THE GAIA ON-BOARD SCIENTIFIC DATA HANDLING

Frédéric Arenou[1]**, Carine Babusiaux**[1,2]**, Fabien Chéreau**[1]**, Shan Mignot**[1]

[1]Observatoire de Paris, GEPI / CNRS UMR 8111, 92195 Meudon cedex, France
[2]Institut d'Astronomie et d'Astrophysique, U.L.B., CP 226, Boulevard du Triomphe, 1050 Bruxelles, Belgium

ABSTRACT

Because Gaia will perform a continuous all-sky survey at a medium (Spectro) or very high (Astro) angular resolution, the on-board processing needs to cope with a high variety of objects and densities which calls for generic and adaptive algorithms at the detection level, but not only. Consequently, the Pyxis scientific algorithms developed for the on-board data handling cover a large range of application: detection and confirmation of astronomical objects, background sky estimation, classification of detected objects, Near-Earth Objects on-board detection, and window selection and positioning. Very dense fields, where the real-time computing requirements should remain within fixed bounds, are particularly challenging. Another constraint stems from the limited telemetry bandwidth and an additional compromise has to be found between scientific requirements and constraints in terms of the mass, volume and power budgets of the satellite. The rationale for the on-board data handling procedure is described here, together with the developed algorithms, the main issues and the expected scientific performances in the Astro and Spectro instruments.

Key words: Gaia; Techniques: Image Processing; Methods: Data Analysis.

1. INTRODUCTION

The motivation of this paper is to describe from the founding principles how the Gaia on-board scientific data processing is handled. After having introduced in this section the main scientific objectives, the technical constraints which are faced lead to adapted technical (hardware and software) solutions whose current status and results are indicated.

The main scientific objective of the Gaia mission is to obtain a complete survey down to $G < 20$ magnitude. The downloaded data should allow for determining unbiased astrometry and photometry on-ground for all objects, whether bright, multiple stars, or asteroids, even in high density regions. Beside completeness, the on-board data handling should thus be versatile, but also efficient and reliable. Efficiency is needed because what is observed but not downloaded is lost. Reliability is mandatory as the on-board data is used to maintain the satellite's attitude, needed to guarantee the high astrometry requirement standards. Of course all these criteria may appear as wishful thinking when the actual sky and instrument properties are taken in account, if no special processing were applied.

2. TECHNICAL CONSTRAINTS

2.1. General Processing

The absence of any all-sky high resolution Catalogue with the same passband as the astrometric instrument of Gaia is in itself a motivation for autonomous on-board detection and data handling. Uploading a Catalogue would thus not be possible or at least suboptimal and at the cost of complicating the on-board processing.

There are a number of other reasons however. The first, as already mentioned, is the need to perfectly synchronise the TDI (time-delayed integration) with the scanning motion of the satellite, which otherwise would cause a blurring detrimental for the astrometric measurements. For the purpose of controlling the attitude along-scan and across-scan motion measurements must be performed, which in turns calls for searching for the stars' *centroids* in some CCDs.

Besides, the whole content of all CCDs cannot be downloaded because of the telemetry bottleneck: the content of the Astrometric focal plane CCDs only would already amount to about 6000 Mbps, while the actual bandwidth reaches a few Mbps only. To give some hint on how the useful information can nevertheless be downloaded, some (illustrative only) numbers for the Astro instrument can be indicated. For more realistic numbers of the telemetry budget accounting for the various window sizes, see Lammers et al. (2005).

Let us assume that about 55 stars/s will be observed on the average in each of the CCDs. *Detecting* then *windowing* each star with say 6×12 pixels of 16 bits each, imply that 10 Mbps would be needed, already gaining a factor 600. Then, because one-dimensional measurements allow for achieving the astrometric performances, a factor 12 can be gained by a 1×12 binning[1]. Besides, this *sampling* greatly improves the signal to noise ratio. Finally, a factor > 2 can perhaps be gained with a lossless compression scheme Portell et al. (2005).

With this overall factor >15 000, Gaia is a very efficient compressing machine which permits to cope with the restricted telemetry bandwidth. But one consequence of this is the need for an on-board computer able to *detect* objects, provide *centroiding* with a high precision, then *window* the objects using some *sampling* and a *propagation* algorithm to correctly track the windows during their traversal of the focal plane because the satellite's motion gives different speeds to the two Fields of View (FoVs).

2.2. CCD Constraints

The afore-mentioned windows are built from individual samples read in the serial register of a CCD. The reading of the CCDs in TDI mode is constrained. First because the astrometric accuracy cannot be achieved without an ultra-stable internal thermal environment. For this reason, the focal plane assembly power dissipation has to be kept constant by the reading of a constant number of samples at each TDI.

A trade-off must be performed for this number of samples per line. It cannot be too high as this would increase the detection noise, yet it should be large enough to fulfill the scientific objectives. Currently, it is such that 3×10^6 (Astro) and about 10^5 (Spectro) stars per square degree can be windowed. Because Gaia will be facing, at least locally, densities larger than the quoted numbers, there is a need for a *selection* algorithm, choosing which star will be windowed.

The samples are produced by binning pixels electronically, i.e., photo-electrons are summed inside the serial register. The reading of the samples is associated to a detection noise (including the read-out noise, the dark noise, analog video chain noise, etc). To maximize the signal to noise ratio and because one dimensional measurements are sufficient for the on-ground data reduction, some *sampling* size must be defined in order to cover most of the object's flux across-scan, taking into account the optical image, the transverse motion, etc. In most of the CCDs the windows are thus made of one across-scan sample only. The AF1 CCDs are, however, an exception because the across-scan centroid of the objects must be measured and the windows are thus composed of so-called 'serial windows' made of 6 samples of 2 pixels. The associated penalty is obviously a worsened signal to noise ratio in the AF1 windows.

[1] We follow here the designation $p \times n$ to indicate a sample made of p pixels along-scan (AL) and n across-scan (AC) electronically binned. As for the CCDs, the used acronyms are ASM=Astrometric Sky Mapper, AF=Astrometric Field CCDs, SSM=Spectro instrument Sky Mapper, MBP=Medium Band Photometer CCDs. A CCD line indicates the AC pixel chain. G and GS are the white light magnitude in respectively ASM/AF and SSM.

While no special windowing problems are foreseen for single stars in non-crowded fields, the desired scientific completeness cannot be achieved for close stars for a technical reason: any given pixel cannot belong to two different samples. Partially overlapping samples are thus not allowed and one only of two (or more) such samples can be read out. Imaging such close systems can then only be achieved through a dedicated *window positioning* algorithm only.

2.3. False Detections

Statistical tests are a compromise between false positives and false negatives and this also applies to any detection algorithm. Poisson and detection noises can produce false detections which can be avoided by adopting conservative *detection* thresholds. This assumes, however, that the CCD characteristics are well-behaved. As a consequence of this, to ensure proper operation throughout the mission duration, *pre-calibration* of the read-out samples needs to be performed.

The radiation environment is more of a problem however. Beside the expected degradation in the sky mappers, more or less accounted for by the pre-calibration step, the cosmic rays and solar protons have other adverse effects. It is currently estimated that in 25% of time, 90% of detected objects in average stellar densities will be cosmic rays, not stars. If not detected as such, particle impacts would increase the on-board processing and telemetry budget. Because of the fixed number of samples which can be read-out by TDI period, cosmic rays would hence decrease the star observation probability. For these reasons, not only is a *confirmation* CCD needed, but some star-cosmic ray *discrimination* should be performed at the detection level to avoid allocating obviously useless samples for the AF1 confirmation step.

2.4. Bright Stars

Saturated bright stars, although in small numbers compared to the billion target stars, are nevertheless the foundation of the most precise science reachable by Gaia. The best precision goes, however, hand in hand with additional observational difficulties: systematic effects, negligible for fainter stars, should be avoided here. One such systematic effect could, for instance, come from a poor across-scan window positioning: because the PSF is asymmetrical, a flux cut-off would lead to biased centroiding during the ground-based data reduction.

In terms of detection, bright stars are extended objects forming large connected components. When the diffraction spike begins to be detected, a connected component begins to be formed which ends at the end of the opposite spike. At this point, the object is formed and considered detected so measurements (centroiding, etc) can be performed. Unfortunately, an upper limit to the time needed for these operations is set by the delay which is available before the star enters the next CCDs. A causality problem can then occur if the observations begin in the subsequent

CCDs before the detection is completed. Another problem stems from the rebounds of the PSF, which may be interpreted as independent objects detections. Then, assuming that all this has been solved, the magnitude must be estimated, which is rendered difficult due to saturated pixels.

In terms of observation, two different strategies have been proposed: either activating CCD gates, which would reduce the effective observing time while preventing pixel saturation or using a special *windowing* of the vertical spikes (Høg et al. 2003). In both cases the *sampling* must, however, be adapted to avoid saturation at the level of the serial register when summing the (non-saturated) pixels.

In summary, bright stars present challenges at the *detection, sampling* and *window positioning* levels.

2.5. Multiple Objects

Double and multiple stars are fundamental objects which bring a wealth of information in stellar and galactic physics. Nevertheless, their handling is far from obvious. As can be seen in Figure 1, a normal windowing would frequently not allow to correctly cover the components. Consequently, the (flux-truncated) one-dimensional measurements would complicate the associated data reduction. Both a larger *sampling* and an adapted *window positioning* are thus required.

Another consequence is that components should be *detected* as close as possible so as to ensure that an adapted windowing can be applied. Below a separation of say 2 pixels, a single window can cover the components' light, but multiple windows may be needed above this threshold.

2.6. Processing Load

Among all the various adverse effects met by the on-board data handling, the most challenging to manage is perhaps the ability to process the data within the available time constraints.

The radiation environment was mentioned above, but its effect on the on-board processing capabilities was not broached. The electronic components can be affected by single event effects (bit flips, short circuits, destructive failures) and by a total ionizing dose (threshold voltage change). Consequently, classical commercial electronic components are not deemed reliable in radiation environments and specially qualified components are required. The consequence of this is that it is illusory to expect the on-board computing performances to be at the level of even our old office PCs. Multiplying the number of processors is not possible either because of the impact on the power, mass and cost budgets. Using programmable integrated circuits (Field-Programmable Gate Array, FPGA) is possible, though, but not for complex algorithms.

It should be mentioned that each of the 10 Video Processing Units in the Astro instrument has to manage more than 1.7 million samples per second, and this, whatever the stellar density. Then, the processing time scales approximately linearly with the number of objects. A large range of densities must be handled: the average density in the ASM is about 25 000 stars per square degree while the maximum density is about 120 times this figure (in the Astro instrument). If we now consider the number of objects per sample, the SSM would have to handle about 50 times more objects than the ASM, suggesting that the processing in Spectro will be much more demanding. Both *detection* and *window algorithms* should be optimised, with a severe *selection* step in SSM.

3. ADOPTED SOLUTIONS

The previous section introduced in a natural way the various functions (written in italics) which are needed to cope with the various constraints. Namely, sequentially: pre-calibration, background estimation, detection, centroiding, classification, sampling, windowing, selection, window positioning, confirmation and propagation of the positions between all successive CCDs. The content of these functions is described below.

The focal planes may be described by three main components: the CCD proximity electronics, the Video Processing Units (VPU) and the Payload Data Handling Unit (PDHU). The role of the VPU is to detect and manage the windows of science data and transmit these to the PDHU for subsequent compression and packetisation. Most of the scientific requirements thus occur at the level of VPU functionalities.

3.1. Astro and Spectro Processing

For what concerns the Astro instrument, detection is performed in the ASM1 and ASM2 CCDs, respectively for the Astro-1 and 2 FoV. Both detections allow, after propagation, for selecting the windows to be observed in AF1 (first selection). Confirmation is then attempted by running the detection on the windows observed in AF1. The objects, if confirmed, then undergo a second selection for the AF2-10 observations. Only confirmed objects will be observed in AF11 and BBP1-5 later on. Some will not be observed in these instruments in cases of high stellar density, as windows are larger and additional window conflicts will then occur.

In the spectro instrument, the objects are detected in SSM1 and in SSM2 then cross-matching is performed between these two detection lists. A selection procedure is then performed and provides the positions of the windows to be observed in the following Spectro CCDs. Beside photometry, one of the CCDs has a special role, which is the prediction of the spectra to observe in the RVS field (see Cropper et al. 2005).

The purpose of confirmation (Astro) or cross-matching (Spectro) is primarily to remove false detections due to cosmic-rays. A cross-matching algorithm is used in Spectro as the use of (a limited number of) windows would otherwise discard too many detections due to frequent crowding. Besides, the cross-matching procedure also tries to detect NEOs, i.e., objects whose motion is significant between SSM1/SSM2 in the first CCD module and SSM3/SSM4 in the second module.

3.2. Detection

Initially, a peak finding algorithm had been developed, SWA, (Babusiaux 1999), which fulfilled most of the detection needs, in terms of detection rate and efficient execution. It was however found safer to develop a segmentation algorithm (GD). First, because extended objects would not be detected, or with a smaller probability with a peak-finding algorithm. Second, because bright and saturated stars would not be handled correctly. Finally because a segmentation algorithm would be more robust with respect to objects smeared by their motion (asteroids), or by the motion of the satellite (in the improbable case of desynchronisation between TDI and scan motion). Finally, multiple objects could perhaps be more easily detected and delineated by the segmentation algorithm.

The GD algorithm has been developed (Chéreau 2002; Mignot 2003a) based on APM (Irwin 1985). The successive steps are the following: computation of the regional sky background, interpolation of it for each pixel, after a denoising step; pixels above noise build connected components which form the objects onto which centroiding and classification can then be performed, after a deblending step. The algorithm was then improved for special objects: double stars (Mignot 2003c) and bright stars. Most of the developments, however, were intended to render a hardware implementation (Mignot 2003d) possible. With this intent, it has been improved for speed, with integer-based operations, a hardware-friendly connected-component search, one-pass deblending and a fast background estimation. The object classification (mostly star/cosmic ray) is currently being implemented.

3.3. Sampling and Windowing

Once a star is detected, a sampling and a windowing has to be applied. Deep, comprehensive, dedicated studies have been performed by E. Høg in successive reports. The sampling has been designed in order to give optimal astrometric and photometric results, accounting for the optical PSF, the detection noise of the CCDs, the motion of the spin axis, the limited telemetry and the presence of double stars.

Every CCD will bring various types of complementary information. The AF1-10 are normal windows for the astrometry, though bright stars have AC resolution for calibration. AF11 has larger windows (2.5″ diameter) for the analysis of the surroundings and the discovery of potential perturbing objects. Wider BBP windows provide the local sky background. The ASMs give the 2 dimensional information though with a degraded resolution (2 × 2 binning).

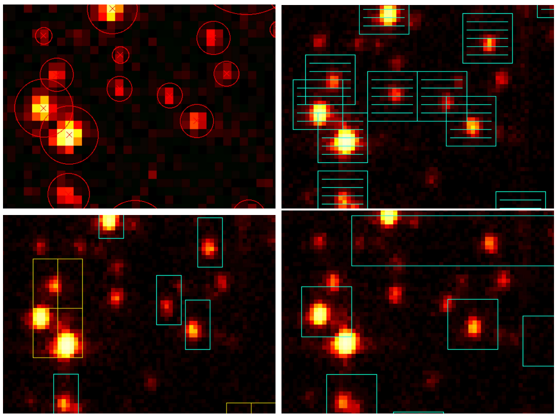

Figure 1. A multiple object in the Baade window as it would be detected in the Astrometric sky mappers (red circles and cross at center, top left), windowed for confirmation in AF1 (top right), windowed in AF2 and AF11 (bottom). Several features can be seen: the overlapping of windows (without overlapping of samples) in AF1 allow to cover several stars; some stars are not covered as the maximum number of samples per line was exceeded. In AF2, a (imperfect) window tiling is used and a large window can be put in AF11; the lack of samples per line is also apparent.

The complex sampling, windowing and, more generally, the scientific requirements for the on-board strategy on which the selection algorithm is based have been described in Høg et al. (2003) and the windowing has further been refined (de Bruijne 2003b) based on studies relative to the properties of the sky (de Bruijne 2003a), in the average and worst-case (the Baade window) stellar densities in Astro. The number of samples per line is based on this maximum number of objects to be handled within the nominal mode and on the size of the windows. This size increases for decreasing magnitudes.

The current version for the sampling and windowing design can be found in Høg (2004) to which the reader is referred.

3.4. Window Positioning and Selection Strategy

It was planned at an early stage that, besides detection, a selection algorithm would be necessary (Babusiaux & Arenou 2001). Performing the windowing may, at first sight, seem something easy but it was realised that a lot of detected stars would be lost simply because of windowing constraints.

In the AF1 CCD there are 6 samples across-scan per window, so two AF1 windows can overlap (see Figure 1b). In all other CCDs however, when two windows overlap, only one can be read, leading to a reduced fraction of the objects which can be downloaded. This is even more obvious when the Spectro instrument is considered (de Bruijne et al. 2004). Less than 80% of the detected stars can have a MBP window in the galactic plane whereas the large pixels windows in AF11 can al-most always be allocated except in very high densities (Figure 2).

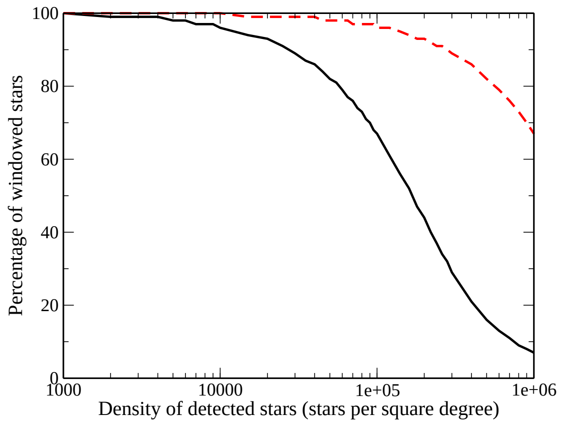

Figure 2. Theoretical proportion (%) of detected single stars of a given magnitude which could subsequently obtain a 10 × 4 window in MBP (full line) or a 68 × 12 window in AF11 (dashed) as a function of their (assumed uniform) density per square degree, if no special windowing positioning strategy were applied.

We could thus be in a circumstance where a star is observed, detected but not windowed because of overlapping. Although single stars only are considered in this argument, the problem applies all the more to double stars. The selection program needs thus to ensure that most of the detected objects can receive a window.

Taking into account the various sizes of the windows with respect to magnitude, the priority to magnitude and the limited number of samples per line, the windowing/selection algorithm has become increasingly complex. Moreover, the processing being performed under stringent real-time constraints, the choice of the windows must be done on a local scale, with no possibility of changing the previous window assignments – this would back-propagate the conflict. Additionally, the processing cannot make use of an image in memory of the pixels lines in order to simulate the areas covered by windows, as was done in a first version of the selection algorithm (Cira & Arenou 2002), as the TDI motion would imply a much too demanding addressing cost.

Maximising the number of observations in crowded fields or for multiple systems, i.e., the scientific return, while respecting these technical constraints has been achieved using several simple ideas: favouring overlapping windows (without overlapping samples, as can be done in AF1), allowing to slightly shift the windows (by a few pixels only) with respect to its normal position, allocating on-the-fly windows of smaller size when normal windows would not fit (AF11, BBP), or tiling several short windows to cover multiple systems. A new selection algorithm has been written (Chaussard 2003), allowing a possible multi-threaded approach and implementing all the complexity outlined above – at the expense of the computing time.

3.5. Payload Data Handling Study

The Gaia Payload Data Handling Electronics (PDHE) contract aims at designing the implementation architecture for the Payload Data Handling System and to develop a representative breadboard (Armbruster 2001).

The main goal of this activity is thus to size the electronics so as to handle the maximum object density in continuous mode. It started early 2003 and the analysis of the Pyxis performances soon showed that the planned computational resources would not be sufficient. A mixed hardware (for pixel processing) and software (for object processing) implementation was first envisaged which lead to an inflection of the Pyxis software development (see Section 3.6). While there were no special problems for most of the sky, it was clear that there were not enough CPU resources, including the needed margin, for the maximum density to be handled.

This contract has now entered its second phase: implementation on the breadboard and a full software solution on a faster processor has been adopted as baseline. Pyxis is being optimised so that a representative estimate of the on-board performances may be achieved.

3.6. Pyxis Software

Pyxis[2] is the software package developed for the data processing on-board Gaia. The Pyxis package encompasses the detection (GD), selection, windowing, propagation and cross-matching algorithms as well as a test environment used in the Astro and Spectro instruments. Its dual use is the evaluation of the resources needed for the Payload Data Handling Electronics contract (Section 3.5) and the simulation of the scientific output for the Gaia GIBIS simulator (Babusiaux 2002).

With the first industrial analysis received in April 2003 and discussions on alternative algorithms (Arenou et al. 2003b) it became clear that it would be difficult to implement the algorithms as they stood within the framework of the foreseen on-board data handling resources – namely to be able to handle the maximum object density in continuous mode. Beside achieving much improved stability and efficiency, most of the development have thus focused on easing migration towards a mixed hardware/software implementation. In-depth analysis of the algorithms led to a thorough rewrite adapted to the target architecture: the hardware part would deal with pixel-based operations, while the object-based operations would be done in software. Demonstrating the feasibility of the pixel-based operations has led to designing original methods devoted to the connected-component search (Mignot 2003b) or to the deblending scheme for overlapping components (Mignot 2003c). A fast estimation

[2]Pyxis is a constellation of the Southern hemisphere also known as the 'Compass', i.e., the instrument which shows the way, one of the functions of the on-board software through the satellite speed measurement. The new Latin word *Pyxis* comes from the Greek *Puxis* (box) and stands here as a reference to the pixel boxes transmitted by the on-board software. The constellation Pyxis was named by the Paris Observatory astronomer Nicolas Louis de La Caille.

of sky background has also been devised (Arenou et al. 2003a). Finally, the software was extensively reviewed in order to provide a realistic approximation of on-board operation: data types were tightly adjusted to suit the underlying scientific requirements, simple logic and low-level implementation have been introduced wherever a hardware implementation might be used, as have threads to fully specify the degree of interdependence of sub-tasks and emulate hardware-based processing. Iterations between the industrial implementation analyses and their evaluation (Mignot & Arenou 2003) for possible trade-offs are still on-going. All these efforts aim at altogether limiting the complexity on the pixel side, ensuring that the data flows on the object side are reduced and providing a realistic framework for both estimating needed on-board resources and expressing scientific requirement.

While, in a first step, the analyses and developments concerned mostly the Astro instrument, a growing involvement now occurs for the sky mappers of the Spectro instrument. Whereas the detection and selection algorithms are developed with the goal to implement them also for the Spectro Sky Mappers processing, the role of the SSM themselves had to be refined. The dual role of SSMs for further selection of spectra and moving objects (NEO/KBO) detection (Chéreau 2003) has led to the development of a dedicated cross-matching algorithm Chéreau (2004).

Beside the evaluation of the processing load on-board, the developed algorithms are directly implemented in the GIBIS pixel simulator (Babusiaux 2005) and used for the assessment of the scientific performances. Arenou et al. (2003c), for instance, compare GD to various other algorithms. As these performances have to be used for further evaluations of telemetry flow and database content, a simplified model and the associated code have been developed, mostly for use within the GASS Gaia simulator (Arenou & Lim 2003).

4. RESULTS

4.1. Detection Rate

The main general scientific objective is completeness. In an average density field, the detection completeness at magnitude 20 is easy to achieve (Figure 3) and the centroiding precision (Figure 3) permits to control the satellite's attitude. In the maximum assumed density, beside small numbers statistical variations, an 85% completeness can be achieved (Figure 5).

The GD detection algorithm uses two thresholds. The first one $T_{\rm pix}$ tests if a pixel is above the noise level and a second threshold $T_{\rm obj}$ decides whether an object made of such connected pixels is above noise. The two thresholds have been chosen so as to ensure that the number of false negative (undetected) is minimised for a number of false positive (false detections) less than 1 for one million samples. Because of the presence of a confirmation step we could afford a much larger number of false positive, so the indicated detection rate is also conservative.

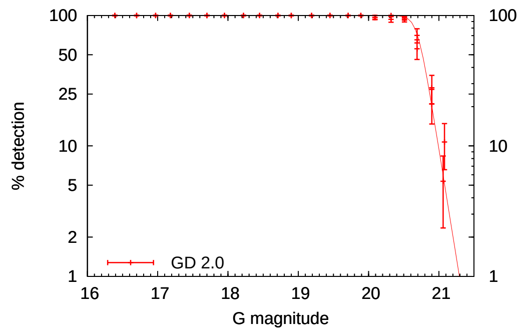

Figure 3. Percentage of detection as a function of magnitude in the ASM for a simulated average stellar density field.

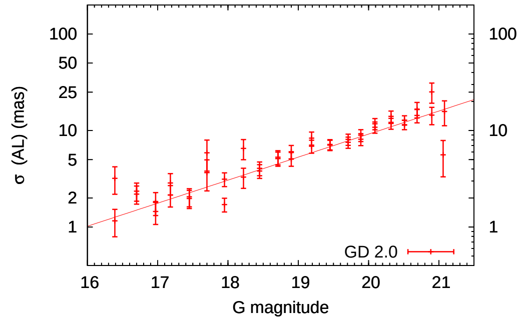

Figure 4. Precision of the on-board centroiding in the Astrometric Sky Mappers for an average density field.

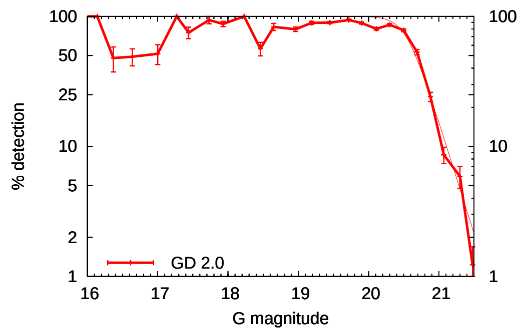

Figure 5. Percentage of detection as a function of magnitude in the ASM for the Baade window. The simulated image has been obtained with a scanned HST image, duplicated AC to cover an ASM CCD. The density is about 4 million stars per square degrees, larger than the maximum assumed density for the on-board processing.

As can be expected from the much higher apparent density and from the much higher level of sky background, completeness cannot be achieved in the Spectro Sky Mappers in a fraction of the Galactic plane (Figures 6 and 7). After the detection step, the fraction of observed objects is further reduced, due to the windowing step. Depending on the implemented complexity at this step (from no special windowing to positioning features, as described at Section 3.4), the total (cumulative) fraction of confirmed, windowed objects can vary between 6% and 25% of $GS < 20.15$ objects in a 400 000 stars per square degree field. This illustrates the gain which can be achieved with the windowing algorithm if enough sam-

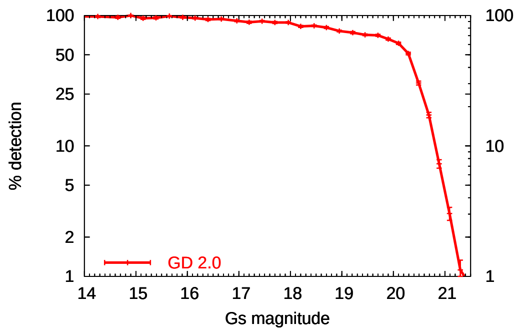

Figure 6. Percentage of detection in SSM for a 100 000 stars per square degree density at $GS = 20.15$.

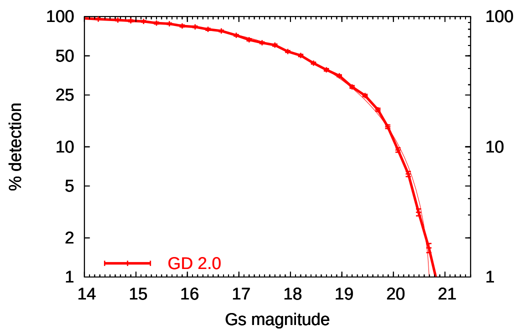

Figure 7. Percentage of detection in SSM for a 400 000 stars per square degree density at $GS = 20.15$.

ples per line are available. As for the undetected objects, most will nevertheless be known from the Astro observations.

Several other points are worth noting: first, completeness should not be evaluated on a single transit because the various transits with different orientations allow to select on one transit what has not been selected on another; an improved completeness will result but with a reduced number of measurements. Second, the quoted percentage refers to a comparison between input stars and detections accounting for only one (the brightest) of possibly multiple components; for close double stars, the fainter secondary would also be windowed, though it is not accounted for here. Third, the presence of multiple systems may well account for the apparent degradation of the centroiding precision; however, what is important onboard is to constrain the satellite's attitude, through the comparison between detection and confirmation centroid measurement, which should remain consistent even for multiple systems; besides, statistical robustness will be introduced for this comparison.

4.2. Double Stars in Astro

Because of the particular attention devoted to the observation of double and multiple stars (including optical doubles in crowded fields) the performances of the detection is worth mentioning here. They are shown in Figures 8 and 9. There is a clear improvement compared to the SWA peak finding algorithm (Figure 10). The improvement is even more noteworthy compared to the other typical detection algorithms used in the astronomical community, rather logically as GD and SWA have been improved for the Gaia case.

To know how this impacts on the Astro telemetry budget, it is however of interest to translate these detection prob-

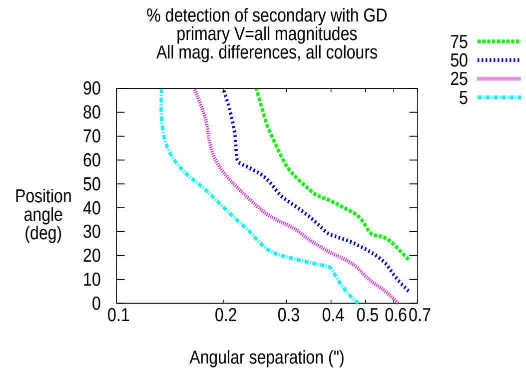

Figure 8. Detection probability (%) of double stars as a function of angular separation (arcsec) and orientation (deg) between components for a uniform distribution of magnitude differences and primary magnitude. The rectangular angular shape of the pixel explains the lower detection probability when components are aligned AC.

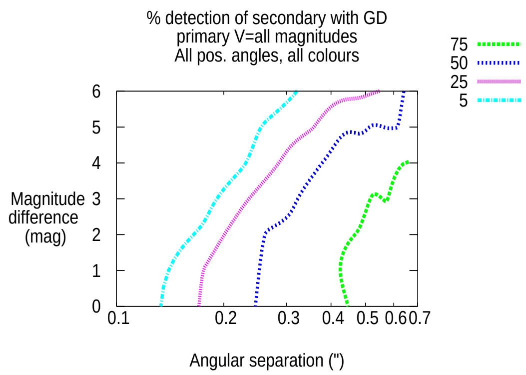

Figure 9. Detection probability (%) of double stars as a function of separation and magnitude difference between components for a uniform distribution of orientation and primary magnitude, with the GD algorithm.

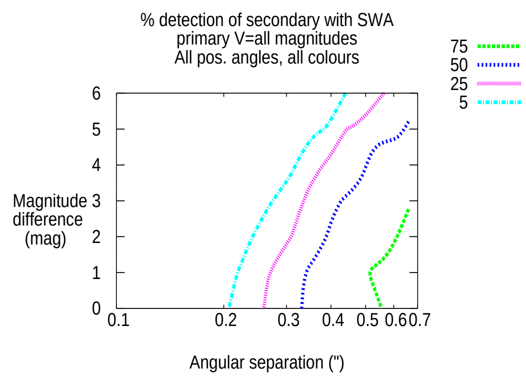

Figure 10. Detection probability (%) of double stars as a function of separation and magnitude difference between components with the SWA peak finding algorithm.

abilities into fractions of measurements, using for this a binary model (Söderhjelm 2004). It is expected that about 5% of stars will be resolved binaries with a separation and magnitude difference within the limits shown in Figure 9, and to which the shown detection probability should be applied. More relevant however is the number of observations which would be binaries with $G < 20$ primaries with a window conflict in AF1, and where the components have a separation larger than 2 pixels AL or AC, otherwise a single window would not be a problem. It is expected that about 3.5% of observations would be in this case, and about 0.6% detected occurrences on-board. This has to be compared to what is expected for optical doubles at magnitude 20, about 0.5%. The Low detection rate for binaries is due to the secondary too faint to be detected. When primaries brighter than 16 are now considered, about 2% of observations may be binaries with window conflict in AF1-10 detected on-board.

5. CONCLUSION

The Gaia cornerstone mission provides a real challenge, not only for the data reduction, but already on board. The complexity of the sky, added to the numerous instrumental constraints, demands a dedicated and complex on-board data handling.

The software development of the scientific algorithms has now reached an encouraging maturity seven years before Gaia launch, thanks to the ≈ 8 (wo)man-year devoted up to now.

On-going studies and development are related to refining the set of measurements and associated methods in order to finely adjust the software part, perform classification, and implement star / cosmic ray discrimination capabilities as well as confirmation. Apart from further adapting the software to the various technical and processing constraints, while maintaining or increasing the scientific return, the general aim of the Pyxis on-board data handling development remains to provide a complete, robust and accurate model of object management in the Astro and Spectro focal planes.

REFERENCES

Arenou, F., Babusiaux, C., Høg, E., Jordi, C., Mignot, S., 2003a, Determination of sky background, Tech. Rep. OBD-CoCo-007, Observatoire de Paris

Arenou, F., Babusiaux, C., Mignot, S., 2003b, Comments on GP-ASG-RS-0001 implementation analysis, Tech. Rep. OBD-CoCo-006, Observatoire de Paris

Arenou, F., Babusiaux, C., Mignot, S., 2003c, Detection performances in the ASM (I), Tech. Rep. OBD-CoCo-005, Observatoire de Paris

Arenou, F., Lim, J., 2003, Simulation of the on-board detection, Tech. Rep. OBD-FAJCL-001, Observatoire de Paris

Armbruster, P., 2001, Payload data handling electronics – SOW and technical requirements, Tech. Rep. GAIA/TADG21/2001, ESA–ESTEC

Babusiaux, C., 1999, SWA - Sliding Window Algorithm - V1.0 users guide, Tech. Rep., Observatoire de Paris

Babusiaux, C., 2002, GIBIS, Gaia Instrument and Basic Image Simulator, Tech. Rep. GAIA-CB-01, Institute of Astronomy, Cambridge

Babusiaux, C., Arenou, F., 2001, Work reference document, Tech. Rep. OBD-CoCo-01, Observatoire de Paris

Babusiaux, C., 2005, ESA SP-576, this volume

de Bruijne, J., 2003a, PDHE load assumptions – properties of the sky, Tech. Rep. GAIA-JdB-009, ESA–ESTEC

de Bruijne, J., 2003b, Windowing and sampling for faint stars in Astro, Tech. Rep. GAIA-JdB-011, ESA–ESTEC

de Bruijne, J., Chéreau, F., Arenou, F., 2004, Windowing and sampling for faint stars in MBP, Tech. Rep. GAIA-JdB-014, ESTEC and Observatoire de Paris

Chaussard, J., 2003, On-board selection algorithm for Gaia, Tech. Rep. OBD-JC-001, Observatoire de Paris

Chéreau, F., 2002, Gaia_Detect documentation, Tech. Rep. OBD-FC-01, Observatoire de Paris

Chéreau, F., 2003, SSM5-6 red-coated CCDs sampling strategy analysis, Tech. Rep. OBD-FC-002, Observatoire de Paris

Chéreau, F., 2004, Cross-matching algorithm description, Tech. Rep. OBD-FC-003, Observatoire de Paris

Cira, H., Arenou, F., 2002, First implementation of the selection algorithm, Tech. Rep. OBD-HC-001, Observatoire de Paris

Cropper, M., et al., 2005, ESA SP-576, this volume

Høg, E., 2004, Summary of Sampling Schemes for ASM, AF, BBP, SSM and MBP, Tech. Rep. GAIA-CUO-151, Copenhagen University Observatory

Høg, E., Arenou, F., Mignot, S., Babusiaux, C., Katz, D., Jordi, C., 2003, Scientific requirements for the on-board processing, Tech. Rep. GAIA-CUO-117

Irwin, M. J., 1985, MNRAS, 214, 575

Lammers, U., et al., 2005, ESA SP-576, this volume

Mignot, S., 2003a, Flowcharts of the SWA 2.5 and GD 1.5 detection algorithms, Tech. Rep. OBD-SM-002, Observatoire de Paris

Mignot, S., 2003b, A hardware-oriented connected-component labeling algorithm, Tech. Rep. OBD-SM-006, Observatoire de Paris

Mignot, S., 2003c, An improved deblending scheme for Gaia's on-board detection, Tech. Rep. OBD-SM-003, Observatoire de Paris

Mignot, S., 2003d, On-board detection – a representativity and feasibility update, Tech. Rep. OBD-SM-007, Observatoire de Paris

Mignot, S., Arenou, F., 2003, Comments on the PDHE data package, Tech. Rep. OBD-SMFA-001, Observatoire de Paris

Portell, J., et al., 2005, ESA SP-576, this volume

Söderhjelm, S., 2004, Theoretical modelling of observational double-star distribution functions, Tech. Rep. DMS-SS-05, Lund Observatory