_{\rm bp}$, 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.

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 produce a total of six independent Monte Carlo Markov chains of samples drawn from the posterior distribution $P(\omega|\boldsymbol{d})$, as described in Sect. 8.2. Each chain has different length but includes at least 225 samples, and we discard the first 25 samples for burn-in. For applications where comparison between chains is useful, we thus retain 200 samples from each chain, for a total of 1200 accepted samples, while for applications where sample order does not matter, we use the full set of accepted samples. With a computational cost of 2.3 wall hours/sample (see Table 2), this set took about three weeks of continuous run time to produce on six nodes, for a total computational cost of 220 000 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

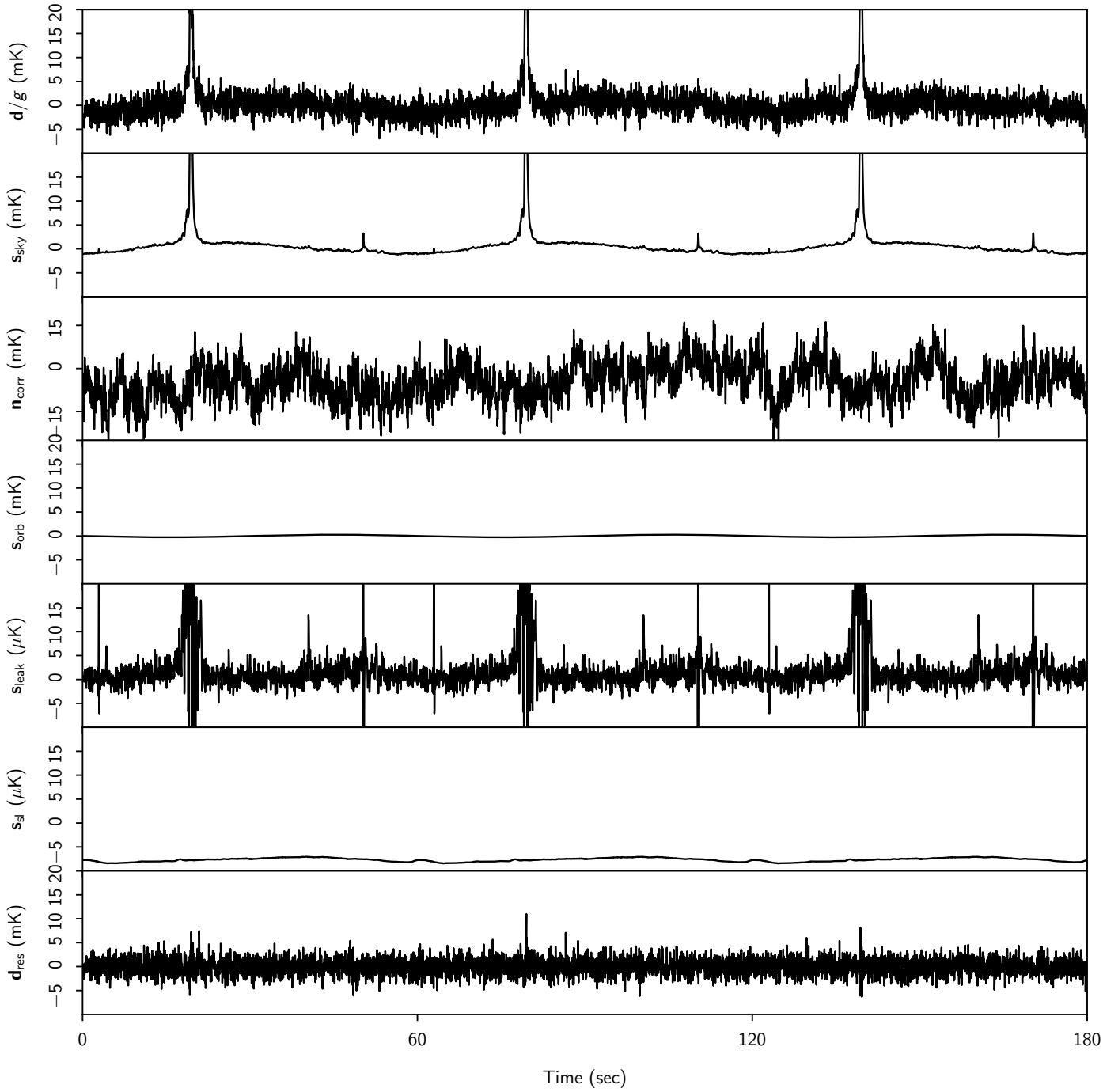

Fig. 10. Time-ordered data segment for the 30 GHz LFI 27M radiometer. From top to bottom, the panels show 1) raw calibrated TOD, $\boldsymbol{d}/g$; 2) sky signal, $\boldsymbol{s}_{\rm sky}$; 3) calibrated correlated noise, $\boldsymbol{n}_{\rm corr}/g$; 4) orbital CMB dipole signal, $\boldsymbol{s}_{\rm orb}$; 5) leakage mismatch correction, $\boldsymbol{s}_{\rm leak}$; 6) sidelobe correction, $\boldsymbol{s}_{\rm sl}$; and 7) residual TOD, $\boldsymbol{d}_{\rm res} = (\boldsymbol{d} - \boldsymbol{n}_{\rm corr})/g - \boldsymbol{s}_{\rm sky} - \boldsymbol{s}_{\rm orb} - \boldsymbol{s}_{\rm leak} - \boldsymbol{s}_{\rm sl}$. Note that the units differ between panels by a factor of 1000, such that each panel spans either a range of 30 mK or 30 μK depending on the respective signal amplitude.

$O(10)$ *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 weeks and months, when more resources are available.

The rest of the section is organized as follows. We first review the various low-level instrumental parameters in Sect. 9.1. In Sect. 9.2 we consider the impact of the various time-domain corrections in the pixel domain, as well as the resulting frequency maps after taking these corrections into account. We discuss goodness-of-fit in Sect. 9.3, before presenting astrophysical sky model parameters in Sect. 9.4 and CMB posterior distributions in Sect. 9.5. We note that this presentation is not intended to be comprehensive, but rather aims to provide a broad overview of the full process, and build intuition regarding the interpretation and interplay between the various components. Individual results are discussed in detail in the companion papers listed in Table 1.

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. 10 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). In addition, for comparison purposes, we scale the entire TOD segment by the inverse gain, converting the signal from volts to μK; we recall, however, that the gain model is constant over this whole section of data, and this does therefore not affect the visual structure, but only the plotting scale. Each panel in Fig. 10 is plotted with a range of either 30 mK or 30 μK, to allow direct visual comparison within groups of similar quantities.

In this plot, we can readily identify by eye many of the features discussed in Sect. 7. First, we see a slow sinusoidal oscillation with a peak-to-peak amplitude of about 3 mK; this is the signature of the solar CMB dipole modulated by the *Planck* scanning strategy. The second most obvious feature is a sharp peak at $t \approx 15$ sec (and periodically repeating with 60 sec intervals), as well as a smaller peak at $t \approx 45$ sec. These correspond to pointings where the satellite points toward the Galactic plane, either near the center or near the anti-center. Beyond that, the only other visually significant fluctuations are due to instrumental noise.

The second panel shows the total sky signal model projected into time-domain, $\mathsf{P}\boldsymbol{s}^{\rm sky}$. Here we can see the same signal features noted above, both the solar dipole and the Galactic plane crossings, only with much higher signal-to-noise, as the total sky model is derived from the full data set, not just the 30 GHz LFI 27M radiometer. Indeed, here we also note some smaller fluctuations superimposed on the dipole signal, and these are primarily higher-order CMB temperature fluctuations. Comparing the amplitude of these fluctuations with the noise level seen in the top panel provides some intuition regarding the amount of data coaddition that is required in order to measure true CMB temperature variations, and the corresponding requirements regarding instrument stability.

The third panel shows the (calibrated) correlated noise component, $\boldsymbol{n}^{\rm corr}/g$; the calibration is performed here only in order to be able to compare its magnitude with the signal components in units of μK. Overall, we see that the correlated noise spans roughly the same range as the sky signal component, but exhibits a distinct $1/f$ type correlation structure in time, and it does not depend on the pointing. In combination, these two facts allow separation of correlated noise from true sky signal.

The fourth panel shows the orbital CMB dipole, caused by the satellite's motion around the Sun, $\boldsymbol{s}^{\rm orb}$. Although its amplitude appears visually small in this plot compared to the signal and correlated noise terms in the second and third panels, it is important that recall that there are no free parameters or uncontrolled uncertainties associated with this particular component. It is therefore possible to correlate its apparent amplitude with the raw data over the entire mission, and thereby derive one absolute calibration value.

The fifth panel shows the bandpass and beam leakage correction, $\boldsymbol{s}^{\rm leak}$, computed as the difference between the sky signal as observed by the 30 GHz LFI 27M radiometer and the mean over all 30 GHz radiometers. Note that the scale of this panel is a factor of one thousand smaller than those of the previous panels. In this plot, we see the impact of the bandpass mismatch in the form of broad spikes centered on the Galactic plane. The magnitude of this effect can reach several tens of μK at low Galactic latitudes, but is almost negligible at high Galactic latitudes. The rapid fluctuations are due to beam mismatch between the various

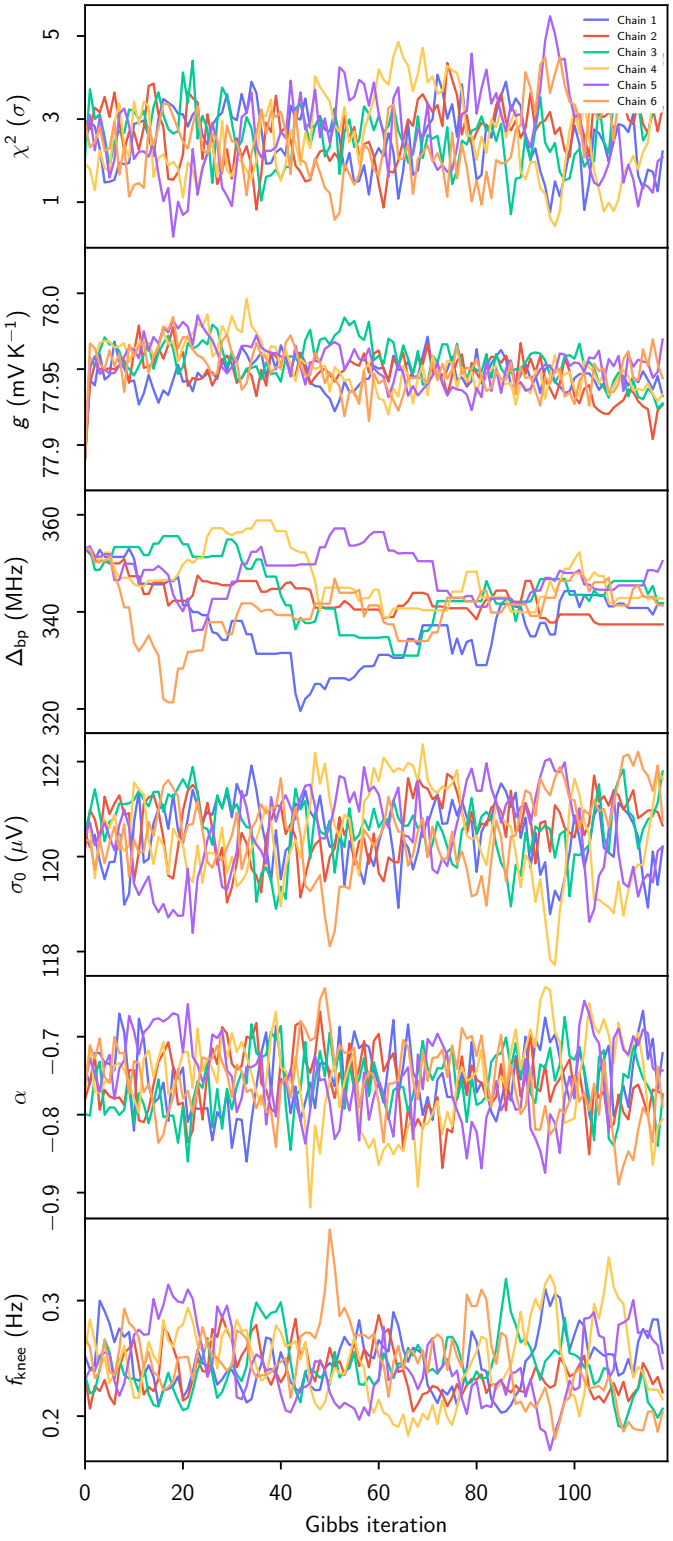

Fig. 11. Example of TOD parameters Gibbs chain for the 30 GHz LFI 27M radiometer. From top to bottom, the panels show normalized reduced χ^2 for a single Pointing Period; gain for the same PID in units of mV K^{-1}; bandpass correction in MHz; white noise level, σ_0; correlated noise slope, α; and correlated noise knee frequency, $f_{\rm knee}$. The six different colored curves correspond to six independent Gibbs chains.

30 GHz detectors, and correspond to CMB fluctuations observed through slightly different beam solid angles. The standard deviation of this component is on the order of 1 μK.

The sixth panel shows the contribution from the far sidelobes, $\boldsymbol{s}^{\rm sl}$. Overall, the two dominant structures in this function are a slow dipole modulation, caused by the solar and orbital CMB dipoles, and a weak imprint of the Galactic plane crossings at about 60 and 120 sec. For this particular combination of radiometer and PID, the overall magnitude of the sidelobe correction is rather limited, although in other cases it may be an order of magnitude brighter than what is seen here.

The bottom panel in Fig. 10 shows the residual resulting from subtracting each of the model terms in Eq. (69) from the raw data. Ideally, this should be uncorrelated stationary Gaussian noise with a standard deviation equal to σ_0. This hypothesis will be tested explicitly in Sect. 9.3, where we compute the χ^2 per PID for each radiometer throughout the mission. In practice, however, correlated features may appear in this residual, and these translate into a bias in the sky map. In this particular example, we see a notable residual near the Galactic plane crossings, which eventually should be masked in higher-level analyses.

Figure 10 represents one single Gibbs sample in the full chain. In contrast, Fig. 11 shows samples from all six Gibbs chains for the instrumental parameters for the same 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 for a single PID (arbitrarily selected to be PID = 5000), as defined by

$$\chi^2 \equiv \frac{\sum_{t=1}^{N_{\rm tod}} \left(\frac{d_t - s_t^{\rm tot}}{\sigma_0}\right)^2 - N_{\rm tod}}{\sqrt{2N_{\rm tod}}}. \tag{132}$$

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.

Regarding the χ^2 values plotted in the top panel of Fig. 11, we see that the mean value is $\chi^2 \approx 2.5$, which thus indicates that the Gaussian white noise null hypothesis is rejected by more than 2σ for this particular PID. This is an early indication that something is indeed wrong with the noise model, and we will return to this observation below.

The second panel shows the gain g for the same PID. In this case, the Markov correlation length appears to be perhaps 20 samples by eye, while the burn-in phase is very short. Considering that we have only 1200 full Gibbs samples available, this implies that the number of independent gain samples per PID is rather limited, and probably around 50. However, this is still sufficient to provide a useful estimate of the true underlying distribution, and the marginal posterior may be summarized as $g = 77.95 \pm 0.02$ mV K^{-1}. Even when adopting a conservative estimate of only 30 independent samples, the sampling uncertainty accounts for less than 20 % of the quoted uncertainty. For further discussion of the gain posteriors, see Gjerløw et al. (2020).

The third panel shows the bandpass shift, $\Delta_{\rm bp}$, 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-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 alternatives. The result is a long correlation length of about 50 samples, resulting in perhaps as few as 20 uncorrelated samples. Still, even with this crude sampler, we do see that the six 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 $\Delta_{\rm bp} = 342 \pm 6$ MHz. However, in this case the sampling uncertainty accounts for at least 30 % of the error bar. For further discussion of the bandpass posteriors, see Svalheim et al. (2020a).

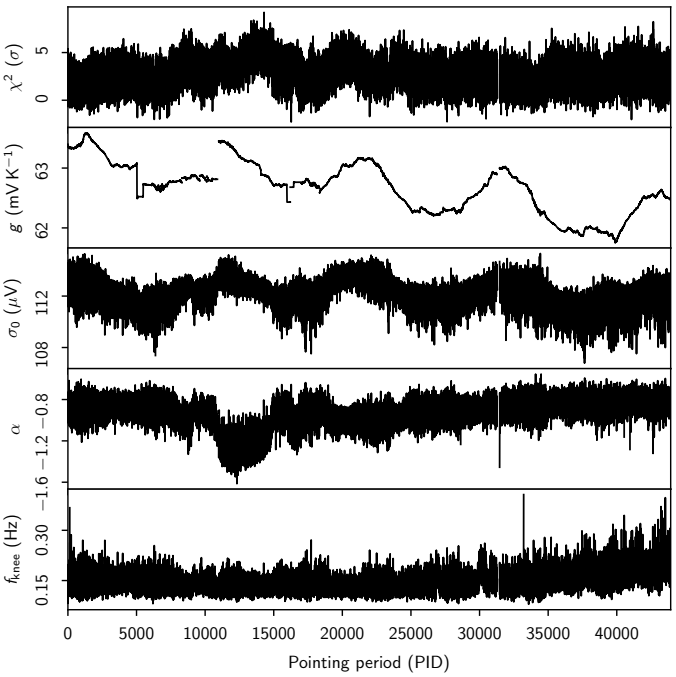

Fig. 12. A single Gibbs sample for the 30 GHz LFI 28M radiometer as a function of PID. From top to bottom the panels show 1) normalized reduced χ^2; 2) gain, g; 3) white noise level, σ_0; 4) correlated noise slope, α; and 5) correlated noise knee frequency, $f_{\rm knee}$.

The three last panels show the three noise PSD parameters, σ_0, α and $f_{\rm knee}$, for the same radiometer. These are much better constrained, because of the large number of samples within each PID. However, there are still obvious internal correlations between the three parameters, such that a low value of σ_0 leads to high values of α and $f_{\rm knee}$, which makes intuitive sense. It is also worth noting that for other parameters in the model, only the actual noise PSD matters, as defined by the combination of these parameters. For further discussion of the noise posteriors, see Ihle et al. (2020).

Figure 12 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. (132). 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. We see that the χ^2 excess noted above for a single 27M PID was not a fluke, but rather a typical example: As discussed by Ihle et al. (2020), all 30 GHz and 44 GHz radiometers exhibit a χ^2 excess per PID at the 2–3 σ

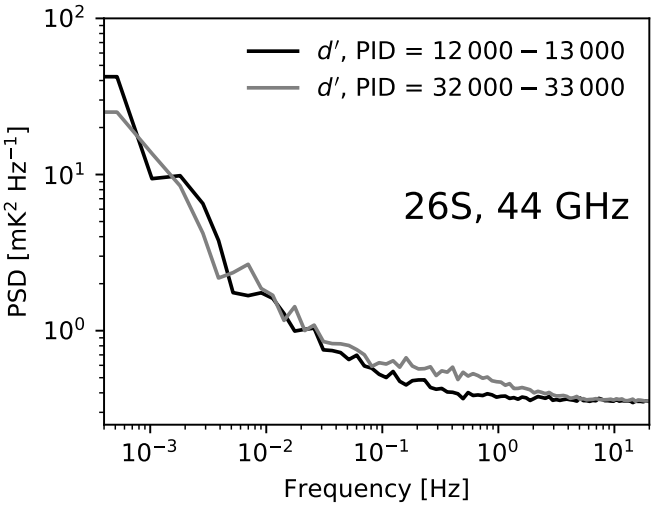

Fig. 13. Comparison of the mean PSD of $\boldsymbol{d} - \boldsymbol{s}^{\rm sky}$ for the 44 GHz LFI 26S radiometer, each of which are averaged over two PID ranges. Note the excess seen in the frequency range between 0.1 and 1 Hz for PIDs between 32 000 and 33 000; this noise contribution cannot be described by a simple $1/f$ model.

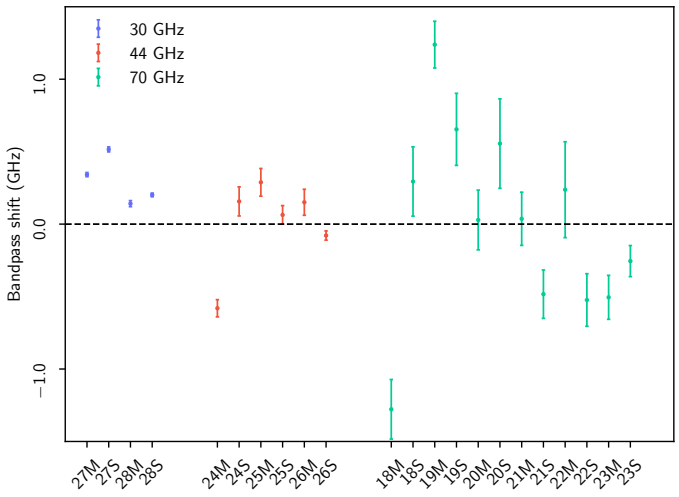

Fig. 14. Estimated bandpass corrections for each LFI radiometer. Error bars indicate $\pm 3\sigma$ uncertainties. Note these parameters are constrained to have vanishing mean within each frequency, and are as such strongly correlated.

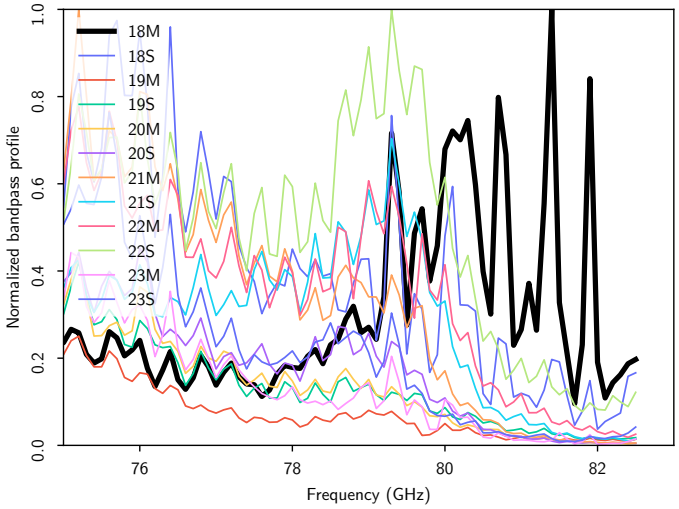

Fig. 15. High frequency tail of individual ground-measured 70 GHz radiometer bandpasses. The 70 GHz LFI 18M radiometer, for which a large negative bandpass correction parameter is derived in the BeyondPlanck Gibbs sampler, is shown as a thick black curve.

level, strongly suggesting a break-down of the assumed noise model for these channels.

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. In addition, we also see evidence for a handful of discrete jumps; these are caused by known events, such as satellite maneuvers or cooler maintenance, as summarized in Planck Collaboration I (2016). We note that this list of discrete jumps is currently inserted by hand into the analysis, and no automatic searches are performed. Undetected jumps may therefore contaminate higher-level analyses, and, as already noted, this issue may represent one of the most important outstanding questions with regards to the 44 GHz channel in the current analysis.

The three bottom panels of Fig. 12 show corresponding plots of the three noise parameters as a function of PID. The same 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. (2020), these can be traced to changes in the thermal environment of the satellite using house-keeping data.

Returning to the systematic χ^2 excess noted above, Fig. 13 compares the power spectral density (PSD) of $\boldsymbol{d} - \boldsymbol{s}^{\rm sky}$ for the 44 GHz LFI 26S radiometer, as averaged over two different PID ranges. Ideally, this should follow a $1/f$ noise profile, and for PIDs between 12 000 and 13 000 this appears to be a reasonable assumption. However, later in the mission there is a clear excess noise contribution at intermediate frequencies, between 0.1 and 1 Hz, that clearly cannot be described as $1/f$ noise. Ihle et al. (2020) shows that this particular radiometer abruptly changes behaviour around PID 20 000, after which this term becomes dominant. For this reason, we choose to exclude the 26S radiometer entirely from the BeyondPlanck analysis during the second half of the mission. Furthermore, to avoid temperature-to-polarization leakage we also exclude the 26M radiometer. In total, this reduces the total available data volume at 44 GHz by 17 %, and the resulting BeyondPlanck 44 GHz frequency map is therefore also correspondingly more noisy than the corresponding DPC and `NPIPE` maps. While this data cut should, at least in theory, result in a more accurate noise characterization for the BeyondPlanck maps, we do emphasize that all other 30 and 44 GHz radiometers also exhibit similar noise excesses, although at quantitiatively lower levels than 26S. In order to actually establish a statistically acceptable noise description of these channels, the assumed $1/f$ noise model will need to be generalized to account for excess noise at intermedate frequencies. This is clearly a top priority for next-generation processing

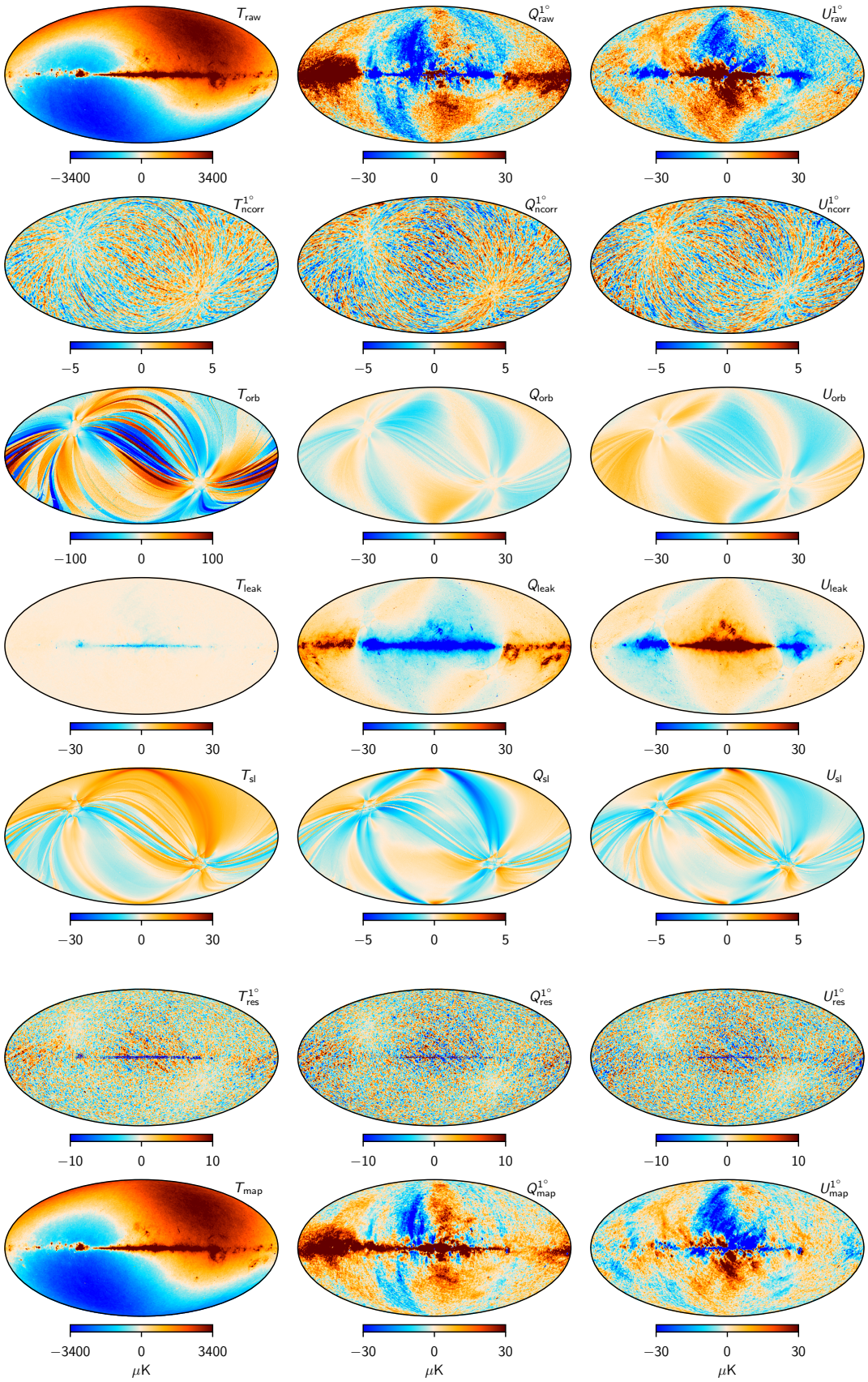

Fig. 16. Comparison between individual individual TOD correction terms and sky maps for the 30 GHz channel for a single Gibbs sample, projected into sky maps. Columns show Stokes T, Q, and U parameters. Rows show, from top to bottom, 1) the raw TOD projected into a sky map; 2) correlated noise, $\boldsymbol{n}_{\rm corr}$; 3) the orbital dipole, $\boldsymbol{s}_{\rm orb}$; 4) bandpass and beam mismatch leakage; and 5) sidelobe corrections, $\boldsymbol{s}_{\rm sl}$. The bottom row shows the final sky map obtained by subtracting all these correction terms from the raw TOD, while the second to last row shows the residual map obtained when also subtracting a model of the sky, $\boldsymbol{s}_{\rm sky}$, from the TOD prior to map binning. Some components, as indicated in the respective labels, have been smoothed to an angular resolution of 1° FWHM. All units are $\mu{\rm K}_{\rm CMB}$.

We now return to our review of the instrumental parameter posterior distributions. In that respect, we note that some low-level parameters are global in nature, and do not map into PID space. The most prominent example of this is the set of bandpass corrections, $\Delta_{\rm bp}$, which are summarized in Fig. 14 in terms of the marginal posterior mean and standard deviation for each radiometer. As described by Svalheim et al. (2020a), the overall shift within each frequency is fixed by comparing the estimates reported by Planck Collaboration X (2016) against the current data set. Doing so, we find that an overall frequency shift of 0.3 GHz is indeed required for the 30 GHz channel, in agreement with Planck Collaboration X (2016), whereas both the 44 and 70 GHz are compatible with no shift. In contrast, the relative distribution of bandpass corrections within a frequency is determined through minimizing the so-called spurious S-maps discussed in Sect. 7.2.2, and this is only mildly dependent on the details of the foreground model. While the absolute bandpass correction for a given channel is directly dependent on the foreground model, the polarization leakage only depends on it to second order, and may therefore be estimated much more directly from a single data set. At the same time, we do note that the uncertainties in the bandpass parameters depend sensitively on the amplitude of the foregrounds. 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.

One visually striking result seen in this plot is the fact that the first 44 and 70 GHz radiometer (24M and 18M, respectively) have a much larger (and negative) correction than other radiometers. To provide some intuition regarding this result, Fig. 15 shows a comparison between the individual 70 GHz radiometer bandpass profiles as measured on ground, zooming in on their high frequency tails. The thick solid black curve shows the 18M radiometer. Here we see that this particular radiometer exhibits large oscillations near the high frequency cutoff, with an amplitude of order unity. No other radiometer shows nearly as strong fluctuations. One possible hypothesis for this result is that spurious standing waves may have been induced by the test equipment itself. Irrespective of its origin, this effect serves as a useful example of the utility of global analysis, by highlighting a previously unnoticed instrumental effect in terms of the final marginal posterior distribution. At the same time, this result also underlines the importance of choosing a physically meaningful parametric model; in this particular case, a parametric model implementing smooth apodization of the tails may be far better motivated than a simple shift model, as adopted in the current analysis. Such models should be explored further in future work.

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 one arbitrarily selected sample (specifically, sample number 101 from the first chain).

The top panel shows the raw TOD binned into a sky map, and provides intuition regarding the overall quality of the data before applying any corrections. Indeed, for the temperature component it is very difficult to spot major artifacts of any kind; the most notable feature is a few correlated noise stripes in the lower left quadrant. For polarization, the dominant effect is the alternating sign along the Galactic plane as a function of longitude, which is due to bandpass mismatch.

The second panel shows the correlated noise component. The most notable features in this map are coherent stripes along the satellite scanning path. It should also be noted that this component is the one that is the least constrained from a-priori considerations among all TOD components, and therefore acts as a "trash can" for possible unmodelled errors; this is the first place one expects to see residuals from modelling errors. And, indeed, we do see hints of artifacts in these maps, perhaps most notably in the Stokes Q parameter: There is a broad bright region near the South Ecliptic Pole. Comparing this with the other panels in the same figure, this feature appears qualitatively similar to all three of the orbital, sidelobe, and leakage corrections. The common denominator among these three terms is their direct dependency on the assumed beam model. As reported by Planck Collaboration II (2020), about 1 % of the *Planck* 30 GHz solid angle has not been accounted for by the current state-of-the-art GRASP model for the LFI 30 GHz channel, and this missing power could at least in principle account for some of the excess structure seen in $\boldsymbol{n}^{\rm corr}$.

The third row shows the orbital dipole. For a single PID, this signal is defined by a perfect dipole along the scanning ring with an amplitude of about 270 μK, convolved with the 4π LFI beam. However, when the same ring is observed six months apart, the phase of the signal is reversed, and the total sum is then both small and difficult to predict. Also, although the intrinsic signal is entirely unpolarized, convolution with far-sidelobes algebraically couples this model to the polarization sector as well.

The fourth row shows the bandpass and beam leakage correction. This effect is clearly the strongest among all polarization corrections, with amplitudes exceeding many tens of μK in the Galactic plane, while still being almost entirely negligible in temperature. Morphologically speaking, the archetypal signature of bandpass mismatch is a variable sign along the Galactic plane, tracing the specific orientation of the detector polarization angles as the different detectors observe at slightly different effective frequencies. At high latitudes, this map is dominated by temperature-to-polarization leakage resulting from different radiometers observing the signal model with different beam FWHMs; large angular scales are dominated by CMB dipole leakage, while small angular scales are dominated by foregrounds and CMB temperature fluctuations.

The fifth row shows the impact of sidelobe pickup. In temperature, the two dominant features are, first, a large-scale pattern broadly aligned with the solar CMB dipole resulting from interactions with the intermediate sidelobes, and, second, individual rings created by the far sidelobes hitting the Galactic plane. The same features are also seen in polarization, but now a more complicated pattern arises due to the additional modulation by the relative orientation of the polarization angles at any given time.

The second to last row shows the TOD residuals binned into a sky map. For most of the sky, this is consistent with white noise, but clear residuals are seen in the Galactic plane, reflecting the structures seen in the bottom panel of Fig. 10. This indicates that the adopted foreground and/or instrument model is not statistically adequate in these very bright regions of the sky. Subsequent CMB-oriented analyses will clearly need to mask these regions prior to power spectrum or parameter estimation.

The last row shows the final co-added frequency map. To assess the overall impact of the various correction terms, this may be directly compared with the raw map shown in the top row.

A complete survey of similar maps for all three LFI frequencies is provided by Suur-Uski et al. (2020), and for the most part these are qualitatively similar to those presented here. However, there is one special case that warrants a dedicated discussion, namely the correlated noise map for the 44 GHz channel. This is shown in Fig. 17, after smoothing to an effective angular reso-

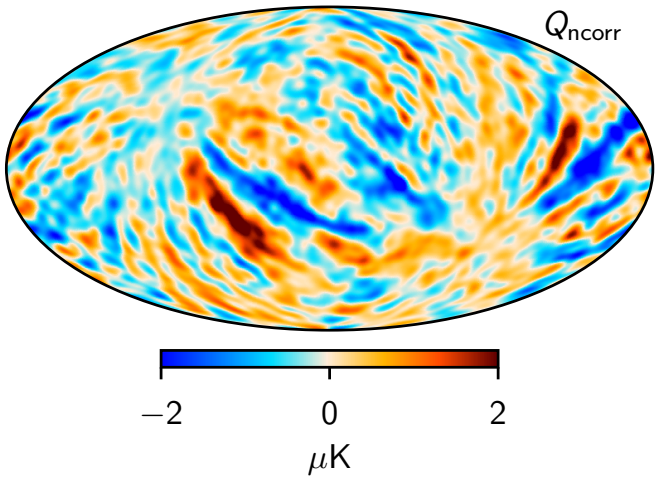

Fig. 17. Correlated noise map, $\boldsymbol{n}^{\rm corr}$, for the 44 GHz frequency channel, smoothed to an effective angular resolution of 5° FWHM. Reproduced from Suur-Uski et al. (2020).

lution of 5° FWHM. The most striking features in this map are strong stripes extending from the right-most edge through most of the Southern Galactic hemisphere. The origin of these stripes has not been robustly identified, and this is an area of intense on-going research. However, we have found that by reducing the gain smoothing window (see Sect. 8.3.1 and Gjerløw et al. 2020) by a factor of two, making the gain model more flexible, these stripes vanish from $\boldsymbol{n}^{\rm corr}$ (Gjerløw et al. 2020). However, that also introduces large gain fluctuations in periods of low CMB Solar dipole amplitude, and this is not an acceptable solution for production analysis. Still, based on preliminary work, it seems likely that these strong 44 GHz stripes are associated with shortcomings in the current gain model, and this will require further work to resolve. Furthermore, excess χ^2 values in the Southern Galactic hemisphere represents the biggest challenge with respect to full-sky CMB polarization reconstruction for BeyondPlanck, and this is thus clearly a top priority issue. In the current analysis, we only use data covering the Northern Galactic hemisphere to derive final low-ℓ CMB polarization constraints.

9.2. Frequency maps

We now turn our attention to co-added frequency maps, as solved for deterministically through Eq. (76). For many users, these represent 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,

$$\hat{\boldsymbol{m}}_\nu = \left\langle \boldsymbol{m}_\nu^i \right\rangle \tag{133}$$

$$\sigma_\nu(p) = \sqrt{\left\langle \left(m_\nu^i(p) - \hat{m}_\nu(p)\right)^2 \right\rangle}, \tag{134}$$

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. (76), $\sigma_\nu^{\rm wn}(p)$. To obtain the full uncertainty, these two terms must be added in quadrature,

$$\sigma_\nu^{\rm tot}(p) = \sqrt{\sigma_\nu(p)^2 + \sigma_\nu^{\rm wn}(p)^2}. \tag{135}$$

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 the posterior mean maps $\hat{\boldsymbol{m}}_\nu$ for each frequency and each Stokes parameter. The polarization maps have been smoothed to an angular resolution of 1° FWHM to reduce noise. Note that the BeyondPlanck temperature maps retain the CMB dipole, similar to `NPIPE` 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.

Figure 19 shows the corresponding posterior standard deviation maps, σ_ν, as defined above, after smoothing to a common angular resolution of 2° FWHM. These maps summarize the combined effect of the various systematic corrections made to the frequency maps, and are are such morphologically rich. The most striking features include:

1. dipole variations in the 30 GHz intensity map, reflecting uncertainties in the absolute calibration of this channel;
2. excess variance for rings aligned with the Galactic plane, 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 2.1 μK, which indicates that these variations are much lower than the intrinsic variance of the CMB temperature fluctuations, which is about 50 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 μK, which is of the same order of magnitude as the expected polarization imprint from cosmic reionization.

Next, Fig. 20 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. They do not account for any coupling to astrophysical foregrounds or instrumental effects beyond those listed above, and, as a result, these maps are structurally much cleaner than the systematic uncertainty maps.

The above figures concern the average frequency map posterior distribution. It is also interesting to consider variations between individual Monte Carlo samples, as they often more easily convey spatial correlations between individual modes. Figure 21 is an example of this, showing 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, $\Delta_{\rm bp}$. 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,

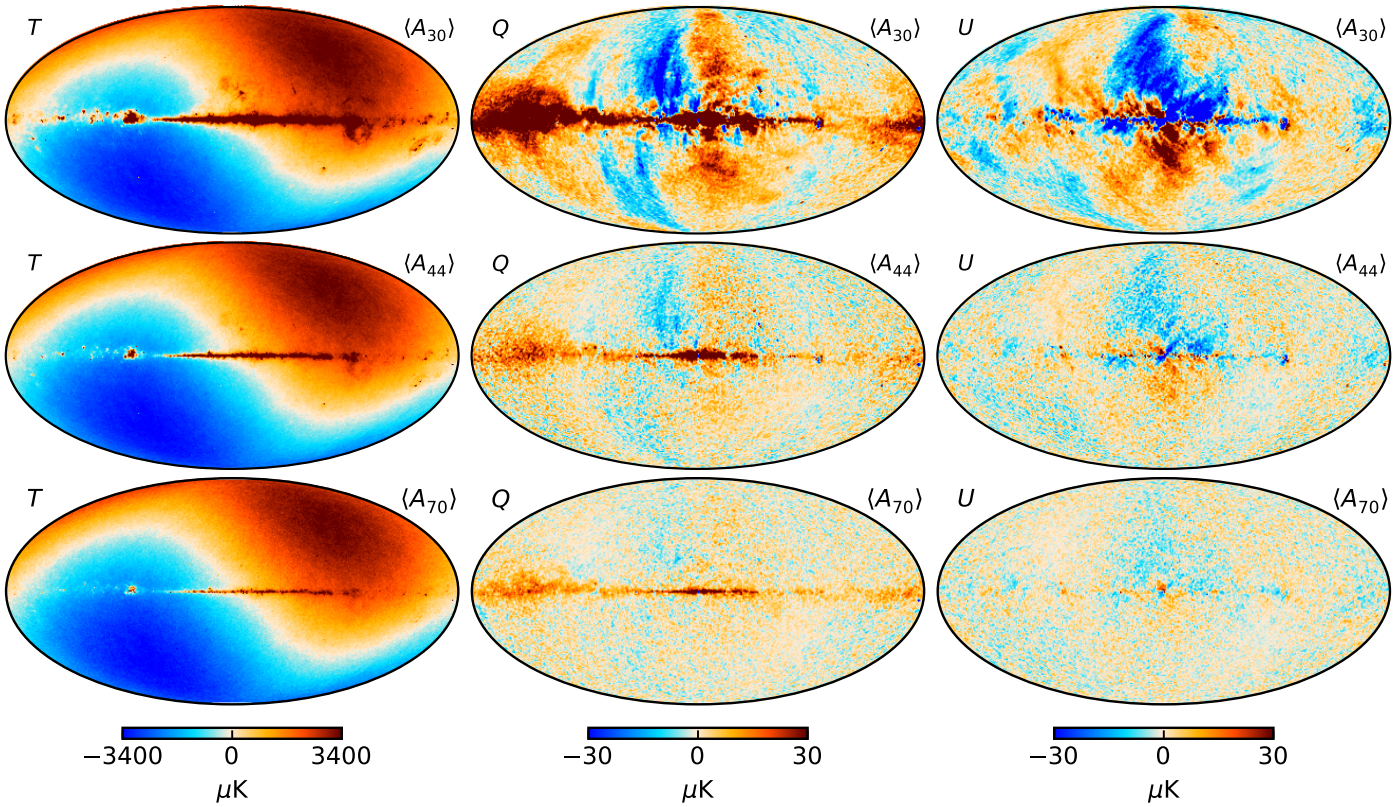

Fig. 18. Posterior mean maps for each LFI frequency. Rows show, from top to bottom, the 30, 44 and 70 GHz frequency channels, while columns show, from left to right, the temperature and Stokes Q and U parameters. The polarization maps have been smoothed to a common angular resolution of 1° FWHM to visually reduce the noise level.

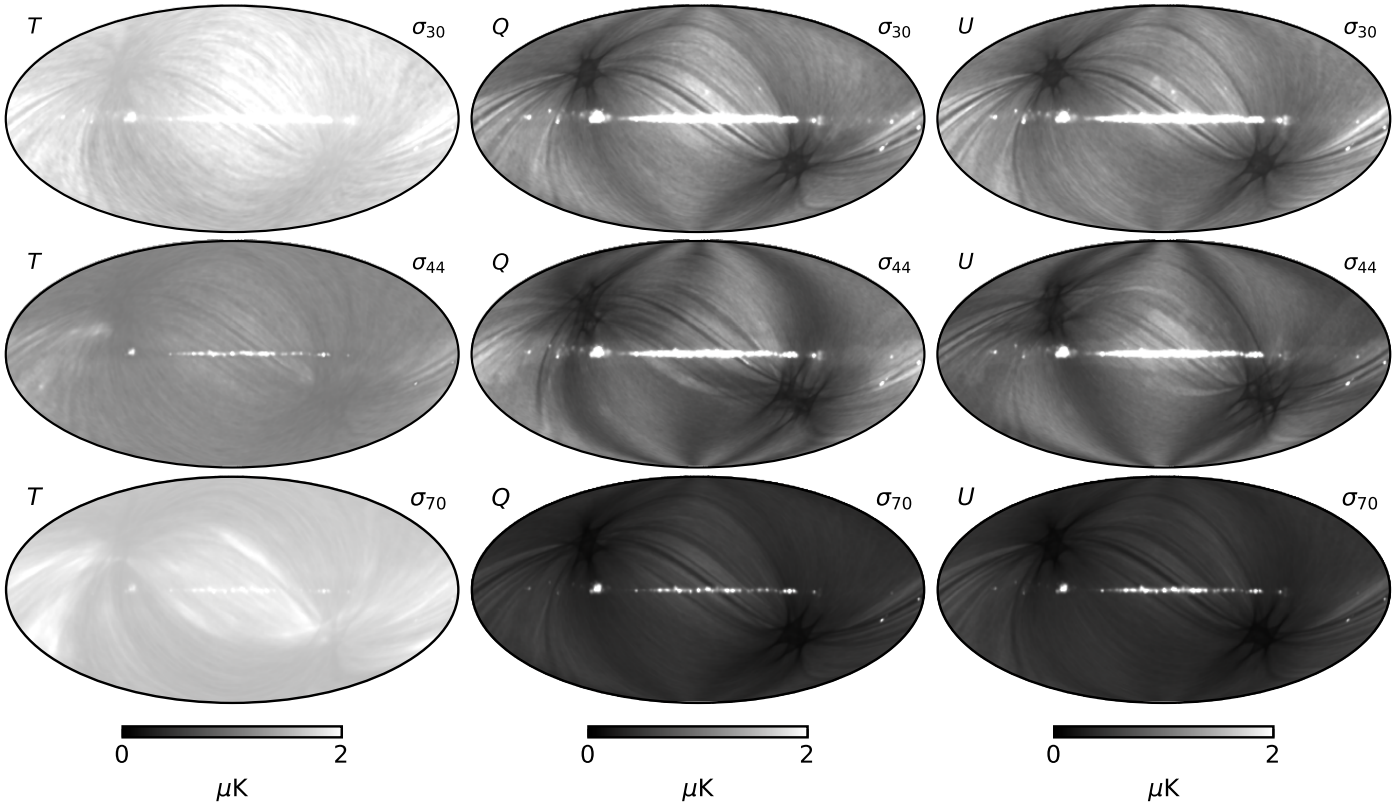

Fig. 19. Posterior standard deviation maps for each LFI frequency. Rows show, from top to bottom, the 30, 44 and 70 GHz frequency channels, while columns show, from left to right, the temperature and Stokes Q and U parameters. Note that these maps do not include uncertainty from instrumental white noise, but only variations from the TOD-oriented parameters included in the data model in Eq. (69).

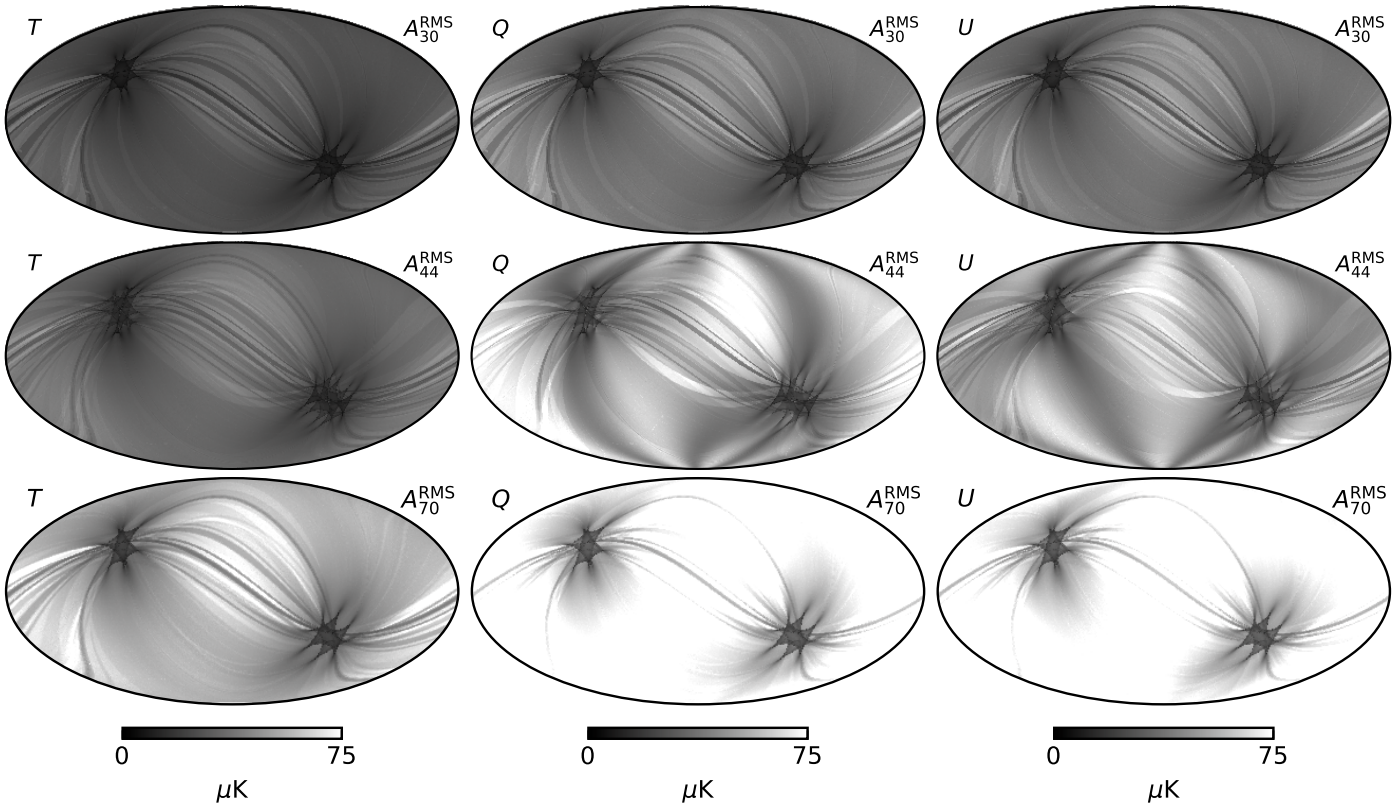

Fig. 20. White noise standard deviation maps for a single arbitrarily selected sample. Rows show, from top to bottom, the 30, 44 and 70 GHz frequency channels, while columns show, from left to right, the temperature and Stokes Q and U parameters. Note that the 70 GHz maps are scaled by a factor of 2, to account for the fact that this map is pixelized at $N_{\mathrm{side}} = 1024$, while the two lower frequencies are pixelized at $N_{\mathrm{side}} = 512$.

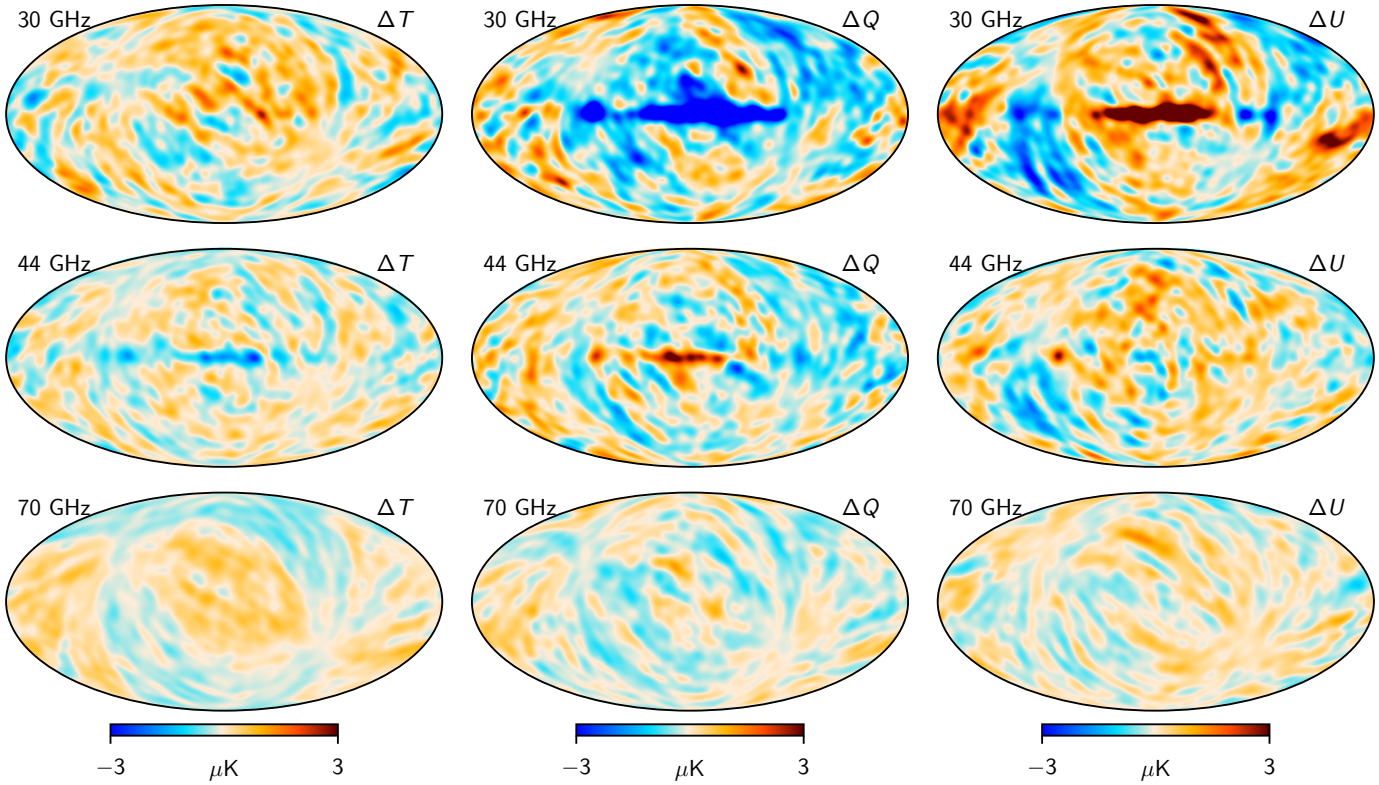

Fig. 21. Difference maps between two frequency map samples, smoothed to a common angular resolution of 7° FWHM. Rows show, from top to bottom, the 30, 44 and 70 GHz frequency channels, while columns show, from left to right, the temperature and Stokes Q and U parameters.

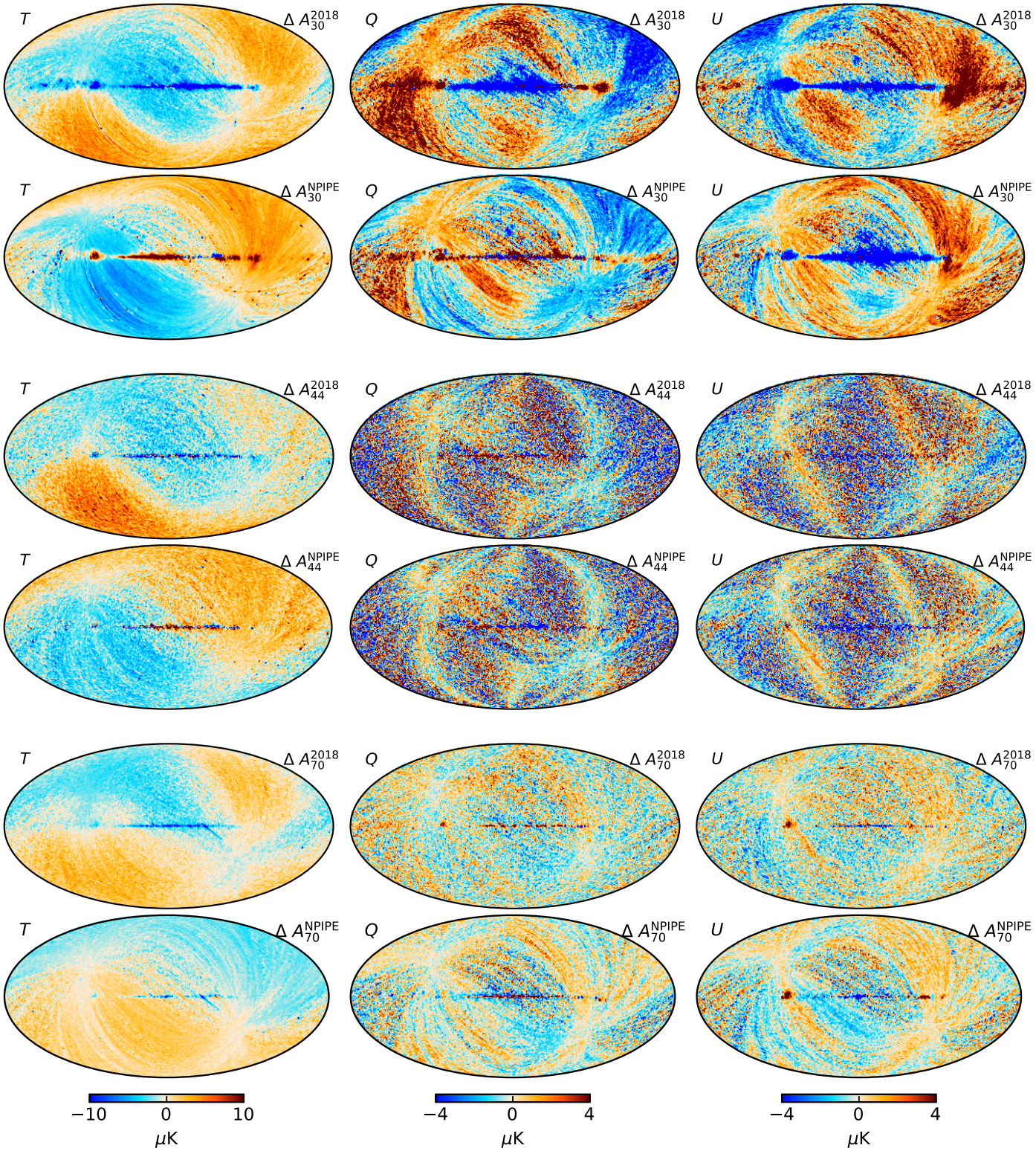

Fig. 22. Differences between BeyondPlanck and 2018 or NPIPE frequency maps, smoothed to a common angular resolution of 2° FWHM. Columns show Stokes T, Q and U parameters, respectively, while rows show pair-wise differences with respect to the pipeline indicated in the panel labels. A constant offset 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.

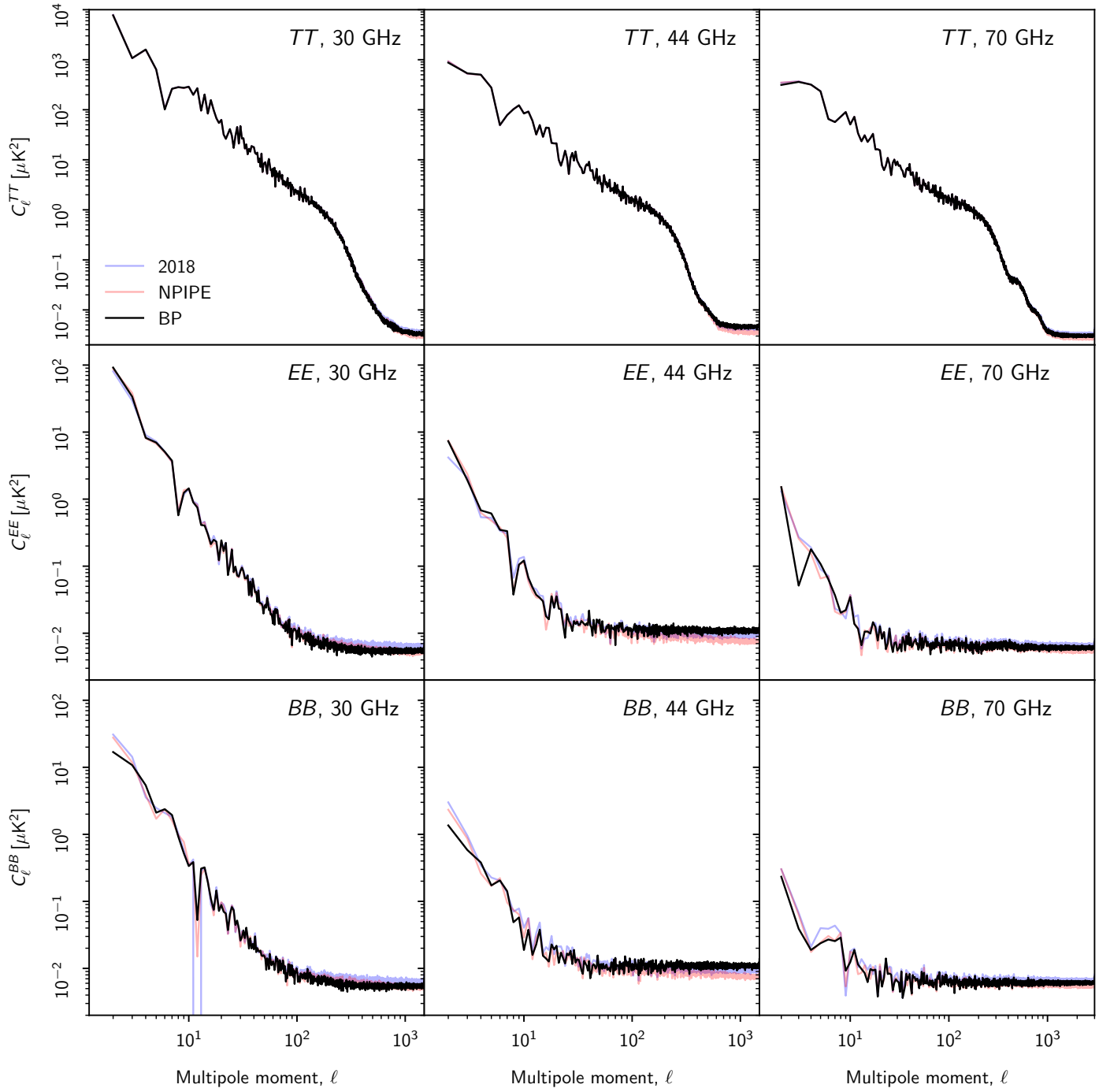

Fig. 23. Comparison between angular auto-spectra computed from the BeyondPlanck (*black*), *Planck* 2018 (*blue*), and NPIPE (*red*) full-frequency maps. Rows show different frequencies, while columns show TT, EE, and BB spectra. All spectra have been estimated with PolSpice using the *Planck* 2018 common component separation confidence mask (Planck Collaboration IV 2020).

again, we strongly recommend analysis of the full map ensemble when using the BeyondPlanck products for external analysis.

Figure 22 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 $10\,\mu$Kin the map-domain. Overall, in temperature BeyondPlanck is thus morphologically similar to NPIPE, but it im-

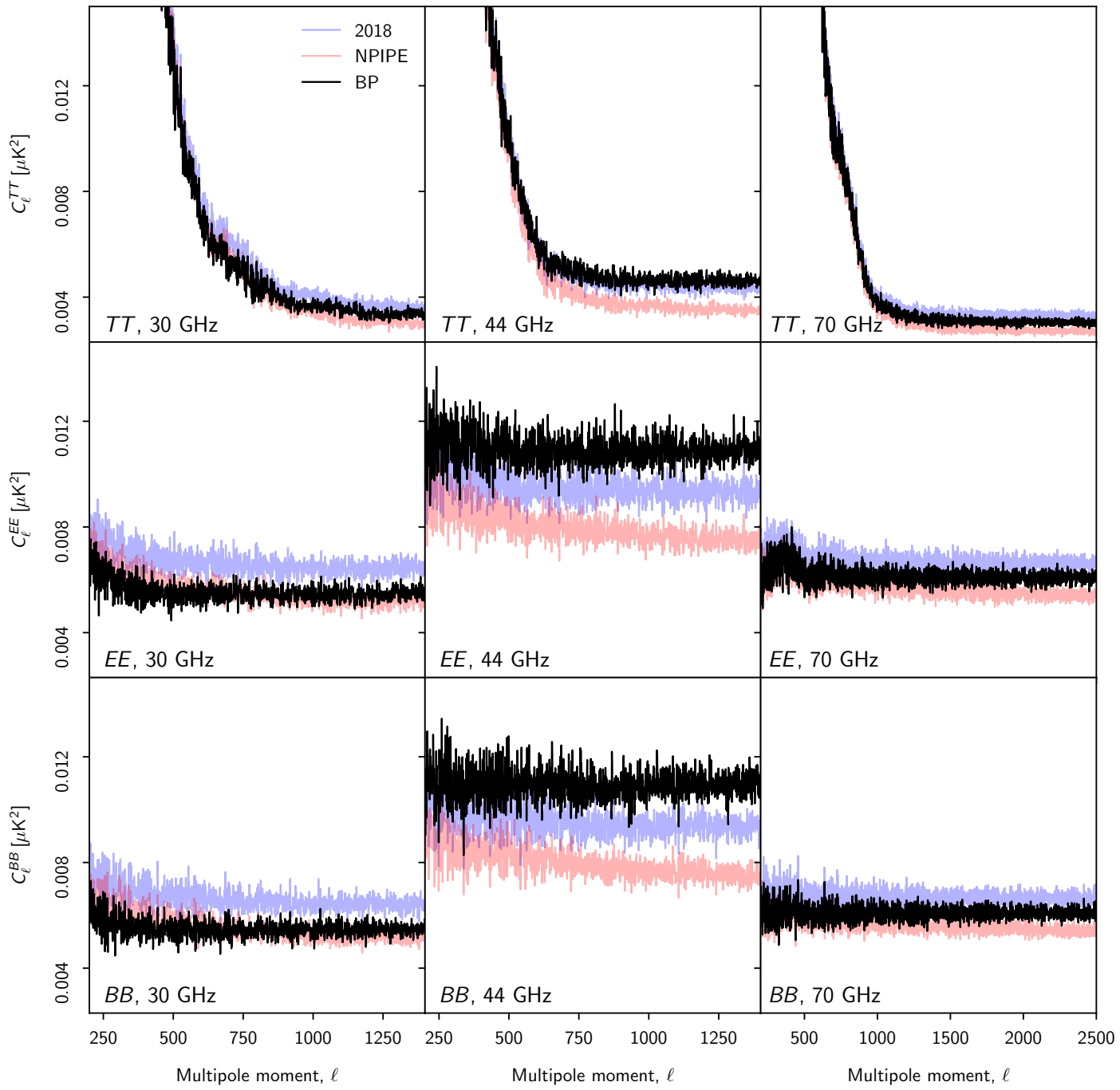

Fig. 24. Same as Fig. 23, but zooming in on the noise-dominated high-ℓ multipole range.

proves 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 modelling 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.

Regarding the 44 GHz channel, two main features stand out in the difference maps in Fig. 22. First, these difference maps obviously exhibit a much higher white-noise level than the corresponding 30 or 70 GHz maps. This is because we exclude data from both the 44 GHz LFI 26M and 26S radiometers for the second half of the mission, as discussed above. Second, we see the same coherent stripes extending through the Southern Galactic hemisphere as seen in the correlated noise map in Fig 17, with clearly different amplitudes in the 2018 and NPIPE differences.

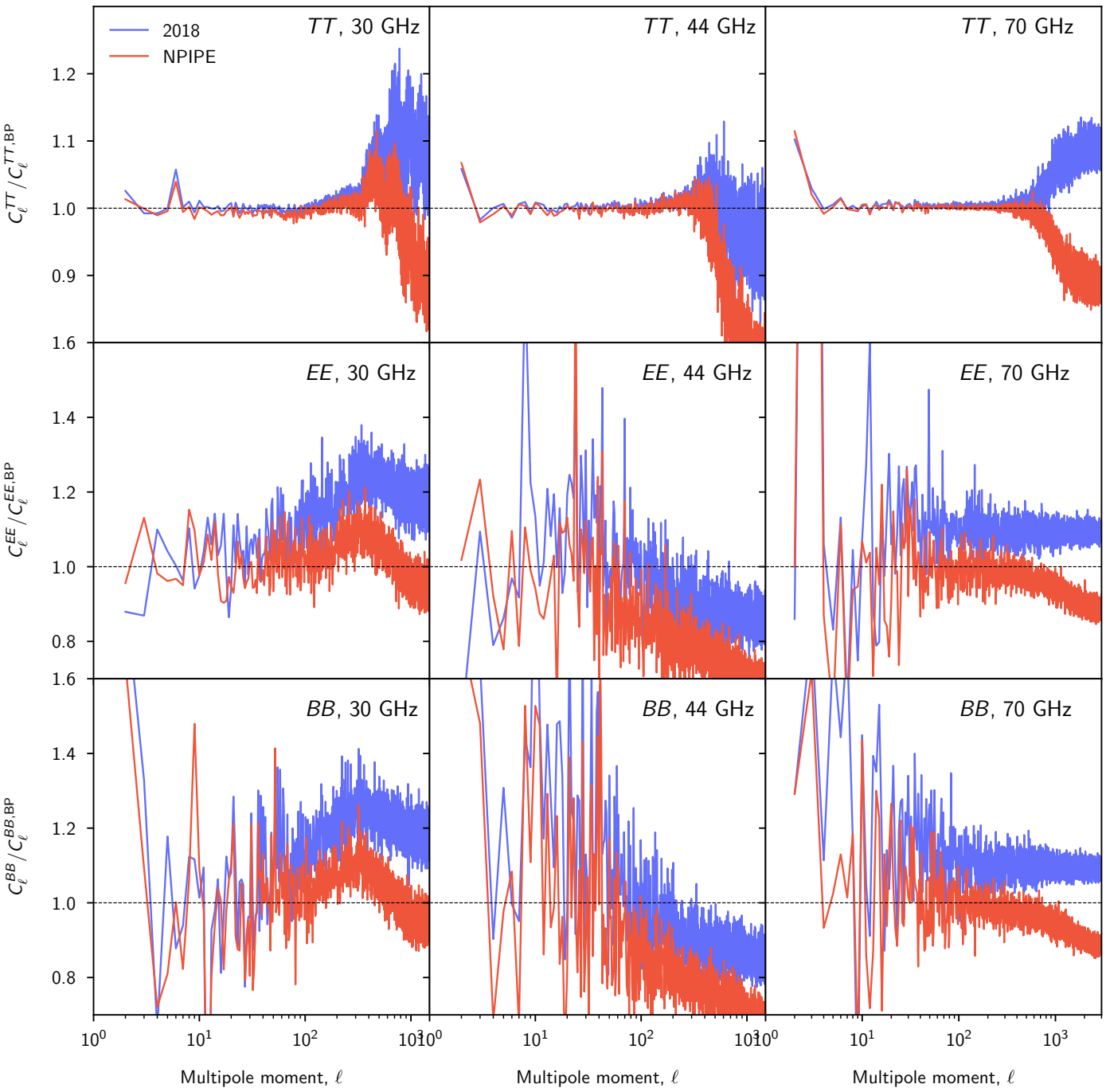

Fig. 25. Ratios between the angular auto-spectra shown in Fig. 23, adopting the BeyondPlanck spectra as reference. *Planck* 2018 results are shown as blue lines, while NPIPE results are shown as red lines. Values larger than unity imply that the respective map has more power than the corresponding BeyondPlanck spectrum.

This shows that at least two out of the three different processing pipelines struggle with this particular problem, whatever its true origin may be.

Finally, and as already noted, the dominant high-latitude structures in the 70 GHz residual maps are morphologically very similar to the 30 GHz differences. This suggesting a foreground-related common mode error in one or more of these maps.

We now turn our attention to the angular power spectrum properties of the BeyondPlanck frequency maps. Figure 23 shows auto-correlation spectra as computed with `PolSpice` (Chon et al. 2004) outside the *Planck* 2018 common component separation confidence mask (Planck Collaboration IV 2020), which accepts a sky fraction of 80 %. All these spectra are clearly signal-dominated at large angular scales (as seen by the rapidly decreasing parts of the spectra at low ℓ's), and noise-dominated at small angular scales (as seen by the flat parts of the spectra at high ℓ's); note that the "signal" in these maps includes both CMB and astrophysical foregrounds. Overall, the three pipelines agree well at the level of precision supported by the logarithmic scale used here; the most striking differences appear to be variations in the high-ℓ plateau, suggesting notably different noise properties between the three different pipelines.

We therefore zoom in on the high-ℓ parts of the spectra in Fig. 24. Here the differences become much more clear, and easy

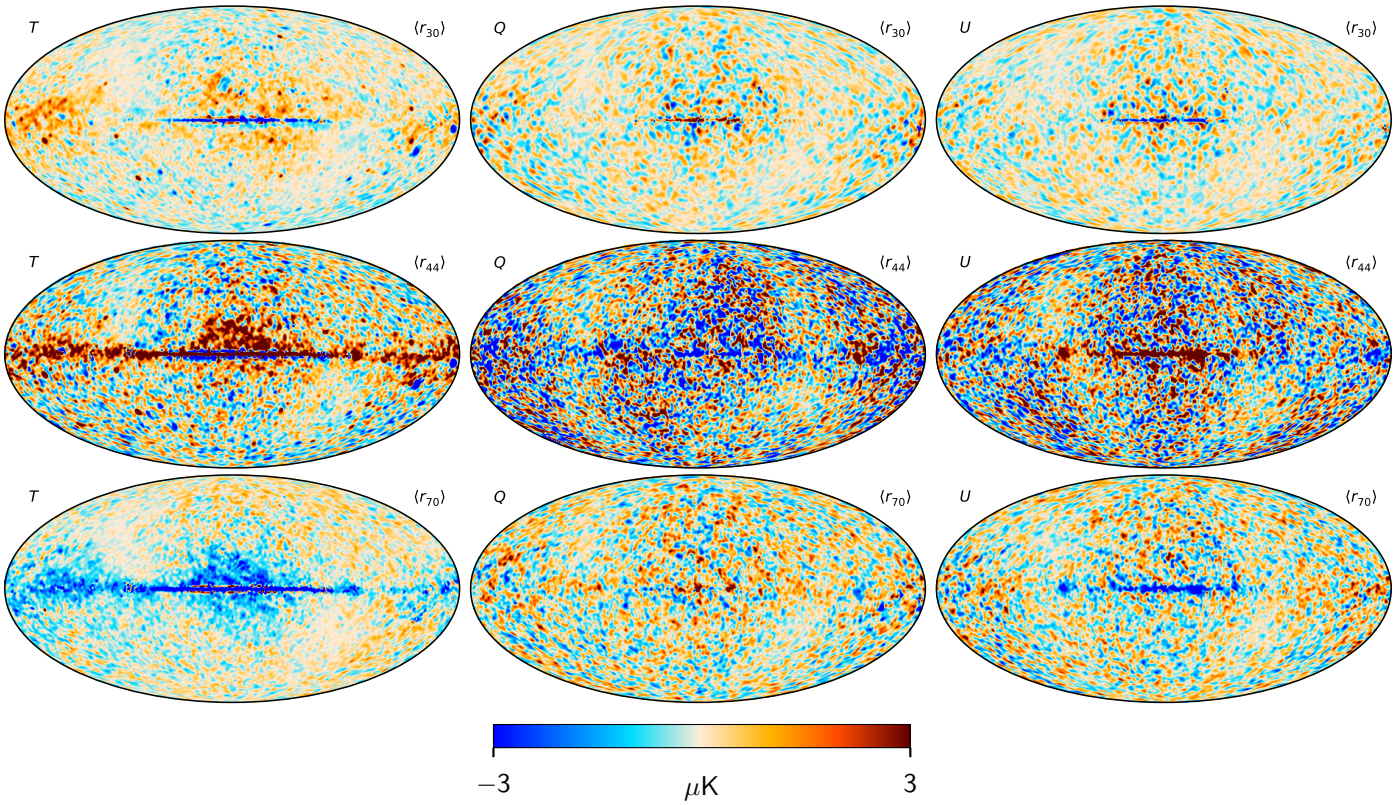

Fig. 26. Posterior mean total data-minus-model residual maps $\boldsymbol{d}_\nu - \boldsymbol{s}_\nu$ for BeyondPlanck LFI 30 (*top*), 44 (*middle*), and 70 GHz (*bottom*). All maps are smoothed to a common angular resolution of 2° FWHM.

to interpret. And in general we note two different trends. First, we note that the overall noise levels of the BeyondPlanck 30 and 70 GHz maps are slightly lower than in the *Planck* 2018 maps, but also higher than NPIPE, although the latter holds less true for 30 GHz than for 70 GHz. The BeyondPlanck 44 GHz map clearly has higher noise than either of the other two, as discussed above. Second, 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.

These differences are further elucidated in Fig. 25, which simply shows the power spectrum ratios between *Planck* 2018 and NPIPE, respectively, and BeyondPlanck. Again, we see that the three codes generally agree to well within 1 % in TT in the signal-dominated regimes of the spectra, but diverge in the noise-dominated regimes. Indeed, at the highest multipoles for 30 and 70 GHz NPIPE typically exhibits 10–15 % less white noise than BeyondPlanck, while BeyondPlanck exhibits 10 % less noise than *Planck* 2018. As discussed in Planck Collaboration Int. LVII (2020), NPIPE achieves lower noise than *Planck* 2018 primarily through three changes. First, NPIPE exploits the so-called repointing periods in the *Planck* scanning strategy, i.e., the periods during which the spin axis of the satellite moves, which account for about 8 % of the total data volume. Second, NPIPE smoothes the reference load LFI data prior to TOD differencing, as described in Sect. 5.1.3, and this results in a similar noise reduction. Third, NPIPE includes data from the so-called "ninth survey" at the end of the *Planck* mission, which accounts for about 3 % of the total data volume. In contrast, BeyondPlanck currently uses the repointing data, but neither smooths the reference load (essentially only because of limited time for implementation and analysis), nor includes the ninth survey. The reason for the latter is that we find that the TOD χ^2 statistics during this part of the mission show greater variation from PID to PID, suggesting less stability of the instrument. To be cautious, these data are therefore also omitted for now, similar to the horn 26 data, but may be included later.

These effects explain the different white noise levels. However, they do not (necessarily) explain the different slopes of the spectra, which instead indicate that the level of correlated noise is significantly lower in the BeyondPlanck maps as compared to the other two pipelines. The main reason for this is as follows: While *Planck* 2018 and NPIPE both destripe each frequency map independently, BeyondPlanck effectively performs joint correlated noise estimation using all available frequencies at once, as described in Sect. 8.3.2. This happens when conditioning on the current sky model during the correlated noise estimation phase, as opposed to applying the destriping projection operator Z independently to each channel. Thereby, the 30 GHz channel is in effect helped by the 70 GHz channel to separate true CMB fluctuations from its correlated noise, while the 70 GHz channel is helped by the 30 GHz channel to separate synchrotron and free-free emission from its correlated noise. And both 30 and 70 GHz are helped by both *WMAP* and HFI to separate thermal and spinning dust from correlated noise. Of course, this also means that the correlated noise component are correlated *between* frequency channels, and it is therefore imperative to actually use the Monte Carlo samples themselves to propagate uncertainties faithfully throughout the system.[20]

[20] It should of course be noted that the traditional pipelines also exhibit a correlated noise component between different frequencies, simply because they use the same foreground sky model to estimate bandpass corrections at different frequencies. This, however, is very difficult to

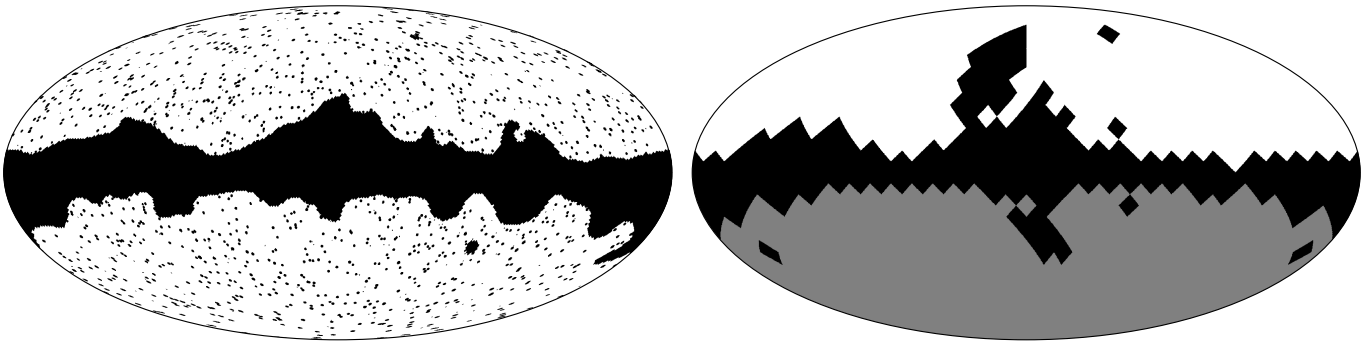

Fig. 27. Temperature (*left*) and polarization (*right*) confidence masks used for BeyondPlanck CMB analysis. In the right panel, the white-plus-gray region defines the BeyondPlanck "full-sky" mask with an accepted sky fraction of 74 %, while the white-only region defines the default BeyondPlanck mask with an accepted sky fraction of 36 %. The sky fraction of the temperature mask is $f_{\rm sky} = 0.69$.

9.3. Residual maps and masking

Having characterized the frequency maps, we will soon turn our attention to the astrophysical and cosmological products as defined by the sky model in Sect. 3.6. However, before doing so, we consider in this section the overall goodness-of-fit of the sky model, potentially unmodelled systematics, and masking required for CMB analysis.

The most detailed and informative summary of goodness-of-fit of a given sky model is provided through pixel-by-pixel residual maps, $\boldsymbol{r}_\nu = \boldsymbol{m}_\nu - \boldsymbol{s}_\nu$. These are shown in Fig. 26 for the LFI channels; for a discussion of all residuals included in the current analysis, we refer the interested reader to Andersen et al. (2020) (for temperature analysis) and Svalheim et al. (2020b) (for polarization analysis).

Starting with the temperature residual maps, we first of all see an imprint of the Galactic plane. This is expected, since the adopted foreground model is quite minimal, and the fits rely on strong priors from external observations; although the current data combination obviously has the statistical strength to identify model errors, it is for instance not strong enough to robustly distinguish between synchrotron spectral index variations from spatial variations in the free-free electron temperature. In particular, the morphology seen in the temperature residuals in Fig. 26 matches well that of Galactic dust emission, which suggests a mild deficiency in the AME model used for the current processing. Further data from experiments like C-BASS (Jew et al. 2019) and QUIJOTE (Génova-Santos et al. 2016) is obviously needed to refine these models, and as a result, a Galactic mask needs to be imposed before performing high-precision CMB analysis. However, we do note that the magnitude of these residuals is modest, with typical peak-to-peak amplitudes smaller than $3\,\mu$K at Galactic latitudes higher than a few degrees.

At high Galactic latitudes, we see clear point source residuals in the 30 GHz channel. We note that many radio sources are intrinsically variable, and such variability is not accounted for in the current model.

Turning our attention to the polarization maps, we note that both the 30 and 70 GHz residual maps appear quite clean, and are dominated by white noise, although some slight hints of large-scale systematics may be seen. We recall that similar structures were seen in the sample-to-sample difference maps shown in Fig. 21, and these uncertainties are thus at least partially accounted for in the model; whether they are fully accounted for, will only be clear after a full χ^2 analysis, which will be presented later.

However, while the 30 and 70 GHz channels appear clean, the 44 GHz channel exhibits some stronger artifacts, with a morphology that was already pointed out in Sect. 9.2; there are some PID ranges with incorrect large-scale corrections that result in stripes seen near the right edge of the map. Although the nature of these stripes is still not understood, we note that their presence did not become visually clear until after the completion of the current BeyondPlanck processing, and only with these results in hand do we now have the necessary tools to track them down. Given that limited progress on understanding the problems regarding the 44 GHz channel was made during the official *Planck* analysis period, we consider this identification an important, if preliminary, success of the current methodology, and a demonstration of usefulness for identifying and isolating low-level systematic effects.

These effects all correspond to unmodelled systematics in the current BeyondPlanck processing, and must be expected to contaminate the final high-level results at some level. At the same time, their actual impact may be small for any given specific application. In general, each higher-order analysis should therefore be accompanied with an appropriately defined goodness-of-fit assessment, typically involving χ^2 calculations that account for the full uncertainties as described by the Markov Chain ensemble.

However, some effects are more striking than others, and require special attention. The most prominent example of such is the Galactic plane, which obviously needs to be masked for precision CMB applications. For temperature, we construct a mask using both the residual map information shown here; combined with pair-wise difference maps evaluated between different algorithms (following Planck Collaboration IV 2020); and finally a point source mask. The full procedure is summarized in Colombo et al. (2020). For polarization, we adopt the product of the *Planck* 2018 common confidence mask and the 9-year *WMAP* polarization analysis mask as our baseline to remove the Galactic plane. However, in recognition of the residuals seen in the 44 GHz channel, covering much of the Southern Galactic hemisphere, we also define a special mask that leaves unmasked only the northern Galactic hemisphere. We will refer to these two masks as the "full-sky" and default masks, respectively, which have accepted sky fractions of 74 and 36 %. These are shown in Fig. 27. Both masks will be considered for CMB analysis, but as we will see, the full-sky mask results in large χ^2 excesses that effectively prohibits robust CMB inference.

both quantify or propagate, because of the substantial cost of including full component separation within a forward simulation pipeline.

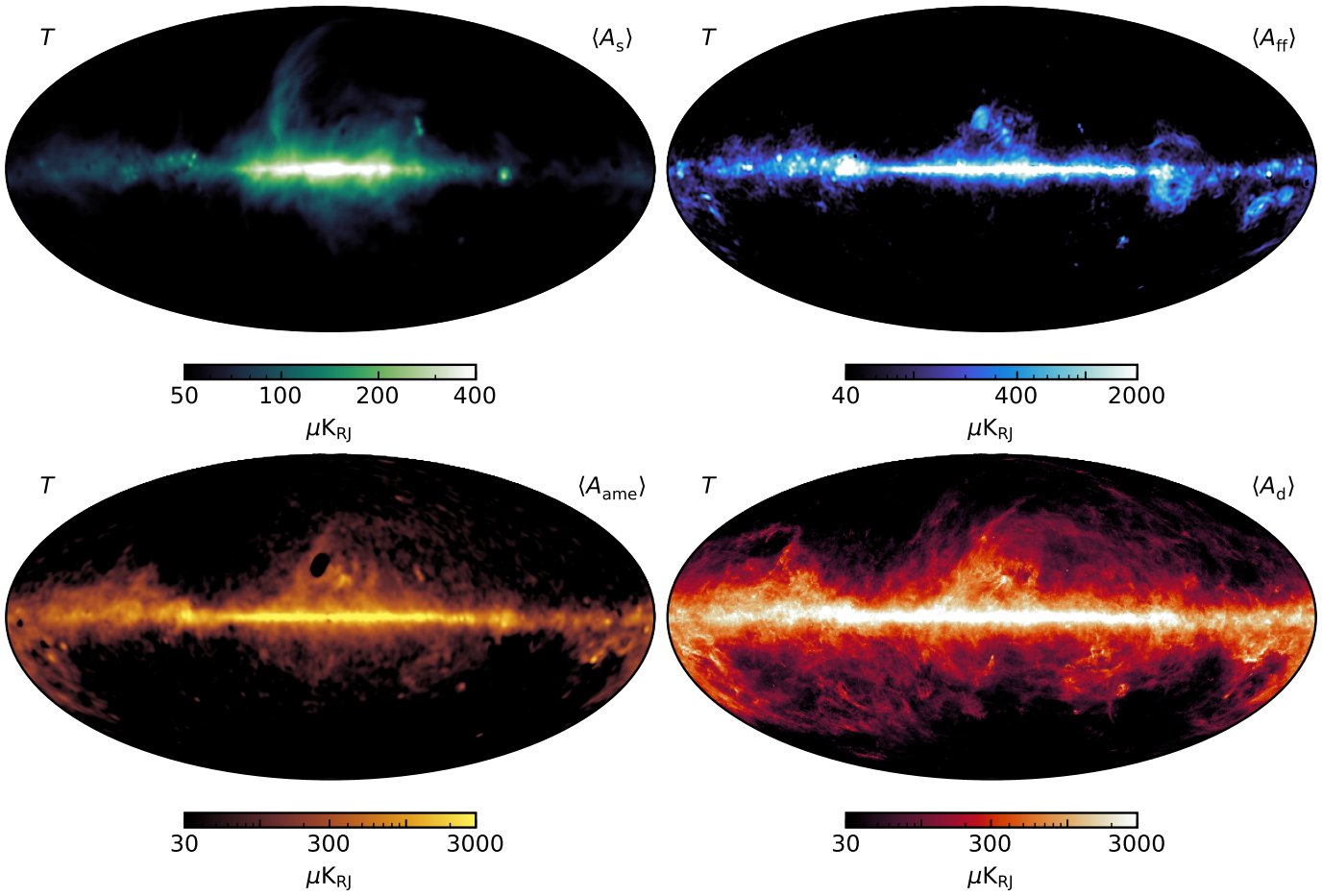

Fig. 28. Posterior mean maps of the aplitude of each of the four intensity foreground components included in the BeyondPlanck analysis. (*Top left:*) Synchrotron amplitude, evaluated at 30 GHz and smoothed to 2° FWHM resolution. (*Top right:*) Free-free amplitude, evaluated at 40 GHz and smoothed to 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. (2020) for further discussion of these maps.

9.4. 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 (2020), and Planck Collaboration Int. LVII (2020). Full joint analysis of all data sets is left for future work.

With that caveat in mind, Fig. 28 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. (2020), 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. 29 shows the posterior mean amplitude for polarized synchrotron emission, and Fig. 30 summarizes the posterior mean (left panel) and standard deviation (right panel) for 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. (2020b), 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 (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

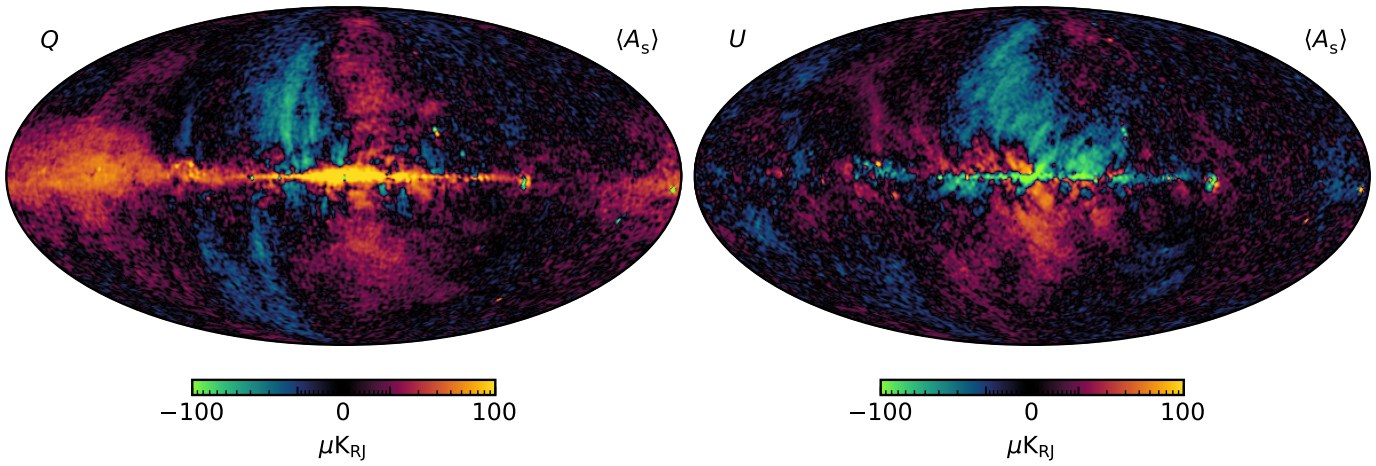

Fig. 29. Posterior mean maps of polarized synchrotron amplitude derived from BeyondPlanck, evaluated at 30 GHz and smoothed to an angular resolution of 1° FWHM. The two columns show Stokes Q and U parameters, respectively; see Svalheim et al. (2020b) for further discussion of these maps.

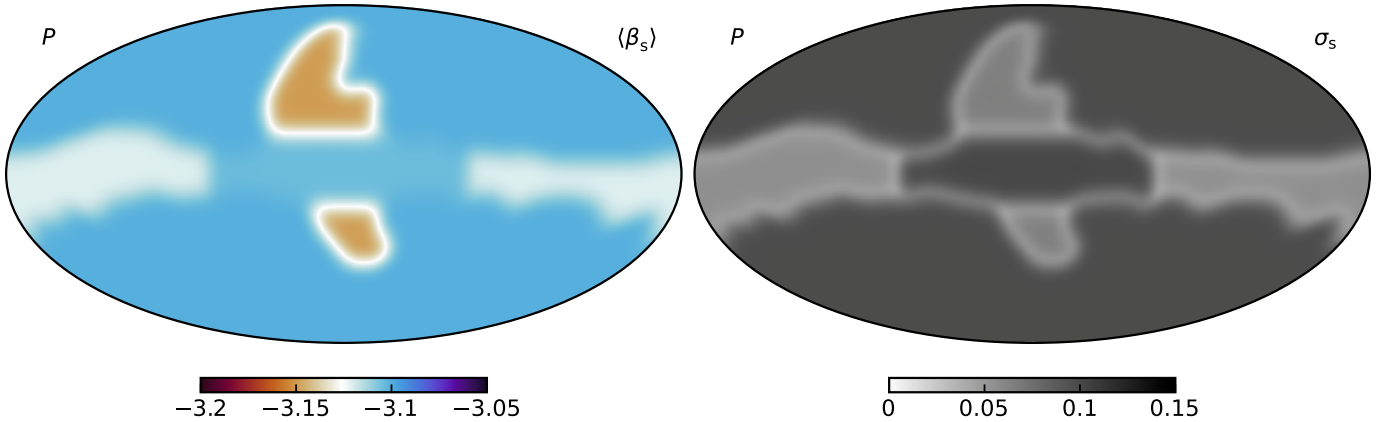

Fig. 30. Posterior mean and standard deviation maps for the spectral index of polarized synchrotron emission, β_s. Note that β_s is fitted in terms of four disjoint regions, each with a constant value but smoothed with a 10° FWHM Gaussian beam to avoid edge effects. The effect of this smoothing is seen in both the mean and standard deviation maps.

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 ($\beta_s^{\rm HGL} \approx -2.5$; see Svalheim et al. 2020b) 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 $\beta_s^{\rm GC} \approx -4$, 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. 30. Specifically, we find $\beta_s^{\rm GP} = -3.14 \pm 0.05$ and $\beta_s^{\rm NGS} = -3.19 \pm 0.05$, respectively (Svalheim et al. 2020b). On the one hand, we note that these values are broadly consistent with previous temperature-only constraints, such as those reported by Planck Collaboration X (2016), who found $\beta_s = -3.1$. On the other hand, our results show no compelling evidence for a significant spectral steepening from low to high Galactic latitudes, as for instance reported by (Kogut 2012) and Fuskeland et al. (2014, 2019). Rather, our results are qualitatively more similar to those derived using *WMAP* polarization data alone by Dunkley et al. (2009), who found a difference of only $\Delta\beta_s = 0.08$ between low and high Galactic latitudes. 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. (2020b).

9.5. CMB posteriors

Finally, we arrive at the main scientific target application of the paper, the CMB posteriors. We start with a discussion of the CMB dipole in Sect. 9.5.1, before presenting the CMB fluctuation maps in Sect. 9.5.2. The BeyondPlanck CMB low-ℓ power spectrum and likelihood are discussed in Sect. 9.5.3, and the high-ℓ power spectrum and likelihood are discussed in Sect. 9.5.4. Finally, cosmological parameters are presented in Sect. 9.5.5.

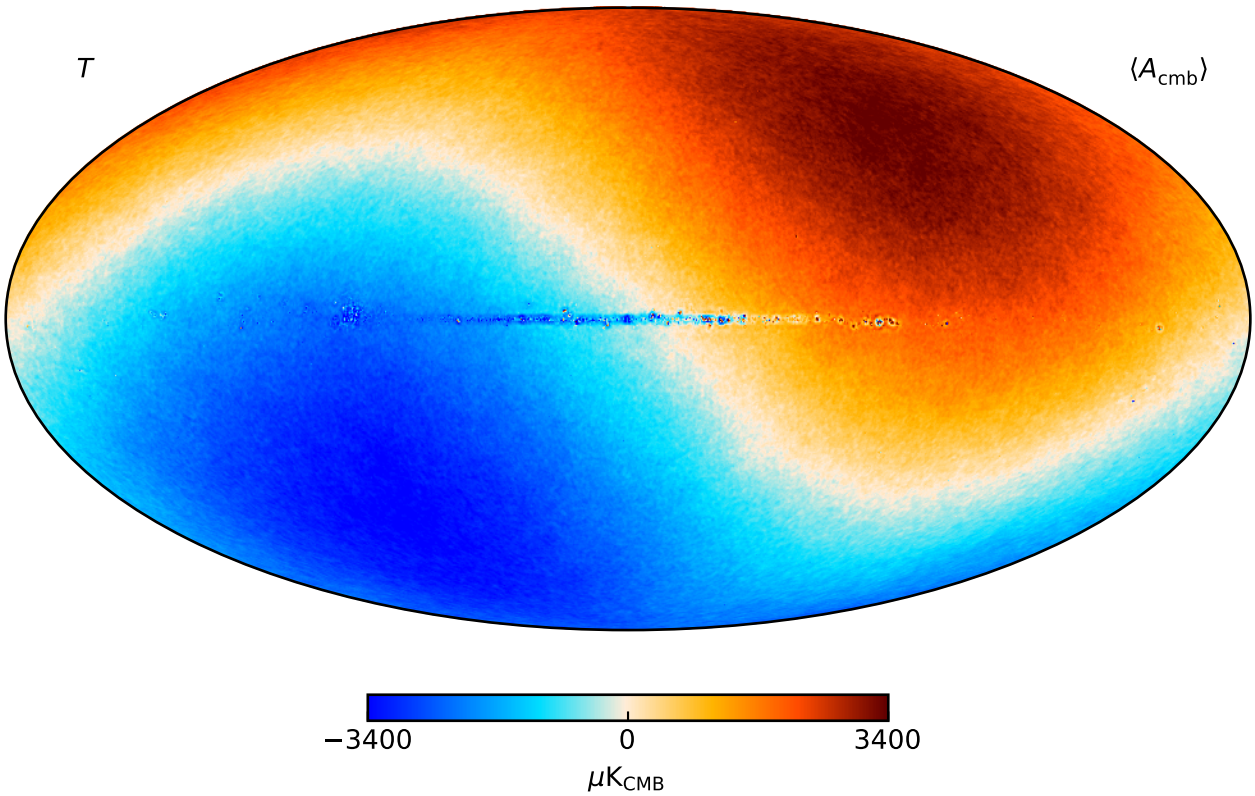

Fig. 31. Posterior mean CMB BeyondPlanck temperature map, smoothed to an angular resolution of 14′ FWHM.

Table 3. Comparison of Solar dipole measurements from *COBE*, *WMAP*, and *Planck*.

		Galactic coordinates		
Experiment	Amplitude [μK_{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)
NPIPE [a,c]	3366.6 ± 2.6	263.986 ± 0.035	48.247 ± 0.023	Planck Collaboration Int. LVII (2020)
BeyondPlanck [e]	**3359.5 ± 1.9**	**263.97 ± 0.09**	**48.30 ± 0.03**	Section 9.5

[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. However, they do not account for prior or model selection uncertainties; see Sect. 9.5.1 for a discussion of these priors.

9.5.1. The CMB solar dipole

In the BeyondPlanck framework, the CMB dipole is in principle 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 $\boldsymbol{s}_{\rm CMB}$. No special-purpose component separation algorithms are applied to derive the CMB dipole, nor does any individual frequency play a more important role than others, except for as dictated by the relative level of instrumental noise in each channel.

However, as discussed by Ihle et al. (2020), Gjerløw et al. (2020), and Suur-Uski et al. (2020), this apparent algorithmic simplicity does not imply that robust CMB dipole estimation is by any means *easy* in the BeyondPlanck procedure. Indeed, the CMB dipole is quite possibly the single most difficult parameter to estimate in the entire model, simply because it both affects, and relies on, a wide range of other parameters in the model. Some of the most important degeneracies are the following:

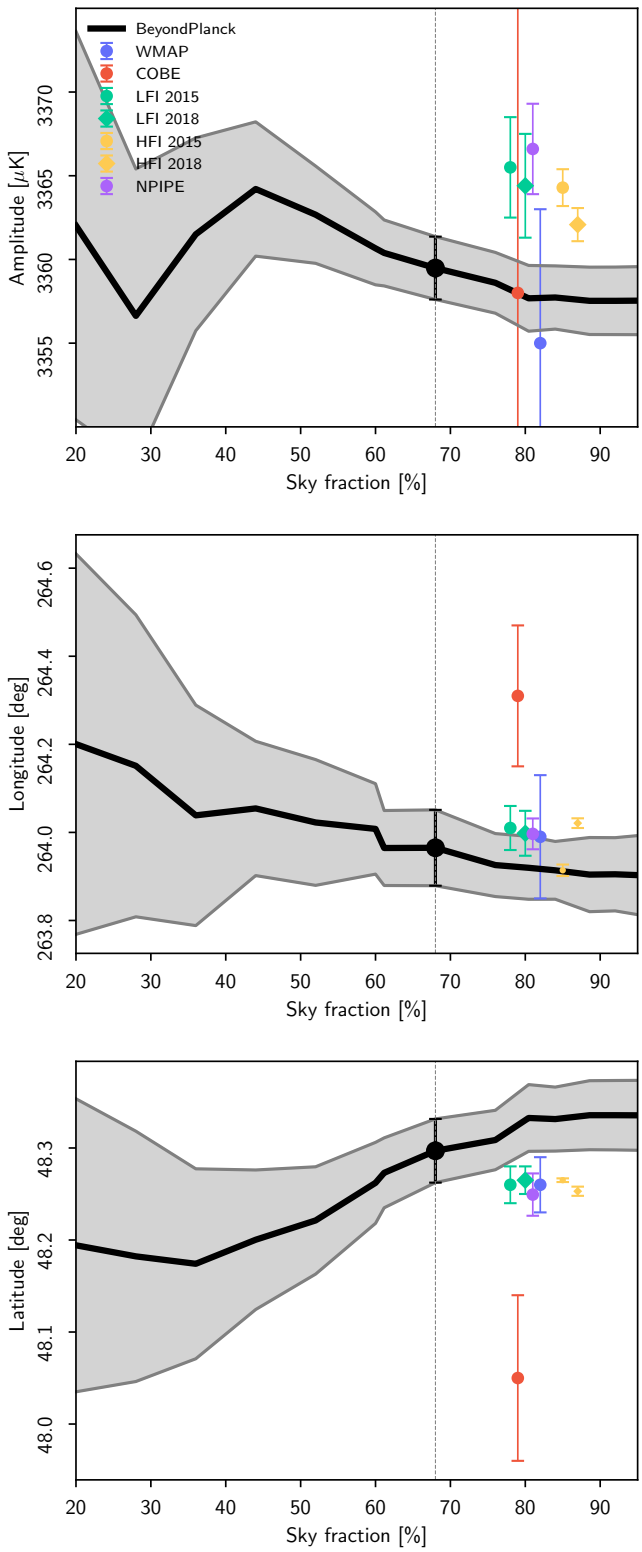

Fig. 32. CMB dipole parameters as a function of sky fraction.Gray bands indicate 68 % posterior confidence regions.

1. *Absolute calibration* – g_0 is the single most important parameter for a robust solar dipole determination, as it directly scales the amplitude of the CMB map. This parameter is itself constrained from the orbital dipole, which is both relatively weak in terms of absolute amplitude, and for significant parts of the mission it is aligned with, and thereby obscured by, the Galactic plane.
2. *Astrophysical foregrounds* – foregrounds have in general both non-zero dipole and higher-order moments with unknown parameters, and must be estimated jointly with the CMB dipole. However, considering that the current data set includes five astrophysical components, each with a free value in each pixel, and we only have eight different frequency maps, the full system is quite poorly constrained; it is therefore possible to add a significant dipole to the CMB map, and subsequently subtract appropriately scaled dipoles from each of the foreground maps, with only a minimal penalty in terms of the overall χ^2. In practice, we see particularly strong degeneracies between the CMB, AME and free-free components, when exploring the full system without priors, leading to massive and obviously nonphysical marginal uncertainties.
3. *Correlated noise* – $n_{\rm corr}$ is only weakly constrained through its PSD parameters, and is therefore able to account for a wide range of modelling errors, including calibration errors. In particular, incorrectly estimated gains will leave a spurious dipole-like residual in the time-ordered data. Since this spurious residual clearly is detector-dependent, it will typically be interpreted by the algorithm as correlated noise, and thereby excite a dipolar structure in $\boldsymbol{n}_{\rm corr}$.
4. *Large-scale CMB quadrupole, foreground and bandpass corrections* – while the CMB polarization quadrupole is predicted by current ΛCDM models to have a very small quadrupole, with a variance of typically less than 0.05 μK^2, there is nothing in the current parametric model that explicitly enforces this. This particular mode therefore opens up a particularly problematic degeneracy for *Planck* through coupling with the gain and bandpass shift as follows: An error in the absolute gain leads to an apparently wrong orbital dipole. However, this can be countered by adding a polarized CMB quadrupole, which has the same SED and nearly the same spin harmonics as the orbital dipole, due to the *Planck* scanning strategy that observes along nearly perfect great circles.[21] Errors in the total polarized sky signal as observed at each frequency can finally be countered by adjusting the combination of relative gains, polarized foreground signals, and bandpass corrections between radiometers, leaving the total χ^2 nearly unchanged.

During the initial test phase of the BeyondPlanck pipeline, the Markov chain was allowed to explore these degeneracies freely, in order to understand their nature, and these runs resulted in a full marginal uncertainty on the dipole amplitude of more than 40 μK, as compared to 3 μK reported by *Planck* LFI for the 70 GHz channel alone (Planck Collaboration II 2020), or 1 μK as reported by HFI (Planck Collaboration III 2020). Although this value by itself could be considered acceptable, given the limited cosmological importance of the CMB dipole, it was also strikingly obvious that all component maps were fully compromised by the poorly constrained calibration, ultimately leading to obviously nonphysical astrophysical component maps.

To break these degeneracies, we instead impose the following effective priors in the analysis pipeline, as already discussed in Sect. 8.3 and by Gjerløw et al. (2020):

1. We estimate the absolute calibration, g_0, using only the orbital dipole as a calibrator; see Sect. 8.3.1. This significantly

[21] This particular degeneracy does not exist for *WMAP*, because of its more complex scanning strategy.

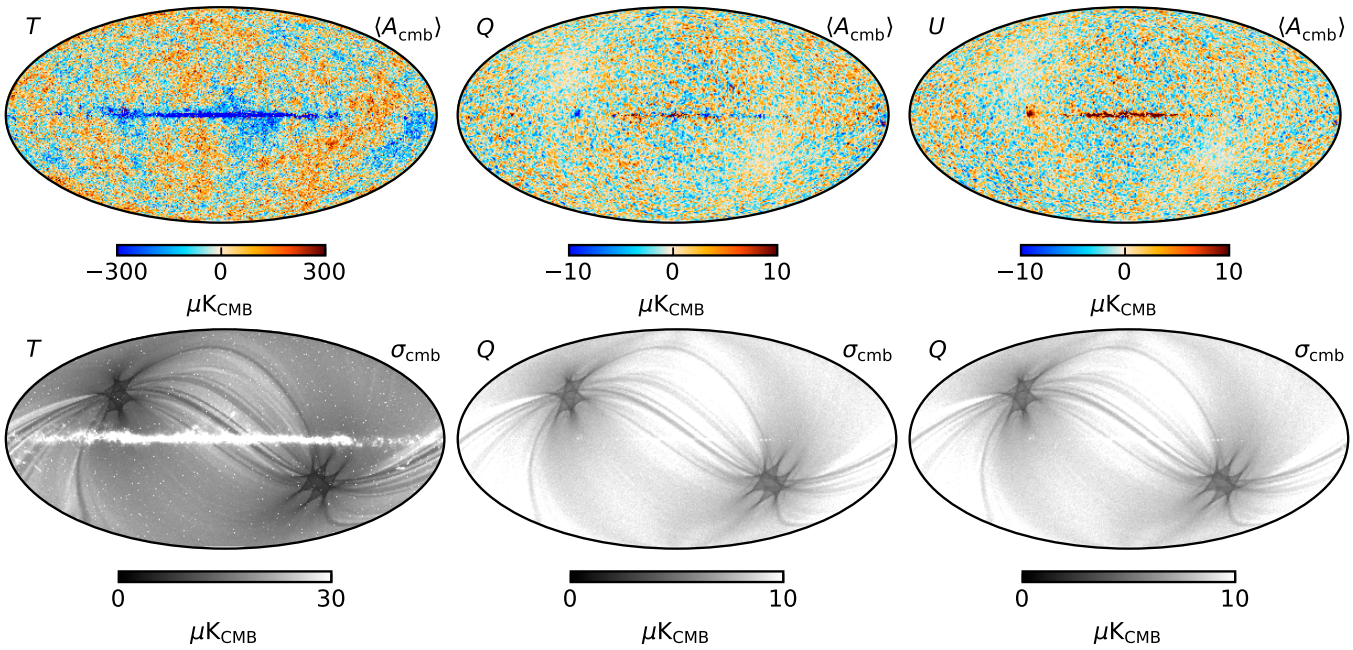

Fig. 33. BeyondPlanck posterior mean (*top row*) and standard deviation (*bottom row*) CMB fluctuation maps. Columns show, from left to right, temperature and Stokes Q and U parameters, respectively. The temperature maps are smoothed to 14′ FWHM resolution, while the polarization maps are smoothed to 1° FWHM.

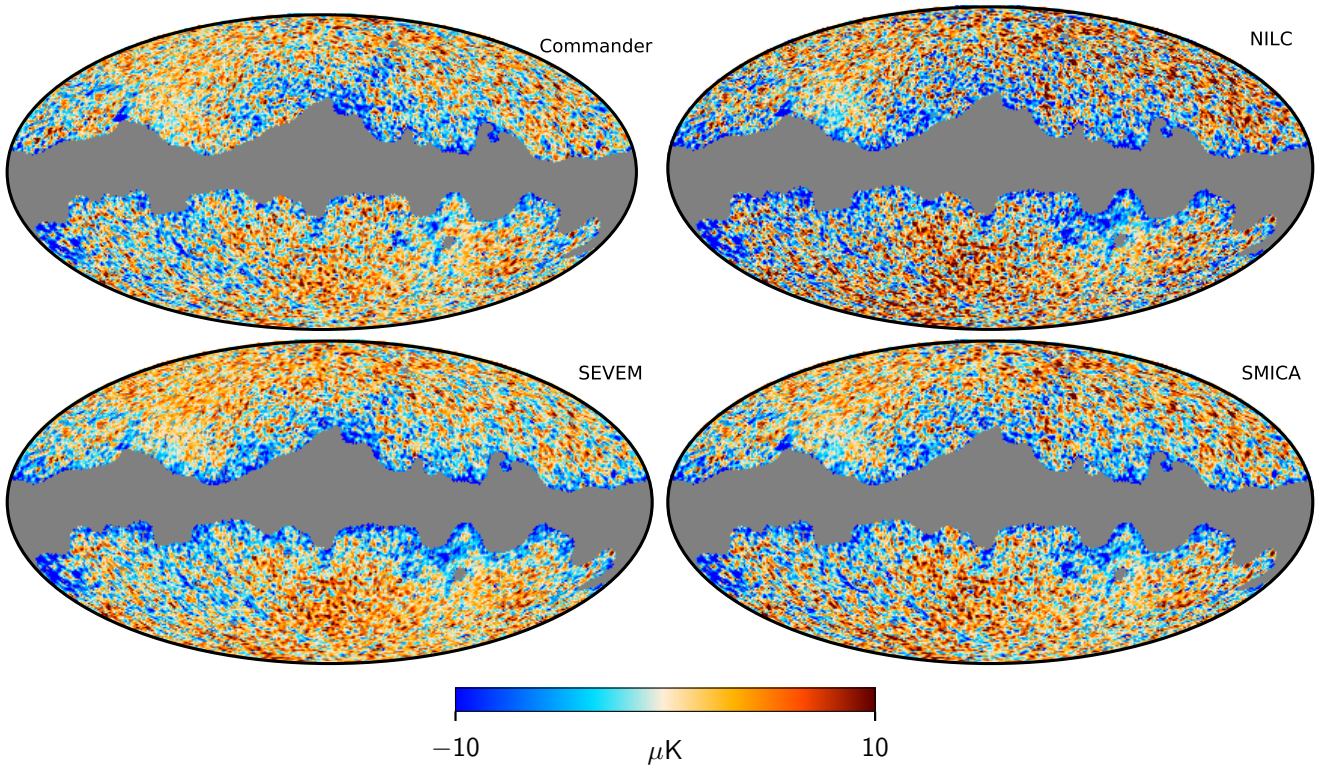

Fig. 34. Difference maps between the BeyondPlanck CMB temperature map and those derived from the full *Planck* 2018 data set (Planck Collaboration IV 2020). From left to right and from top to bottom, the various panels show differences with respect to `Commander`, `NILC`, `SEVEM`, and `SMICA`. All maps are smoothed to a common angular resolution of 1° FWHM.

reduces the degeneracy between the foregrounds and the absolute calibration.

2. We enforce active spatial priors on the free-free and AME amplitude maps, as discussed by Andersen et al. (2020). Specifically, we use (an appropriately scaled version of) the *Planck* 857 GHz map as a spatial prior for AME (Planck Collaboration Int. LVII 2020), and the *Planck* 2015 free-free map (Planck Collaboration X 2016) as a prior for free-free emission. In effect, we thus incorporate external information

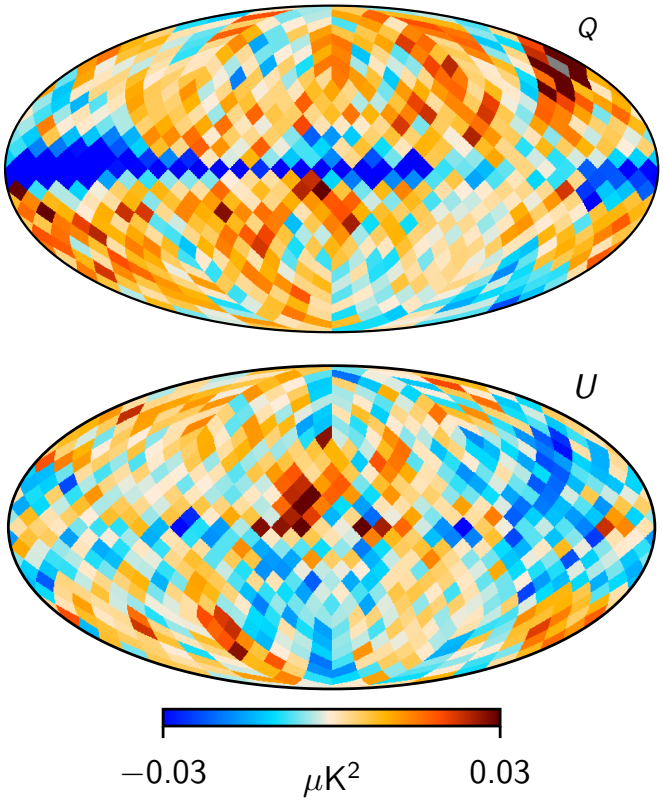

Fig. 35. Single column of the low-resolution CMB noise covariance matrix, constructed as the mean outer-product of Monte Carlo samples. The column corresponds to Stokes Q pixel number 100, which is marked in gray, and located in the top right quadrant near the 'Q' label. Note that non-zero correlations are detected nearly across the full sky in both Stokes parameters.

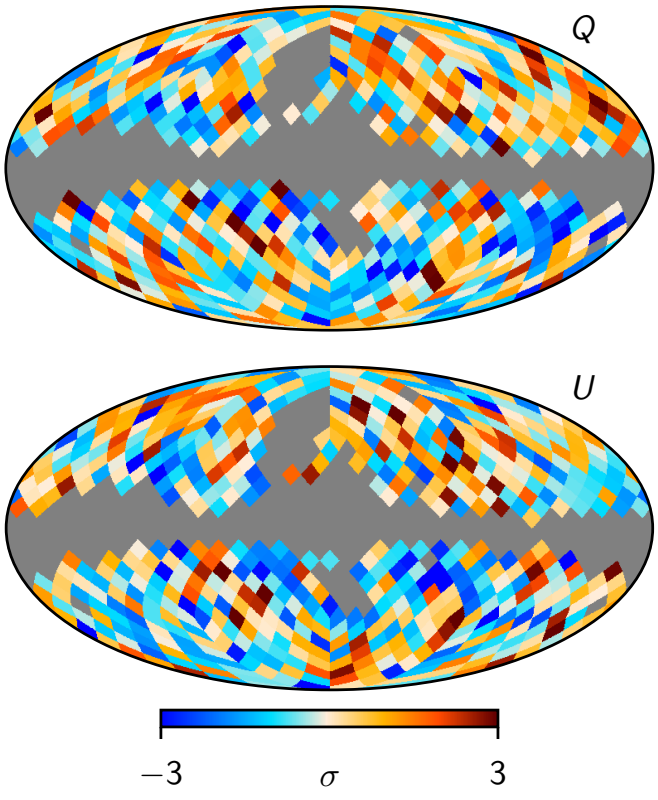

Fig. 36. BeyondPlanck low-resolution and "whitened" CMB polarization map, as defined by $\mathsf{N}_{\rm CMB}^{-1/2}\boldsymbol{s}_{\rm CMB}$ at a `HEALPix` resolution of $N_{\rm side} = 8$ and masked with the BeyondPlanck "full-sky" mask. Top and bottom panel shows Stokes Q and U parameters, respectively, and the color scales span $\pm 3\sigma$.

from *Planck* HFI to help stabilize the model, but without introducing data that directly constrains the CMB component.

3. We set the CMB polarization quadrupole to zero during the relative gain calibration phase. This is less strict than the LFI DPC and `NPIPE` procedures, both of which set the entire CMB polarization signal to zero in this step (Planck Collaboration II 2020; Planck Collaboration Int. LVII 2020). As discussed by Gjerløw et al. (2020), the strong coupling between the gain and the CMB polarization is associated with the quadrupole alone, and not higher-order modes. As a formal justification of this prior, one may argue that this corresponds to imposing a ΛCDM prior to that particular mode. In principle, this does bias the resulting posterior distribution; however, as long as ΛCDM is a reasonable approximation to the true sky, then the effect is negligibly small, because of the very small predicted amplitude of the CMB quadrupole. Furthermore, the true CMB polarization quadrupole is likely to be even smaller than that predicted by ΛCDM, since the CMB temperature quadrupole is already known be low at the 1 % level, and these two are correlated through the TE power spectrum. For additional verification, we will in the following estimate cosmological parameters both with and without the EE $\ell = 2$ mode included, and we find no significant difference.

Figure 31 shows the marginal CMB temperature fluctuation posterior mean map as derived in BeyondPlanck, given both 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.

Because of the small but non-negligible Galactic plane, we must impose an analysis mask before estimating final dipole parameters. For this purpose, we use the Wiener filter estimator described by Thommesen et al. (2020), which in-paints the Galactic mask with a constrained realization prior to parameter estimation; this is necessary in order to account for, and marginalize over, coupling to higher-order CMB fluctuations. This method was also adopted for the dipole estimates presented in Planck Collaboration Int. LVII (2020), although we introduce one significant difference to that analysis: In the current analysis we estimate the magnitude of systematic uncertainties directly from the BeyondPlanck Gibbs samples, as opposed to putting in it by hand. Specifically, instead of producing 9000 constrained realizations from a single maximum likelihood map, as done in Thommesen et al. (2020) and Planck Collaboration Int. LVII (2020), we now produce 100 constrained realizations from each of the 900 available Gibbs samples. Since each of these realizations have different gain, correlated noise, and foreground residuals, the full ensemble therefore now accounts for these uncertainties automatically. The only additional term we put by hand into to the error budget is a contribution of 0.7 μK from the CMB monopole uncertainty itself (Fixsen 2009).

Using this methodology, we estimate the CMB dipole parameters over a series of Galactic masks, ranging in sky fraction from 20 to 95 %. The results from these calculations are shown in Fig. 32. Overall, we see that the posterior distributions are quite

stable with respect to sky fraction. Furthermore, we note that the uncertainties do not decrease after $f_{\rm sky} \approx 0.7$, as they would if the full error budget could be described in terms of white noise and sky fraction. Rather, the weight of the additional sky coverage is effectively reduced when marginalizing over the various systematic contributions, as desired. We adopt a sky fraction of $f_{\rm sky} = 0.77$ to define our final dipole estimates, corresponding to the sky fraction used for the main CMB temperature analysis. The resulting values are plotted as black points in Fig. 32, and tabulated together with previous estimates in Table 3.

Several points are worth noting regarding these results. First, we see that the reported best-fit BeyondPlanck dipole amplitude is 3359.5 ± 1.9 μK, which is lower than the latest LFI 2018 estimate of 3364.4 ± 3.1 μK, which in turn is lower than the NPIPE estimate of 3366.6 ± 2.6 μK. In fact, it is even lower than the latest HFI estimate of 3362.08 ± 0.99 μK. On the other hand, it is higher than the *WMAP* estimate of 3355 ± 8 μK. Algorithmically speaking, this makes intuitively sense, considering that *WMAP* observations actually are used actively in the BeyondPlanck analysis, while it is not in the *Planck*-only analyses. At the same time, we note that the BeyondPlanck estimate is statistically consistent with any one of these estimates, as measured in terms of their error bars; the biggest outlier is NPIPE, which still only represents a 2.7σ variation.

Regarding the directional parameters, two observations are worth pointing out. First, we see that the BeyondPlanck uncertainties are substantially larger than any of the previous *Planck*-dominated results. Here it is worth recalling again that no additional systematic error contributions are added by hand to the BeyondPlanck directional uncertainties, and the reported values are thus the direct result of degeneracies within the model itself. Perhaps the biggest algorithmic difference in this respect is the fact that the current algorithm explicitly marginalizes over the full foreground model, while most other approaches condition on external constraints. The second observation is that the BeyondPlanck latitude is higher than any of the previous results. The statistical significance of this difference is modest, only about 1–$2\,\sigma$, but compared with the remarkable internal agreement between *Planck* and *WMAP*, it is still noteworthy. In this respect, we recall that we are currently using the *Planck* 2015 free-free map as an informative prior in the current processing, and CMB and and free-free emission are known to be strongly correlated for the current data set; see Andersen et al. (2020). Performing a joint analysis of LFI, HFI, and *WMAP* without an external free-free prior might be informative regarding this point.

9.5.2. CMB maps

Next, we consider the CMB fluctuation maps, as shown in Fig. 33. The top row shows the posterior mean (after subtracting the dipole from the temperature component), while the bottom row shows the posterior standard deviation. We see a narrow Galactic plane imprint in both temperature and polarization, and both in the mean and standard deviation maps. Fortunately, at least at a visual level, the obviously offending features are well covered by the analysis masks shown in Fig. 27.

Figure 34 shows difference maps between the BeyondPlanck temperature map and the four official *Planck* 2018 foreground-cleaned CMB maps (Commander, NILC, SEVEM, and SMICA; see Planck Collaboration IV 2020). A best-fit monopole and dipole have been removed outside the *Planck* 2018 common confidence mask in each case, and all maps are smoothed to a common resolution of 1° FWHM. Several notable feature may be seen in these difference maps. First of all, we notice a significant amount of random noise at high Galactic latitudes, which is due to the fact that the BeyondPlanck CMB map does not include HFI data.

At low Galactic latitudes, the main difference is a blue Galactic plane. From the top left panel in Fig. 33, we see that this is indeed coming from the BeyondPlanck map, indicating that the current BeyondPlanck signal model over-subtracts foregrounds in the Galactic plane. At the same time, by comparing this figure to Fig. 7 in Planck Collaboration IV (2020) (which shows pairwise differences between each of the four *Planck* 2018 CMB maps), we see that the absolute internal differences within the *Planck* 2018 maps are of comparable order-of-magnitude as the difference between BeyondPlanck and *Planck* 2018, although generally covering a slightly smaller sky fraction.

9.5.3. Low-ℓ CMB likelihood

Our main scientific goal with the BeyondPlanck pipeline is to constrain the CMB power spectrum and cosmological parameters through end-to-end analysis. Both of these operations are most conveniently facilitated through the CMB power spectrum likelihood, $\mathcal{L}(C_\ell)$. As discussed in Sect. 8.3.8, it is most convenient to split this function into two components, one low-dimensional and low-ℓ component that employs dense matrix operations, and one high-dimensional and high-ℓ component that employs faster operations; a similar split is used by both *WMAP* (Hinshaw et al. 2013) and *Planck* (Planck Collaboration V 2020).

As described in Sect. 8.3.8, we employ a brute-force Gaussian pixel-based likelihood estimator at low-ℓ, as has been a standard procedure since *COBE*-DMR,

$$P(C_\ell \mid \hat{\boldsymbol{s}}_{\rm CMB}) \propto \frac{\mathrm{e}^{-\frac{1}{2}\hat{\boldsymbol{s}}^t_{\rm CMB}(\mathsf{S}(C_\ell)+\mathsf{N})^{-1}\hat{\boldsymbol{s}}_{\rm CMB}}}{\sqrt{|\mathsf{S}(C_\ell)+\mathsf{N}|}}. \tag{136}$$

The only noteworthy variation is that we employ an optimized basis set that reduce the number of basis vectors in $\hat{\boldsymbol{s}}_{\rm CMB}$ from $N_{\rm pix}$ to the number of modes with a relative signal-to-noise ratio larger than 10^{-6} over $2 \leq \ell \leq 8$; see Gjerløw et al. (2015) for details. We choose an upper truncation of $\ell \leq 8$ as a compromise between minimizing the dimensionality of the noise covariance matrix and retaining as much constraining power with respect to τ as possible. As shown by Colombo et al. (2020), all main results shown in the following are robust against variations in this cut-off.

We construct the inputs to the low-ℓ likelihood as follows:

1. For each of the 900 full-resolution Gibbs samples discussed in the introduction to this section, we draw 50 new samples from $P(\boldsymbol{s}^{\ell\leq 64}_{\rm CMB} \mid \boldsymbol{d}, \omega \setminus \boldsymbol{s}^{\ell\leq 64}_{\rm CMB})$, resulting in a total of 45 000 low-ℓ CMB map samples. Note that each of these samples represents a complete sample from the full posterior; the only difference is that this particular sub-volume of ω is sampled 50 times more densely than for other parameters.
2. Next, each sample is individually downgraded to $N_{\rm side} = 8$. In temperature, we additionally smooth to 20° FWHM before downgrading to avoid sub-pixel effects from high signal-to-noise high-ℓ fluctuations, while for polarization we simply boxcar average over each $N_{\rm side} = 8$ pixel.
3. We then compute the posterior mean CMB map, $\hat{\boldsymbol{s}}_{\rm CMB}$ and covariance matrix $\mathsf{N}_{\rm CMB}$ simply by averaging over samples, as given by Eq. (122). Additionally, we add $1\,\mu$K of random Gaussian regularization noise to the temperature component, both in $\hat{\boldsymbol{s}}_{\rm CMB}$ and $\mathsf{N}_{\rm CMB}$, to make the noise covariance matrix non-singular, and we set all off-diagonal temperature-

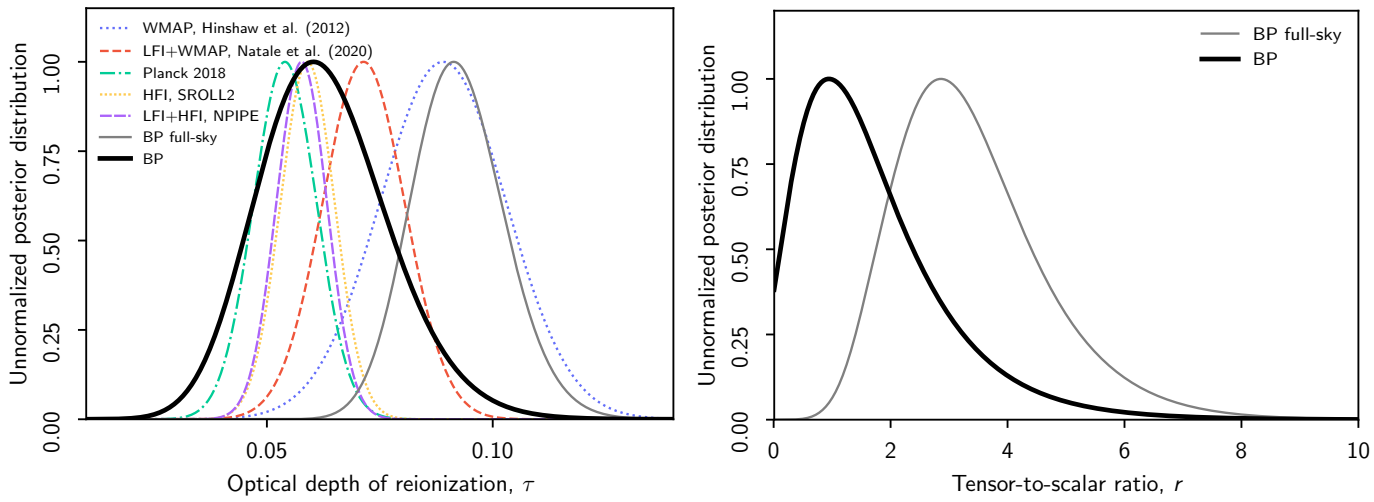

Fig. 37. (*Left panel*:) Comparison of (un-normalized) marginal posterior distributions of the reionization optical depth from *Planck* 2018 (green; Planck Collaboration VI 2020), 9-year *WMAP* (red; Hinshaw et al. 2013), *WMAP Ka–V* and LFI 70 GHz (orange; Natale et al. 2020); and BeyondPlanck *EE*-only using multipoles $\ell = 2$–8 (gray/black; Paradiso et al. 2020). The thin gray line shows BeyondPlanck constraints derived using the full-sky mask ($f_{\rm sky} = 0.74$), while the solid black line shows constraints using the northern mask only ($f_{\rm sky} = 0.36$). (*Right panel*:) Corresponding marginal BeyondPlanck tensor-to-scalar ratio posteriors derived using *BB* multipoles between $\ell = 2$–8, including either the Northern Galactic hemisphere (*black*) or the full sky (*gray*).

Table 4. Summary of cosmological parameters dominated by large-scale polarization, and goodness-of-fit statistics. Note that the BeyondPlanck full-sky case has an unacceptable goodness-of-fit, and should not be used for cosmological analysis. The main science result from the current analysis are summarized in the top two entries, and are evaluated with a small polarization sky fraction. For completeness, the third row shows results evaluated with nearly full-sky data, but these are strongly contamined by systematic errors, as indicated by the high χ^2 value.

Analysis Name	Data Sets	$f_{\rm sky}^{\rm pol}$	τ	$r_{95\,\%}^{BB}$	χ^2 PTE	Reference
BeyondPlanck, $\ell = 2$–8	LFI, *WMAP Ka–V*	0.36	$0.060^{+0.015}_{-0.013}$	< 4.3	0.16	Paradiso et al. (2020)
BeyondPlanck, $\ell = 3$–8	LFI, *WMAP Ka–V*	0.36	$0.061^{+0.015}_{-0.014}$	< 5.4	0.16	Paradiso et al. (2020)
BeyondPlanck, $\ell = 2$–8, full-sky . .	LFI, *WMAP Ka–V*	0.74	$0.091^{+0.010}_{-0.098}$	$2.9^{+1.3}_{-1.0}$	$5 \cdot 10^{-4}$	Paradiso et al. (2020)
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	0.071 ± 0.009			Natale et al. (2020)
Planck 2018	HFI 100×143	0.50	0.051 ± 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. (2020)

polarization elements in $\mathsf{N}_{\rm CMB}$ to zero, noting that temperature is massively signal-dominated.

4. We apply the Galactic analysis mask (as defined in Fig. 27) by removing the relevant rows and columns from both $\hat{\mathsf{s}}_{\rm CMB}$ and $\mathsf{N}_{\rm CMB}$.
5. Finally, we eigenvalue decompose the $\mathsf{S}(C_\ell) + \mathsf{N}_{\rm CMB}$ matrix, where C_ℓ is defined by the best-fit *Planck* 2018 ΛCDM spectrum for $2 \le \ell \le 8$, and define the projection operator P as the set of all columns corresponding to an eigenvalue, ϵ_i, larger than $10^{-6}\max(\epsilon_i)$. This multipole range includes all values for which the LFI data has a significant signal-to-noise ratio, while the eigenvalue threshold is chosen to avoid issues with numerical precision.

To build useful intuition regarding $\hat{\mathsf{s}}_{\rm CMB}$ and $\mathsf{N}_{\rm CMB}$, we first show an single column of $\mathsf{N}_{\rm CMB}$ prior to eigen-mode compression in Fig. 35; this column corresponds to pixel 100 in the Stokes Q map, which is marked by gray in the figure (located near the Q label). Several interesting features may be seen here. First, we see that there is a broad stripe (or region) extending from the pixel in question in both Q and U. This stripe corresponds to correlated noise and gain fluctuations modulated by the *Planck* scanning strategy. Second, we see that the Galactic plane is negatively correlated with the pixel in question in Q, and positively correlated in U; these correlations result from variations in both gain and bandpass, coupled to the foreground components. Third, we also see that there are significant correlations at further distances in the form of broad extended regions; these are at least partially due to the *WMAP* horn separation of 141°, which in effect couples pixels across the entire sky. The complexity of this map illustrates a significant advantage of the Monte Carlo sampling approach: Constructing this full matrix analytically would be extremely difficult, which is why it has never been done in the CMB literature until now. To date, only correlated noise, absolute calibration, and linear template corrections have been accounted for in this matrix; see, e.g., Hinshaw et al. (2013) and Planck Collaboration V (2020).

Figure 36 shows the "whitened" CMB polarization map, as defined by $\mathsf{N}_{\rm CMB}^{-1/2}\boldsymbol{s}_{\rm CMB}$. This essentially measures the local signal-to-noise ratio in each pixel (taking into account the full covariance matrix structure), and should be consistent with Gaussian random noise with vanishing mean and unit standard deviation for a strongly noise-dominated map. Overall, this map

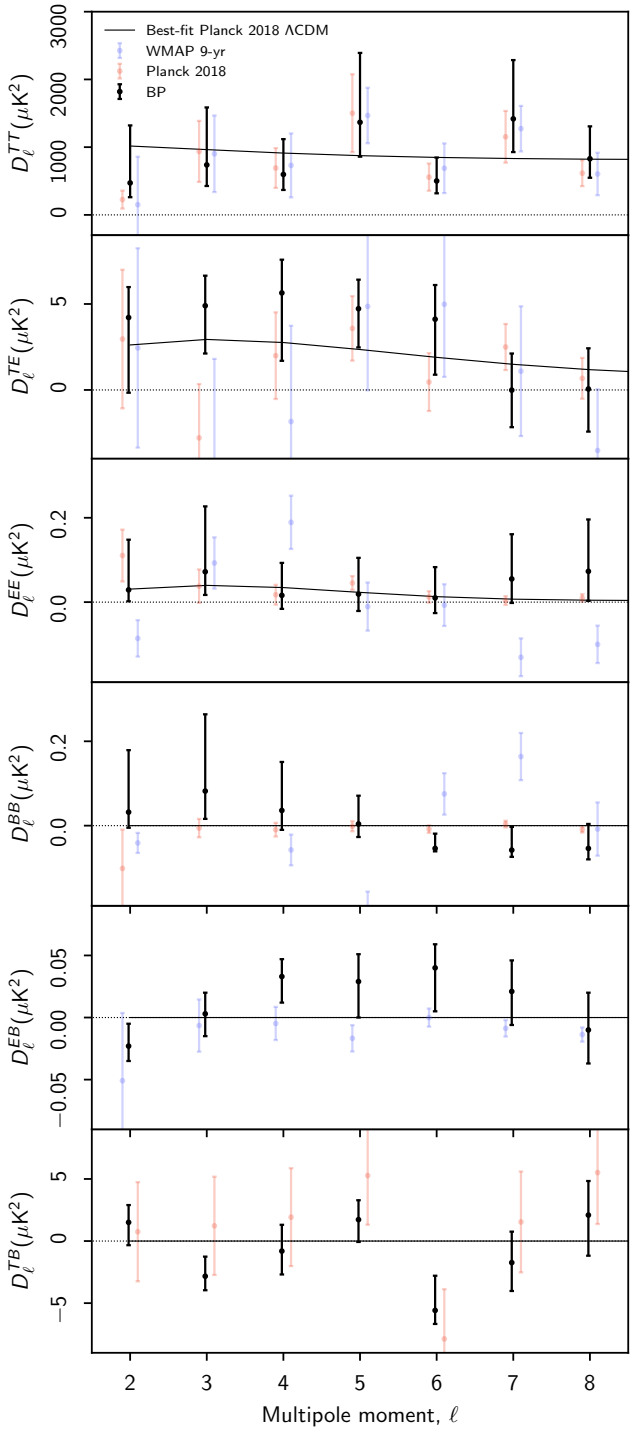

Fig. 38. Comparison between low-ℓ angular CMB power spectra, as derived by the *Planck* collaboration using both LFI and HFI data (blue points; Planck Collaboration V 2020); by the *WMAP* team using just *WMAP* data (red points; Hinshaw et al. 2013); and by BeyondPlanck using both LFI and *WMAP* data (black points; Colombo et al. 2020). Thin black lines indicate the *Planck* 2018 best-fit ΛCDM spectrum (Planck Collaboration VI 2020). The BeyondPlanck data points are evaluated by conditionally slicing the posterior distribution ℓ-by-ℓ with respect to the best-fit ΛCDM model, by holding all other multipoles fixed at the reference spectrum while mapping out $P(C_\ell|\boldsymbol{d})$, to visualize the posterior structure around the peak.

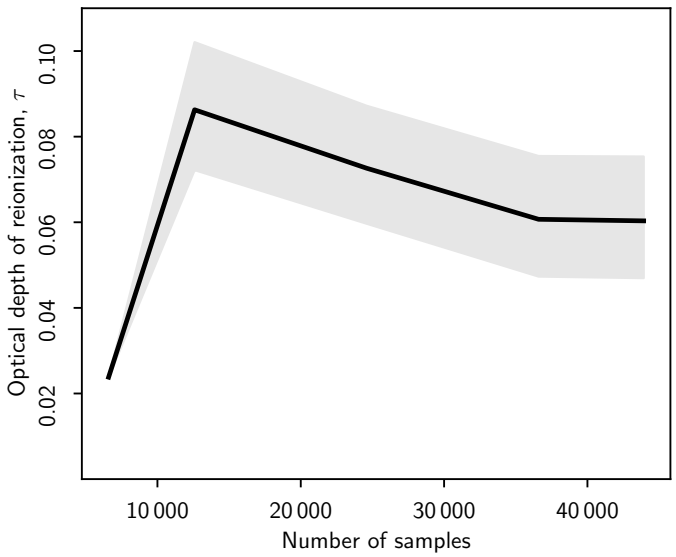

Fig. 39. Convergence of constraints of the reionization optical depth as a function of the number of posterior samples used to construct the CMB mean map and covariance matrix. The solid black line shows the posterior mode for τ, and the gray region shows the corresponding 68 % confidence region.

does appear largely noise dominated, but there are also some intriguing coherent large-scale features that could indicate excess signal. Also, there is fortunately no compelling evidence of residual foreground contamination in this map, and the chosen mask appears to perform well.

As a first quality assesment, we start by slicing the marginal posterior distribution in Eq. (136) with respect to the two main low-ℓ polarization parameters, τ and r. Specifically, we adopt the best-fit *Planck* 2018 ΛCDM model (Planck Collaboration V 2020) to define a reference model, but vary in turn τ and r. Each model spectrum, C_ℓ, is computed with `CAMB` (Lewis et al. 2000). When constraining τ we vary only the EE spectrum, while when constraining r we vary only the BB spectrum. In all these calculations, we keep $A_s e^{-2\tau}$ fixed at the best-fit *Planck* value (Planck Collaboration V 2020) to break a strong degeneracy with the amplitude of scalar perturbations, A_s, essentially adopting the high-ℓ HFI information as a effective prior on the overall amplitude of the spectrum. These calculations are repeated for both the full-sky and the Northern Galactic hemisphere masks as defined above.

To assess the internal consistency of the resulting models, we evaluate the χ^2 of that model with respect to the data in question,

$$\chi^2 = \hat{\mathbf{s}}_{\rm CMB}^t \left(\mathsf{S}(C_\ell^{\rm bf}) + \mathsf{N}_{\rm CMB}\right)^{-1} \hat{\mathbf{s}}_{\rm CMB}. \tag{137}$$

For a Gaussian and isotropic random field, this quantity should be distributed according to a $\chi^2_{n_{\rm dof}}$ distribution, where $n_{\rm dof}$ is the number of degrees of freedom, which in our case is equal to the number of basis vectors in $\hat{\mathbf{s}}_{\rm CMB}$. The signal covariance matrix is in each case evaluated using the best-fit value of τ derived from the parameter exploration described above. The results are summarized in Fig. 37 and Table 4.

Starting with the full-sky case, we first note that the χ^2 probability-to-exceed (PTE) for this region is $5 \cdot 10^{-4}$, indicating a poor fit. Furthermore, the best-fit cosmological parameters are $\tau = 0.091^{+0.010}_{-0.098}$ and $r = 2.9^{+1.3}_{-1.0}$, both of which are in strong tension with previous results from *Planck* and other experiments (e.g., BICEP2/Keck Array and Planck Collaborations 2015; Planck Collaboration V 2020; Tristram et al. 2020).

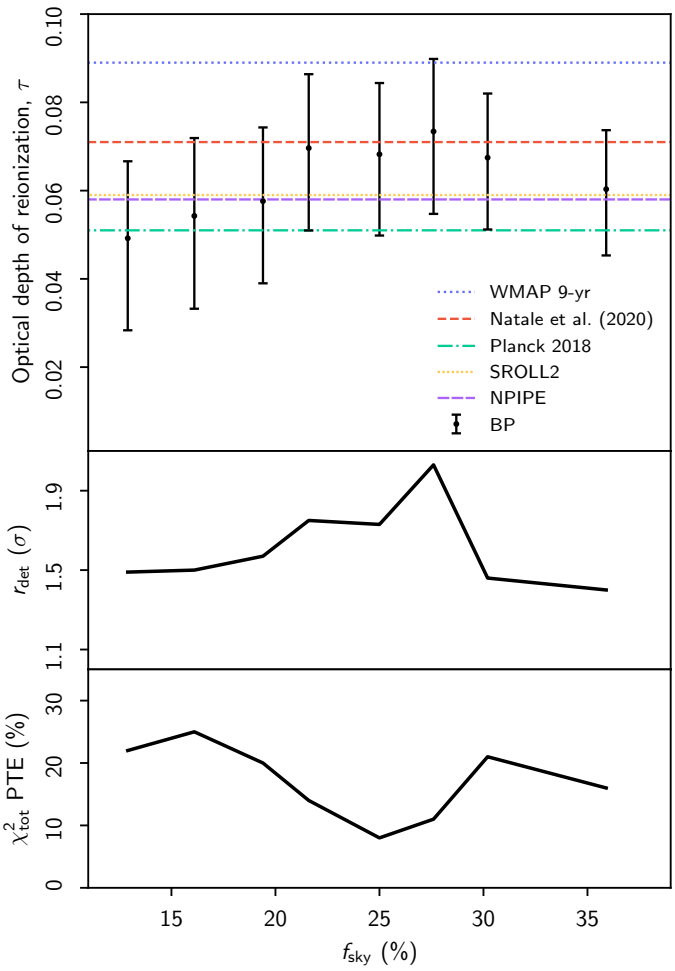

Fig. 40. Estimates of τ (*top panel*), r (*middle panel*), and χ^2 (*bottom panel*) as a function of sky fraction, $f_{\rm sky}$, using only low-ℓ polarization information. The tensor-to-scalar ratio is given as a detection level in units of σ, as measured by the likelihood ratio relative to a model with vanishing B-mode signal.

Thus, even with clearly nonphysical cosmological parameters, this data set fails a basic goodness-of-fit test, and this can only be explained by a break-down of some fundamental assumption, namely either Gaussianity, isotropy, or the basic ΛCDM framework itself. Of course, at this point, we already know about at least one plausible candidate to explain this violation of Gaussianity and statistical isotropy, namely the strong stripes in the 44 GHz channel that are spatially mostly confined to the Southern Galactic hemisphere. Because of this χ^2 excess, we conclude that the BeyondPlanck processing is not yet sufficiently mature to allow analysis of full-sky data.

We therefore turn our attention to the Northern hemisphere alone, for which no strong stripes are seen in the 44 GHz channel. In this case, we see that the χ^2 PTE is 16 %, which is statistically fully acceptable. Furthermore, the best-fit tensor-to-scalar ratio is consistent with zero within a statistical significance of $1.4\,\sigma$, and with an upper 95 % confidence limit of $r < 4.3$. The fact that the peak location of this distribution is slightly positive could be related to the breakdown of the $1/f$ noise model for the 30 and 44 GHz channels reported by Ihle et al. (2020), but the impact of this effect must clearly be small compared to the much larger statistical uncertainties. The best-fit value of the optical depth of reionization is $\tau = 0.060^{+0.015}_{-0.013}$, which is in excellent agreement with previous results. Of course, the uncertainty of this measurement is relatively large, since the accepted sky fraction is small ($f_{\rm sky} = 0.36$).

To visualize the low-ℓ CMB angular power spectrum, we once again employ Eq. (120), and simply compute slices through $\mathcal{L}(C_\ell)$ along each dimension, while fixing all other elements at the best-fit *Planck* 2018 ΛCDM spectrum.[22] For each slice, we report the conditional posterior maximum value as a point estimate, and an asymmetric 68 % confidence interval as its uncertainty. Figure 38 shows a comparison of the resulting BeyondPlanck low-ℓ power spectra evaluated from the Northern Galatic hemisphere with those published by *WMAP* (Hinshaw et al. 2013) and *Planck* 2018 (Planck Collaboration V 2020); the latter two are simply reproduced from tabulated values without reprocessing.

Before concluding this section, we consider two internal consistency tests. First, Fig. 39 shows constraints on τ as a function of number of samples used to generate the low-ℓ likelihood inputs. As discussed by Sellentin & Heavens (2016), it is not sufficient that $\mathsf{N}_{\rm CMB}$ itself converges to some specified precision to obtain robust results, but also that $\mathsf{N}^{-1}_{\rm CMB}$ and $|\mathsf{N}_{\rm CMB}|$ reach a corresponding precision. We see that about 35 000 samples are required to achieve a precision that results in a Monte Carlo uncertainty that is significantly smaller than the posterior width. Regarding the detailed behaviour of this function, we note that the sharp increase around 10 000 samples happens when the final covariance matrix first becomes well-conditioned, and there are no longer any algebraically degenerate modes. However, at this point only a sub-space of the systematic parameter volume has been probed by the Monte Carlo Markov chains, and the corresponding predicted variance is therefore too low, resulting in a positive bias in τ. Only around 35 000 has the full distribution become stationary, and no further improvements are made with more samples.

In the second consistency test we plot τ, r and the χ^2 PTE as a function of $f_{\rm sky}$, adopting the same suite of masks as Planck Collaboration Int. LVII (2020), but after multiplying each with our Northern Galactic mask. The results from these calculations are summarized in Fig. 40. We see that all values of τ are consistent with the *Planck* HFI results to better than $1\,\sigma$. For the tensor-to-scalar ratio, we find that the overall detection level varies between 1.4 and $2.0\,\sigma$, and this variation is also reflected in the χ^2, for which the PTE varies between 8 and 25 %. Overall, all results are statistically consistent for all sky fractions.

9.5.4. High-ℓ CMB likelihood

We now turn our attention to the high-ℓ CMB power spectrum and likelihood. In this respect, we note that the LFI and *WMAP* polarization measurements have very low CMB signal-to-noise ratios at multipoles $\ell \gtrsim 10$, and we therefore restrict ourselves in the following to temperature alone. This has two main advantages. First, the cost of producing a single high-ℓ constrained CMB sample is reduced by a factor of three, since we no longer need to perform polarized spherical harmonics transforms. Second, this also allows us to employ the so-called Blackwell-Rao estimator (Chu et al. 2005), which is by far the preferred estimator for sampling-based likelihood estimators due to its high level of precision combined with excellent computational speed.

[22] A more common convention is to fix other spectra at the joint maximum likelihood solution. However, this sometimes leads to pathological likelihood shapes for strongly noise-dominated modes, and the resulting spectra are in such cases poor representations of the likelihood shape that is actually relevant for parameter estimation.

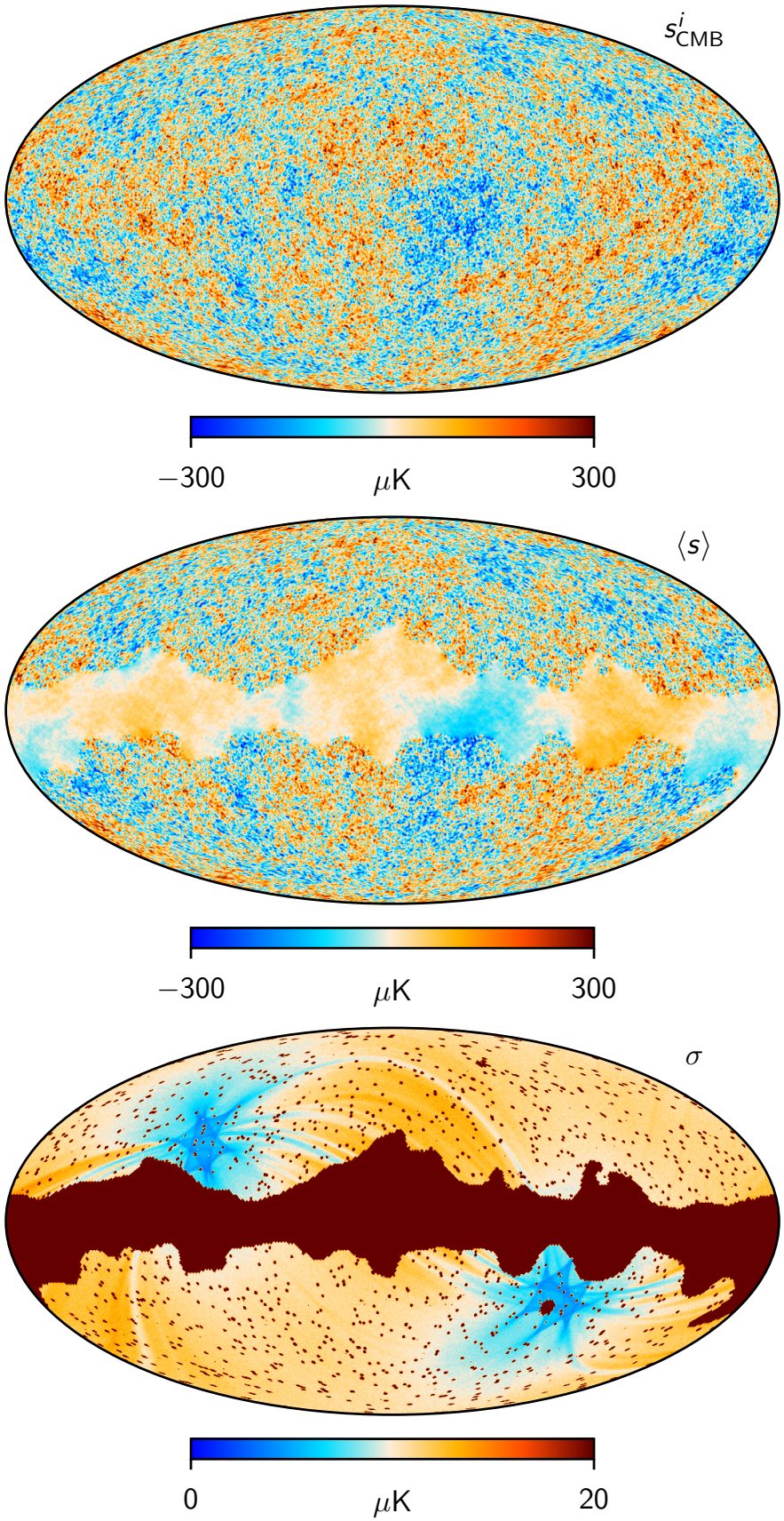

Fig. 41. Full-resolution CMB temperature constrained realization maps. (*Top*:) Single constrained realization, $\boldsymbol{s}^i$, drawn from $P(\boldsymbol{s} \mid \boldsymbol{d}, C_\ell, \ldots)$. (*Middle*:) Posterior mean map, $\langle \boldsymbol{s} \rangle$, as evaluated from the ensemble of constrained CMB realizations; note that the small-scale signal amplitude inside the mask decreases smoothly to zero with increasing distance from the edge of the mask. (*Bottom*:) CMB posterior standard deviation map, as evaluated pixel-by-pixel from the ensemble of constrained CMB realizations. This map is dominated by instrumental noise outside the mask, and by random fluctuations informed by the assumptions of isotropy and Gaussianity inside the mask.

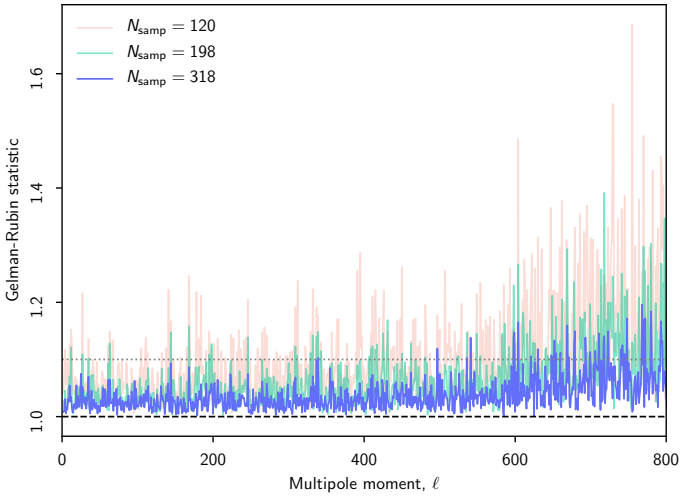

Fig. 42. Gelman-Rubin convergence statistic for the BeyondPlanck TT angular power spectrum, as evaluated from eight σ_ℓ chains with each N samples. A value lower than 1.1 (dotted gray line) typically indicates acceptable convergence.

However, it does require a substantial signal-to-noise ratio per multipole in order to converge, which is not the case for current polarization data. In practice we employ the Gaussianized Blackwell-Rao estimator, as presented by Rudjord et al. (2009) and used by the *Planck* 2015 and 2018 likelihoods (Planck Collaboration XI 2016; Planck Collaboration V 2020), in order to reduce the number of samples required for convergence at high multipoles.

The inputs to this likelihood estimator are generated as follows. For each full-resolution Gibbs sample[23] discussed in the introduction to this section, we draw a new sample from $P(\boldsymbol{s}_{\rm CMB} \mid \boldsymbol{d}, \omega \setminus \boldsymbol{s}_{\rm CMB})$, as we did for the low-$\ell$ estimator, but with the following important differences:

1. The re-sampling step is performed at full angular resolution.
2. The noise level of each input frequency map is set to infinity for each pixel that is excluded by the BeyondPlanck temperature mask, giving these zero weight in the fit.
3. We condition on a CMB power spectrum prior, $\mathsf{S}(C_\ell)$, when solving the Wiener filter equation, which results in a sample, $\boldsymbol{s}^i$, that has the Galactic plane in-painted with a Gaussian constrained realization.
4. The power spectrum, C_ℓ^i, is Gibbs sampled over between each sky sample, $\boldsymbol{s}^i$, such that $\{\boldsymbol{s}^i, C_\ell^i\}$ explore the full corresponding posterior distribution. Multipoles higher than $\ell > 800$ are fixed at the best-fit *Planck* ΛCDM power spectrum, while all multipoles below $\ell \leq 800$ are sampled ℓ-by-ℓ.

For each sky sample, we compute the observed power spectrum, σ_ℓ^i, as defined by Eq. (6), which serves as the actual input to the Blackwell-Rao estimator. Note that this procedure is very nearly identical to that described by Chu et al. (2005), and later adopted by both *WMAP* (Hinshaw et al. 2013) and *Planck* (Planck Collaboration V 2020), but with one fundamental difference: While all previous analyses drew samples from one fixed set of frequency sky maps, we now marginalize over a whole ensemble of frequency sky maps. Thus, this is the first time that low-level instrumental systematic errors are also propagated through the

[23] At the time of writing, only 400 high-resolution samples have been produced after burn-in. The wall-time cost per sample is 5 hours.

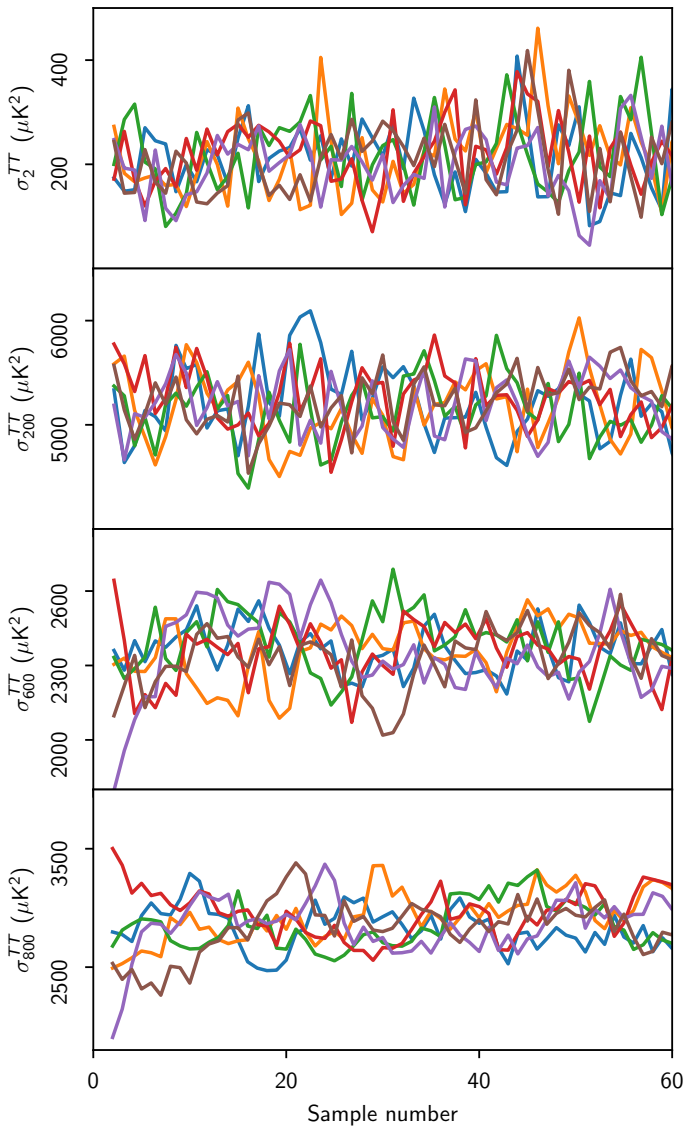

Fig. 43. Trace plots for four representative angular power spectrum multipole coefficients, σ_ℓ^{TT}. From top to bottom, each of the four panels show $\ell = 2$, 200, 600, and 800. Each curve shows one independent Gibbs chain.

Blackwell-Rao estimator into high-level cosmological parameters, in addition to instrumental noise, sky cut, and foreground uncertainties (Planck Collaboration V 2020).

To provide some useful visual intuition, Fig. 41 summarizes the properties of the constrained realizations that feed into the Blackwell-Rao estimator. The top panel shows a single CMB sky map sample from the full posterior distribution. This represents one perfect full-sky CMB map that is *consistent* with the underlying data. Note that it is impossible to see the analysis mask in this plot; these pixels are in effect replaced with a Gaussian random realizations with a power spectrum given by C_ℓ^i and phases that are constrained by the high-latitude information. The middle panel shows the corresponding posterior mean map. At high Galactic latitudes, this map is virtually identical to the single sample (only with very slightly lower noise), while inside the mask it is much smoother. Only modes that can be meaningfully estimated from high latitudes under the assumption of statistical isotropy have non-zero value. Finally, the bottom panel show the posterior standard deviation map. At high latitudes, this is dom-

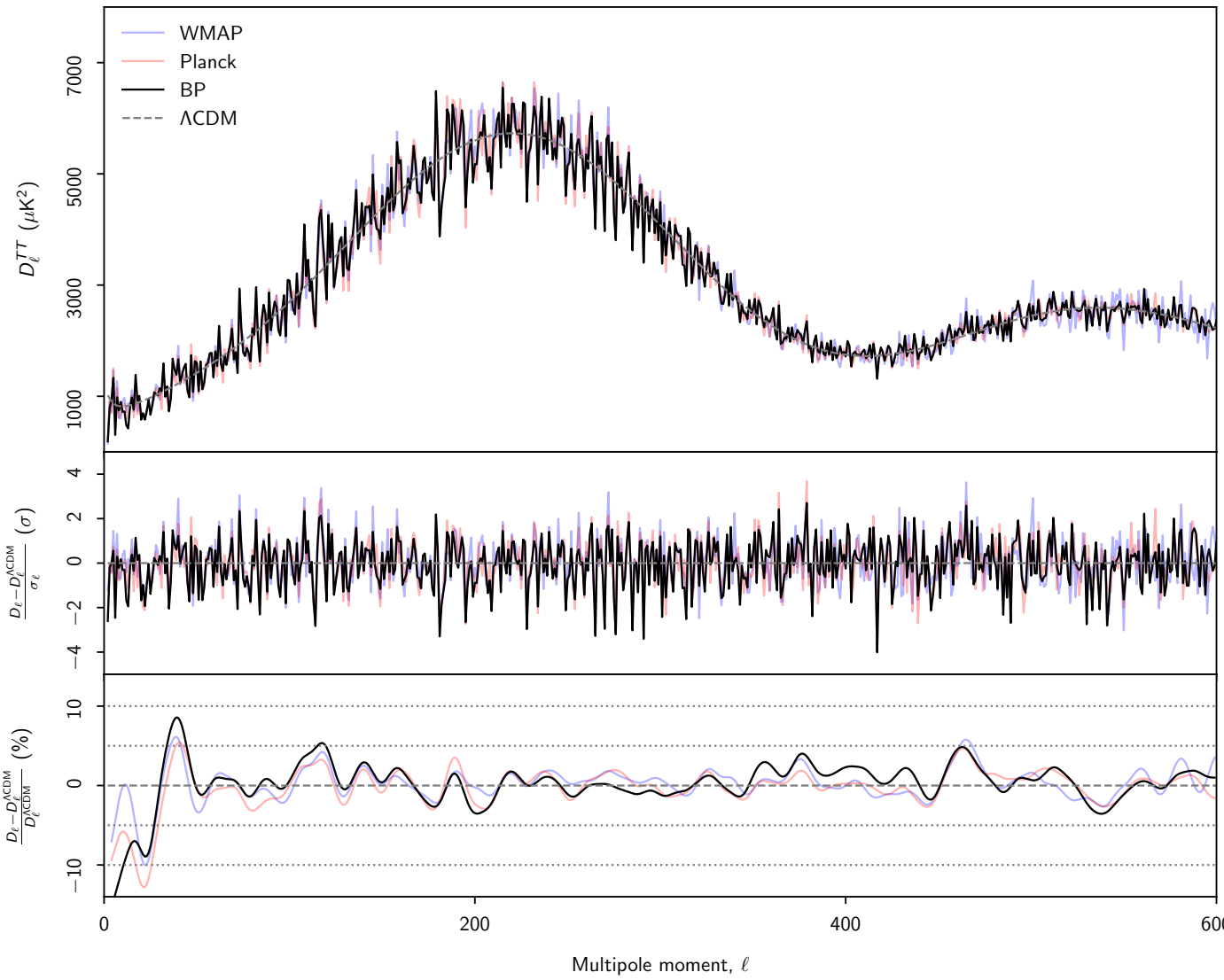

Fig. 44. (*Top*:) Angular CMB temperature power spectrum, D_ℓ^{TT}, as derived by BeyondPlanck (black), *Planck* (red), and *WMAP* (blue). The best-fit *Planck* 2018 ΛCDM power spectrum is shown in dashed gray. (*Middle*:) Residual power spectrum relative to ΛCDM, measured relative to full quoted error bars, $r_\ell \equiv (D_\ell - D_\ell^{\Lambda\mathrm{CDM}})/\sigma_\ell$. 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.

inated by instrumental noise and systematic effects, while inside the mask it is dominated by CMB sample variance.

The intuitive interpretation of the Blackwell-Rao estimator is now very simple: Assume first that we happened to know the exact CMB sky, without noise, foregrounds or systematic effects. In that case, the CMB power spectrum likelihood is given analytically by the inverse Wishart distribution shown in Eq. (12). However, in reality, we do of course not know the true sky perfectly, and we therefore have to marginalize over all possible CMB skies that are consistent with all observed data. In other words, we have to average the Gaussian likelihood over the ensemble of $\boldsymbol{s}^i$ samples—and that is precisely the definition of the Blackwell-Rao estimator.

As shown by Wandelt et al. (2004), the Blackwell-Rao estimator is guaranteed to converge to the true, exact CMB likelihood in the limit of an infinite number of samples. Of course, in practice we do not have an infinite number of samples, and a common way of assessing convergence is through the so-called Gelman-Rubin R statistic (Gelman & Rubin 1992), which compares the sample variance as measured *within* each chain to the variance *between* chains. With the appropriate scaling factors, R should typically be lower than 1.1 for acceptable convergence. Figure 42 shows the R statistic for the σ_ℓ^i ensemble described above, which suggests that acceptable convergence is achieved up to $\ell \approx 600$–700 with the current sample set.

These numbers are, however, only general rules-of-thumb, and should always be combined with human visual inspection of the actual chains as well. A selection of four representative multipoles are therefore shown in Fig. 43, in which each color represents one independent Gibbs chain. Here we see that $\ell = 2$ and 200 mix very well, and the Markov chain correlation length is essentially zero. As a result, robust convergence will be achieved with a relatively low number of samples. At $\ell = 600$, the correlation length is notably longer, but a good estimate of the full distribution may still be derived from the sample set shown. At $\ell = 800$, it is even longer, and the mixing is about to break down.

We conclude from these results that the current sample set has converged well below $\ell \lesssim 700$. Still, to be conservative we

Table 5. Comparison of basic 6-parameter ΛCDM model parameters as derived by BeyondPlanck (Paradiso et al. 2020), *Planck* 2018 (Planck Collaboration VI 2020), and *WMAP* (Hinshaw et al. 2013). The second column shows results for BeyondPlanck only, using only TT multipoles below $\ell \leq 600$ and polarization below $\ell \leq 8$. The third column shows similar results when also adding TT multipoles between $600 < \ell \leq 2500$ from *Planck* 2018. For *Planck*, we show results from the `Plik` pipeline using the TT+TE+EE+lowE+lensing data combination, while for *WMAP* we show results based on C^{-1}-weighted 9-year *WMAP*-only data. Note that *Planck* and *WMAP* adopt slightly different conventions for some parameters, and we report both where applicable. Columns marked with "Δ" show differences with respect to BeyondPlanck-only results, as measured in units of σ.

	BeyondPlanck		*Planck* 2018		*WMAP*	
Parameter	$\ell \leq 600$	+*Planck* $\ell > 600$	Estimate	$\Delta\,(\sigma)$	Estimate	$\Delta\,(\sigma)$
$\Omega_b h^2$	0.02226 ± 0.00088	0.02230 ± 0.00022	0.02237 ± 0.00015	−0.1	0.02243 ± 0.00050	−0.2
$\Omega_c h^2$	0.115 ± 0.016	0.1227 ± 0.0025	0.1200 ± 0.0012	−0.3	0.1147 ± 0.0051	0
Ω_Λ	⋯	⋯	⋯	⋯	0.721 ± 0.025	⋯
$100\theta_{\rm MC}$	1.0402 ± 0.0048	1.04064 ± 0.00048	1.04092 ± 0.00031	−0.2	⋯	⋯
τ	0.067 ± 0.016	0.074 ± 0.015	0.054 ± 0.007	0.8	0.089 ± 0.0014	−1.4
$10^9 \Delta_{\mathcal{R}}^2$	⋯	⋯	⋯	⋯	2.41 ± 0.10	⋯
$\ln(10^{10} A_{\rm s})$	3.035 ± 0.079	3.087 ± 0.029	3.044 ± 0.014	−0.1	⋯	⋯
$n_{\rm s}$	0.962 ± 0.019	0.9632 ± 0.0060	0.9649 ± 0.0042	−0.1	0.972 ± 0.013	−0.5

only include multipoles between $9 \leq \ell \leq 600$ in the final BeyondPlanck high-ℓ TT likelihood for cosmological parameter estimation. The resulting power spectrum is shown in Fig. 44, and compared with those 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 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.

9.5.5. Cosmological parameters

Finally, we are ready to present cosmological parameters from the BeyondPlanck analysis pipeline. In the following, we use `CosmoMC` (Lewis & Bridle 2002) to explore different cosmological models, coupled to a likelihood of the following form,

$$\ln \mathcal{L}_{\rm BP}(C_\ell) = \ln \mathcal{L}^{\ell=2-8}_{\rm low-}\ell(C_\ell) \quad (138)$$

$$+ \ln \mathcal{L}^{\ell=9-600}_{\rm BR}(C_\ell) \quad (139)$$

$$+ \ln \mathcal{L}^{\ell=601-2500}_{Planck}(C_\ell) \quad \text{(optional)}, \quad (140)$$

where the likelihood in line 138 is given by Eq. (120); the likelihood in line 139 is given by Eq. (7) in Rudjord et al. (2009) (which, intuitively, is also given by averaging Eq. (128) over the ensemble of available Gibbs samples); and the likelihood in line 140 is defined by the Gaussian TT-only *Planck* 2018 likelihood (Planck Collaboration V 2020).

The low-ℓ likelihood in Eq. (138) is defined directly in terms of a full multi-variate pixel-based Gaussian distribution with mean and covariance tuned using the posterior samples. Similarly, the Blackwell-Rao estimator in Eq. (139) is defined simply by averaging the inverse Wishart distribution (as defined in Eq. (128)) over all available Gibbs samples. Intuitively, each Gibbs sample represents one possible full-sky and noiseless CMB realization that is consistent with all observed data, and the the Blackwell-Rao estimator is thus simply equivalent to averaging the appropriate distribution for cosmic variance (i.e., the inverse Wishart distribution) over all possible CMB realizations that are consistent with the measurements. Finally, the high-ℓ *Planck* likelihood is simply defined as a multi-variate Gaussian distribution in terms of angular power spectra. We estimate parameters both with and without the *Planck* high-ℓ likelihood.

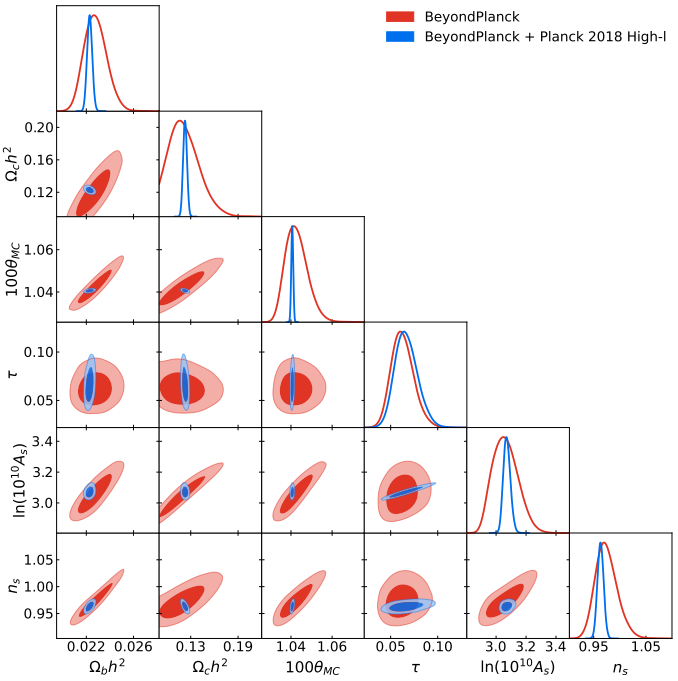

Fig. 45. Marginal cosmological parameter posterior distributions for a basic 6-parameter ΛCDM model as estimated from the BeyondPlanck-only CMB likelihood using TT multipoles up to $\ell \leq 600$ and polarization multipoles up to $\ell \leq 8$ (blue distributions), and also when including $\ell > 600$ temperature information from *Planck* 2018 (red distributions).

For a complete discussion regarding the parameter constraints that is derived from this likelihood, and comparisons with previously published results, we refer the interested reader to Paradiso et al. (2020). Here we only show one single case, namely that corresponding to the basic six-parameter ΛCDM model. These are summarized in terms of posterior means and

standard deviations in Table 5, and in terms of 1- and 2-dimensional marginal distributions in Fig. 45.

Overall, we see that the agreement between BeyondPlanck, *Planck* 2018 and *WMAP* is very good. For *Planck* 2018, which is statistically dominated by HFI measurements, the parameter showing the biggest difference is τ, with a positive shift of $0.8\,\sigma$. We note however that both SROLL2 (Pagano et al. 2020) and NPIPE (Tristram et al. 2020) report slightly higher values of τ than *Planck* 2018; as measured relative to these, the BeyondPlanck excess is only $0.4\,\sigma$.

From Table 5, we also see that the BeyondPlanck uncertainties are generally larger than those of *Planck* 2018, even when combined with high-ℓ information from *Planck*. This is most likely due to the fact that BeyondPlanck does not exploit HFI measurements below $\ell \leq 600$, and therefore observes notably larger uncertainties between the first and third accoustic peaks in the CMB power spectrum. In addition to these differences in raw data volume, it is also important to note that the BeyondPlanck framework marginalizes over a significantly larger model of nuisance parameters than either of the two previous pipelines, and these also contribute to larger uncertainties. For a detailed breakdown of statistical and systematic uncertainties, and their impact on cosmological parameters, we refer the interested reader to Colombo et al. (2020); Paradiso et al. (2020).

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-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. A concrete example of such synergies has already been demonstrated in the current paper, where information contained in the *WMAP* observations is used to support the calibration and component separation of LFI, and, as a result, we are now, for the first time, able to fit for the spectral index of polarized synchrotron emission across the two experiments.

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. (2020), 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 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[24] source code, as well as various pre- and post-processing tools,[25] 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. Preliminary documentation is provided,[26] 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.
4. *Community-based support environment* – we have established a web-based discussion forum[27] dedicated to end-to-end analysis where interested parties may share experiences and discuss issues. All participation in this forum is of course voluntary, with no expectations of either commitments or guarantees from any participant, but the hope is that the forum will grow into a useful and active discussion platform for anything from bugs and code development issues to high-level scientific questions.

In addition, all main results (both full chain files and selected post-processed posterior mean and standard deviation maps) are available from the BeyondPlanck homepage,[28] and eventually through the *Planck* Legacy Archive.[29] For further details regarding the reproducibility aspects of the work, we refer the interested reader to Gerakakis et al. (2020).

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. (2020a): The full BeyondPlanck analysis, as presented here, has a computational cost of 220 000 CPU hours, which is roughly equivalent to the cost of producing $O(10)$ 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. 2020a), and now requires only about 1.5 TB of RAM to run. Computers with this amount of memory and clock cycles are now widely available, and a full *Planck* LFI analysis therefore no longer requires the use of expensive supercomputers – 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

[24] http://beyondplanck.science
[25] https://github.com/cosmoglobe/c3pp
[26] https://docs.beyondplanck.science
[27] https://forums.beyondplanck.science
[28] http://beyondplanck.science
[29] https://pla.esac.esa.int/

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. (2020a). As two real-world demonstrations of the extendability of the framework, we present preliminary applications to both *LiteBIRD* and *WMAP* in two companion papers (Aurlien et al. 2020; Watts et al. 2020).

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,[30] 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. Due to limited resources and time, BeyondPlanck only considered the *Planck* LFI data in the time domain, although some preliminary work has also been done on *WMAP* and simulated *LiteBIRD* observations.

Regrettably, an application for continued funding for integrating HFI and *LiteBIRD* into the same framework was recently rejected. The referee noted that *the advantages of the proposed methodology over other available methodologies (. . .) are not sufficiently explained*, and *the use of the stakeholder knowledge, such as instrument modellers and astrophysicists, is not sufficiently considered.* These are very reasonable comments, and a busy Principal Investigator of a given experiment who reads this paper may ask very similar questions: "Why should I care about Bayesian statistics? What's in it for me?" To answer these questions, we make the following points:

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 primarily 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. This work has already started for *LiteBIRD*, *SPIDER*, and *WMAP*, but we 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.

[30] 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.

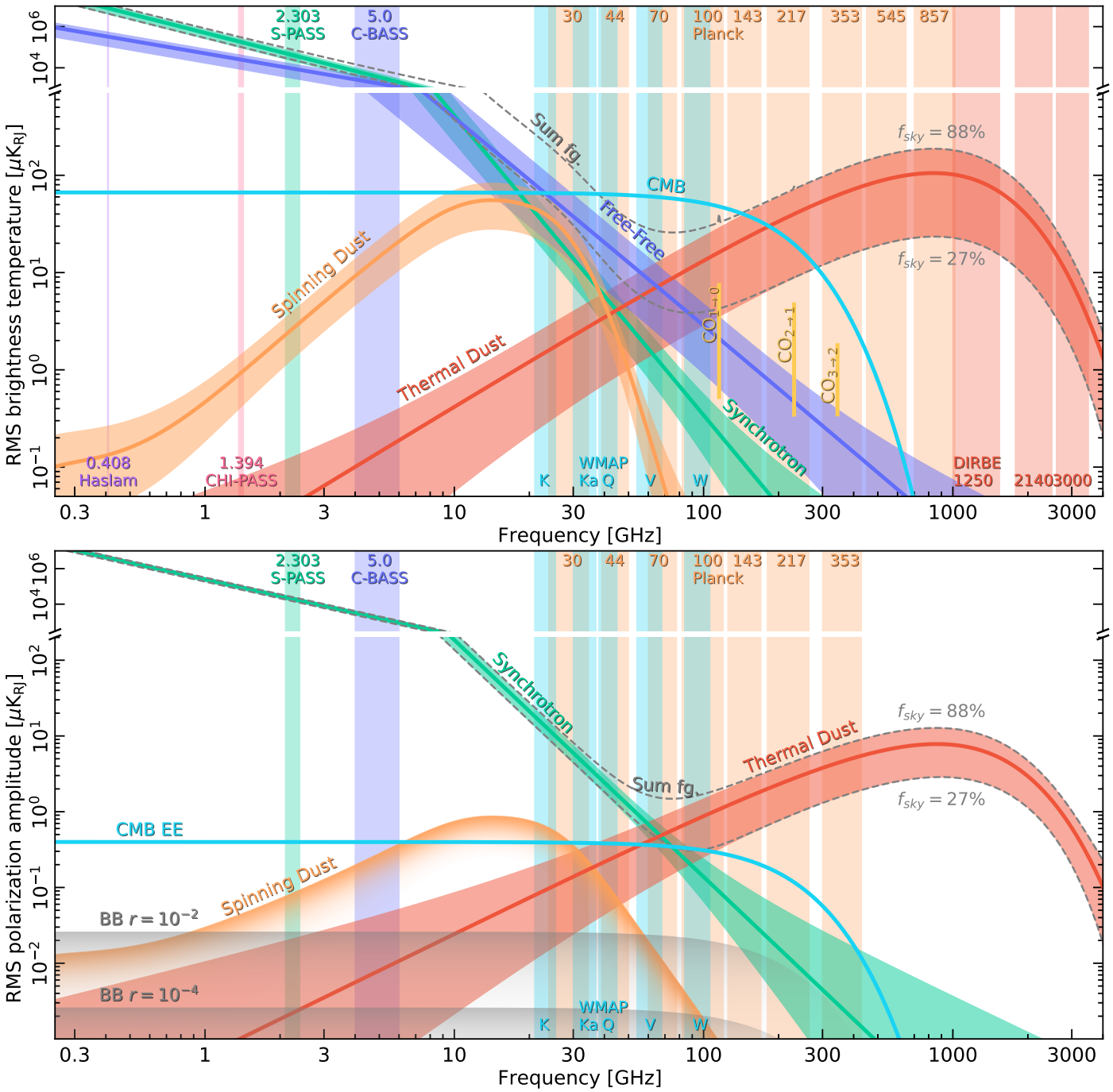

Fig. 46. Brightness temperature (*top panel*) and polarization amplitude (*bottom panel*) RMS as a function of frequency and astrophysical component, and as derived from the BeyondPlanck and *Planck* sky models. Vertical bands indicate the frequency ranges of various experiment. All components have been 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. For polarization, the spinning dust component (orange curve) indicates an upper limit as presented by Herman et al. (2020), not a detection. In the current BeyondPlanck analysis, only the three LFI channels are modelled in the time-domain. A long-term future goal is to include all publicly available and relevant data (for instance *WMAP* and *Planck* HFI) into this model; preferrably 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 Cosmoglobe project.

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 (Herman et al. 2020). Experiments of this type should also be relatively straighforward 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 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, at least partially, succeeded in integrating the LFI 44 GHz channel into a statistically viable low-ℓ CMB likelihood. In the process, we have identified two important breakdowns of the current 44 GHz instrument model, namely a limited range of scanning rings for which the gain model appears to break down, and a general short-coming of a simple $1/f$ model to describe correlated noise in both the 30 and 44 GHz channels. Understanding the nature of these systematic errors, and mitigating them, is an important goal for the immediate future (and may be possible even before the current suite of papers goes into press!). For now, however, only the Northern Galactic hemisphere may be used for cosmological low-ℓ polarization analysis.
2. We have for the first time constructed a full, dense, low-resolution CMB covariance matrix that accounts for marginalization over a wide range of important systematic time-ordered effects, including gain, bandpass, and foreground corrections, in addition to the usual 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 $\tau = 0.060^{+0.015}_{-0.013}$. 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 $3359.5 \pm 1.9\,\mu$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.

Before concluding, we reemphasize that this is not the end. While the BeyondPlanck project itself contractually ends 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 46 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; there are still known bugs and memory leaks in the code (Galloway et al. 2020a); 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.

Acknowledgements. We thank Prof. Pedro Ferreira for useful suggestions, comments and discussions, and Dr. Diana Mjaschkova-Pascual for administrative support. We also thank the entire *Planck* and *WMAP* teams for invaluable support and discussions, and for their dedicated efforts through several decades without which this work would not be possible. The current work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement numbers 776282 (COMPET-4; BeyondPlanck), 772253 (ERC; bits2cosmology), and 819478 (ERC; Cosmoglobe). In addition, the collaboration acknowledges support from ESA; ASI and INAF (Italy); NASA and DoE (USA); Tekes, Academy of Finland (grant no. 295113), CSC, and Magnus Ehrnrooth foundation (Finland); RCN (Norway; grant nos. 263011, 274990); and PRACE (EU).

References

Ali-Haimoud, Y. 2010, SpDust/SpDust.2: Code to Calculate Spinning Dust Spectra
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. 2020, A&A, in preparation [arXiv:201x.xxxxx]
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
Aurlien et al. 2020, A&A, in preparation [arXiv:201x.xxxxx]
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 Collaboration. 2020, A&A, in preparation [arXiv:201x.xxxxx]
BICEP2/Keck Array and Planck Collaborations. 2015, Phys. Rev. Lett., 114, 101301
Brilenkov et al. 2020, A&A, in preparation [arXiv:201x.xxxxx]
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. 2020, A&A, in preparation [arXiv:201x.xxxxx]
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
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. & Hensley, B. S. 2020, arXiv e-prints, arXiv:2009.11314
Draine, B. T. & Lazarian, A. 1998, ApJ, 494, L19
Dunkley, J., Komatsu, E., Nolta, M. R., et al. 2009, ApJS, 180, 306
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. 2019, arXiv e-prints, arXiv:1909.05923
Fuskeland, U., Wehus, I. K., Eriksen, H. K., & Næss, S. K. 2014, ApJ, 790, 104
Galeotta et al. 2020, A&A, in preparation [arXiv:201x.xxxxx]
Galloway et al. 2020a, A&A, in preparation [arXiv:201x.xxxxx]
Galloway et al. 2020b, A&A, in preparation [arXiv:201x.xxxxx]
Gelman, A. & Rubin, D. B. 1992, Statist. Sci., 7, 457
Geman, S. & Geman, D. 1984, IEEE Trans. Pattern Anal. Mach. Intell., 6, 721
Génova-Santos, R., Rubiño-Martín, J. A., Rebolo, R., et al. 2015, MNRAS, 452, 4169
Gerakakis et al. 2020, A&A, in preparation [arXiv:201x.xxxxx]
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. 2020, A&A, in preparation [arXiv:201x.xxxxx]
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
Génova-Santos, R., Rubiño-Martín, J. A., Peláez-Santos, A., et al. 2016, Monthly Notices of the Royal Astronomical Society, 464, 4107
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
Hensley, B. S. & Draine, B. T. 2020, ApJ, 895, 38
Herman et al. 2020, A&A, in preparation [arXiv:201x.xxxxx]
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. 2020, A&A, in preparation [arXiv:201x.xxxxx]
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
Jew, L., Taylor, A. C., Jones, M. E., et al. 2019, Monthly Notices of the Royal Astronomical Society, 490, 2958
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. 2020, A&A, in preparation [arXiv:201x.xxxxx]
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
Krachmalnicoff, N., Carretti, E., Baccigalupi, C., et al. 2018, A&A, 618, A166
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, arXiv e-prints, arXiv:2005.05600
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. 2020, A&A, in preparation [arXiv:201x.xxxxx]
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 XIV. 2014, A&A, 571, A14
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. 2020, 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, in press [arXiv:2007.04997]
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
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
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
Sugai, H., Ade, P. A. R., Akiba, Y., et al. 2020, Journal of Low Temperature Physics, 199, 1107
Sunyaev, R. A. & Zeldovich, Y. B. 1972, Comments on Astrophysics and Space Physics, 4, 173
Suur-Uski et al. 2020, A&A, in preparation [arXiv:201x.xxxxx]
Svalheim et al. 2020a, A&A, in preparation [arXiv:201x.xxxxx]
Svalheim et al. 2020b, A&A, in preparation [arXiv:201x.xxxxx]
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
Thommesen, H., Andersen, K. J., Aurlien, R., et al. 2020, arXiv e-prints, arXiv:2007.06250
Tristram, M., Banday, A. J., Górski, K. M., et al. 2020, arXiv e-prints, arXiv:2010.01139
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. 2020, A&A, in preparation [arXiv:201x.xxxxx]
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

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 $\left\langle(x-\mu)^2\right\rangle = \sigma^2$ because $\left\langle\eta^2\right\rangle = 1$.

A sample from a multi-variate normal distribution $N(\bar{\mu},\mathsf{C})$ with mean vector $\bar{\mu}$ and covariance matrix C may be produced in a fully analogous manner,

$$\boldsymbol{x} = \bar{\mu} + \mathsf{C}^{\frac{1}{2}}\bar{\eta}, \tag{A.2}$$

where now $\bar{\eta}$ 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 ($\mathsf{C} = \mathsf{L}\mathsf{L}^t$ for positive definite matrices, where $\mathsf{C}^{\frac{1}{2}} = \mathsf{L}$) and singular-value decomposition ($\mathsf{C} = \mathsf{V}\Sigma\mathsf{V}^t$ 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 eigen-vector 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 $\boldsymbol{d}$: $\boldsymbol{d}' \rightarrow \boldsymbol{d}-\boldsymbol{b}$, 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)

$$\begin{aligned}-2\ln P(\boldsymbol{x}\mid\boldsymbol{d}) &= \boldsymbol{x}^t\mathsf{S}^{-1}\boldsymbol{x} + (\boldsymbol{d}-\mathsf{T}\boldsymbol{x})^t\mathsf{N}^{-1}(\boldsymbol{d}-\mathsf{T}\boldsymbol{x})\\ &= \boldsymbol{x}^t\mathsf{S}^{-1}\boldsymbol{x} + \boldsymbol{d}^t\mathsf{N}^{-1}\boldsymbol{d} + \boldsymbol{x}^t\mathsf{T}^t\mathsf{N}^{-1}\mathsf{T}\boldsymbol{x} -\\ &\qquad \boldsymbol{d}^t\mathsf{N}^{-1}\mathsf{T}\boldsymbol{x} - \boldsymbol{x}^t\mathsf{T}^t\mathsf{N}^{-1}\boldsymbol{d}\\ &= \boldsymbol{x}^t(\mathsf{S}^{-1}+\mathsf{T}^t\mathsf{N}^{-1}\mathsf{T})\boldsymbol{x} - 2\boldsymbol{x}^t\mathsf{T}^t\mathsf{N}^{-1}\boldsymbol{d},\end{aligned} \tag{A.5}$$

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, 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 a $O(n^3)$ operation. To circumvent this problem, we instead consider the same equation in the form

$$(\mathsf{S}^{-1}+\mathsf{T}^t\mathsf{N}^{-1}\mathsf{T})\boldsymbol{x} = \mathsf{T}^t\mathsf{N}^{-1}\boldsymbol{d}. \tag{A.9}$$

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 simply add one random zero-mean covariance term for each element in the covariance matrix to the right-hand side of the equation,

$$(\mathsf{S}^{-1} + \mathsf{T}^t\mathsf{N}^{-1}\mathsf{T})\boldsymbol{x} = \mathsf{T}^t\mathsf{N}^{-1}\boldsymbol{d} + \mathsf{T}^t\mathsf{N}^{-1/2}\boldsymbol{\eta}_1 + \mathsf{S}^{-1/2}\boldsymbol{\eta}_2. \tag{A.10}$$

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.9),

$$(\mathsf{S}^{-1} + \mathsf{T}^t\mathsf{N}^{-1}\mathsf{T})\boldsymbol{x} = \mathsf{T}^t\mathsf{N}^{-1}\boldsymbol{d} + \mathsf{S}^{-1}\boldsymbol{m} + \mathsf{T}^t\mathsf{N}^{-1/2}\boldsymbol{\eta}_1 + \mathsf{S}^{-1/2}\boldsymbol{\eta}_2. \tag{A.11}$$

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., $\mathsf{N}^{-1} \rightarrow 0$), $\langle \boldsymbol{x} \rangle = \boldsymbol{m}$, 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 $\mathsf{S} \rightarrow 0$ (or, equivalently, $\mathsf{S}^{-1} \rightarrow \infty$), 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.