The eLISA/NGO Data Processing Centre

V. Beckmann, A. Petiteau, E.K. Porter, G. Auger, E. Plagnol, P. Binétruy

9^{th} *LISA Symposium, Paris*
ASP Conference Series, Vol. 467
G. Auger, P. Binétruy and E. Plagnol, eds.

The eLISA/NGO Data Processing Centre

Volker Beckmann, A. Petiteau, E. Porter, G. Auger, E. Plagnol, and P. Binétruy

[1]*François Arago Centre, APC, Université Paris Diderot, CNRS/IN2P3, CEA/Irfu, Observatoire de Paris, Sorbonne Paris Cité, 10 Rue Alice Domon et Léonie Duquet, 75205 Paris Cedex 13, France*

Abstract. Data analysis for the eLISA/NGO mission is going to be performed in several steps. The telemetry is unpacked and checked at ESA's Science Operations Centre (SOC). The instrument teams are providing the necessary calibration files for the SOC to process the Level 1 data. The next steps, the source identification, parameter extraction and construction of a catalogue of sources is performed at the Data Processing Centre (DPC). This includes determining the physical and astrophysical parameters of the sources and their strain time series. At the end of the processing, the produced Level 2 and Level 3 data are then transferred back to the SOC, which provides the data archive and the interface for the scientific community. The DPC is organised by the member states of the consortium. In this paper we describe a possible outline of the data processing centre, including the tasks to be performed, and the organisational structure.

1. Introduction

Preparation for the first space-based observatory for gravitational waves, the eLISA/ NGO (evolved Laser Interferometer Space Antenna / New Gravitational Wave Observatory) project (Jennrich & al. 2012), requires also to define the data analysis framework within the Satellite Ground Segment (SGS). An important aspect to keep in mind when discussing the NGO data is the fact that NGO observes the whole sky at any given time, i.e. the telescope is not pointing in any direction and its observations are truly omni-directional. Most of the sources detectable will be only visible in the data stream after months to years of observation time, and it is therefore necessary to analyse the accumulated data in order to achieve the highest sensitivity. The main measurements that will be used for extracting scientific information are the interferometric measurements performed by the phasemeter: two laser beams are combined on a photodiode and the phasemeter analyses the photodiode signal to measure the phase difference between the two incoming beams.

There are 3 phase measurements in each optical bench of the daugther spacecrafts and 4 in each mother optical bench. In addition, there are measurements of the time travel along each arm in each direction. In total, this corresponds to 14 phase measurements and 4 distance measurements. These "raw measurements" contain the necessary information for "cleaning" the phase measurements. Added to information from other on-board instruments as for example the Gravity Reference Sensor (6 channels for each test mass) and attitude of the spacecraft, they constitute the data level L0.

In the phase measurements between incoming and local laser beam, the gravitational wave signal is overlaid by laser noise and clock noise which is several orders of magnitude stronger than the signal we are interested in. But the noise can be reduced by the time delay interferometry (TDI) technique. The main idea is based on the fact that each noise is present in at least two phase measurements. By combining these measurements using appropriate delays, this noise can be identified and reduced.

The TDI output signal contains all gravitational wave information and constitutes with spacecrafts ephemerids (and other potentially relevant information) the data level L1. All gravitational wave information on astrophysical and cosmological sources are extracted from this TDI signal.

2. Data products

- **Level 0: raw data**. Level 0 data are the raw data streams necessary to construct TDI channel and all relevant information about subsystems as Gravity Reference Sensor (GRS) and Interferometry Measurement System (IMS). The L0 data contain all phasemeter measurements, all data from GRS and the complete science and payload housekeeping data. Level 0 data are those downlinked from the spacecraft to the ground station. For a 2 year mission, the total volume of Level 0 data is about 135 Gbyte (330 Gbyte if the duration is extended to 5 years).
- **Level 1: calibrated data**. Level 1 data are the fully processed data streams, including spacecraft ephemerids, that are needed to isolate individual gravitational wave signals. In addition, they include software and models used for producing Level 1 data from Level 0 data, in particular the relevant time shifts, the GRS dynamical model, and other information concerning potential change in effective optical path length. Level 1 data are produced by the ESA NGO Science Operations Centre with support of the instrument teams. The volume of Level 1 data at the end of the mission is about 7 Gbyte (2.4 Gbyte for TDI channel plus 4.2 Gbyte for spacecraft ephemerids).
- **Level 2: reconstructed data**. Level 2 data are the reconstructed and noise reduced TDI measurements and include the results of source identification and parameter extraction. L2 data consist of the model parameters and their posterior probability density function of identified gravitational wave sources and other identifiable events. They include reconstructed waveforms and associated detector signals used for identification. They also include regular update on critical parameters for transients like the time at coalescence and sky position.
- **Level 3: source catalogue**. The Level 3 data are the result of scientific evaluation. L3 data consist of a catalogue of sources, their (astro)physical parameters and their strain time series (gravitational wave amplitude as a function of time). L3 data also include “residual” data stream, i.e. L1 data stream with the contribution of identified sources removed (the residual data contains only instrument and gravitational wave noise). L3 data will be regularly updated during the mission.

A safe upper limit for the total data volume of the entire eLISA/NGO mission up to the Level 3 data products is that it will not exceed 10 Tbytes.

3. Ground Segment and Data Analysis Organisation

The structure proposed by ESA for the ground segment is shown in Figure 1. Here, the ground stations and the Mission Operations Centre (MOC) as well as the Science Operations Centre (SOC) are under the responsibility of ESA. The Instruments Operations Teams (IOTs) and the Data Processing Centre (DPC) are under the responsibility of the member states of the NGO consortium.

The data processing is partly under the responsibility of ESA which operates the Science Operations Centre and which does the processing from the Level 0 to the Level 1 data and which is in charge of the archive and distribution of data products and event notices. The Instrument Operations Teams, under the responsibility of the member states of the NGO consortium, provides the calibration files for the L0 to L1 processing. The Data Processing Centre, also operated by the member states, provides the L1 to L3 processing and the software which can be used by the community in order to perform independent scientific data analysis.

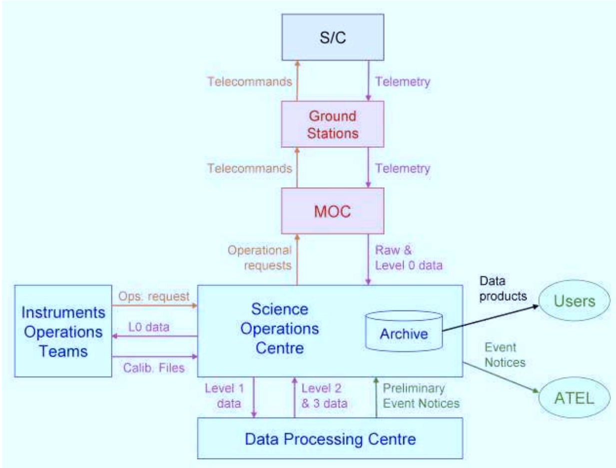

Figure 1. The organisational structure of the eLISA/NGO Satellite Ground Segment. Ground stations, mission operations centre (MOC) and science operations centre (SOC) are managed by the European Space Agency (ESA). The instrument operations teams (IOTs) and the data processing centre are under the responsibility of the eLISA/NGO consortium and under management of the member states.

4. The Data Processing Centre

In the following a French solution for the Data Processing Centre is presented. It has to be noted that contributions by other NGO consortium member states to the Data Processing Centre are possible, either through provision of FTEs to the DPC, or through execution of certain tasks of the DPC in other member states. We first outline the data processing strategy in the next subsection and then describe the structure of the Data Processing Centre.

4.1. Processing Strategy

Standard Analysis Pipeline. The main goal of this pipeline will be to extract as many sources, with as much astrophysical information as possible from the data. The output of this pipeline will be the estimation of source parameters with statistically interesting confidence intervals, posterior density functions for each parameter, and the time series best fit template with its associated correlation level with the data. This information will be used to construct the source catalogue defined in the L3 data.

The pipeline will run continuously maintaining list of sources and their best estimate parameters corresponding to the best estimate. When a new block of data is available, it is integrated into the search and both the search for new sources and the refinement of existing sources will be conducted. It is conceivable, as is currently done in the ground-based community, that this pipeline will apply many different waveform models and search algorithms. The cross-correlation of results from different waveforms and algorithms will then provide the confidence level for a detection. The standard analysis is also able to predict possible transients and can thus trigger event notices which are communicated to the SOC. This pipeline may be updated during the mission lifetime (adhering to a certain standard to be pre-determined by the community) and will exist until the first agreed version of the source catalogue and the L3 data is compiled.

Quick-Look Analysis Pipeline. The main goal of this pipeline is to perform a very fast search in order to extract as quickly as possible transient events such as black hole binary mergers, as well as the known verification binaries. In the case of the verification binaries, a quick detection of these sources will also serve to calibrate the detector and ensure the proper functionality of the instrument. This is going to be done in close collaboration with the instrument teams. It is conceivable that after any interruption in the data, the quick search pipeline will once again search for the verifications binaries to assure continuity of the data stream.

In the case of transient events, this is particularly important for predicting when the most interesting GW emissions will occur and therefore plan an extended Ground Station communication during that time. The information gained from the quick search pipeline will also be particularly important in deciding on the duration of protected periods where no ground or spacecraft maintenance will occur in order to maximise the detection of a transient source. Notice of transient event are published (and updated) through standard astronomy alert services in order to make joint observations with other observatories.

4.2. Data Centre Structure

The management of the DPC focuses on the main tasks to be performed by the centre, i.e. on providing reliable pipelines which integrate the algorithms provided by the eLISA/NGO consortium, and their timely execution and provision of results at the L2 and L3 level to the eLISA/NGO consortium, and software distribution and support. i.e., the focus is on the data processing, not on the scientific kernel of the pipelines, nor on the scientific interpretation of results. Nevertheless, the management of the DPC needs a strong background in eLISA/NGO science in order to understand the required tasks.

The management of the DPC consists of two positions: The project scientist (PS) has a strong background in eLISA/NGO science and provides thus the main link to the eLISA/NGO consortium. The PS engages in the discussion with the eLISA/NGO consortium about the algorithms to be implemented and establishes a strong link to the corresponding partners at the Instrument Operations Team and at the Science Operations Centre. The project manager (PM) is in charge of the technical implementation of the DPC and of organizing the tasks of software implementation, s/w management, quality control, and pipeline execution. Both, project scientist and project manager, begin their work at and for the data processing centre 6 years before launch (L-6yr), i.e. one year before the first s/w engineering activities commence.

The standard analysis (SA) pipeline receives Level 1 data from the SOC and processes them up to Level 3. The results are transmitted back to the SOC. The main task at the DPC are the design of the pipeline, integration of the algorithms provided by the eLISA/NGO community, testing and bug-fixing. Software engineering activities for the SA start 5 years before launch with 3 engineers. Each of the 3 engineers is given a closer attachment to one of the following scientific topics: Galactic Binaries, stochastic background, Super Massive Black Hole Binaries, and EMRIs. Thus, each of the engineers is especially knowledgeable about the subtleties of the given scientific topic in order to optimize the algorithm implementation. Two years before launch additional 2 engineers join the team in order to help with the documentation and bug-fixing effort. A system administrator is in place to support the activities starting 5 years before launch. After launch, the weekly processing of the SA is going to put high pressure on the system, especially considering the high processing demand by the SA pipeline. Thus, as second system administrator will start at this moment.

The quick-look analysis pipeline presents a simplified version of the standard analysis pipeline. Therefore, coding effort is relatively light with respect to the SA. An important aspect is though the visualization of the incoming data, i.e. the possibility to verify completeness of data, available data, access to house-keeping data. Another important part of the work here is also the close collaboration with the IOTs. One Full-Time Equivalent (FTE) engineering is starting 5 years before launch. Another 0.5 FTE are added 2 years before launch in order to thoroughly test the quick look pipeline and to improve the visualization tools. With launch, another engineer is hired in order to ensure a 5 days/week coverage of the quick look analysis activities. 2 FTE are included in order to monitor the data stream and the processing in near-real time. The system administration for the QA pipeline is provided through the system engineer of the SA pipeline described in the previous section.

While strict version control and quality control is in place from the beginning of the activities at the DPC, especially toward launch (and beyond) a high demand for testing, problem reports, version management, platform testing, etc. will occur. Two engineers are starting their work 2 years before launch in order to cope with these issues. The quality control engineer is going to maintain also the data base for the software version handling.

The software of the Standard Analysis (SA) pipeline (but not the Quick look analysis chain) is made publicly available to the eLISA/NGO community. Together with dedicated 0.5 FTE the software integration team shares the activities of the user support, as they are the most knowledgeable concerning the s/w structure and specific problems in the processing. It has to be pointed out that the expertise concerning the scientific kernel of the SA pipeline remains with the eLISA/NGO consortium.

References

Jennrich, O., & al. 2012, NGO Assessment Study Report, ESA