Beyond PLANCK

Deliverable 2.1:

Flagging module

Authors Samuele Galeotta
Daniele Tavagnacco
Andrea Zacchei

Date May 31st, 2018

Work Package WP2 - Data Selection and Flagging

DocId [INAF_BP02-01-1.0]

Revision History

Version	Authors	Date	Changes
1.0	Samuele Galeotta Daniele Tavagnacco Andrea Zacchei	May 31st, 2018	Initial Version

Contents

1 Introduction

The main goal of the BeyondPlanck project is to build an end-to-end Gibbs sampler for the Planck LFI data, and use it to improve the overall calibration and fidelity of the final LFI sky maps.

Working Package 2, *Data Selection and Flagging*, serves as starting point to select at the *timelines* level, which data to use in the Gibbs sampler. This working package is responsible to flag the data that should be excluded from analysis, according to a predefined criterion such as: maneuver periods, gain changes in the data acquisition electronics that caused saturation, abrupt changes in voltage outputs caused by gain fluctuations, etc. The main objective of this WP is to provide to the subsequent WPs input data clean from instrumental effects.

Working Package 2, *Data Selection and Flagging*, is composed by two modules:

- *Data extraction and anomalies flagging,* responsible for the extraction of the data timelines from the Planck LFI DPC database into a portable data format, and the flagging of anomalous data and planets transits in the observations.
- *Data selection,* in charge of preparing data as input to the Gibbs sampler, transforming raw data in engineering value and computing pointing information.

2 Data extraction and anomalies

In the Planck project data are preserved at LFI DPC with a predefined structure inside many Oracle databases organized as follows:

- *Level1 db* containing the raw data reformatted from telemetry packets into time ordered streams.
- *Level2 db* containing time ordered data streams in engineering value before and after calibration process: this will be the input to the Gibbs sampler.
- *Test db* containing all the output of the pipeline tests.

This structure is not easily portable because is highly dependent by the database engine (Oracle). For this reason, a different approach is needed for the scope of BeyondPlanck.
After some analysis, a file based approach resulted to be more portable. In this framework, the first module of Data Selection and Flagging will extract all the *Level1* raw data into HDF5 format files. This file format is well known in the Astronomical community, is free and has I/O function and graphical interfaces already available.
The document describing the contents of each HDF5 file typology is in preparation.
During the data extraction procedure, all the anomalies in the data and Solar system planets observations are identified and flagged according to the predefined bit mask structure inherited from Planck definition.
Figure 1 shows an example of a HDF5 extracted file. The extraction software, in its draft form, is available in the BeyondPlanck GitLab repository, even if the module is strongly dependent from the Trieste Planck DPC software environment and its interface to the Oracle databases.

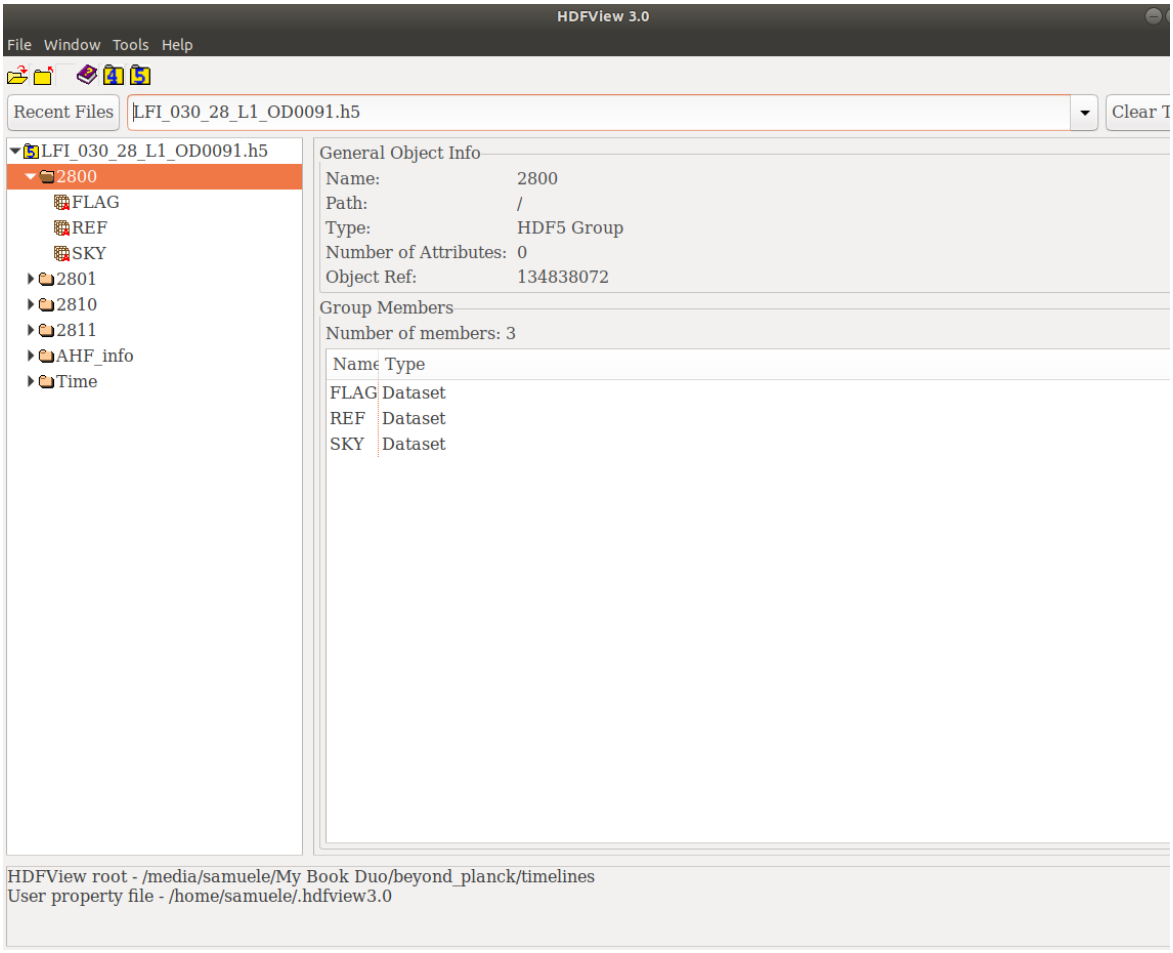

Figure 1: HDF5 Raw Level1 data

3 Data selection

The *data selection* module uses the Level1 raw data as input and apply an algorithm for the removal of all the known instrument systematics directly from the data streams. The resulting clean datasets are available for the calibration work package (WP 3).
In Figure 2 an overview of the *data selection* module is reported.

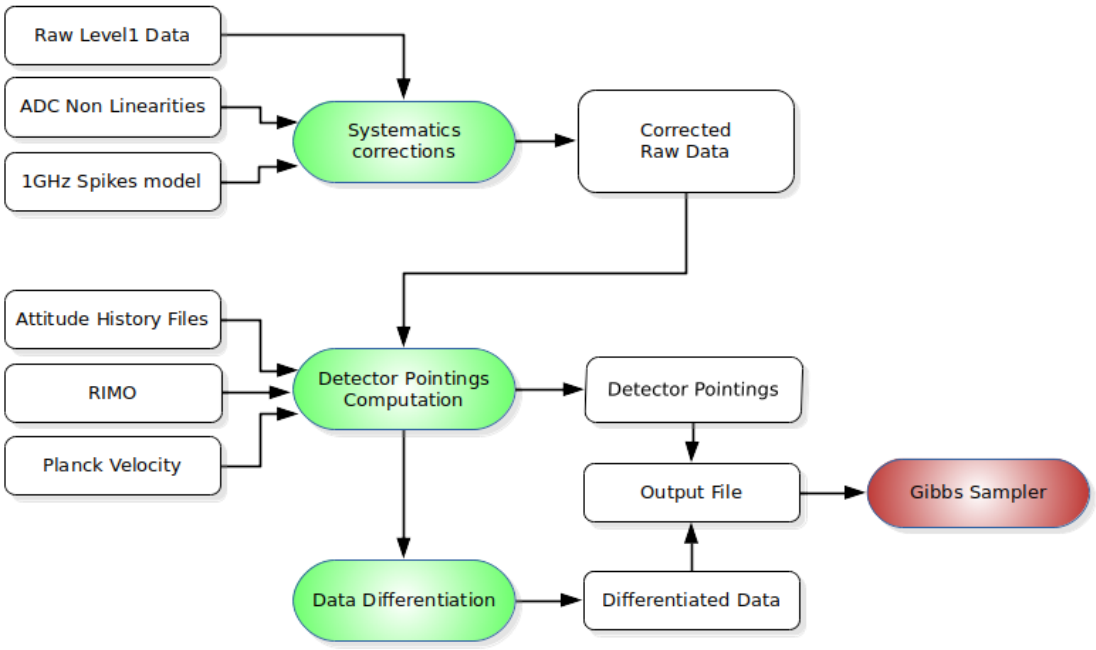

Figure 2: Data selection overview

The *Data* Selection pipeline is divided into three steps:

1. *Systematics correction*: this step transforms raw *Level1* input data to engineering value data using the instrument models built during the Planck project and correct the data for the known instrument systematic effects.
2. *Detector pointing computation*: using the satellite attitude information, the instrument model and the Planck velocity file, this step computes detector pointing associated to each acquired sample.
3. *Data Differentiation*: this step is necessary to reduce the 1/f instrument noise by combining the data signal coming from the sky and the corresponding data signal coming from a reference source in the Planck satellite, see Planck Collaboration II.

2014, A&A, 571, A2 and Planck Collaboration II. 2016, A&A, 594, A2 for further details.

The output of *Data Selection* pipeline is a set of HDF5 files each one containing both detector pointing and differentiated data. The file format is still in definition and tests are ongoing to define the optimal configuration. The document describing the contents of each HDF5 file typology is in preparation.
The *Data Selection* software prototype is available in the BeyondPlanck GitLab repository and it can be executed with the following command:

```
mpirun -n N git_repository/calib/cmake.build/DataSelection param.txt
```

where *N* is the number of processing nodes required and *param.txt* is the parameter file. A README file is available in the repository with the description of the keywords used in the parameter file.
The software is completely written in C++ language and uses most of the algorithms developed during the Planck project. The base algorithms have been extended by adding specific classes to handle FITS and HDF5 files to provide a portable and computing environment independent code.