Baltic Astronomy, vol. 9, 511–517, 2000

DATA MANAGEMENT FOR THE LOW FREQUENCY INSTRUMENT OF THE ESA PLANCK MISSION

F. Pasian[1], M. Bersanelli[2] and N. Mandolesi[3]

[1] *Astronomical Observatory, Via Tiepolo 11, 34131 Trieste, Italy*
[2] *University of Milano, Via Celoria 16, 20133 Milano, Italy*
[3] *TESRE/CNR, Via Gobetti 101, 40129 Bologna, Italy*

Received September 8, 2000.

Abstract. Planck is the M3 mission of ESA's Horizon 2000+ programme, expected to be launched in early 2007. In this paper, the data management aspects for the mission are discussed. They include processing and archiving of ground-test data, simulations of instrumental effects on observations, production of simulated data, sampling and quantization effects, data compression, handling of telemetry and telecommands, calibration of raw data, creation of frequency maps, separation of astrophysical components.

Key words: space vehicles – astronomical databases

1. THE PLANCK MISSION AND THE LFI INSTRUMENT

Planck is the M3 mission of ESA's Horizon 2000 programme, due to be launched in March 2007. During its nominal lifetime (14 months), Planck will carry out two full sky surveys at nine frequencies between 30 and 900 GHz, with high calibration accuracy, freedom from systematic errors, stability and sensitivity. The main scientific goal of the mission is to map the anisotropy of the Cosmic Microwave Background (CMB) at all angular scales $\geq$ 10 arcmin with an accuracy set by the astrophysical limits, and to extract the main cosmological parameters with an accuracy as great as a few percent. In addition, the Planck surveys will produce a wealth of astrophysical information in the microwave-submillimiter range.

The spacecraft will be placed on an orbit around the second Earth-Sun Lagrange point (L2), and will rotate with a ~1 rpm spin period, observing all of the sky in near-great circles.

Two instruments constitute Planck's payload: the Low Frequency Instrument (LFI) and the High Frequency Instrument (HFI). The LFI is a radiometer array with 28 feed horns operating at 30, 44, 70 and 100 GHz with angular resolution of 33, 23, 14 and 10 arcmin respectively. All channels are polarization sensitive.

The LFI is based on pseudo-correlation receivers: a differential measure is made between signals coming from the sky and a reference load at 4 K. The LFI will operate in a steady and continuous mode throughout the mission survey. The real-time data streams will be dominated by detector noise while the final maps, with high signal to noise ratio, will be produced by co-adding long-term redundant observations.

2. LFI DATA HANDLING DURING OPERATIONS

The sampling rate of the analog-to-digital converter (ADC) has been chosen to be 8192 Hz; all channels are sampled by the ADC at the same rate, thus the resulting total raw data rate is 21 Mbps. The on-board Signal Processing Unit (SPU) averages together the samples and produces the bins to be sent to the ground. The quantization step needs to be accurately chosen to avoid saturation of the detectors with bright sources, but be sensitive enough to detect the CMB variations. The baseline quantization step q has been currently set such that $\sigma/q = 2$ (where σ is the sigma of the white noise), which allows to reach a loss-less compression factor of ~ 3.8 (Maris et al. 2000). The total data rate for the LFI is expected to be 31 Kbps with a 20% contingency and adding 4 Kbps of instrument housekeeping. The LFI compressed science data and the housekeeping are included in fixed-length self-descriptive telemetry packets, which are sent to ground with the ones produced by the other instrument (HFI).

The baseline transmission bandwidth between the Planck spacecraft and the Ground Station (Perth) corresponds to an equivalent transmission rate of 60 Kbps over 24 hours. An increase of such rate to accomodate the LFI and HFI requests (~90 Kbps total) is currently being studied. Transmission of data to ground is not continuously distributed over 24 hours, due to the limited visibility windows of the Ground Station (from 8 to 13 hours/day depending on the season), and to possible radio frequency interference between the transmitter and the radio channels of the instrument, which could affect the quality of science data. A whole day of data is therefore sent from the spacecraft to the Ground Station during a period of only 2–3 hours a day. RT (real-time data, i.e. those being acquired during transmission) are sent to the ground interspersed with those acquired in the preceding acquisition period and stored on-board on a solid-state recorder.

In the currently expected scenario, Planck data are transmitted from the Ground Station to the Mission Operations Centre (MOC, at ESOC) using a redundant dedicated 128 Kbps line. RT data are sent as they come, while data recorded on-board are transmitted within 16 hours. MOC staff checks spacecraft and instrument housekeeping information, detecting out-of-range telemetry parameters, to guarantee the safety and health of Planck subsystems. No scientific packets are opened. This baseline set-up would allow the MOC to react within the visibility period of the spacecraft if anomalies are detected in the RT housekeeping data, within the subsequent visibility period otherwise. Spacecraft and instrument telemetry are stored in a telemetry archive from which the instrument Data Processing Centres (DPCs) will downlink the data. MOC will also produce auxiliary data, such as reconstructed pointing information. "Consolidated" telemetry (e.g. with holes in the flow due to ground transmission problems filled) and auxiliary information will be available in the MOC telemetry archive for the DPCs to downlink a couple of days after the related observations have actually taken place.

The lines from the MOC to the DPCs are currently expected to be dedicated 256 Kbps lines; this would allow all data to reach the DPCs within 8 hours, i.e. within 24 hours from their acquisition by the Ground Station. All telemetry will be piped by both DPCs, thus allowing to have a replicated raw telemetry archive. In the hypothesis of having a total of 90 Kbps of science data (30 for LFI, 60 for HFI) the volume of compressed raw telemetry would be ~1 GB/day.

3. LFI DATA PROCESSING

The tasks of the Planck DPCs have been described in detail elsewhere (Pasian & Gispert 2000, Pasian et al. 1999). To summarize, the DPCs are expected to produce calibrated time-series for all detectors, maps of the sky in the instruments frequencies, maps of the sky for diffuse astrophysical components (CMB, dust, synchrotron and free-free emission) and catalogs of galactic and extragalactic sources.

While the first two data products are the responsibility of each individual instrument DPC, the other two rely on the merging of data and analysis from both DPCs. A prototype Integrated Data and Information System (IDIS) is used (Bennett et al. 2000) to allow inter-Consortia and intra-Consortium information and data exchange.

When data reach the LFI DPC, routine telemetry checks will be performed by the Real-Time Assessment (RTA) and Quick-Look Analysis (QLA) systems on housekeeping and science telemetry, respectively. The purpose is to identify or predict anomalous behaviour, or to monitor in time the response of the detectors. Correspondingly, requests will be sent to the MOC to change the instrumental set-up. Besides this routine checking-feedback activity, three basic processing steps can be defined:

– telemetry processing: packets received from MOC are opened, decompressed, and organised as timelines containing sequences of housekeeping or science data (TOI – Time-Ordered Information);

– data calibration and frequency maps making: the raw timelines are used to reconstruct, through an iterative process, calibrated scans per each detector, as well as instrumental performances and properties, and maps of the sky for each channel;

– separation of components: the frequency maps produced by both instruments are transformed into maps of the underlying astrophysical components; data from both HFI and LFI need to be analyzed jointly to reach the final expected result.

The above processing steps define three logical levels (1–3), which for the LFI will be performed in Trieste supported by OAT and SISSA. A Level 4 in common between the HFI and LFI Consortia and supported by MPA (Garching) can additionally be defined, and will collect, archive, and prepare the release of all the public material at the end of the Planck mission.

A rough estimation of the amount of data to be handled by the LFI DPC is reported in the following. The raw compressed telemetry archive ingests 1 GB/day, of which 330 MB are LFI science data. Once decompressed, transformed from bytes into floating-point data, and adding pointing information from MOC auxiliary data, 6 GB/day of raw science TOI are produced. Level 2 processing results in the production of 6 GB/day of calibrated science TOI. In total, telemetry plus LFI TOI totals 13 GB/day. Considering that being maps ~120 MB each, thus almost negligible, the total amount of LFI data at the DPC is expected to be ~4 TB over the 2 nominal full-sky surveys. In addition to this, HFI calibrated TOI and maps are needed to perform component separation (another ~4 TB), thus totaling ~8 TB of storage for the whole mission. In case of a requirement for storing data products from intermediate processing steps, additional storage would be required. Part of this storage could be kept off-line.

Most of the above considerations refer to Level 2 processing. Level 3 requirements are almost negligible in terms of storage, while they demand careful planning of computing resources.

Some of the processing tasks to be performed by the DPC and for the scientific analysis of the data are quite expensive from the computational point of view. This is due to the fact that general algorithms for solving such problems perform matrix operations globally on the whole data set (time-series or maps), yielding $O(M^3)$ complexity (Bond et al. 1999). Examples are map-making, the CMB data maximum likelihood analysis and the separation of components, where simplifications, approximations and possibly the use of a new class of algorithms will be needed. As an example, a preliminary approach to components separation using Independent Component Analysis (ICA) provides interesting results (Baccigalupi et al. 2000) while being much less computationally intensive.

To solve the storage and computational challenges mentioned above, and to cope with different requirements, the DPC needs to be structured as follows: a subsystem having data handling capabilities (CPU, RAM, on-line storage and archive, all substantially conventional) to perform interaction with the MOC, storage of instrument parameters in a proper database, calibration of scientific time lines; a subsystem with a high degree of parallelism to perform creation of maps, handling of noise correlation matrices, and possibly additional scientific analysis; efficient network connections with the various levels of the HFI DPC and with possible LFI secondary sites (e.g. in the US) procuring if necessary dedicated links.

4. DATA HANDLING BEFORE OPERATIONS

Already during the studies assessing the feasibility of the project (Bersanelli et al. 1996), checking instrument behavior through simulated data has been essential. From a model of the sky including CMB and foregrounds, a noise model, a beam model and a scanning strategy, simulated output from LFI is obtained using state-of-the-art numerical methods. The purpose of this exercise is two-fold: in an initial phase, to assess the impact of technical solutions on the scientific objectives of the mission; subsequently, to define prototype data reduction algorithms, to be eventually integrated in the DPC data reduction pipeline. Some of the modeling and simulation activity (the one mainly dealing with spacecraft and scanning strategy aspects) is being shared with the HFI Consortium; instrument-dependent peculiarities are handled by a dedicated LFI team.

The routine generation of simulated LFI data is considered to be a pre-operations DPC activity, carried out by a Level S running at the main DPC site in Trieste. In principle, Level S could produce a whole set of different Planck/LFI complete surveys, differing by some specific parameter representing a specific LFI setting. This implies building for Level S a system having appropriate computational power and the data handling capabilities of the complete DPC.

An additional pre-launch activity involving data management is related to ground-based instrument tests. The amount of data produced will be small compared with the actual operation data, but with high throughput, since no transmission limitation will be in place and the full LFI data rate will be tested. The DPC will provide storage for the data produced during the tests, and will build an initial version of the RTA and QLA software based on the ESA standard system SCOS-2000. The developed RTA and QLA will be re-used for the DPC Levels 1 and 2, while this activity will give a chance of testing processing algorithms on "real" data.

ACKNOWLEDGMENTS. We wish to acknowledge J. Tauber, J. Dodsworth, P. Estaria of ESA, J-F. Sygnet, C. Mercier and F. Bouchet of the HFI Consortium, and all LFI Consortium members, on whose behalf this presentation was made. A grateful thought is dedicated to the memory of our colleague and friend R. Gispert.

REFERENCES

Baccigalupi C. et al. 2000, MNRAS, in press

Bennett K. et al. 2000, in *Advanced Global Communications Technologies for Astronomy*, eds. R. I. Kibrick & A. Wallander, Proc. SPIE, Vol. 4011, p. 2

Bersanelli M. et al. 1996, ESA Report D/SCI(96)3

Bersanelli M., Mandolesi N. 2000, Astroph. Lett. Comm., 37, 171

Bond J. R., Crittenden R. G., Jaffe A. H., Knox L. 1999, Computing in Science and Engineering, Vol. 1, No. 2, p. 21–35

Maris M., Maino D., Burigana C., Pasian F. 2000, A&AS, in press

Pasian F., Gispert R. 2000, Astroph. Lett. Comm., 37, 247

Pasian F., Mandolesi N., Bersanelli M. 1999, SAIt AstroTech Journal, Vol. 2, No. 2