*Highlights of Spanish Astrophysics VI, Proceedings of the IX Scientific Meeting of the Spanish Astronomical Society held on September 13 - 17, 2010, in Madrid, Spain. M. R. Zapatero Osorio et al. (eds.)*

# Reduction of the Gaia astrometric data

**C. Fabricius[1,2], J. Torra[1,2,3], J. Portell[1,2,3], and J. Castañeda[2,3]**

[1] Institut d'Estudis Espacials de Catalunya
[2] Institut de Ciències del Cosmos de la Universitat de Barcelona
[3] DAM, University of Barcelona, Martí i Franquès 1, E-08028 Barcelona

## Abstract

In two years, the European Space Agency (ESA) will launch the Gaia mission. Gaia will observe all point-like sources in the sky between magnitude 6 and 20, a total of more than a thousand million stars and planets. For the brightest sources we will obtain parallaxes and yearly proper motions with a precision of 10 microarcseconds, while the precision will degrade to 300 microarcseconds for the faintest stars. The reduction of the astrometric data is a very complex and extensive task, which only can be undertaken in a collaboration between many groups within the ESA member states. We describe the role of Spain and of the national supercomputing centre (CNS-BSC) in this work, and describe some of the most critical points in modelling the instrument and in the observations. We describe the initial data treatment which must be carried out daily during the mission, and the process of repeated improvement of the intermediate data which is carried out several times, before reaching the final values. The fulfilment of these tasks, especially the latter, is a technological challenge for the mission.

## 1 Introduction

Gaia is an ESA mission which will make an astrometric survey of the sky between magnitudes 6 and 20, in order to improve our understanding of the Galaxy. The reason for doing this from space is to avoid the disturbance of the Earth's atmosphere on the measurements and to keep the instruments in very high thermal stability.

Gaia builds on the experience with the very successful Hipparcos mission, but is far more ambitious. Hipparcos observed about 120 000 pre-selected stars, obtaining parallaxes and yearly proper motions with an accuracy of 1–2 milliarcseconds. This was supplemented with a lower precision survey from the Tycho instrument, giving positions and two colour photometry for 2.5 million stars down to 11–12 mag. Since the publication in 1997, the

Hipparcos and Tycho catalogues have had a massive impact on Galactic astronomy and have formed the reference for the various optical sky surveys.

Gaia, coming 25 years after Hipparcos, will survey the sky to magnitude 20, i.e. more than a thousand million stars, and is aiming for a precision for parallaxes and yearly proper motions better than 10 microarcseconds for stars brighter than 10 mag, and below 25 microarcseconds at 15 mag. For the faintest stars the goal is 300 microarcseconds. In addition, Gaia will do low resolution spectro-photometry for all the sources, and radial velocities for sources brighter than 16–17 mag.

Gaia will be launched at the end of 2012, and will go to the vicinity of the Lagrangian point L2 in order to have a stable environment. It will start routine observations at mid 2013 and work for 5 or 6 years. When it has finished, a few more years will be needed before the final data release, but before then, there will be several preliminary data releases, the first one maybe as soon as one year after the start of observations.

While ESA will launch and operate Gaia, the data processing will be carried out by a Data Processing and Analysis Consortium (DPAC), involving universities and research institutions in almost all ESA member states. DPAC has more than 300 members, of which almost 40 are based in Spain apart from the some 25 ESA employees at ESAC in Villafranca del Castillo.

The Spanish groups are involved in many aspects of the Gaia mission, cf. [2], like simulations, photometric data processing, cf. [1], stellar classification, variability, etc., but we will here only consider the astrometric data reduction.

## 2 The observational principles

Gaia is designed to do global astrometry, and like Hipparcos it therefore has two fields of view combined in a common focal plane. Gaia will be rotating once every 6 hours and the two fields of view will scan a band around the great circle defined by the axis of rotation. The fields are separated 106.5 degrees, and Gaia is thereby continuously connecting areas on the sky separated by this angle. The rotation axis is slowly precessing around the direction to the Sun and always at a distance of 45 degrees. The whole sky is thereby covered many times and scanned in many different directions. During the five years, a source will on average transit a field of view 80 times. In each transit the overlapping field will be a different one, and as a result it is in fact possible to build a rigid astrometric system on the sky. Only the overall orientation and rotation are in principle undetermined, and must be settled using the Gaia observations of extra-galactic sources, primarily quasars.

The focal plane, cf. Fig. 1, includes 14 sky mapper CCDs, 62 astrometric CCDs, 14 spectro-photometry CCDs, and 12 CCDs for radial velocities. The lower part of the figure shows Gaia with Solar panels deployed (Sun shining from below); the 3-meter diameter torus carrying the mirrors and the focal plane assembly; and finally a CCD with its charge injection structure at the leading (left) edge, gate structures to shorten the exposure time for bright sources, and the readout register at the trailing (right) edge.

Transitting stars will follow almost horizontal lines and be observed in one of the seven

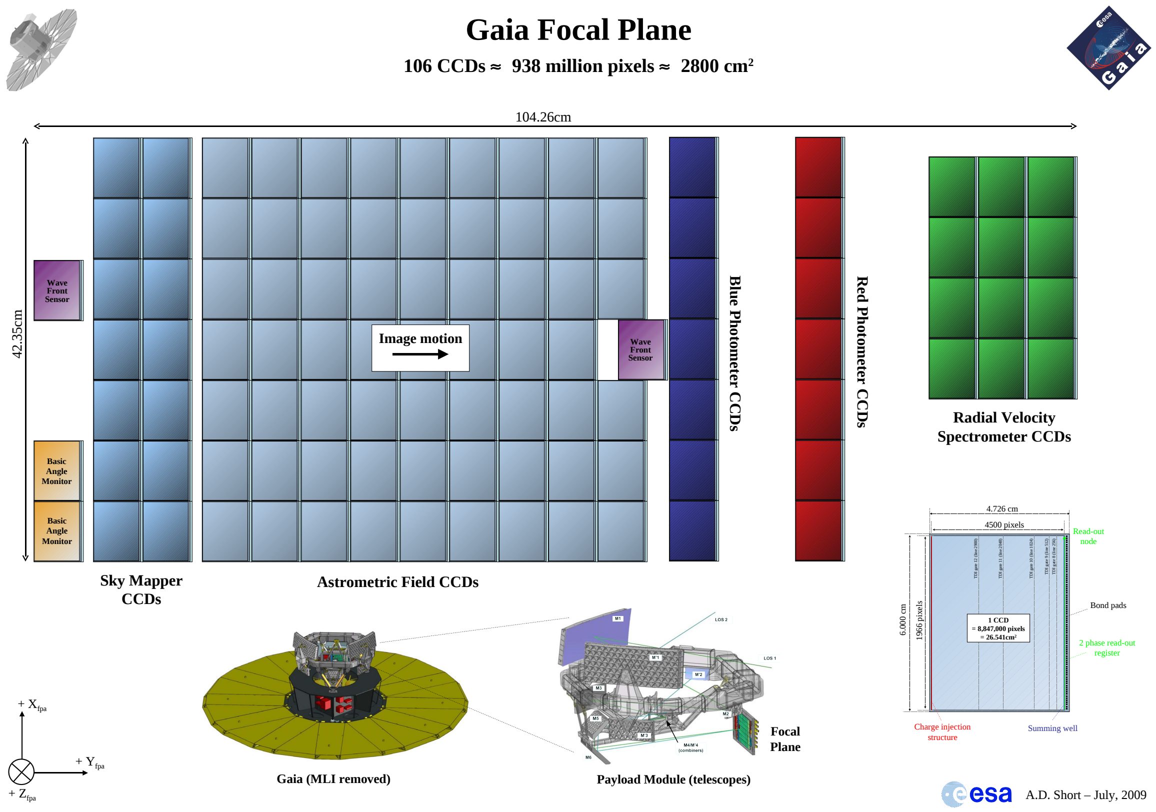


Figure 1: The Gaia focal plane. See text for details.

rows. As they enter the field, the stars are detected by the Sky Mapper CCDs to the left, where one set of CCDs (leftmost) see only the preceding field of view, and the other set of CCDs only the following field. The next 9 CCDs are used for astrometry, the following two for photometry, and the last three, where present, are used for the radial velocities. The first astrometric CCD has the additional functionality of censoring false detections, e.g. due to cosmic rays. If the detection is not confirmed, the observation is abandoned, thereby saving valuable telemetry capacity.

The images cross a CCD in 4.5 seconds, and all CCDs are read out continuously, at the same speed as the image motion, giving about 2 million pixels per second in each CCD. Reading at such a speed would, however, give an unacceptably large data set and an unacceptably high readout noise. The full set of pixels is therefore not read, but only small areas around the predicted position for the detected sources, while remaining pixels are rapidly *flushed*. Normally, the pixels of these small areas are collapsed during the readout process itself, reducing further the readout noise and leaving us with one dimensional measurements. The important information is contained in the along scan direction, so this is in principle quite acceptable, and in fact similar to the situation in Hipparcos, where all measurements were also one dimensional.

Fundamental for the astrometry is now that we can determine the transit time for each source at each CCD from these typically 6–12 samples, and with a sufficient accuracy to reach the mission goals. In the brightest end we aim for 7 microarcseconds, and comparing this with the pixel size in the scan direction of 59 milliarcseconds and with the number of CCDs and transits, it is clear that we need to understand all instrumental effects at a level of a couple of thousands of a pixel.

# 3 How to reach microarcseconds

Trying to reach microarseconds, and trying to reach a few thousands of a pixel for image positions, is clearly a challenge. One thing is to have favourable photon statistics, and quite another thing to have a sufficiently stable instrument, and to be able to model all instrument effects at that level. What in the end will make this possible, is the huge amount of observations of suitable calibration sources.

Many effects must be addressed. The temperature must be extremely stable in order for the optical configuration to remain constant. This will be achieved partly through the stable environment at L2, and partly through a very careful management of the electric power spent on-board. Even so, the basic angle between the two fields of view is monitored constantly with a special laser device.

Another concern is impacts from micro-meteorites, or tiny displacements of the solar panels, that may disturb the smooth rotation around the spin axis.

We will, however, not here enter into the full range of issues to calibrate or all the actions being taken to mitigate them, but focus on problems directly related to the image shape. The PSF itself depends on the colour (or rather SED), the field of view, and the location in the focal plane. A detailed calibration of the PSF is mandatory, and this is one of the motivations for including spectro-photometry in the Gaia instrument. However, one thing is the optical PSF, and another how charge builds up on the CCD, how it is transferred down the columns, how it passes the readout register, and how it is finally converted into a digital reading.

In a real world CCD, charge transfer is not perfect. A particular problem is radiation damage, which can show up as so-called *traps* which may keep a passing electron captive for a while before releasing it again, and thereby distort the charge image. In Gaia, one way of mitigating the effects of the traps is through frequent injections of charge at the leading edges of the CCDs. The price of these injections is of course that no data can be acquired immediately after such an injection. The amount of distortion to the images depends in a complex way on the brightness of the signal, on recently passing images, etc., apparently killing any hopes of accurate astrometry. At the same time, the amount of damage will increase during the mission and depends critically on the Solar activity. For this reason detailed studies are in progress, including numerous experiments in the laboratory, to fully understand the details of the distortion. Preliminary results are promising, but even if they should not be perfect, the good thing is that as the scan directions vary on the sky, the residual effect will have wildly varying influence on the astrometric coordinates, and may be corrected with an empirical calibration.

When the charge has reached the readout register, another distortion is introduced during the serial transfer, depending on the distance from the readout node and on the brightness. This effect must also be understood in detail, but is less serious as the principal measuring direction is along the scan. Finally, the bias level has a dependence on the number of pixels that have just been flushed, and on other details of the readout, adding the necessity to know the detailed readout history of every sample. Adding issues like saturation, non-linearities in the amplifier, and variations in sensitivity among the pixels, it is clear that a very detailed instrument model is needed, that the calibration task will be considerable, and the computational task overwhelming.

## 4 The astrometric processing chain

For practical reasons, the astrometric processing has been subdivided into a number of individual processes, e.g. analysing the raw CCD data, reconstructing the satellite attitude, calculating astrometric parameters, calibrations, etc. A main division line has been put between the analysis of the CCD data, giving fluxes and positions in the CCDs, and the further analysis of these intermediate data.

### 4.1 IDT: the Initial Data Treatment

The Initial Data Treatment, IDT, is a process developed by the University of Barcelona, with contributions from partners in several other countries. It will run daily during the mission at the science operations centre at ESAC. It takes care of all the data that arrived over the last day, converts and combines the telemetry packets into a user friendly form. It does a first reconstruction of the attitude, a first identification of the observed sources, and a first processing of the astrometric observations, getting image parameters, i.e. the positions and fluxes. The quality must be sufficiently good to allow a detailed check of the performance and health status of the instruments. In other words, IDT sets the grounds for the more detailed processing of astrometry, photometry, and spectroscopy that follows.

### 4.2 AGIS: the Astrometric Global Iterative Solution

The Gaia astrometry is derived in the Astrometric Global Iterative Solution, AGIS, which also runs at ESAC. It iteratively improves the astrometric solution, various calibration parameters, and the satellite attitude. It does not see the CCD images and therefore depends entirely on the image parameters provided by IDT and later updated by IDU (Sect. 4.3). AGIS is developed by the University of Lund in collaboration with the ESAC staff.

### 4.3 IDU: the Intermediate Data Updating

The other Spanish contribution to the Gaia astrometry is the so-called Intermediate Data Updating, IDU. It revisits the CCD image data, reconsiders the identification of the observations, and looks after all the image calibrations. It is, in short, responsible for achieving the highest possible precision of Gaia. IDU is developed by the University of Barcelona with contributions from partners in other countries. It will run at the national super-computing facility Mare Nostrum (Barcelona Supercomputing Centre - *Centro Nacional de Supercomputación*). IDU forms an iterative loop with AGIS and with the photometric process, and this iteration will have to continue after the end of the mission, to reach the final convergence.

The first IDU task is to reconsider the identifications of the observations. It is in principle simple, and the main challenge is to manage some $10^{11}$ observations and more than $10^9$ source catalogue entries in a consistent and efficient manner.

The second major task is the image calibrations. This includes determining the effective PSF as a function of colour and as a function of location in the focal plane. This may at first seem relatively simple, but a number of non-linear effects in the CCDs, as discussed in Sect. 3, make the detailed calibrations highly complex. Input is here needed from AGIS to determine chromatic displacements, and from the photometric process to know the precise colours.

The final task is to recompute the image parameters for each of the nearly $10^{12}$ CCD sub-images. It is here necessary to know the detailed history of each CCD, not only the illumination history, but also the detailed readout history of each sample. The illumination history is needed for knowing in what state are the traps in the CCD, and thereby the amount of distortion an image of the brightness in question may suffer. The detailed readout history is needed because the electronic bias level changes from sample to sample as a function of the precise locations of the CCD sub-images that were read.

### 4.4 Complex cases

The full iteration loop works best for isolated stars, while other sources (double or multiple stars, solar system objects, galaxies, etc) will need a dedicated treatment. There are also plans for using the data to look for fainter components around each star, but this may prove so demanding, that it is only done if a source appears disturbed. These processes will run at CNES in Toulouse, and at the University of Cambridge.

## Acknowledgments

This work was supported by the MICINN (Spanish Ministry of Science and Innovation) - FEDER through grant AYA2009-14648-C02-01 and CONSOLIDER CSD2007-00050.

## References

[1] Jordi, C., Carrasco, J. M., Fabricius, C., & Voss, H. 2010, in *Highlights of Spanish Astrophysics V*, eds. J. M. Diego et al., 147

[2] Torra, J. 2007, in *Highlights of Spanish Astrophysics IV*, eds. F. Figueras et al., 255