GAIA

Composition, Formation and Evolution of the Galaxy

Report on the Concept and Technology Study

compiled by the

GAIA Science Advisory Group

Abstract

The GAIA mission will provide unprecedented positional and radial velocity measurements with the accuracies needed to produce a stereoscopic and kinematic census of about one billion stars in our Galaxy and throughout the Local Group. This amounts to about 1 per cent of the Galactic stellar population. Combined with astrophysical information for each star, provided by on-board multi-colour photometry, these data will have the precision necessary to quantify the early formation, and subsequent dynamical, chemical and star formation evolution of the Milky Way Galaxy.

Additional scientific products include detection and orbital classification of tens of thousands of extra-solar planetary systems, a comprehensive survey of objects ranging from huge numbers of minor bodies in our Solar System, through galaxies in the nearby Universe, to some 500 000 distant quasars. It will also provide a number of stringent new tests of general relativity and cosmology.

A complete satellite design is presented, including the proposed payload, corresponding accuracy assesments, and results from a prototype data reduction development. The satellite can be launched in 2009, within the specified budget for the next generation ESA Cornerstone missions.

http://astro.estec.esa.nl/GAIA

Contents

Executive Summary

The GAIA mission has been under study by ESA over the past three years. It is a candidate for the next cornerstone mission of the science programme, arising from the recommendations of the Horizon 2000+ Survey Committee in 1994.

GAIA will rely on the proven principles of ESA's Hipparcos mission to solve one of the most difficult yet deeply fundamental challenges in modern astronomy: to create an extraordinarily precise three-dimensional map of about one billion stars throughout our Galaxy and beyond. In the process, it will map their motions, which encode the origin and subsequent evolution of the Galaxy. Through comprehensive photometric classification, it will provide the detailed physical properties of each star observed: characterizing their luminosity, temperature, gravity, and elemental composition. This massive stellar census will provide the basic observational data to tackle an enormous range of important problems related to the origin, structure, and evolutionary history of our Galaxy.

GAIA will achieve this by repeatedly measuring the positions of all objects down to $V = 20$ mag. On-board object detection will ensure that variable stars, supernovae, burst sources, micro-lensed events, and minor planets will all be detected and catalogued to this faint limit. Final accuracies of 10 microarcsec at 15 mag, comparable to the diameter of a human hair at a distance of 1000 km, will provide distances accurate to 10 per cent as far as the Galactic Centre, 30 000 light years away. Stellar motions will be measured even in the Andromeda galaxy.

GAIA's expected scientific harvest is of almost inconceivable extent and implication. Its main goal is to clarify the origin and history of our Galaxy, by providing tests of the various formation theories, and of star formation and evolution. This is possible since low mass stars live for much longer than the present age of the Universe, and therefore retain in their atmospheres a fossil record of their detailed origin. The GAIA results will precisely identify relics of tidally-disrupted accretion debris, probe the distribution of dark matter, establish the luminosity function for pre-main sequence stars, detect and categorize rapid evolutionary stellar phases, place unprecedented constraints on the age, internal structure and evolution of all stellar types, establish a rigorous distance scale framework throughout the Galaxy and beyond, and classify star formation and kinematical and dynamical behaviour within the Local Group of galaxies.

GAIA will pinpoint exotic objects in colossal and almost unimaginable numbers: many thousands of extra-solar planets will be discovered, and their detailed orbits and masses determined; tens of thousands of brown dwarfs and white dwarfs will be identified; some 100 000 extragalactic supernovae will be discovered and details passed to ground-based observers for follow-up observations; Solar System studies will receive a massive impetus through the detection of many tens of thousands of new minor planets; inner Trojans and even new trans-Neptunian objects, including Plutinos, may be discovered. GAIA will follow the bending of star light by the Sun and major planets, over the entire celestial sphere, and therefore directly observe the structure of space-time—the accuracy of its measurement of General Relativistic light bending may reveal the long-sought scalar correction to its tensor form. The PPN parameters γ and β, and the solar quadrupole moment J_2, will be determined with unprecedented precision. New constraints on the rate of change of the gravitational constant, $\dot{G}$, and on gravitational wave energy over a certain frequency range, will be obtained.

The Concept and Technology Study demonstrates that these goals are feasible by means of an ESA-only mission, technically achievable on the time-scale of a 2009 launch, and within a budget profile consistent with the current cornerstone envelope. GAIA will carry the demonstrated Hipparcos principles into orders of magnitude improvement in terms of accuracy, number of objects, and limiting magnitude, by combining them with available technology. This will yield a huge scientific output at an affordable cost for the ESA member state community.

GAIA will be a continuously scanning spacecraft, accurately measuring one-dimensional coordinates along great circles, and in two simultaneous fields of view, separated by a well-defined and well-known angle (these one-dimensional coordinates are then converted into the astrometric parameters in a global data analysis, in which distances and proper motions 'fall out' of the

processing, as does information on double and multiple systems, photometry, variability, metric, planetary systems, etc.). The payload is based on a large but feasible CCD focal plane assembly, with passive thermal control, and a natural short-term (3 hour) instrument stability due to the sunshield, the selected orbit, and a robust payload design.

The telescopes are of moderate size, with no specific design or manufacturing complexity. The system fits within a dual-launch Ariane 5 configuration, without deployment of any payload elements. The study identifies a 'Lissajous' orbit at L2 as the preferred operational orbit, from where about 1 Mbit of data per second is returned to the single ground station throughout the 5-year mission. The 10 microarcsec accuracy is evaluated through a comprehensive accuracy assessment programme; this remarkable accuracy is possible partly by virtue of the (unusual) instrumental self-calibration achieved through the data analysis on-ground. This ensures that final accuracies essentially reflect the photon noise limit for localisation accuracy, exactly as achieved with Hipparcos.

In addition to assisting in the direction and monitoring of the industrial Concept and Technology Study, the Science Advisory Group has studied the main elements of an 'end-to-end' programme. The report demonstrates that star selection can be effectively undertaken autonomously on-board, which has the far-ranging scientific implications noted earlier, and which also eliminates the need for a complex and costly pre-launch programme of observation definition: the Science Operations Centre activities associated with the mission will also be correspondingly greatly simplified.

The studies include an assessment of the resulting satellite data stream and its analysis, and have supplied confidence in its feasibility. Simulation results are presented which demonstrate that the massive GAIA data volume, representing some 20 Tbytes of raw data, can be handled with appropriate early planning, and that the data set can be reduced using realistic projections of recent developments in storage devices and state-of-the-art concepts of object-oriented data bases.

The present report proposes that GAIA is undertaken as a purely ESA mission, which will maintain Europe's lead in fundamental positional astronomy, and place an enormous astrophysical advance in the hands of a very large number of European scientists.

It is recommended that the spacecraft and payload are procured together by the industrial prime contractor. This approach is considered appropriate for GAIA due to the very advanced technology (including metrology) aspects of the payload, the integrated nature of the entire system, and the complexities and scale of the projected data analysis tasks. Options for contributions by member states are identified, should this organisational approach be preferred.

All scientific elements can be undertaken by the ESA scientific community. It is proposed that the mission is established as an 'observatory' facility: community scientists will be able to request results of observations on any of the objects observed by GAIA, at rather short notice.

While challenging, the entire GAIA design is within the projected state-of-the-art, and the satellite can be developed in time for launch in 2009. With such a schedule, a complete stereoscopic map of our Galaxy will be available within 15 years. The successful completion of this programme will characterise the structure and evolution of stars and our Galaxy in a manner completely impossible using any other methods, and to an extent inconceivable even a decade ago.

The document is organised as follows:

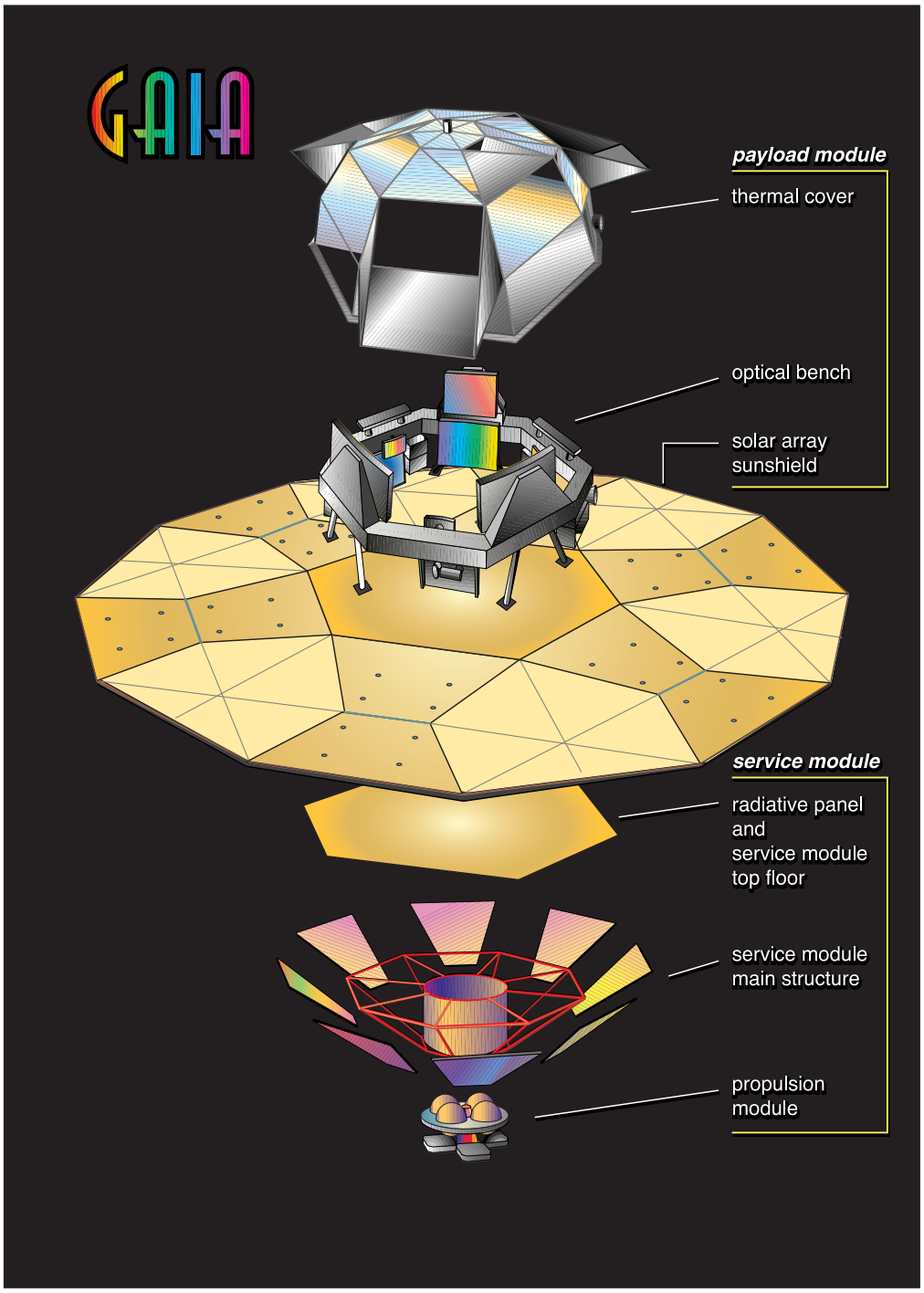
Part I: Science Case Section 1 presents the science case for GAIA. The central theme is to derive the formation and evolution of the Galaxy through a massive stellar population census covering all types of object.

Part II: Technical Design This covers the requirements on the payload and spacecraft (Section 2), the resulting payload design (Section 3), spacecraft design (Section 4), and orbit selection and spacecraft operations (Section 5).

Part III: Mission Performance Assessment This comprises assessment principles (Section 6), methodology (Section 7), and accuracy results for astrometry, photometry and radial velocities (Section 8).

Part IV: Data Analysis An overview of the foreseen data analysis methods is presented, along with results of specific experiments carried out to demonstrate the feasibility of the proposed approach (Section 9). An outline management programme for the satellite is also given (Section 10), including scientific institute involvement, and for the scientific aspects, including possibilities for an 'outreach' programme based on GAIA.

Appendices These cover (A) related ground and space programmes; (B) details of the interferometric option also evaluated during the study phase; (C) acronyms; and (D) scientific and technical contributors to this study of the GAIA mission.



Frontispiece: Exploded view of the GAIA satellite.

Summary of the GAIA Science Capabilities

Objectives: Galaxy origin and formation; physics of stars and their evolution; Galactic dynamics and distance scale; solar system census; large-scale detection of all classes of astrophysical objects including brown dwarfs, white dwarfs, and planetary systems; fundamental physics

Measurement Capabilities:

- **catalogue:** ~ 1 billion stars; 0.34×10^6 to $V = 10$ mag; 26×10^6 to $V = 15$ mag; 250×10^6 to $V = 18$ mag; 1000×10^6 to $V = 20$ mag; completeness to about 20 mag
- **sky density:** mean density $\sim 25\,000$ stars deg^{-2}; maximum density $\sim 3 \times 10^6$ stars deg^{-2}
- **accuracies:** median parallax errors: 4 μas at 10 mag; 11 μas at 15 mag; 160 μas at 20 mag
- **distance accuracies:** from Galaxy models: 21 million better than 1 per cent; 46 million better than 2 per cent; 116 million better than 5 per cent; 220 million better than 10 per cent
- **tangential velocity accuracies:** from Galaxy models: 44 million better than 0.5 km s^{-1}; 85 million better than 1 km s^{-1}; 210 million better than 3 km s^{-1}; 300 million better than 5 km s^{-1}; 440 million better than 10 km s^{-1}
- **radial velocity accuracies:** 1–10 km s^{-1} to $V = 16 - 17$ mag, depending on spectral type
- **photometry:** to $V = 20$ mag in 4 broad and 11 medium bands

Scientific Goals:

- **the Galaxy:** origin and history of our Galaxy — tests of hierarchical structure formation theories — star formation history — chemical evolution — inner bulge/bar dynamics — disk/halo interactions — dynamical evolution — nature of the warp — star cluster disruption — dynamics of spiral structure — distribution of dust — distribution of invisible mass — detection of tidally disrupted debris — Galaxy rotation curve — disk mass profile
- **star formation and evolution:** *in situ* luminosity function — dynamics of star forming regions — luminosity function for pre-main sequence stars — detection and categorization of rapid evolutionary phases — complete and detailed local census down to single brown dwarfs — identification/dating of oldest halo white dwarfs — age census — census of binaries and multiple stars
- **distance scale and reference frame**: parallax calibration of all distance scale indicators — absolute luminosities of Cepheids — distance to the Magellanic Clouds — definition of the local, kinematically non-rotating metric
- **Local group and beyond:** rotational parallaxes for Local Group galaxies — kinematical separation of stellar populations — galaxy orbits and cosmological history — zero proper motion quasar survey — cosmological acceleration of Solar System — photometry of galaxies — detection of supernovae
- **Solar system:** deep and uniform detection of minor planets — taxonomy and evolution — inner Trojans — Kuiper Belt Objects — disruption of Oort Cloud
- **extra-solar planetary systems:** complete census of large planets to 200–500 pc — orbital characteristics of several thousand systems
- **fundamental physics:** γ to $\sim 5\times10^{-7}$; β to $3\times10^{-4}-3\times10^{-5}$; solar J_2 to $10^{-7}-10^{-8}$; $\dot{G}/G$ to $10^{-12}-10^{-13}$ yr^{-1}; constraints on gravitational wave energy for $10^{-12} < f < 4\times10^{-9}$ Hz; constraints on $\Omega_{\rm M}$ and Ω_Λ from quasar microlensing
- **specific objects:** $10^6 - 10^7$ resolved galaxies; 10^5 extragalactic supernovae; 500 000 quasars; $10^5 - 10^6$ (new) solar system objects; $\gtrsim$ 50 000 brown dwarfs; 30 000 extra-solar planets; 200 000 disk white dwarfs; 200 microlensed events; 10^7 resolved binaries within 250 pc

Summary of the GAIA Mission Profile

Payload:

- two identical astrometric telescopes:
  - fully-reflective 3-mirror Silicon Carbide optics
  - separation of viewing directions: 106°
  - monolithic primary mirrors: 1.7×0.7 m^2
  - field of view: 0.32 deg^2
  - focal length: 50 m
  - wavelength range: 300–1000 nm
- payload and focal plane assemblies:
  - thermally and mechanically decoupled from service module
  - operating temperature: ~ 200 K
  - focal plane detectors: CCDs operating in TDI mode
  - pixel size along-scan: 9 μm
- spectrometric instrument:
  - spectrometer for radial velocity measurements + medium-band photometry
  - entrance pupil: 0.75×0.70 m^2
  - field of view: 4 deg^2
  - focal length: 4.17 m
  - focal plane detectors: CCDs operating in TDI mode

Orbit and System Aspects:

- orbit: Lissajous-type, eclipse-free, around L2 point of Sun/Earth system; 220–240 day transfer orbit
- sky scanning: revolving scanning with scan rate = 120 arcsec s^{-1}, precession period = 76 days
- spacecraft: 3-axis stabilized; autonomous propulsion system for transfer orbit; electrical (FEEP) thrusters for operational attitude control; 6 deployable solar panels (space between them filled with multi-layer insulation to form the sun shield)
- science data rate: 1 Mbps sustained, 3 Mbps on down-link, using electronically steerable high-gain phased array antenna
- launch mass: 3137 kg (payload = 803 kg, service module = 893 kg, system margin (20%) = 339 kg, fuel = 1010 kg, launch adaptor = 92 kg)
- power: 2569 W (payload = 1528 W, service module = 641 W, harness losses = 76 W, contingency (10%) = 224 W
- dimensions: payload: diameter = 4.2 m, height = 2.1 m; service module: diameter = 4.2 m (stowed)/8.5 m (deployed), height = 0.8 m
- launcher: Ariane 5, dual launch, SPELTRA, 1666 mm interface
- lifetime: 5 years design lifetime (4 years observation time); 6 years extended lifetime

Summary of Required Technology Developments:

- high-performance, small-pixel (9 μm) CCD for payload focal planes
- high-speed, high integration, low-noise detector chain
- further validation of SiC technology for large mirrors and toroidal structure
- high data rate phased array antenna for L2 operation
- ground data base/archiving/processing architecture for $\sim$ 20 TB of raw data

Payload Aspects Capable of Further Optimisation:

- fainter magnitude ($V > 17$ mag) limit for radial velocities
- derivation of astrophysical parameters from photometric measurements
- photometric quality for faint stars in complex regions
- detailed telemetry budget

Preface

The scientific success of Hipparcos demonstrated the feasibility of accurate global astrometry from space, and its fundamental impact on many branches of astrophysics. It naturally led to a proposed M3 mission, Roemer (Høg 1993; Lindegren et al. 1993a; Høg & Lindegren 1994), to measure parallaxes and proper motions for ∼100 million stars, complete to $V \sim 15.5$ mag, with accuracies in the range 0.1 mas ($V \lesssim 11$ mag) to 1.5 mas ($V \sim 17$ mag). Roemer was rated highly, but not accepted in view of its relatively modest accuracy gain over Hipparcos.[1]

A concept for making observations with significantly improved accuracies, GAIA, was proposed in response to a call for ideas for a cornerstone mission (Lindegren et al. 1993b; Lindegren & Perryman 1994; **??**; Lindegren et al. 1995). The proposal was based on a scanning mission with two viewing directions, each comprising a small non-deployable interferometer which could be accommodated within the Ariane 5 fairing. A combination of the enlarged baseline and collecting area, and use of CCD detectors in the focal plane, led to projected accuracies of around 10 μas at 15 mag, and a target list of more than 50 million objects. The substantial science case for a μas astrometric mission was given a high rating by the Horizon 2000+ Survey Committee, and GAIA was recommended for adoption as the next cornerstone mission assuming that the 10 μas accuracy goal could be demonstrated (Battrick 1994).

Approval of a 'Concept and Technology Study' for GAIA (along with the other three 2009 cornerstone mission candidates) was given in 1996. An *ad hoc* Science Advisory Group was established under the responsibility of the ESA Director of Science; amongst other tasks this advisory group prepared a Payload Definition Document, which formed the basis of a one-year industrial study, which took place between mid-1997 and mid-1998. Three industrial proposals were submitted in June 1997. The contract subsequently was awarded to Matra Marconi Space (MMS) in July 1997.

The study objectives were threefold: (i) to establish a feasible mission design, satisfying the scientific requirements formulated by the Science Advisory Group, at the lowest cost and with minimum development risks; (ii) to identify further investigations and key areas of technological studies and validation needed to carry out the programme; (iii) to assess the cost of the mission, on the basis of a cost-effective development and procurement approach.

During the study, regular industrial progress meetings, and regular meetings of the Science Advisory Group took place. Three working groups were also established: a Science Working Group, a Photometry Working Group, and an Instrument Working Group. Together, many details of the overall system design were evaluated and iterated. The industrial study was divided into two parts: Phase 1 (4 months), devoted to system and instrument trade-offs, leading to a proposed concept further evaluated during Phase 2 (8 months). The most important results of this second phase are: (i) a technical solution which yields the highest possible astrometric accuracy, and allows measurement of all objects down to faint magnitudes; (ii) a spacecraft and payload design satisfying the targeted scientific goals, which is realistic and feasible for launch in 2009 within the projected state-of-the-art; (iii) detailed simulation and accuracy analysis tools, allowing the expected performances to be stated with confidence; (iv) a feasible approach to the data analysis, demonstrated through simulations, providing confidence in handling the data volume expected from GAIA, and carrying out the complex data analysis on competitive time scales.

As now proposed, GAIA will extend the achievable astrometric accuracies by 2–3 orders of magnitude over Hipparcos, to better than 10 μas at 15 mag. It will extend the limiting magnitude of the astrometric observations from 12 to 20 mag, and the completeness limit from around 7–9 mag to a uniform faint limit of about 20 mag. As a result, GAIA will measure about 1 billion (10^9) stars in a global stereoscopic stellar census of our Galaxy and its nearest neighbours. It will also obtain multi-epoch, multi-colour, photometry as crucial diagnostic data for all stars observed, along with radial velocities for the brighter objects to complete the kinematical data.

[1] The subsequently proposed German DIVA and the recently approved US FAME missions (see Appendix A) may fill this post-Hipparcos regime while the ESA community aims for the dramatic advance offered by GAIA.

Part I
Scientific Case

1 Science Goals

1.1 Overview

Understanding the Galaxy in which we live is one of the great intellectual challenges facing modern science. The Milky Way contains a complex mix of stars, planets, interstellar gas and dust, radiation, and the ubiquitous dark matter. These components are widely distributed in age (reflecting their birth rate), in space (reflecting their birth places and subsequent motions), on orbits (determined by the gravitational force generated by their own mass), and with chemical element abundances (determined by the past history of star formation and gas accretion). Astrophysics has now developed the tools to measure these distributions in space, kinematics, and chemical abundance, and to interpret the distribution functions to map, and to understand, the formation, structure, evolution, and future of our entire Galaxy. This potential understanding is also of profound significance for quantitative studies of the high-redshift Universe: a well-studied nearby template underpins analysis of unresolved galaxies with other facilities, and at other wavelengths.

Understanding the structure and evolution of the Galaxy requires three complementary observational approaches: (i) a census of the contents of a large, representative, part of the Galaxy; (ii) quantification of the present spatial structure, from distances; (iii) knowledge of the three-dimensional space motions, to determine the gravitational field and the stellar orbits. That is, one requires complementary astrometry, photometry, and radial velocities. Astrometric measurements uniquely provide model independent distances and transverse kinematics, and form the base of the cosmic distance scale. Photometry, with appropriate astrometric and astrophysical calibration, gives a knowledge of extinction, and hence, combined with astrometry, provides intrinsic luminosities, spatial distribution functions, and stellar chemical abundance and age information. Radial velocities complete the kinematic triad, allowing determination of gravitational forces, and the distribution of invisible mass. The GAIA mission will provide all this information.

GAIA will provide an immense quantity of extremely accurate astrometric and photometric data from which all branches of astrophysics will benefit directly. Current understanding of the physics and the early and late evolution of individual stars will be revolutionized: very accurate absolute information covering the whole Hertzsprung–Russell diagram, from pre-main sequence to stellar death, will be available for even short-lived stellar evolutionary phases, not only for single stars, but also for large numbers of binaries and multiple systems. The spatial distribution of dark matter will be determined. A census of the minor bodies in the Solar System, together with measurement of the numbers of planets around stars as a function of spectral type, will quantify planetary system formation modeling, and optimize searches for life elsewhere. A large, well-defined all-sky catalogue of galaxies and quasars will quantify studies of the structure of the local Universe, and of much larger-scale structures at high redshift. The stars, galaxies, active galactic nuclei and quasars mapped by GAIA are the natural complement to surveys at other wavelengths, from ground-based radio data to high-energy photons observed by XMM and INTEGRAL. Fundamental physics will benefit from local metric mapping, which will test General Relativity to unprecedented accuracy. Any large survey of new domains in discovery space also brings the new and unpredicted.

GAIA is timely as it builds on recent intellectual and technological breakthroughs. Current understanding and exploration of the early Universe, through microwave background studies (e.g., Planck) and direct observations of high-redshift galaxies (HST, NGST, VLT) have been complemented by theoretical advances in understanding the growth of structure from the early Universe up to galaxy formation. Serious further advances require a detailed understanding of a 'typical' galaxy, to test the physics and assumptions in the models. The Milky Way and the nearest Local Group galaxies uniquely provide such a template.

Through its global survey nature, GAIA is ideally suited to provide an extremely dense and accurate optical reference system. The limiting magnitude is sufficiently faint to allow a direct link to the extragalactic system by observation of quasars. This is critical for multi-wavelength astrophysics (VLBI, ALMA, FIRST, HST, XMM, DeNIS), as is maintaining the system into the future.

1.2 Structure and Dynamics of the Galaxy

The primary objective of the GAIA mission is the Galaxy: to observe the physical characteristics, kinematics and distribution of stars over a large fraction of its volume, with the goal of achieving a full understanding of the Galaxy's dynamics and structure, and consequently its formation and history. GAIA will make this goal possible by providing, for the first time, a catalogue which will sample a large and well-defined fraction of the stellar distribution in phase space from which significant conclusions can be drawn for the entire Galaxy. Hipparcos did this for one location in the Galaxy, the Solar neighbourhood; GAIA will accomplish this for a large fraction of the Galaxy.

1.2.1 Galactic Structure

The most conspicuous component of the Milky Way is its flat disk which contains nearly 10^{11} stars of all types and ages orbiting the Galactic Centre. The Sun is located at 8.5 kpc from the Galactic Centre. The disk displays spiral structure, and also contains interstellar material, predominantly atomic and molecular hydrogen, and a significant amount of dust. The inner kpc of the disk also contains the bulge, which is less flattened, may contain a bar, and consists mostly of fairly old stars. At its centre lies a supermassive black hole of $\sim 2.9 \times 10^6 M_\odot$. The disk and bulge are surrounded by a halo of about 10^9 old and metal-poor stars, as well as ∼140 globular clusters and a small number of satellite dwarf galaxies. The entire system is embedded in a massive halo of dark material of unknown composition and poorly known spatial distribution.

The distributions of stars in the Galaxy over position and velocities are linked through gravitational forces, and through the star formation rate as a function of position and time. The initial distributions are modified, perhaps substantially, by small and large scale dynamical processes: these include instabilities which transport angular momentum (bars and warps), and mergers.

It is still unclear, for example, whether spiral arms are density wave enhancements in the background stellar distribution, or whether they are predominantly gas-dynamical, being regions of enhanced star formation. If a density enhancement exists, it will affect the stellar motions in a characteristic way, and this could be tested on the relatively nearby Perseus arm, about 2 kpc distant. Stellar motions would be determined for stars near the arm, both foreground and background (distinguished by their parallaxes), with proper motion accuracy requirements of some 100 μas yr^{-1}, corresponding to about 1 km s^{-1} in space velocity.

Understanding our Galaxy requires measurement of distances and space motions for large and unbiased samples of stars of different mass, age, metallicity, and evolutionary stage. GAIA's global survey of the entire sky to $V \sim 20$ is the ideal approach to define and measure such samples. The huge number of stars, impressive accuracy, and faint limiting magnitude of GAIA will quantify our understanding of the structure and motions within the bulge, the spiral arms, the disk and the outer halo, and will revolutionize dynamical studies of our Galaxy.

Sections 1.2.2–1.2.8 describe GAIA's contributions to the many open issues regarding the structure of the Galaxy in detail. A schematic overview of the main Galaxy components and sub-populations, covering all Galactic distances, is given in Table 1.1, together with the requirements on astrometric accuracy and limiting magnitude. For each sub-population, one or more representative kinematic tracers are indicated. These can be identified reliably in the GAIA data, and their absolute magnitude can be determined from the GAIA astrometry and photometry. Columns 1–5 present the typical absolute magnitude and location in the Galaxy, Columns 6–8 give relevant photometric qualifiers, and Columns 9–12 list the relevant astrometric parameters. Column 11 demonstrates that the proper motions for single stars of all listed sub-populations in the Galaxy can be measured by GAIA with a significant, and often a superb, precision (for details, see Høg et al. 1999a).

It is appropriate to emphasize how important it is to refer the proper motions in the Galaxy to a non-rotating reference frame (see Section 1.8.11). If this were not done and if the GAIA reference system would simply be that of Hipparcos, it would have a residual uncertainty of 250 μas yr^{-1}. This would mean a systematic error of 6 km s^{-1} at 5 kpc and 12 km s^{-1} at 10 kpc. The Galactic rotation speed at these distances is of the order of 220 km s^{-1}. It is possible, with data from tens of thousands stars, to determine the galactic rotation curve with an internal precision of 1 km s^{-1} (a few μas yr^{-1}). The addition of an error rotation vector in a random direction of several km s^{-1} would introduce biases not only in the rotation speed which is an indicator of the dynamical properties of the Galaxy, but also in the position of the plane of rotation, equivalent to producing a fictitious warp (see also Section 1.8.10).

Table 1.1: Summary of selected Galactic kinematic tracers, and the limiting magnitudes and astrometric information necessary to study them (adapted from Høg et al. 1999a). Columns 7–8 demonstrate that the faint magnitude limit of GAIA is essential for probing these different Galaxy populations, while columns 10–11 demonstrate that GAIA will provide data of adequate precision to meet the scientific goals.

(1) Tracer	(2) M_V mag	(3) ℓ deg	(4) b deg	(5) d kpc	(6) A_V mag	(7) V_1 mag	(8) V_2 mag	(9) ϵ_T km/s	(10) σ_{μ_1} μas/yr	(11) σ'_{μ_1} –	(12) σ'_{π_1} –
Bulge:											
gM	–1	0	< 20	8	2–10	15	20	100	10	0.01	0.10
HB	+0.5	0	< 20	8	2–10	17	20	100	20	0.01	0.20
MS turnoff	+4.5	1	−4	8	0–2	19	21	100	60	0.02	0.6
Spiral arms:											
Cepheids	–4	all	< 10	10	3–7	14	18	7	5	0.03	0.06
B–M Supergiants	–5	all	< 10	10	3–7	13	17	7	4	0.03	0.05
Perseus arm (B)	–2	140	< 10	2	2–6	12	16	10	3	0.01	0.01
Thin disk:											
gK	–1	0	< 15	8	1–5	14	18	40	6	0.01	0.06
gK	–1	180	< 15	10	1–5	15	19	10	8	0.04	0.10
Disk warp (gM)	–1	all	< 20	10	1–5	15	19	10	8	0.04	0.10
Disk asymmetry (gM)	–1	all	< 20	20	1–5	16	20	10	15	0.14	0.4
Thick disk:											
Miras, gK	–1	0	< 30	8	2	15	19	50	10	0.01	0.10
HB	+0.5	0	< 30	8	2	15	19	50	20	0.02	0.20
Miras, gK	–1	180	< 30	20	2	15	21	30	25	0.08	0.65
HB	+0.5	180	< 30	20	2	15	19	30	60	0.20	1.5
Halo:											
gG	–1	all	< 20	8	2–3	13	21	100	10	0.01	0.10
HB	+0.5	all	> 20	30	0	13	21	100	35	0.05	1.4
Gravity, $\mathcal{K}_Z$:											
dK	+7–8	all	all	2	0	12	20	20	60	0.01	0.16
dF8-dG2	+5–6	all	all	2	0	12	20	20	20	0.01	0.05
Globular clusters (gK)	+1	all	all	50	0	12	21	100	10	0.01	0.10
internal kinematics (gK)	+1	all	all	8	0	13	17	15	10	0.02	0.10
Satellite orbits (gM)	–1	all	all	100	0	13	20	100	60	0.3	8

Key: (1) tracer of a specific sub-population. HB: signifies blue horizontal branch, RR Lyrae, red horizontal branch, or clump giants, as appropriate; for globular clusters, the cluster itself is used as tracer, with its proper motion the mean of many cluster star motions; globular cluster kinematics refers to the internal kinematics of globular clusters of our own Galaxy; (2) absolute visual magnitude of a typical tracer star; (3–4) appropriate Galactic coordinates; (5) optimal distance, or upper distance limit; (6) typical visual extinction along the line of sight; for low latitudes the extinction in a Galactic window is given; (7–8) typical range of apparent visual magnitudes; (9) expected velocity dispersion for tracer stars of the sub-population, in the proper motion direction; (10) expected GAIA proper motion standard error, σ_{μ_1}, for a single star at representative magnitude; (11) relative proper motion error; (12) expected relative error on GAIA astrometric distances at representative magnitudes.

1.2.2 The Bulge

Bulge stars are predominantly moderately old, unlike the present-day disk; they encompass a wide abundance range, peaking near the Solar value, as does the disk; and they have very low specific angular momentum, similar to stars in the halo. Thus the bulge is, in some fundamental parameters, unlike both disk and halo. What is its history? Is it a remnant of a disk instability? Is it a successor or a precursor to the stellar halo? Is it a merger remnant? It is not clear whether the formation of the bulge preceded that of the disk, as predicted by the Larson (1990) 'inside-out' scenario; or whether it happened simultaneously with the formation of the disk, by accretion of dwarf galaxies; or whether it followed the formation of the disk, as a result of the dynamical evolution of a bar (see review by Wyse et al. 1997). Large-scale surveys of proper motions and photometric data inside the bulge can cast light on the orbital distribution function. Knowing the distance, the true space velocities and orbits can be derived, thus providing constraints on current dynamical theories of formation (Wyse 1997).

To date, studies of the bulge are hampered by the superposition of much foreground structure, by extreme reddening, by the presence of the central parts of the disk and the halo, by what may well be an unrelated dense central stellar cluster (Section 1.10), and by the presence of continuing star formation, whose population provenance remains unclear. Given this complexity, it is evident that very considerable data sets, mapping as much of phase space as is possible, are required. One must determine distribution functions over age, metallicity, element ratios, location, angular momentum, and orbital characteristics. In fact, ideally, what is demanded is a large astrometric survey (including radial velocities), complemented by detailed chemical abundance studies of a suitable representative subsample. Addressing this important problem requires data obtained in fields which are both crowded and highly reddened, imposing severe constraints on the observational techniques adopted.

The highly accurate parallaxes, proper motions and magnitudes acquired by GAIA for more than 10^6 stars per square degree, will allow the vast majority of red and asymptotic giant branch stars, and a significant fraction of the clump stars in Baade's Window to be measured with a precision higher than 10–15 per cent. With $V = 20$ as the limiting magnitude, red and asymptotic giant branch stars can be detected over a range of 5 mag.

The numbers quoted here are based on analysis of HST images, and on simulations of the stellar populations in the bulge by means of the so-called HRD-Galactic Software Telescope (Ng et al. 1995; Bertelli et al. 1995; Schmidtobreick et al. 1998), which provides detailed predictions about the colour-magnitude diagram and luminosity function that one would observe along any line of sight. The synthetic H-R diagrams are calculated by means of the Padova library of stellar tracks (Girardi et al. 1996 for Z = 0.0001; Bertelli et al. 1990 for Z = 0.001; Bressan et al. 1993 for Z = 0.020; Fagotto et al. 1994 for Z = 0.0004, 0.004, 0.008, 0.05, and 0.10). Three different stellar populations are used to simulate the stellar content of the bulge. Either an old population, 11–12 Gyr, with metal content Z = 0.001–0.05, or an old, 11–12 Gyr, but more metal-rich population with Z = 0.005–0.05, or a younger but metal-rich component with ages from 7–10 Gyr and Z = 0.008–0.05. Various mass distributions are considered. The simulations take into account the contamination by disk and halo stars. The Sun galacto-centric distance is assumed to be 8 kpc, and the reddening law is taken from Ng et al. (1996).

Shape There is substantial evidence that the bulge is not axisymmetric, but instead has a triaxial shape seen nearly end-on. Indications for this come from the asymmetric near-infrared light distribution, star counts, the atomic and molecular gas morphology and kinematics, and the large optical depth to micro-lensing (Gerhard 1996; Kuijken 1996; Unavane et al. 1998; Zhao & de Zeeuw 1998; Freudenreich 1998; Vauterin & Dejonghe 1998; Dehnen 1998). The shape, orientation, and scale-length of the bulge, and the possible presence of an additional bar-like structure in the disk plane, remain a matter of debate. The GAIA proper motions to faint magnitudes, in particular in a number of low-extinction windows, will allow unambiguous determination of the shape, orientation, tumbling rate and mass profile of the bulge. The large-scale kinematics of the Galaxy also contains imprints of the non-axisymmetric central potential (see Section 1.2.3).

Several models for the mass density distribution in the bulge have been proposed, including: (i) a simple power law with index $n = 3$ (Ng et al. 1997), in which the possible flattening of the spheroid perpendicular to the Galactic plane is a free parameter; (ii) the COBE-DIRBE oblate spheroidal G0-model of Dwek et al. (1995), which is a Gaussian-type function, with axial ratios $1:1:0.56$, similar to the values derived from the infrared maps by Kent et al. (1991) and Weiland et al. (1994); (iii) the COBE-DIRBE constrained triaxial model of Kent et al. (1991), as modified according to model G2 by Dwek et al. (1995), which is a boxy Gaussian model with axial ratios $1:0.22:0.16$; (iv) the triaxial model B1, introduced by Binney et al. (1997b) to fit COBE-DIRBE data with the best-fit parameters by Bissantz et al. (1997), and axial ratios $1:0.6:0.4$; (v) the exponential triaxial function according to model E2 by Dwek et al. (1995), based on IRAS data for Mira variables (Whitelock et al. 1991), with axial ratios $1:0.18:0.39$, considered by Dwek et al. (1995) to give the best fit to the COBE-DIRBE observations.

The angle between the Sun-centre line and the major axis of the bar varies from 25° (Bissantz et al. 1997) to 20° as in the G2 model of Dwek et al. (1995), or even to 33° as for model E2 by the same authors. The reason why it is so difficult to derive the shape of the Galactic bar is that three-dimensional distributions cannot be uniquely recovered from projected surface brightness distributions such as the COBE/DIRBE maps. In addition, bars with the same density distribution could have different pattern speeds. No unique solution can be found using only one-velocity component diagrams, unless the gravitational potential is known, since the velocity dispersion in the star motions smears out the effects of the bar on the distribution function (Vauterin & Dejonghe 1998). Groundbased proper motion studies in Baade's Window (Spaenhauer et al. 1992) to $V \sim 16$ are contaminated by foreground stars, but suggest a velocity structure consistent with a tumbling triaxial bar (e.g., Zhao et al. 1994; Zhao et al. 1996; Zhao 1996). Significant constraints on the bulge shape will require proper motions for many thousands of stars.

Metal Content Rapid formation of the bulge would imply that the bulk of its stellar content ought to display α-enhanced abundances as expected when only Type II supernovae contribute to the enrichment process. In contrast, slow formation would imply that the vast majority of stars

Table 1.2: Expected turnoff magnitude of the bulge population (age = 7–10 Gyr, 0.008≤ Z ≤0.05).

Mass Model	Axial ratios	Major Axis (kpc)	Bar Position Angle (°)	Turn-off Magnitude
Power law	1 : 1 : 0.8	2.4	0	19.45
Oblate G0	1 : 1 : 0.56	0.70	0	19.45
Triaxial G2	1 : 0.22 : 0.16	2.0	20	19.45
Triaxial E2	1 : 0.18 : 0.39	0.75	33	20.05
Triaxial B1	1 : 0.6 : 0.4	1.8	25	19.35

have [α-elements/Fe] = 0.0 as expected when both Type I and Type II supernovae are at work (e.g., Matteucci 1998). GAIA will provide estimates of the metallicity for $\sim 10^4 - 10^5$ stars per square degree inside Baade's Window. These provide significant constraints, which can be further improved by follow-up high-resolution ground-based spectroscopy.

To date, the abundances of α-elements in bulge stars are known for a few dozen stars (McWilliam & Rich 1994; Barbuy et al. 1999). The mean [Fe/H] is about –0.25, with a large spread, ranging from [Fe/H] = –1 to 0.45. For an old population, an uncertainty as large as Δ[Fe/H] = 0.1 results in a difference of about $\Delta = 0.07$ mag for the turn-off luminosity. If no information about the metal content is available, this magnitude difference can be misinterpreted as being due to age, giving an age error of about 1 Gyr.

Age There is no general agreement on the age and age spread for the stars in the bulge (e.g., Rich 1992; Wyse et al. 1997). GAIA will substantially reduce these uncertainties.

While all determinations of ages agree that the bulge ought to be older than 5–8 Gyr, they diverge on whether the presence of a component substantially younger than 15 Gyr is required (Renzini 1993; Rich 1988; Terndrup 1988; Holtzman et al. 1993; Ortolani et al. 1995; Bertelli et al. 1995). Several causes make it difficult to derive firm ages for the stars in the Galactic bulge by means of colour-magnitude diagrams and luminosity function in the interval $15 \leq V \leq 17$. These include the age-metallicity degeneracy, the photometric errors on magnitudes, the presence of binary stars blurring the turn-off, contamination by disk stars, and the uncertainties on reddening along the line of sight. An additional complication is the spatial distribution of the bulge stars along the line of sight due to the intrinsic shape of the bulge itself. Current estimates of the age are based on the implicit assumptions that the bulge is spherically symmetric, and that all stars are at the same distance.

Simulations of the colour-magnitude diagram for stellar populations with different spatial distributions show that the magnitude of the turn-off and main sequence termination stage depend on the assumed shape of the bulge. Table 1.2 illustrates this by presenting the expected turn-off magnitude for different shapes, limited to the case of a young population of 7–10 Gyr. A population of ∼12 Gyr would put the turnoff-magnitude 0.35 mag fainter. Changing the mass distribution changes the turn-off magnitude, e.g., from 19.45 (power law) to 20.05 (triaxial E2).

1.2.3 Large-Scale Structure of the Disk

Star formation has been reasonably continuous in the disk of the Galaxy over the past 12 Gyr and, as a result, the disk contains stars with a range of metallicities, ages and kinematics. In the past decades radio and mm observations, combined with kinematic models, have revealed the distribution and kinematics of the interstellar gas for nearly the entire Galaxy. It has delineated the spiral structure, and mapped the warping of the Galactic disk outside the Solar orbit. Very little is known about the stellar disk beyond 1–2 kpc from the Sun. This is due to significant interstellar extinction towards the central regions of the Galaxy at optical wavelengths, and by our inability to determine accurate distances and space motions. The GAIA parallaxes, proper motions, radial velocities and photometry will allow derivation of the structure and kinematics throughout the stellar disk for a large fraction of the Milky Way.

Spiral Arms Spiral arms are a distinguishing feature of disk galaxies with an appreciable gaseous component, and are clearly evident in the radio and far-infrared emission of our Galaxy. They are an important component, as they have associated streaming motions, redistribute angular momentum and are the primary locations of star formation, funneling mass from one component (the gas) to another (the stars).

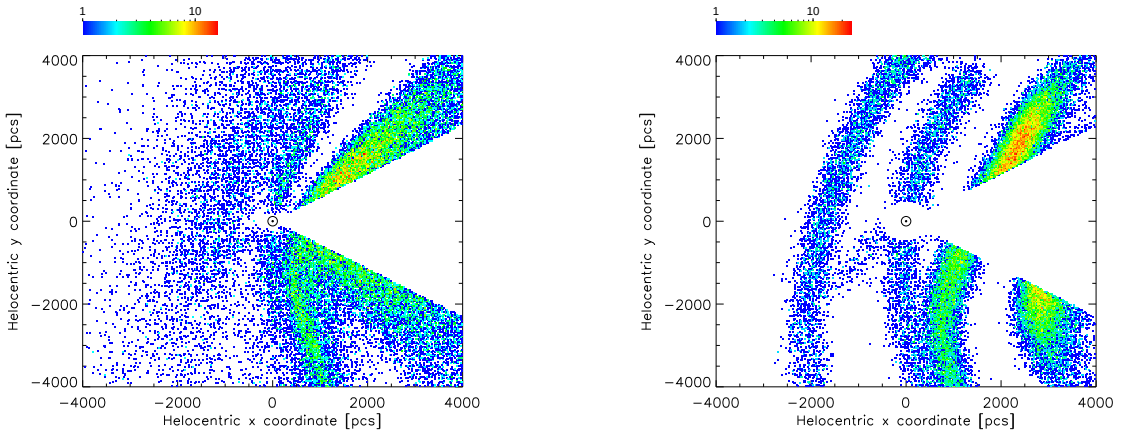


Figure 1.1: Distribution in the Galactic plane of a hypothetical spectroscopic and photometric statistically complete survey of 50 000 OB stars (coordinates are heliocentric, with the Galactic Centre at 8000,0) with photometric distances (left) and with GAIA parallaxes (right), as generated by a synthetic catalogue generator (from Drimmel et al. 1997).

Currently our understanding of the large-scale dynamics and structure of the Galactic disk derives mainly from 21-cm observations of HI, but these observations only provide the density as a function of radial velocity in any given direction, that is, a single velocity component. To infer the actual distribution of the gas, and its kinematics, we rely upon the assumption of circular rotation. Even with this assumption distances within the Solar circle are ambiguous. Much more problematic is the validity of the assumption itself, which certainly does not hold in the presence of non-axisymmetric structure, such as the spiral arms and the triaxial structure of the bulge (which makes our Galaxy barred). It is also possible that the galactic disk is elliptic and/or lopsided, as is seen in many other disk galaxies (Rix & Zaritsky 1995; Rudnick & Rix 1998). All of these structures will introduce non-circular motions (see below).

The current map of the spiral arms relies on both photometric and kinematic distances to bright HII regions, though primarily on the latter (Georgelin & Georgelin 1976). Distance ambiguity is resolved by assuming that larger regions are closer. The kinematic distances follow from the assumption of circular motions, yet these regions are located where the streaming motions associated with the spiral arms are strongest (Figure 1.1).

The Hipparcos Catalogue has provided an accurate spatial and kinematic map of the local Orion arm in the vicinity of the Sun. GAIA will provide a direct map of all the major arms on our side of the Galaxy, using the location of young tracer populations. It will identify what constitutes an arm, its kinematic signature, and its stellar population mix.

Because star formation occurs along a moving front, and because stellar lifetimes are a strong function of initial mass, the luminosity function at the bright end will vary strongly with location; the distribution of more luminous stars will be offset from the less luminous stars. GAIA will be able to directly observe the relative offsets of various stellar populations (or types), from which the pattern velocity of the arms may be inferred. In addition, the initial luminosity function (or initial mass function) for the bright stars may be determined at different locations in the Galaxy. And finally, as these short-lived stars approximately preserve the bulk motions of the gas from which they were recently born, they may be compared with the observed streaming motions of the gas derived from HI and CO measurements. These streaming motions are of the same order as the velocity dispersion observed for young stars ($\sim$10 km s^{-1}), and will vary according to location within the arms. Therefore, to look for such systematic motions in these stars, one must densely sample the stellar distribution function at each position in the Galaxy. Only a survey mission such as GAIA will be capable of doing this.

Kinematics of the Galactic Disk The models used to describe the stellar kinematics of the Galactic disk have become ever more sophisticated as the quality of the observations improved, starting from the ellipsoidal and Gaussian model for the velocity distribution in the Solar neighbourhood (Schwarzschild 1907), and the description of the well-known longitudinal wave-like dependence of the mean proper motion and radial velocity as due to differential Galactic rotation (Oort 1928). However, it was known early-on that this picture is too simple: Eddington (1914)

already argued that such a model cannot describe equilibrium, while Kapteyn (1905) and others pointed out that the distribution of stellar velocities in the Solar neighbourhood is not smooth but clumpy, in particular for early-type stars. Clumps in the velocity distribution correspond to moving groups and stellar streams, which Eggen (1965) hypothesized to be the debris of stellar associations and open clusters (see Section 1.2.4).

Hipparcos measured the velocity distribution in the Solar neighbourhood (Dehnen & Binney 1998) for large stellar samples free of kinematic biases, which in the past plagued studies of local stellar kinematics. These data revealed that the local stellar velocity distribution is indeed much more structured and complex than Schwarzschild's simple model: there is substantial sub-structure even for late-type, and hence on average old, stars. The interpretation of these structures is not yet complete, but it is clear already that the properties of both the stellar warp outside the Solar circle (Dehnen 1998) and the bar in the inner Galaxy (Dehnen 1999) have left their imprints.

GAIA will measure not only the local kinematics with much improved accuracy, but the full six-dimensional stellar distribution function throughout a large part of the Galactic disk. This will allow not only a determination of the gravitational potential of the Galaxy and its distribution function, but also reveal how much a given stellar population deviates from dynamical equilibrium. This in turn will constrain the formation history of the Galactic disk and its components, e.g., the past variations of pattern speed and strength of the central bar and spiral arms.

Age-Metallicity-Velocity Relation for the Galactic Disk As the Galactic disk settled within its dark halo, star formation began in the denser inner regions and gradually extended outwards into the disk of gas. The massive stars evolved and chemically enriched the interstellar medium from which the next generation of stars formed. Dynamical evolution also took place: disk stars were perturbed from their circular orbits by passing giant molecular clouds (Spitzer & Schwarzschild 1953) and spiral arms. As the disk evolved, it became chemically enriched and its stellar velocity dispersion increased. If this picture is correct, we would also expect to see that the mean age of disk stars is younger at larger distances from the centre. Evolution of the disk is still going on, in our Galaxy and in other disk galaxies (e.g. Bell & de Jong 2000). Clearly, we still do not completely understand the mechanism by which the velocity dispersion of disk stars grow. Although both molecular clouds and spiral arms appear to be capable of heating the disk at about the observed rate, neither mechanism can explain all of the features of the observed velocity dispersion versus age relation. Other mechanisms, such as black holes in the galactic halo, have also been discussed (see also Velazquez & White 1999; Sanchez-Salcedo 1999).

To understand the evolution of disks in a more quantitative way, we need to know the full multivariate distribution of stars over the five variables: age, metallicity and the three components of velocity. Other disk galaxies give only a qualitative indication of the evolutionary processes, because we can only estimate the average age, metallicity and motions from the spectra of integrated disk light, and they are of low accuracy.

In our Galaxy, we can measure accurate ages, metallicities and kinematics of individual stars in the disk and thick disk, and so construct the full multivariate distribution. These observations are very difficult from the ground: precise stellar ages and velocities can be measured only for stars in the nearest few hundred parsecs. A seminal study by Edvardsson et al. (1993) showed how complex the situation is in the Solar neighbourhood. The thin disk near the Sun shows a chemical abundance spread of almost a factor 10, with little change over the age range 3–10 Gyr. The thick disk stand out as an old population (age $>$ 10 Gyr) with a vertical velocity dispersion of about 40 km s^{-1}, double that of the oldest stars in the thin disk, and with significantly lower chemical abundance. The Edvardsson sample of $\sim$200 stars is in no sense complete and does not provide the distribution of stars with age, even in the Solar neighbourhood. It just gives an indication of how the disk properties change with stellar age.

With GAIA, the ability to measure precise distances of distant stars provides the breakthough. The most accurate ages for individual older stars come from the subgiant G stars which are evolving across from the main sequence turnoff to the bottom of the red giant branch. Their absolute magnitude and metallicity give their ages directly. The stellar ages are needed to a precision of about 25 per cent, so their distances must be measured to about 7 per cent. Away from the obscuring galactic plane, GAIA will give a complete sample of G subgiants with accurate kinematics and distances, out to about 3 kpc from the sun. The GAIA sample will cover a very significant

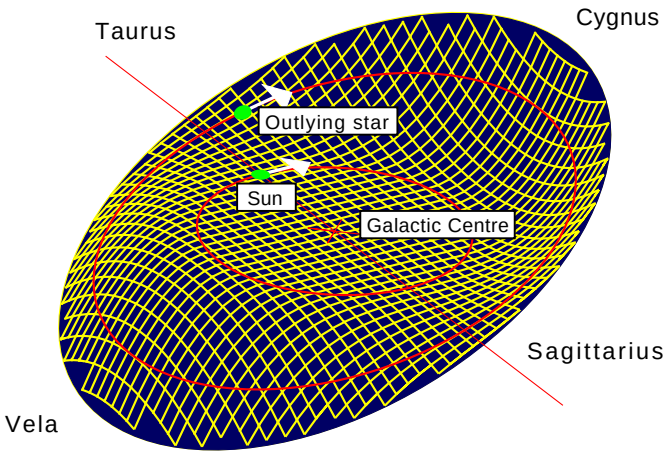


Figure 1.2: The disk of the Galaxy based on HI observations (from Smart et al. 1998). The vertical axis is exaggerated by a factor of 10. The arrows show the motion of the Sun and an outlying star on their orbits in the Galaxy. The outlying star has an upward motion as seen from the Sun. Directions in the sky, as seen from the Earth, are indicated by the constellation names. The line indicates the locus of zero vertical displacement for the warped disk.

radial zone in the galactic disk, from about 2 to 4 scale-lengths from the centre. Ground-based metallicity data will complement chemical abundance estimates from GAIA photometry. From this sample of subgiants, it will be possible to construct the full multivariate distribution of kinematics, age and metallicity to unravel the star formation history of the disk, and its chemical and dynamical evolution, over a wide interval in Galactocentric radius.

Warp Galactic disks are thin, but they are not flat. Approximately one-half of all spiral galaxies have disks which warp significantly out of the plane defined by the inner galaxy (Burton 1998; Binney 1992). Remarkably, there is no realistic explanation of this common phenomenon, though the large-scale structure of the dark matter, and tidal interactions, must be important, as the local potential at the warp must be implicated. Neither the origin nor the persistence of galaxy warps is understood, and insufficient information exists to define empirically the relative spatial and kinematic distributions of the young (OB) stars which should trace the gas distribution, and the older (gKM) stars which define a more time-averaged gravitational field.

It has been known since 1957, following the first 21-cm sky surveys, that our Galaxy is strongly warped (see Figure 1.2). The line of nodes is close to the Sun-centre line. In the north ($\ell = 90°$) the HI disk curls steadily up to over 3 kpc above the plane at $R \sim 16$ kpc. In the south ($\ell = 270°$) the disk curves about 1 kpc below the plane, before turning back towards the plane at $R \sim 15$ kpc, nearly reaching the plane again at $R \sim 20$ kpc.

The kinematic data from Hipparcos (Smart et al. 1998), limited to within $\sim$3 kpc of the Sun, provides confusing and conflicting information. To construct a spatial and kinematic map of the warp to the edge of the (optical) disk, sufficient distance resolution is needed in order to define the disk plane beyond the Solar circle, independently for young and older tracers. At the probable disk edge, $R = 15$ kpc, 7 kpc from the Sun, the warp induces a mean offset out of the plane of ~ 1 kpc, which one needs to observe in a large sample of early-type and late-type stars with apparent magnitudes near $I = 15$ mag.

The expected kinematic pattern (at least, in existing plausible models) is most strongly constrained by the straightness of the line of nodes: these should wind up in at most a few rotation times, typically less than 2 Gyr. A relevant shear pattern corresponds to systematic motions dependant on warp phase and galactocentric distance superimposed on Galactic rotation. A plausible velocity amplitude associated with the warp at the optical disk edge is significantly less than 0.1Ω, with Ω the disk rotation angular velocity. This will be distributed between latitude and longitude contributions depending on the local geometric projection.

At $R = 15$ kpc, for a flat rotation curve, the systematic disk rotation corresponds to 6 mas yr^{-1}. The kinematic signature from a 1 kpc-high warp corresponds to a systematic effect of ~ 90 μas yr^{-1} in latitude and ~ 600 μas yr^{-1} in longitude. For such a signal to be detected the reference frame

must be rigid to better than a few microarcsec on scales of $\sim 10^\circ$ (i.e. matching the high-frequency warp structure) and on scales of 2π radians, requirements well within the GAIA capabilities. The corresponding distance requirements are more demanding: at the warp a mean parallax is less than 100 μas, so that resolution of the warp within 10 per cent implies distance accuracies of 10 μas at $I \sim 15$ mag. Along lines of sight with typical reddening, the study of the Galactic warp will be within the limits of GAIA's performance.

Although HI warps are most pronounced near the optical 'edge' of disk galaxies, recent surveys show some 50 per cent of all optical disks are warped in their outer parts (Reshetnikov & Combes 1998). In the Local Group the stellar disk does follow the HI warp, most dramatically in M31 (Section 1.8.4), but also in our own Galaxy. These provide the opportunity to define the structure of the warp in apparently representative spiral galaxies.

Extinction and the Distribution of Galactic Interstellar Matter The combination of GAIA parallaxes with GAIA photometry over a large part of the visual spectrum will provide a database of unprecedented size and accuracy to investigate the distribution of interstellar matter (Chen et al. 1999b). The dust embedded in the gas causes extinction of star light, both in terms of dimming, A_λ, and as a colour excess, e.g., $E(B-V)$. The inferred $E(B-V)$ can then be used, through the known correlation of extinction and column density of neutral gas $N(\mathrm{H\,I})$, to estimate the amount of gas along the pathlength to the star. The power of this method was demonstrated by Knude & Høg (1999) using Hipparcos data. Important topics which can be addressed with GAIA data are the optical thickness of the Milky Way disk and the scale length of the dust distribution.

Gas column densities follow from H I 21-cm line emission, but three aspects complicate their interpretation: (i) for large column densities the 21-cm emission may be self-shielded, so that the integrated intensity is not representative of the column of gas. Since the effect of extinction on intensity is as $e^{-\tau_\lambda}$, the true amount of dust (and thus gas) can be determined; (ii) emission can come from gas at any distance in the direction measured, while the derived extinction is only due to gas in front of the star. Using many stars at different distances in almost the same direction allows mapping of the location of the gas; (iii) the optical and H I measurements are made with different beamwidths.

1.2.4 Small-Scale Structure of the Disk

The Galactic disk contains much small-scale structure, including star forming regions, open clusters, expanding associations, and stellar streams. The GAIA data will allow a vastly improved identification of member stars, and, in many cases, a study of the internal group dynamics. This is also of key importance for the calibration of stellar luminosities, and the calibration of the main sequence as a function of age, helium content, and metallicity (see Section 1.4).

Finding Sub-Structure The internal velocity dispersions in open clusters, associations, and streams are at most a few km s^{-1}. Members of such a 'moving group' therefore have nearly the same space velocity, and form a coherent structure in velocity space. A variety of methods is available to identify such groups in phase space; these use the measured positions and velocities.

Members of moving groups which cover at most a few square degrees on the sky (such as open and globular clusters beyond 200 pc) can be identified in the distribution of proper motions in the vector-point diagram (Vasilevskis et al. 1958), as illustrated in Figure 1.8 for NGC 6397. In this case the spread in proper motions is mostly due to measurement errors and the internal velocity dispersion of the cluster. Selecting members of groups that cover larger areas on the sky, such as the Hyades and Pleiades, or the large nearby associations which cover ∼1000 square degrees, is more difficult. The constant space velocity of such groups projects to a proper motion and radial velocity which vary significantly across the field. This perspective effect will wash out the group signature in the vector-point diagram. In this case the classical convergent point method (e.g., de Bruijne 1999) can be used, as well as methods which include parallax measurements (Hoogerwerf & Aguilar 1999) and/or radial velocities. Streams of stars that have escaped from open clusters due to the Galactic tidal field also form coherent structures in velocity space, but may cover even larger areas on the sky, with the space velocity of the member stars varying along the stream. Such structures can be identified by looking for concentrations in the space of the integrals of motion (Comerón et al. 1998a; Helmi et al. 1999b).

The Hipparcos data have allowed extension of the volume where kinematic member selection over large fields can be carried out reliably to about 500 pc from the Sun, well beyond the range where individual parallaxes resolve the groups. Examples include a comprehensive census for nearby OB associations (de Zeeuw et al. 1999), see Figure 1.3, detection of nearby moving groups (Chereul

Table 1.3: Statistics for clusters like Praesepe, as observed by GAIA at several distances. Praesepe, like the Hyades and the Pleiades, is a quite rich cluster with more than 1000 known members.

Distance	180 pc	300 pc	1 kpc	3 kpc	5 kpc
Number of clusters	6	10	120	∼1000	∼3000
$(m-M)_0$	6.3	7.4	10	12.4	13.5
A_V	< 0.5	< 0.5	0.5–2.5	1–4	2–8
$V = 15$:					
Number of stars	450	350	50–150	0–20	0
Depth resolution (pc)	0.3	1	10	90	250
Accuracy of V_T (km s^{-1})	<0.01	0.015	0.05	0.15	0.25
σ_{m-M}	0.004	0.007	0.02	0.1	0.2
$V = 20$:					
Number of stars	> 1000	1000	350–550	50–300	0–150
Depth resolution (pc)	10	25	250	–	–
Accuracy of V_T (km s^{-1})	0.25	0.4	1.4	4	7
σ_{m-M}	0.12	0.3	0.5	1.5	2

et al. 1999), and the identification of a stellar stream in the distribution of Population II stars in the Solar neighbourhood (Helmi et al. 1999a; see also Section 1.2.6). GAIA will allow similar kinematic selection in a much larger volume, which includes all the globular clusters (see Section 1.2.7), as well as the dwarf spheroidal satellites of the Galaxy (see Section 1.8.2). This will provide a complete and reliable list of members to faint magnitudes. Combining the individual parameters of members, the cluster mean distances and velocities will be improved. The precision of mean astrometric parameters of groups closer than 1 kpc will be better than 1 μas (see Section 1.4.9).

Open Clusters An open cluster contains a few tens to a few thousand stars, born in the same molecular cloud and gravitationally bound. Members of an open cluster are coeval, have the same chemical composition as the parental cloud, share the same space motion to within the internal velocity dispersion of the cluster (a few tenths of a km s^{-1}), and are distributed across the whole mass spectrum, from brown dwarfs to A, B or O stars, depending on the age of the cluster. GAIA will observe all known open clusters, and will discover thousands of new ones. Simulations of fairly rich clusters at different distances show that GAIA will be able to identify members of the vast majority of clusters closer than 5 kpc (Table 1.3).

More than one thousand open clusters are presently known (LyngAstron. Astrophys. 1987). Only half of them have an estimated distance, mostly smaller than 2 kpc. They are tracers of the young and intermediate-age disk components of the Galaxy, with ages ranging from a few Myrs to a few Gyrs, and metallicities [Fe/H] from -0.5 dex to $+0.2$ dex. The possibility of detecting a cluster depends on several parameters: the number of members brighter than $V = 17$ mag (for which GAIA will measure radial velocities and medium-band photometry) or brighter than the limiting magnitude $V = 20$ mag, the magnitude of the brightest stars, the interstellar absorption, the differential velocity with respect to that of the local standard of rest. These parameters are connected to the intrinsic parameters of the cluster, namely its age, the total number of members (which also depends on its age), and its space velocity. A young rich cluster such as NGC 6231, containing WR and O stars, has more than 30 stars brighter than $V = 17$ mag even at the distance of the Large Magellanic Cloud. An old cluster such as NGC 752, in a region of high absorption ($A_V = 5$) has very few stars brighter than $V = 17$ mag at only 2 kpc.

OB Associations An OB association is a gravitationally unbound group of stars, including a number of early spectral type (O- and B-stars). Associations are thought to originate as a group born in a relatively small region of space inside a giant molecular cloud. After the removal of gas and dust the association is left unbound, and expands. This scenario is supported by the large physical sizes (up to several tens of parsecs in diameter) and low densities of associations. As much as 90 per cent of the stars in the Galaxy may have formed in OB associations.

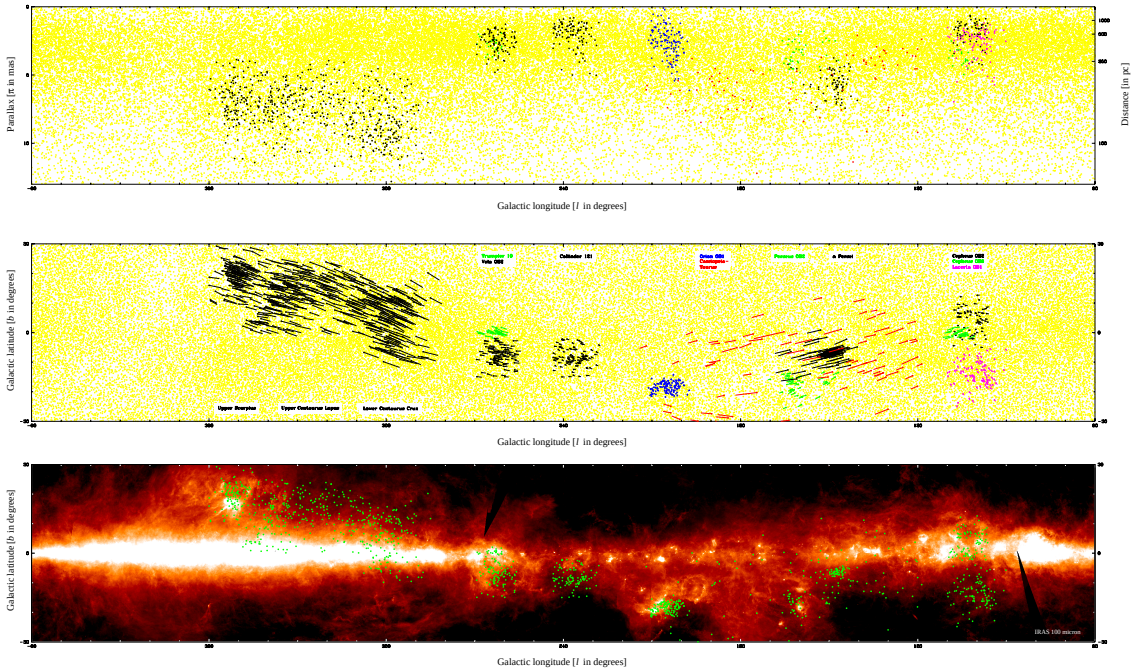


Figure 1.3: Kinematic selection of nearby associations using Hipparcos positions, parallaxes, and proper motions: (a) parallaxes of the OB association members, superimposed on all stars observed by Hipparcos in the range –30° < b <30°; (b) positions and proper motions of the members; (c) positions of the members superimposed on the IRAS 100 μm background.

The large extent of nearby OB associations on the sky has traditionally prevented accurate kinematic membership determination for any but the brightest stars. Hipparcos provided a major step forward for stars more massive than the Sun in about a dozen OB associations nearer than 500 pc (de Zeeuw et al. 1999). GAIA will detect associations to much larger distances. For those nearer than 2 kpc, reliable member selection will be possible down to the smallest stellar masses, and so will include the objects that are still on their way to the main sequence, including objects bright in X-rays (see Section 1.3.3).

The current positions and velocities of stars in nearby OB associations can be used to trace the stellar orbits back in time to their formation location. This requires accurate knowledge of the Galactic potential in the Solar neighbourhood including granulary structures caused by, e.g., giant molecular clouds or massive open clusters. Current data do not provide such a detailed description of the local potential, and are not sufficiently accurate to measure the expansion (Brown et al. 1997), but GAIA will: detailed N-body simulations of realistic simulated OB associations observed with GAIA accuracy show that the age and initial spatial configuration of expanding associations can be reconstructed out to distances of ∼1.5 kpc for ages younger than ∼20 Myr (Reuland et al. 2000).

Superclusters and Moving Groups Eggen (1994) defined a 'supercluster' as a group of gravitationally unbound stars that share the same kinematics and may occupy extended regions of the Galaxy, and a 'moving group' as the part of the supercluster that can be observed from the Earth. Two factors act against their persistence: Galactic rotation and disk heating. It is therefore striking that some classical moving groups are several hundred million years old.

Superclusters can originate in a number of ways. Stellar evolution, removal of the parental molecular material by stellar winds and supernovae, internal dynamical evolution, and the Galactic tidal field together cause open clusters to gradually become unbound, resulting in a 'corona' of co-movers, and star streams that are mixed with the general field star population. They may also result from the dissolution of larger star agglomerations, such as star complexes (see below). Alternatively, they could represent orbit structure caused by large-scale non-axisymmetry of the Galactic disk, as described in Section 1.2.3.

Hipparcos has: (i) confirmed the existence of moving groups (e.g. Chereul et al. 1999; de Zeeuw et al. 1999), detecting members up to distances of about 500 pc; (ii) established that the fraction of moving group members in the field star population is around 10–30 per cent; (iii) determined that some moving groups have larger velocity dispersions (4–8 km s^{-1}) than previously thought (not explained by the propagation of Hipparcos observational errors, which lead to contributions of less than about 2 km s^{-1} for samples at $r < 300$ pc, Figueras et al. 1997); (iv) detected sub-structures inside classical moving groups, probably related to their formation process (Asiain et al. 1999; de Zeeuw et al. 1999). The GAIA data will allow detection of this sub-structure out to a few kpc.

The Gould Belt Most stars in the Solar neighbourhood younger than about 60 Myr are located in a flattened structure a few hundred parsecs in size, with the Sun inside it. This is known as Gould's Belt (see the review by Pöppel 1997), and contains many of the nearby OB associations, many young low-mass stars which are bright in X-rays, and much interstellar atomic and molecular gas. Its total stellar mass is estimated to be a few times $10^5 M_\odot$. Remarkably, Gould's Belt is inclined by nearly 20° relative to the general stellar population of the Galactic plane, which has an average age of several billion years. The kinematics of the Gould Belt stars cannot be explained by differential galactic rotation alone, but shows evidence for an additional expansion and rotation of the system as a whole (Lindblad et al. 1997; Comerón 1999).

The distribution and kinematics of the atomic and molecular gas associated with the Gould Belt is well studied. Much less is known about the stars in the Belt, despite the major contributions by Hipparcos for the massive stars (de Zeeuw et al. 1999), and by ROSAT for low-mass stars (Guillout et al. 1998). GAIA will allow measurement with high accuracy of (i) the distance of the young active stars which form the bulk of the Gould Belt, mapping the details of the structure; (ii) the kinematics of the Gould Belt stars; and (iii) determine individual stellar ages through medium-band photometry, which will be of great importance in defining the age of this structure (Torra et al. 1997). These measurements will allow a full understanding of the Gould Belt, and a reconstruction of the sequence of star formation in the Solar neighbourhood in the past 60 Myr.

A large number of X-ray selected candidate Gould Belt members will come from new X-ray telescopes such as CHANDRA and XMM, which will probe smaller areas of sky with much higher X-ray sensitivity. The portion of the Gould Belt nearest to the Sun has a distance of ~ 100 pc, while the more distant part should be at a maximum distance smaller than 800–1000 pc. Up to this distance GAIA will be able to measure all the stars down to early-M spectral types, and to obtain spectra (and hence activity indicators) down to mid-K (see Section 1.3.3).

Star Complexes and Star-Forming Regions The Gould Belt system is an example of a 'star-complex': a grouping of stars hundreds of pc in size and up to 10^8 yr in age (Efremov 1988). It is unlikely that the Gould Belt system is unique in our Galaxy, so that an understanding of this system will help the detection of comparable systems at larger distances (Efremov 1995). Detailed knowledge of the local star formation history is also of great value in the study of extragalactic Gould Belt analogs (i.e., massive-star formation in spiral disk galaxies) and starburst galaxies.

The concept of star complexes has strong theoretical support (Elmegreen & Elmegreen 1983). Regular spacing between star-gas complexes along spiral arms in a number of distant galaxies was first discovered by Elmegreen & Elmegreen (1983), and arguments have been given that the complexes are not random agglomerations of unrelated entities, but that they are more likely the product of large-scale instabilities in galactic star-gas disks (Efremov 1994). Star complexes are probably characteristic of all disk galaxies which contain sufficient amounts of gas-dust matter (Efremov 1993). Knowledge of the dynamics, structure and location of the stellar complexes inside spiral arms are essential for understanding both the nature of the complexes and the origin of the arms themselves (Efremov & Chermin 1995). To date, star complexes are only poorly known in our own Galaxy, with the Gould Belt, and two complexes in the Carina-Sagittarius arm (defined through Cepheid data) being the best candidates.

GAIA will: (i) measure the distribution of young OB stars out to significant distances, yielding aggregates detectable out to distances of 3 kpc; (ii) provide clues to the process of fragmentation of a HI-supercloud into Giant Molecular Clouds; (iii) give new insights into the rotation of Galactic star complexes; (iv) provide a physical link with the grand design structure (e.g. spiral arms); (v) establish whether they are regularly spaced inside spiral arms.

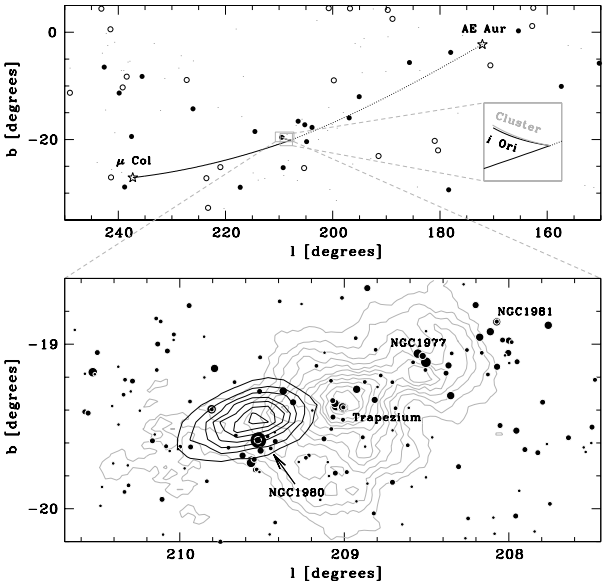


Figure 1.4: Top: orbits, calculated back in time, of the runaways AE Aur (dotted line) and μ Col (solid line) and the binary ι Ori projected on the sky. The starred symbols depict the present position of the two runaways. The stars met $\sim$2.2 Myr ago. Using the conservation of linear momentum principle the orbit of the parent cluster (grey solid line) is calculated from the time of the assumed encounter to the present. The large symbols denote all stars in the Hipparcos Catalogue brighter than $V = 3.5$ mag; filled symbols denote O and B stars, open symbols denoted stars of other spectral type. The small symbols denote the O and B type stars brighter than $V = 5$ mag. Bottom: the predicted position of the parent cluster (contours) together with all stars in the Tycho Catalogue in the field down to $V = 12.4$ mag. The size of the symbols scales with magnitude; the brightest star is ι Ori. The clusters NGC 1980, NGC 1981, NGC 1977, and the Trapezium are indicated. The grey contours display the IRAS 100 μm map, and mainly outline the Orion Nebula.

Run-Away Stars Having mapped the local sub-structure, it will be possible to identify the origin of the many known high-velocity OB stars, which are often found at very significant distances from young stellar groups. They are thought to originate by dynamical ejection from dense young clusters, or in the unbinding of a binary after a supernova explosion of one component.

Nearly 200 run-away stars are known, but the current accuracy of the distances and motions are so modest that it is not possible to retrace their orbits reliably, except for the nearest candidates (Blaauw 1961; Hoogerwerf et al. 2000). The mean motions of the parent associations or clusters are not known very accurately either. The formidable precision of the GAIA distances and motions will allow deriving the orbits of clusters and associations, and establishing the site of the events that produced the run-aways, both the known ones and new ones that will be discovered by GAIA. Figure 1.4 shows the dramatic situation for the run-away stars AE Aur and μ Col, where the Hipparcos results have permitted the origin of these two stars to be traced back to a dynamic ejection encounter with the massive, highly-eccentric binary ι Orionis, 2.2 Myr ago (Hoogerwerf et al. 2000). An example of the detection, by Hipparcos, of a new runaway in CMa, and some further candidates in Cygnus OB2, can be found in Comerón et al. (1998b).

Internal Structure and Dynamics of Stellar Aggregates The insight into the structure and dynamics of the Hyades provided by Hipparcos (Perryman et al. 1998; Brown et al. 1997; de Bruijne & Hoogerwerf 2000) demonstrates dramatically the effect that GAIA astrometry will have on our understanding of the internal dynamics of open clusters (Platais et al. 1995). Only two open clusters (Hyades and Coma Ber) lie within the 100 pc distance horizon yielding Hipparcos distances (from individual objects within the clusters) to better than 10 per cent accuracy.

With GAIA, individual parallax accuracies will be 1 per cent or better for at least 30 clusters within 500 pc. Details of mass segregation, the occurrence of binaries, internal dynamical evolution, cluster evaporation, evidence for tidal distortion and signatures for encounters with giant molecular clouds (Terlevich 1987), and any indications of missing mass, can all be studied for these objects. It will also be possible to reconstruct the Initial Mass Function of clusters younger than a few times 10^8 yr, down to a mass limit ranging from the brown dwarf limit for the Hyades to less than 1 $M_\odot$ at 3 kpc. This will provide a test for star formation scenarios (see Section 1.3.5), and can be compared with mass determinations through binaries (see Section 1.5.6).

N-body computations with accurate treatment of binaries demonstrate that the rapidly decreasing binary proportion, its radial dependence and the form of the period distribution, together with structural and kinematical data, are very useful diagnostics for the present and past dynamical state of a young cluster (Kroupa et al. 1999). The internal dynamics of globular clusters is discussed in Section 1.2.7, and that of the dwarf satellites in Section 1.8.2.

Mass segregation and the velocity dispersion as a function of radius and mass are keys to understand whether the presently observed mass segregation in open clusters is the result of equipartition of energy among members (in which case the velocity dispersion should decrease with the mass of members), or if the mass segregation was established at the time of cluster formation (see e.g., Bonnell & Davies 1998).

The observational errors of the most accurate astrometric data available today (Hipparcos) are, at best, of the same order of magnitude as the velocity dispersions in moving groups, even for the most nearby ones. Thus, the selection of their members is not severely hampered by our understanding of the detailed processes which establish the velocity structure of the groups. This situation will change with the GAIA data. Out to a few kpc, the velocity dispersion of moving groups will be resolved, even for dispersions as small as 0.1 km s^{-1}, so that any pre-conceptions on the characteristics of moving groups might affect/influence the selection of their members. For example, due to equipartition of energy, velocity dispersion depends on stellar mass, massive stars having the smallest and low-mass stars the largest values. Therefore, if one determines the velocity dispersion using bright stars only, some low-mass stars are liable not to be selected. Furthermore, (initial) mass segregation, placing the massive low-dispersion stars in the centre and the low-mass high-dispersion stars in the outer parts, might lead to, e.g., incorrect determination of size, total mass, or total binding energy. Hence member selection and the analysis of the characteristics of the moving group are intricately connected, and will require simultaneous treatment.

1.2.5 The Thick Disk

The oldest stars in the Milky Way are made up of at least the following contributions: any stars which formed *in situ* in what is now the stellar halo and thick disk; any stars which formed *in situ* in the thin disk, and were scattered and/or heated into what is now the thick disk; any old stars in small galaxies which later merged into the Milky Way; and the stars in the intermediate-mass galaxy whose merger created the thick disk, if this is indeed its formation mechanism. Clearly, given this probable diversity of histories, it is naive to imagine the distributions of chemical abundances and kinematics in the present halo and thick disk will be simple, smooth, or gaussian. It has proved surprisingly hard to discover robust deviations from simple, smooth distributions. Conversely, only when deviations from simple distribution functions have been quantified will it begin to be possible to measure the relative importances of these various processes.

Thick Disk Formation The specific questions GAIA addresses concerning the origin of the thick disk are: the evolutionary connection between the halo and the thick disk, and role that sub-structure played in the evolution of the halo and disk. GAIA will do this by determining the kinematics and chemical abundance distributions functions to test if they are smoothly varying and overlapping, or contain significant structure. The merging of sub-structure plays an important role in essentially all models of the formation and evolution of a disk galaxy like the Milky Way Galaxy. Merging is inherent in the popular Cold-Dark-Matter-dominated theory for the origin of structure in the Universe, where larger scales form by the hierarchical clustering of smaller scales. The destruction and assimilation of the small-scale structure is not so well-understood; indeed, the relatively high density of the smaller-mass systems poses a potentially serious problem for CDM-models, since these small, dense systems are predicted to survive to the present day in significantly larger numbers than the observed satellite galaxies (Klypin et al. 1999).

The modern manifestations of 'monolithic collapse' models (cf. Eggen et al. 1962) study the dissipative collapse and spin-up of gaseous baryons within the gravitational potential of a dominant dark halo (Fall & Efstathiou 1980; Dalcanton et al. 1997; Mao et al. 1998). In this latter scenario, fragmentation occurs and stars form in clusters (not necessarily bound) with scale set by the Jeans mass, leading to perhaps proto-globular clusters in the halo phase (Fall & Rees 1985) and open clusters in the disk phase. Field stars are then created from disrupted clusters, as evident in the moving groups seen clearly in the younger stars in the disk (e.g. in the Hipparcos Catalogue, Dehnen 1998), and perhaps still persisting in the halo. These smooth collapse models are attractive in providing an explanation for the observed large scale-lengths of disk galaxies, and for their small scale-heights. The thick disk could have been formed during the initial dissipative collapse, perhaps when some critical metallicity is reached.

Interactions with satellite galaxies are now well identified (e.g. the Sagittarius dwarf spheroidal galaxy, Ibata et al. 1994), and the occurence of a 'minor merger' between the Milky Way and a small satellite galaxy provides an attractive explanation for the thick disk (see Gilmore & Wyse 1985; Gilmore et al. 1989; Freeman 1993; Majewski 1993). The old age derived for the thick disk, of around 10–12 Gyr, or as old as globular clusters of the same metallicity (Gilmore & Wyse 1985; Carney et al. 1989; Gilmore et al. 1995), then requires that any interactions since this look-back time have been mild.

The masses, mass density profiles (and hence potential well depths) and orbits of sub-structure formed by fragmentation will differ greatly from those of sub-structure reflecting the small-scale inherent in the power spectrum of primordial density fluctuations. It is likely that these differences will be manifest in a difference in stellar populations also, such as higher velocity dispersions, and a larger spread in stellar ages and metallicities within a 'primordial' dark-matter-dominated system. However, large N-body and hydrodynamic simulations are only now achieving high enough dynamic range to start to model the Local Group in a cosmological context (Klypin et al. 1999). Further, analytic techniques and smaller-scale N-body simulations investigating the dissolution of sub-structure have been limited to the case of motion within a fixed, time-independent, Galactic potential (e.g. Johnston 1998; Helmi & White 1999), with obvious limitations if one wishes to investigate Cold-Dark-Matter hierarchical clustering scenarios. The data that GAIA will obtain will be the guide as to how to improve upon these early theoretical investigations.

Kinematic Signatures The stellar halo is probably a multi-component entity (e.g. Norris 1994); GAIA will characterise those components, and their inter-relationships, and their relation to the other components of the Galaxy. The kinematics of stars retain memory of the initial conditions at the time of star formation, through the fact that, as expressed in the collisionless Boltzmann equation, phase space density is conserved for collisionless particles (see Barnes 1996). This also allows us to relate the kinematics observed on one part of the orbit to other, more distant parts (cf. May & Binney 1986). The symmetries of the underlying potential in which the stars move determine which quantities, such as energy and angular momentum, are conserved along the orbit (the integrals of motion). The survival of sub-structure inside a larger-scale potential is governed to first order by its density relative to that of the larger body interior to its orbit (Binney & Tremaine 1987); a factor of three is a good rule-of-thumb for a system to survive. The relative mass of the sub-structure is also important, as this determines the timescale for orbital decay by dynamical friction, which brings the satellite galaxy into the denser inner regions, on a time scale of $(M_{\rm galaxy}/M_{\rm satellite}) \cdot t_{\rm cross}$, where $t_{\rm cross}$ is the crossing time of the 'parent' galaxy.

If coupling to a disk is important in removing orbital energy from a satellite, then the sense of the orbit with respect to the disk angular momentum vector is also important; prograde orbits couple more strongly than do retrograde orbits, and thus satellites on prograde orbits are more strongly affected; see e.g. Velazquez & White (1999). Stars removed from a satellite galaxy by tidal effects will remain on orbits close to that of the centre of mass of the satellite, with an energy difference only slightly larger than that corresponding to the internal potential well depth of the satellite galaxy; thus the velocity dispersion of newly-unbound stars is very close to that of the original satellite member stars. The slight differences will cause a tidal stream to form over time (e.g. the analytic treatment of Tremaine 1993). Most mass-loss from a satellite will occur at the pericentric distance of its orbit, its closest approach to the denser central regions of the larger galaxy, due to the relative-density criterion for tidal stripping discussed above. Thus for an elliptical orbit, mass loss can be approximated as occuring in distinct events, at each pericentre passage. A tidal stream so formed will then have low stellar density, and a velocity dispersion of order that of the original satellite. Intersecting streams can give the appearance of a higher velocity dispersion stream.

Most of the mass of the stellar halo is interior to the Solar circle, reflecting the steeper density profile of the halo compared to the disk, so surveys of the stellar halo should, if they wish to trace the typical halo, target the halo within a few kiloparsecs of the Sun (the standard de Vaucouleurs fit to the stellar halo has a de-projected half-mass radius of ~ 3.7 kpc, while that of the stellar disk is approximately 1.7 exponential scale-lengths, or 5–7 kpc). These stars will be on orbits that pass through more distant regions. While all timescales, including mixing timescales, are shorter in the inner galaxy, where densities are higher, than in the outer galaxy, kinematical signatures of sub-structure in the halo may remain even after 10 Gyr (Helmi & White 1999). A key aspect of the GAIA survey is that it naturally quantifies, to very high precision, the old stellar populations to several kpc from the Sun, and will determine the time-integrated structure.

In the merger event that may have formed the thick disk, some of the orbital energy of the satellite goes into the internal degrees of freedom of the disk, and heats it. Depending on the mass, density profile and orbit of the satellite, some 'shredded-satellite' stars will be left behind at the scene, and again may leave a kinematic signature, distinct from the thick disk that results from the heated thin disk. Since satellites on prograde orbits are favoured to form the thick disk (as mentioned above), one might expect a signature to be visible in the mean orbital rotational velocity of stars, lagging the Sun by more than does the canonical thick disk. The relative number of stars in the 'shredded satellite' versus the heated-thin disk (now the thick disk) depends on the details of the shredding and heating processes, and may well vary strongly with location. Unfortunately, despite much recent interest in simulations of disk heating, which have become more sophisticated, including more physics and more particles (e.g. Hernquist & Mihos 1995; Walker et al. 1996; Huang & Carlberg 1997; Velazquez & White 1999), they still do not include initial conditions that in any model of galaxy evolution describe the Milky Way disk 12 Gyr ago.

Metallicity Distributions The metallicity distributions of long-lived stars contain the integrated chemical evolution over the life time of the Galaxy, and are central to studies of the evolution of the Milky Way (Tinsley 1980). The combination with kinematics provides unique constraints on the early stages of Galactic evolution; for example Gilmore et al. 1995 and Wyse & Gilmore (1995) have studied the thin disk/thick disk interface utilising combined kinematic and chemical abundance survey data. The different Galactic components of thin disk, thick disk and halo that are probed by GAIA have overlapping kinematics and metallicity distributions, so that robust statistical assignment of a given star to a specific component requires a measurement of each property. The means and a rough dispersion of these distributions are reasonably well known, at least for high-velocity stars in the Solar neighbourhood (Ryan & Norris 1991). Their detailed shapes are not, despite the fact that deviations from a smooth distribution, plus the populations in the overlapping wings of these distributions, contain strong constraints on the processes of Galaxy formation and evolution. Perhaps most importantly, metallicities are essential for reliable analysis of probable remnant sub-structures in phase space.

The metal-rich stars in the halo studied thus far are clearly not part of a smooth transition to the thick disk/thin disk (Gilmore & Wyse 1998), showing distinctly different elemental abundances than do stars in the metal-poor disk (Nissen & Schuster 1997). Many of these metal-rich halo stars, selected by their high proper-motion (Carney et al. 1994), are on extremely elliptical orbits; we do not know if this is a general characteristic of the metal-rich halo. Indeed, the overall normalisation of metal-rich halo stars is unknown. The metal-rich stellar halo has a metallicity similar to that of a typical star in the most massive of the companion (gas-poor) galaxies to the Milky Way (Unavane et al. 1996), and thus is of particular interest. The extremely metal-poor tail of the halo is the focus of much attention (e.g. Beers et al. 1999) but these surveys are incapable of providing the normalisation relative to the more typical halo star, which requires a survey such as that of GAIA. Some additional constraint, such as the metallicity distribution, or the spectral types of the stars, is necessary to place kinematic anomalies in context. For example, the phase-space structure found by Dehnen (1998) in the Hipparcos dataset (Section 1.2.3) is unrelated to star streams in the halo. A narrow spread of metallicities within any kinematic anomaly/moving group would imply that it is a real association with a smaller-scale structure, rather than simply a statistical fluctuation within the normal halo, which has a broad metallicity distribution. GAIA is optimised to provide the relevant information, and the use of *in situ* tracers will avoid the kinematic bias inherent in local samples of high-velocity stars, allowing investigations of possible correlations between kinematics and chemistry.

There may be observational evidence for remnant phase-space sub-structure associated with formation of the thick disk, albeit with low statistical significance, in the Hipparcos results. We note that in axisymmetric potentials, the angular momentum about the $z-$axis is an (isolating) integral of the motion, is a conserved quantity, and plays a special role in the kinematic analysis. This is approximately the V velocity for stars near the Sun. The V-components of the space motions of local halo stars from an almost unbiased subset of the Carney et al. (1994) sample have been studied by Fuchs et al. (1998) (see Figure 1.5). They find an excess population, relative to an assumed symmetric halo with a Gaussian velocity distribution, but importantly, only in a restricted metallicity range, that of the metal-rich halo. An intriguing result, but with very low level of significance.

1.2.6 The Stellar Halo

The stellar halo of the Galaxy contains only a small fraction of its total luminous mass, but the kinematics and abundances of halo stars, globular clusters, and the dwarf satellites contain imprints of the formation of the entire Milky Way (e.g., Eggen et al. 1962; Searle & Zinn 1978). The most metal-deficient stars, with [Fe/H] < -3.5 (Ryan et al. 1996) represent a powerful tool to understand primordial abundances (e.g. Molaro et al. 1997) and the nature of the objects which produced the first heavy elements.

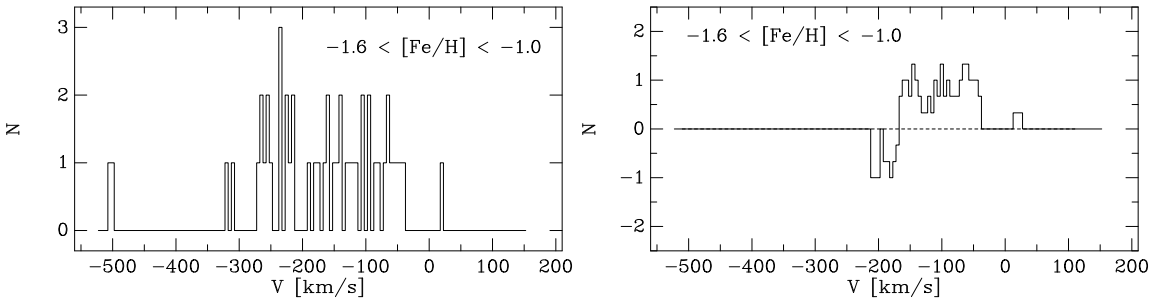


Figure 1.5: A sample of local subdwarfs with Hipparcos parallaxes and proper motions from Figure 3 of Fuchs et al. (1998), with metallicities and radial velocities from Carney et al. (1994). Left: all stars in the metallicity range indicated. Right: the 'excess population' over a symmetric distribution function, identified by folding the distribution in the left panel about V = –220 km s^{-1}, and then subtracting the left-hand side from the right-hand side.

At higher Galactic latitudes one may study particularly the disk gravitational potential, and hence the distribution of dark matter in the Galaxy (Section 1.2.8), the kinematics of the old disk and the thick disk, and of the halo. Photographic surveys have led the way here, since the low star density requires wide fields, and faint limiting magnitudes. However, the data remain confusing and contradictory. It is not possible as yet to know robustly just what are the kinematic and spatial distribution functions of the various populations, but GAIA will be uniquely able to provide these.

One example of an outstanding problem is the determination of the stellar velocity ellipsoid far from the Galactic plane with sufficient precision to determine its orientation reliably. This is a key parameter in the determination of the disk potential, and correspondingly the local distribution of dark matter. In the halo, comparison of the local velocity ellipsoid with the stellar density distribution similarly determines potential gradients. In the halo, however, the potential is purely dark matter dominated. The question of whether the dark matter is flattened, or triaxial, provide tests of galaxy formation models, and of the nature of dark matter.

Abundances, Ages, and Chemical Evolution Metallicities of stars in the outer halo may differ from those in the inner halo if the origin sites of such stars is different. Metallicity histograms of halo stars as a function of radius would allow models of the halo formation to be developed and tested. G and K giants would be particularly useful for this because they form at all metallicities. Large numbers of red/blue horizontal branch stars in the inner/outer halo would be found by GAIA and may be used to measure the age gradient in the halo.

The merger formation of the thick disk implies a merger of the proto-Milky Way with a substantial satellite. Such an event would leave a large number of the satellite's stars in the thick disk, which would thus be a mix of stars from two galaxies. In the outer Galaxy, systematic differences in the phase-space distribution functions of these two populations are expected. If the two populations could be segregated (statistically) in this way, then a direct test of the merger model is possible, and the GAIA chemical abundance data will allow reconstruction of the star formation and chemical histories of both the disk of the early Milky Way, and also of the now-lost satellite intruder.

Halo Streams The halo of the Milky Way is likely to be the most important component that may be used to distinguish among competing scenarios for the formation of our Galaxy. The classical picture of inner monolithic collapse (Eggen et al. 1962) plus later accretion in the outer Galaxy (Searle & Zinn 1978) predicts a smooth distribution both in configuration and velocity space for our Solar neighbourhood, which is consistent with the available observational data (see however below, and Majewski et al. 1996). The currently popular hierarchical cosmologies propose that big galaxies are formed by mergers and accretion of smaller building blocks, and many of its predictions seem to be confirmed in high-redshift studies.

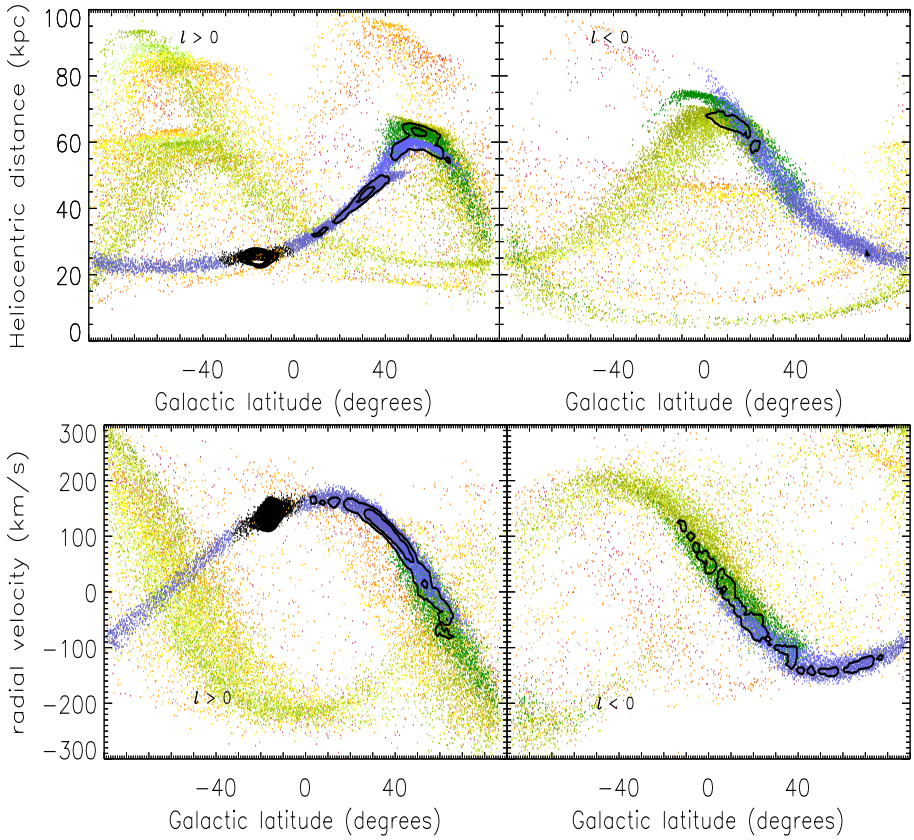


Figure 1.6: An example of the phase-space distribution of stars originating in a disrupted satellite for the case of Sagittarius (Helmi & White 2000). Top: the predicted distribution in heliocentric distance versus latitude. Bottom: the predicted distribution in radial velocity versus latitude. Different colours indicate material lost in different pericentric passages. Streams as coherent structures in phase-space are clearly visible even after 10 Gyr of evolution, and will be easily picked out by GAIA.

Those merging and accretion events leave signatures in the phase-space distribution of the stars that once formed those systems (Figure 1.6a). Helmi & White (1999) have shown that, after 10 billion years, the spatial distribution of stars in the inner halo should be fairly uniform, whereas strong clumping is expected in velocity space. This clumping appears in the form of a very large number of moving groups (several hundred in a 1 kpc^3 volume centered on the Sun, if the whole stellar halo would have been built in this way) each having very small velocity dispersions and containing several hundred stars. The required velocity accuracies to detect individual streams are less than a few km s^{-1}.

Samples of a few thousand stars with measurement errors as large as 15 km s^{-1} would allow a distinction to be made between the (classical) hybrid Eggen et al. (1962) plus Searle & Zinn (1978) and the hierarchical galaxy formation pictures. A volume of 1 kpc^3 centred on the Sun contains of the order of $10^5\ M_\odot$ in M subdwarfs. For uncertainties in the parallax and proper motion in the range 4–26 μas and 3–16 μas yr^{-1} respectively, the uncertainty in two of the velocity components is in the range 1–5 km s^{-1}, whereas for the radial velocity (third component) it is 3–10 km s^{-1}. The clumpy nature of the distribution will thus be easily distinguishable in the GAIA data. The primary evidence in the Local Group for mergers as important aspects of Galaxy evolution are the existence of the Galactic thick disk (Gilmore & Reid 1983), which is probably the remnant of the last major Galactic merger, some 10 Gyr ago, and the existence of the Sagittarius dwarf spheroidal galaxy (Sgr), currently deep inside the Galactic potential well, and probably being tidally destroyed today (Ibata et al. 1994). The next two most distant Galactic satellites, the Large and the Small Magellanic Clouds, are mutually interacting, with the Small Magellanic Cloud in particular probably being partially tidally disrupted, and the Magellanic Stream of HI being pulled from the Large Magellanic Cloud (Section 1.8.1).

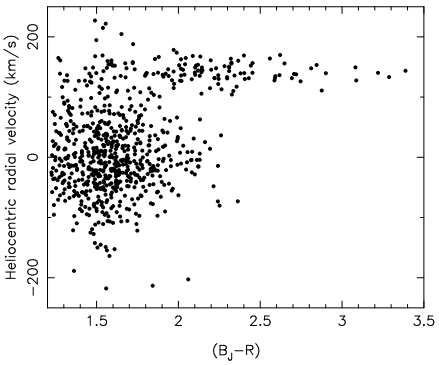


Figure 1.7: The discovery colour-velocity phase-space slice for Sgr, illustrating the need for, and efficiency of, the combination of kinematics and photometry in isolating Sgr and foreground bulge members (Ibata et al. 1995).

The space of adiabatic invariants allows better identification of the different merging events. Here clumping should be stronger since all stars originating from the same progenitor have very similar integrals of motion, resulting in a superposition of the corresponding streams. The plane defined by the total angular momentum and its z-component is suitable for finding such sub-structures, both because lumps remain coherent even after complete phase-mixing and also because of GAIA's accuracy. Helmi et al. (1999a) have applied this succesfully to Hipparcos data for the Solar neighbourhood, to discover for the first time a disrupted satellite in the inner Milky Way.

Sagittarius A stellar system which merged with the Milky Way will leave a low-contrast stream in phase space, only detectable from a suitably precise and suitably large volume map of phase space. How much debris has been left behind already by our closest neighbours, the Sgr dwarf and the Magellanic Clouds? How many former neighbours strayed too close? Consideration of the Sgr dSph illustrates the complexities. It lies behind the Galactic bulge, so that isolation of its member stars is difficult. Hence its discovery only recently, in phase space (Figure 1.7). By combining their radial velocity and preliminary proper motion data (derived from APM photographic astrometry and distances from photometry), Ibata et al. (1997) deduced an orbit for Sgr with period about 1 Gyr, so that Sgr has completed 10–15 orbits. Early models predicted complete tidal destruction of Sgr in 2–3 orbits. This led to suggestions that Sgr's internal kinematics is dominated by relatively dense cold dark matter (Ibata et al. 1997), or that it experienced a recent triple encounter involving also the two Magellanic Clouds which lowered Sgr to its present orbit (Zhao 1998).

Helmi & White (2000) showed that Sgr may in fact be the remnant of a galaxy with original mass $\sim 6\times10^8 M_\odot$ which fell onto the Milky Way about 12.5 Gyr ago. By now, the tidal tails of Sgr have wrapped around the Galaxy a number of times, resulting in multiple overlapping streams over a substantial fraction of the sky, as shown in Figure 1.6b. GAIA is ideal to find every Sgr member: the measured proper motion (APM and HST, cf. Ibata et al. 1997) is 2100 μas yr^{-1} (Table 1.13), easily measureable even at the GAIA limiting magnitude. Dynamical mapping however requires kinematics which resolve the potential gradients. The observed internal dispersion is 12 km s^{-1}. GAIA will provide individual three-dimensional motions with this precision for all stars down to the horizontal branch, allowing mapping of the potential well with several thousands of stars: a unique determination of a resolved galactic potential well (see also Section 1.2.8).

The parallax of individual Sgr giants ($I = 14 - 15$ mag) is 40 μas. While the measuring error (10–20 μas) will not allow single-star distances, the orientation and line-of-sight depth of Sgr, important parameters in the dynamical modelling, and sensitive tests of tidal disruption models, can be determined for groups of stars, providing the necessary additional data. The red clump mean V magnitude is about 18.2, with $V-I \sim 1.1$. With a hundred such stars observed by GAIA, a mean distance known to better than 10 per cent would be obtained, with 2000 stars, 2 per cent.

The metallicity distributions for Sgr stars will allow reliable determination of its star formation history, in the same way as for the Milky Way Galaxy (Section 1.3). Thus, the future population structure of the Milky Way, after completion of Sgr disruption, can be determined. In this way, the GAIA results will determine the stellar population mix of a future post-merger spiral galaxy.

The Outer Halo GAIA will find several million individual stars in the outer halo (defined here as galactocentric distance $R > 20$ kpc). These will mostly be G and K giants and red and blue horizontal branch stars. G and K giants are intrinsically bright, they form in all known old stellar population types, they have easily measurable radial velocities, and they are historically well studied because they are the most easily accessible stars in the globular clusters. Horizontal branch stars have been the preferred tracer stellar type for the outer halo to date, because they can be much more easily identified amongst field stars than G and K giants. In particular, blue horizontal branch stars have been very easy to locate, since almost all faint ($14 < V < 19$ mag), blue ($0.0 < B - V < 0.2$) stars are halo blue horizontal branch stars (Sommer-Larsen et al. 1997). However these stars are a biased tracer of the halo population in the sense that they do not always form in old metal weak populations (viz. the second parameter problem in globular clusters). The bluest horizontal branch stars (sdB stars) seem, however, to have rather disk-like kinematics, while the A and RR Lyr stars predominantly have halo-like kinematics (Altmann & de Boer 199).

Redder horizontal branch stars and G and K halo giants are drowned out by the huge numbers of foreground turnoff and dwarf stars in the Galactic disk, and so sophisticated and very time consuming methods must be used to remove this foreground component. This is usually done by semi-automated spectroscopic techniques (e.g. Ratnatunga & Freeman 1985; Beers et al. 1992; Flynn & Morrison 1990). GAIA will circumvent all these difficulties. The late-type foreground dwarfs are much closer than the background late-type giants, so that at faint magnitudes ($V <$ 19 mag) the dwarfs have a measurable parallax while the background giants do not. It will be possible to lift the veil of foreground stars and reveal of order millions of background halo stars, on the giant branch, and the red and blue horizontal branch.

Good radial velocities for outer halo stars are feasible to $V \sim 17$ mag to an accuracy of 10 km s^{-1}. Proper motions of such objects should be good to about 20 μas, but the parallax cannot be measured at all accurately. Therefore to derive full space motions for these objects the distance must be derived via classical photometric indicators. For halo blue or red horizontal branch stars this is straightforward because they have a well known absolute magnitude of $M_V \sim 0.6$. The local number density of such stars is about 50 kpc^{-3} (Morrison 1993), yielding about 25 000 within 5 kpc for which parallaxes accurate to a few percent will be measurable. As a result the absolute magnitude versus colour relation will be well determined, and distant outer halo blue horizontal branch or red horizontal branch halo stars for which the parallax is negligible will be assigned photometric distances on this basis. GAIA will provide the best calibration yet for such stars. GAIA photometry will have sufficient accuracy to identify blue horizontal branch stars unambiguously from the more-nearby halo blue stragglers down to $V = 20$ mag.

The local volume density of halo red giants ($M_V < 0.5$) is 37 ± 7 kpc^{-3} (Morrison 1993), yielding of order 20 000 within 5 kpc of the Sun. For all these inner halo giants very accurate parallaxes and hence absolute magnitudes will be measured. As is well known, giant luminosity is a well-behaved function of colour and metallicity. Good distances to outer halo giants for which the GAIA parallax is of modest value can be derived from the colour and metallicity indicators measured by GAIA. Most (> 70 per cent) of the outer halo stars found will be at high enough galactic latitude that reddening will not be the main source of uncertainty in establishing the distance. The remaining outer halo stars will be passing through or beyond the reddening layer of the disk, and reddenings will not be easy to obtain directly for each star, although statistically the reddening in such fields could be determined via photometry of the blue horizontal branch stars, leading to better distance estimates of these stars and the red giants as well.

1.2.7 Globular Clusters

GAIA's astrometric capabilities will allow kinematic member selection (Section 1.2.4), of stars in Galactic globular clusters, to $V \lesssim 20$ mag (see Section 6.4.6 for as assessment of the observational limitations for globular clusters). This will separate the field halo stars from the cluster members, allow internal dynamical studies, derivation of very accurate distances and space motions for the clusters, and will also provide much-improved colour-magnitude diagrams. An example is shown in Figure 1.8 for one of the nearest globular clusters, NGC 6397, for which members have been separated from field stars based on HST proper motions (King et al. 1998).

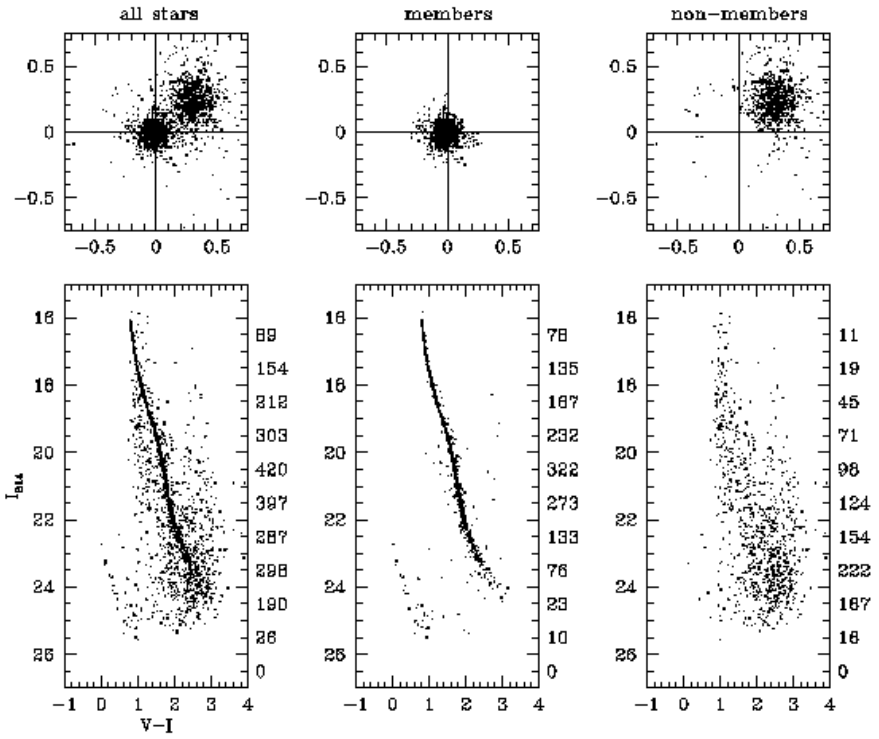


Figure 1.8: Kinematic member selection illustrated by Hubble Space Telescope data for NGC 6397, from King et al. (1998).

Globular clusters play a key role in the calibration of fundamental stellar parameters. Many observational effects and theoretical complications make interpretation of globular cluster properties far from straightforward; but cluster age determinations essentially require absolute magnitude calibrations of the main-sequence and, in particular, the turn-off point as a function of chemical composition (see Section 1.4). For absolute ages to be accurate to a billion years, essential for a resolution of the age conflict, the distance of the cluster must be determined with an accuracy of better than 3 per cent.

Internal Dynamics and Distances The one-dimensional internal velocity dispersion in globular clusters is 10–20 km/s. This can be resolved easily by GAIA. Combined with radial velocities, this will provide the full phase space structure of the clusters. Dynamical modeling will then provide very accurate distance estimates, by comparing the tangential velocity distribution (in μas s^{-1}) with that in the radial velocities (in km s^{-1}). The observed velocity distributions also constrain models of globular cluster formation and evolution. Within 47 Tuc, for example, proper motions of 10 μas yr^{-1} correspond to transverse velocities of 0.2 km s^{-1}. This accuracy exceeds that of any ground-based astrometric study of the internal motions in a globular cluster by a large margin—even those with more than 50 years baseline.

There are dozens to hundreds of stars in each of the ∼20 nearest globular clusters brighter than $V = 15$ mag (including 47 Tuc, ω Cen, M3, M5, and M15). The individual distances would be accurate to 10–20 percent, resulting in a mean trigonometric distance good to a few percent or better.

Spectroscopic binaries have been detected in globular clusters with amplitudes of tens of km s^{-1} and periods of years, corresponding to separations of order 1 mas. Parallaxes and annual proper motions at the level of 50 μas or better would provide distances and orbital data necessary to clarify the formation and evolution of these binary systems, and their role in the formation of the milli-second pulsars now known to exist in the cores of globular clusters (Tucholke & Brosche 1995).

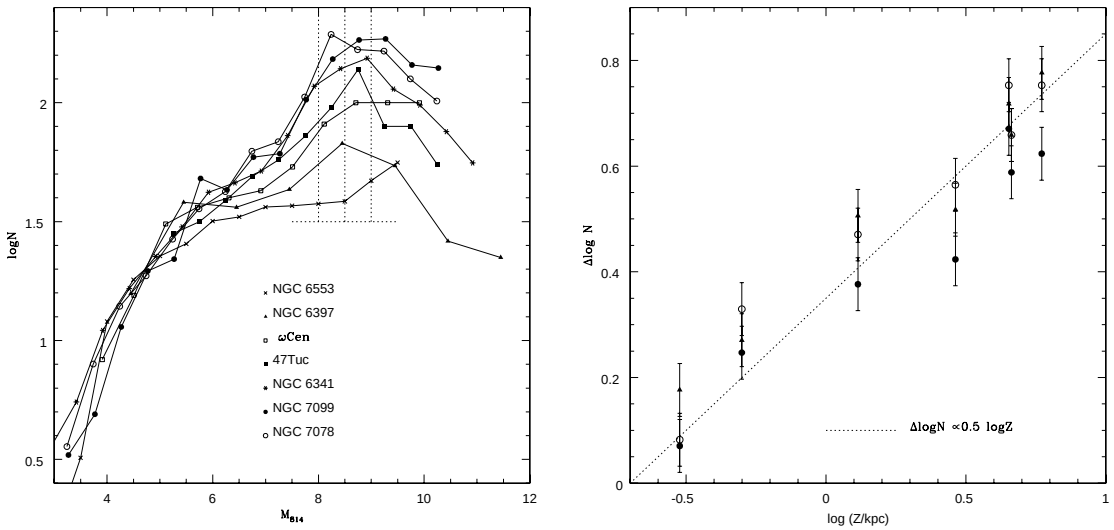


Figure 1.9: (a) HST luminosity functions for seven Galactic globular clusters; (b) relative flattening of the faint end of the luminosity functions in (a), measured at M_{814} =8.0, 8.5, 9.0 (filled circles, open circles, triangles) plotted against distance above the Galactic plane.

Tidal Disruption The dynamical evolution and eventual disruption of a stellar cluster is dominated by three processes: mutual gravitational interactions of the stars; the tidal field of the host galaxy; and mass loss by stellar evolution. In addition, gas loss, primordial mass segregation, etc, can affect the early evolution, while the binary fraction will affect late core evolution. Of the three dominant processes affecting cluster evolution and dispersal, the least well described in most modeling is the effect of Galactic tidal fields, both steady-state and time dependent, especially involving disk and bulge shocking in the time-dependent case. Tidal effects are however important, and possibly dominant, in cluster evolution: it is even possible that a substantial fraction of the Galactic globular cluster population has been destroyed (Gnedin & Ostriker 1997), now contributing a significant part of the inner halo and bulge field star population.

An additional motivation for an improved understanding of tidal effects on clusters is the ambition to infer the stellar initial mass function over a very wide range of ages and metallicities from direct observations of globular clusters. The mutual gravitational interactions between the constituent stars in a cluster drive them towards energy equilibrium. This inevitably generates some mass segregation, with the lower-mass stars rising to the outer radii, and some escaping from the cluster. This normal mass-segregation induced evaporation is enhanced in a steady tidal field, and can be substantially enhanced in a time-dependant (shock) tidal field. While in principle it is possible to model such an effect, and thereby correct for it in a derived IMF, in practice this is extremely uncertain. Not only is the Galactic orbit of any specific cluster at best crudely known, but the present, and especially the past, gravitational field of the inner Galactic disk and bulge are insufficiently known.

Direct evidence for the crucial role of Galactic tides in both cluster survival and in derivations of the stellar initial mass function for Milky Way globular clusters has recently become available. Figure 1.9a shows the stellar luminosity function for seven Galactic globular clusters from HST (Elson et al. 1996; Santiago et al. 1996; Piotto et al. 1997). While the luminosity functions agree well at brighter magnitudes, the faint ends differ markedly. Such differences have been attributed to some combination of metallicity effects, evaporation of low mass stars, and/or stripping of low mass stars by tidal shocking (Piotto et al. 1997). One may quantify the difference among the luminosity functions as the increment $\Delta \log N$ at $M_{814} = 8.0, 8.5, 9.0$. Figure 1.9b, from Elson et al. (1998), shows $\Delta \log N$ plotted against the distance of each cluster from the Galactic plane. There is a striking correlation, which suggests that tidal shocking is primarily responsible for the differences among the luminosity functions. A plot of $\Delta \log N$ against metallicity ($-0.3 <$ [Fe/H] < -2.5 for this sample) shows no correlation.

Open questions include: Does the luminosity function in any Galactic globular cluster reflect an unmodified IMF? Has the effect of tidal evaporation saturated for the clusters at present farthest from the Galactic plane? What is the real IMF? To answer these questions require an improved understanding of the effects of tides on globular cluster dissolution, and accurate determination of the orbits of the individual globular clusters, and their current tidal signatures, in environments where disk and bulge shocking has a range of effects, from minimal to dominant.

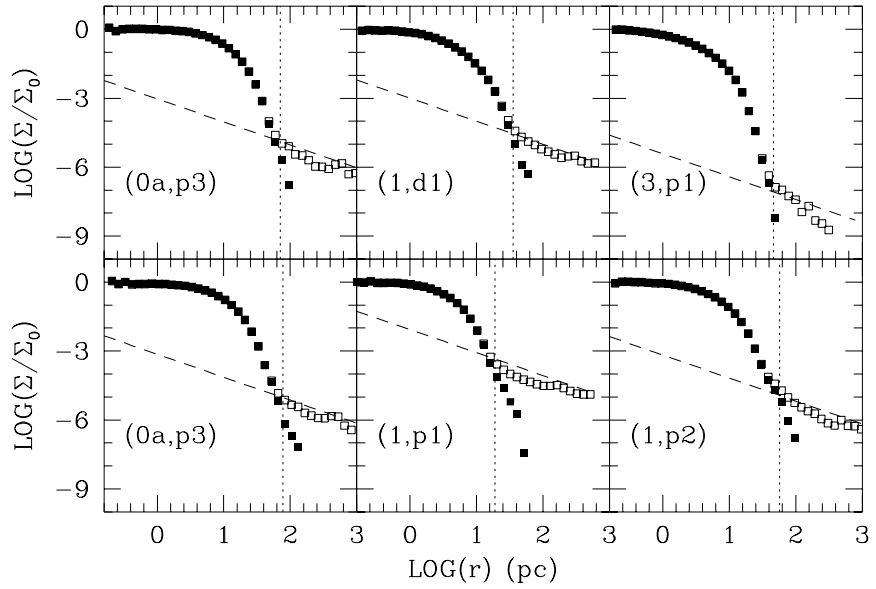


Figure 1.10: Simulations of a variety of realistic globular clusters evolving in the Galactic tidal field. In each case the solid points indicate stars bound to the cluster, open squares stars which have been unbound by the Galactic tide, and the vertical dotted line what would be classified as the 'tidal radius'. The characteristic r^{-1} tail, indicative of tidal disruption, is evident, and could be detected by GAIA (Johnston et al. 1999a).

Additionally, they require careful kinematic mapping of the stellar distributions around the clusters, near their present tidal radii. There is direct observational evidence for tidal evaporation of some outer Galactic clusters from star count studies (Grillmair et al. 1995). GAIA will be able to find the tidal streams, and to measure their kinematics.

Detailed simulations (e.g. Figure 1.10) show that the concept of a 'tidal radius' is complex, with no simple boundary beyond which stars are unbound. A characteristic feature of steady tidal disruption is however an extended r^{-1} tail of stars with kinematics only slightly different than those of the parent. Such stars would be readily distinguishable by GAIA. Since the clusters with the largest mass-loss rates are those nearest the bulge and disk, identifying such stars from the crowded foreground Galactic populations has proved impossible from present data.

Suitable globular clusters exist, and are amenable to study, in the Galaxy, in the Large and Small Magellanic Clouds, and in the Sagittarius dwarf spheroidal. For the LMC there is a disk and a bar, of uncertain present and past dynamical significance. The dynamical effects of the LMC disk and bar on the LMC clusters are however substantially less than those of the very much more massive Galactic disk and bulge on Galactic clusters. The LMC thus provides an ideal intermediate test case, where tidal effects are significant, but probably have a relatively larger static contribution and a relatively smaller time-dependant (shocks) term. The SMC has neither disk nor bar, though perhaps has a chequered dynamical history.

For the Magellanic Clouds, the radial velocity dispersion of the cluster system is ~ 20 km s^{-1}; for Sgr the radial velocity dispersion is 11 km s^{-1}, while the inner Galactic disk dispersion is ~ 50 km s^{-1}. A single point accuracy of < 5 km s^{-1} is then sufficient to distinguish cluster and field stars. This corresponds to 40 μas yr^{-1} for Sgr, 20 μas yr^{-1} in the LMC and SMC, and 100 μas yr^{-1} in the inner disk clusters. This is all within the GAIA capabilities.

1.2.8 Dark Matter and the Mass of the Milky Way

Visible stars and gas appear to make up only a small part of the total mass of galaxies. Most of the mass is thought to be in some kind of dark form, which can only be detected through its gravitational effects and not through emitted light. The primary evidence for this conclusion was uncovered in the 1970s when it was found at that time that disk galaxies rotate much faster along their outer edges than any reasonable model for the matter distribution (which is assumed to be proportional to the emitted light) would allow. In the halo, the Galactic potential is purely dark-matter dominated. Is the dark matter flattened, or triaxial? These questions are direct tests of galaxy formation models, and the nature of the dark matter.

The Galactic rotation curve is well established inside a galactocentric radius of 20 kpc, and constrains the mass profile fairly accurately. Outside this radius the main constraints are distances and radial velocities of the globular clusters, the dwarf satellites, and M31 (Section 1.2.8). A variety of methods and arguments show that all measurements to date are consistent with a model in which the mass distribution is essentially an isothermal sphere with a constant circular velocity $v_c \sim 180$ km s^{-1} (Zaritsky 1999). Out to a distance of ~ 300 kpc—nearly halfway to M31—this corresponds to a mass of about $2 \times 10^{12}\ M_\odot$. The average mass-to-light ratio is over 100 in Solar units, so most of this matter is dark, or at least severely underluminous, and most of it is at large radii.

Dark Matter in the Disk The distribution of mass in the Galactic disk is characterized by two numbers, its local volume density ρ_o and its total surface density $\Sigma(\infty)$. They are fundamental parameters for many aspects of Galactic structure, such as chemical evolution (is there a significant population of white dwarf remnants from early episodes of massive star formation?), the physics of star formation (how many brown dwarfs are there?), disk galaxy stability (how important dynamically is the self-gravity of the disk?), the properties of dark matter (does the Galaxy contain dissipational dark matter, which may be fundamentally different in nature from the dark matter assumed to provide flat rotation curves, and what is the local dark matter density and velocity distribution expected in astroparticle physics experiments?), and non-Newtonian gravity theories (where does a description of galaxies with non-Newtonian gravity and no dark matter fail?).

The most widely referenced and commonly determined measure of the distribution of mass in the Galactic disk near the Sun is the local volume mass density ρ_o, i.e. the amount of mass per unit volume near the Sun, which for practical purposes is the same as the volume mass density at the Galactic plane. This quantity has units of $M_\odot$ pc^{-3}, and its local value is often called the 'Oort limit' in honour of the early attempt at its measurement by Oort (1932). The contribution of identified material to the Oort limit may be determined by summing all local observed matter – an observationally difficult task. The uncertainties arise in part due to difficulties in detecting very low luminosity stars, even very near the Sun, in part from uncertainties in the binary fraction among low mass stars, and in part from uncertainties in the stellar mass–luminosity relation. All these quantities will be determined directly, to extremely high precision, by GAIA.

The remaining information required is the volume density of the interstellar medium. This is scientifically interesting, since the physically important quantity (for dynamical purposes) is the appropriately averaged volume density of the patchily distributed interstellar medium at the Solar galactocentric distance.

The second measure of the distribution of mass in the Solar vicinity is the integral surface mass density. This quantity has units of $M_\odot$ pc^{-2}, and is the total amount of disk mass in a column perpendicular to the Galactic plane. It is this quantity which is required for the deconvolution of rotation curves into 'disk' and 'halo' contributions to the large-scale distribution of mass in galaxies. To date, both ρ_o and $\Sigma(\infty)$ are known with modest accuracy. Significant improvement over the current measurements requires accurate distances and velocities for a large sample of tracer stars to faint magnitudes. GAIA's exquisite astrometric precision will make this possible.

Pre-Hipparcos determinations of $\Sigma(\infty)$ using K giants derived values in the range $\sim 50 M_\odot$ pc^{-2} (Kuijken & Gilmore 1989b; Kuijken & Gilmore 1989a; Bahcall et al. 1992). The best recent determination of ρ_o, using Hipparcos data, puts it at $\sim 0.1 M_\odot$ pc^{-3} (Crézé et al. 1998; Bienaymé 1999), consistent with all the Galactic dark matter being distributed in the halo. The implications for direct searches for dark matter particles are also considerable. The contribution of this local mass density to the local circular velocity, assuming an exponential disk with the Sun 2.5 radial scale lengths from the Galactic centre, is $v_{\rm c,disk} \sim 150\,(\Sigma_{\rm local}/60 M_\odot {\rm pc}^{-2})^{1/2}$ km s^{-1}. The local circular velocity is ~ 220 km s^{-1}, and the contributions to this circular velocity from the various components generating the Galactic potential add in quadrature. Thus the Galactic disk is far from dominating the local potential well.

The Hipparcos analyses are restricted to nearby young stars, which need not be in gravitational equilibrium with the larger-scale Galactic potential. Indeed, Crézé et al. (1998) and Dehnen (1998) have shown that these stars are not in equilibrium, thus restricting their validity as tracers. If one knew both the local ρ_o and $\Sigma(\infty)$, one could immediately constrain the scale height of any contribution to the local volume mass density which was not identified. That is, one could measure directly the velocity dispersion, i.e., the temperature, of the 'cold' dark matter.

Although $\Sigma(\infty)$ and ρ_o are different measures of the distribution of mass in the Galactic disk near the Sun, both are derived from the vertical Galactic force field $K_z(z)$ (Binney & Tremaine 1987). Out to about 1 kpc from the Sun, GAIA will determine positions and space motions for bright K giants and A stars, the latter of particular importance because accurate ages can be estimated. Distances will be accurate to a few percent. This will permit an accurate measurement of the density distribution $\rho(R, z)$ as a function of Galactocentric radius R and vertical height above the plane z, as well as the vertical velocity distribution $f(w)$ as a function of R and z, which is all the information needed to reconstruct the gravitational potential field of the disk. Such observations will be compared to models of the disk's gravity in order to establish if unseen disk matter is required to explain the disk's gravity away from the Solar neighbourhood. These models will be well constrained because GAIA will also measure the amount of visible matter within this same region (within about 1 kpc of the Sun) since it will probe the luminosity function/mass function to the end of the hydrogen-burning sequence and beyond (Section 1.3.4).

Within 1 kpc, the distances of tracer stars will be very accurate. Assigning an absolute magnitude to the tracers will therefore depend on an accurate reddening estimate, particularly at low Galactic latitudes ($|b| < 30^\circ$). For early-type tracers, photometric methods exist and the GAIA filter system will be sufficient to measure the reddenings. For late-type stellar tracers (which are to be preferred since they are older than the early-type stars, and have had more time to become kinematically mixed in the Galactic potential) the reddening is a challenge to measure individually. However, since one would be interested in isolating ensembles of tracer stars in a certain absolute magnitude/colour range as a function of (R, z) position in the disk, one could use features in the colour-magnitude diagram to establish the reddening along the lines-of-sight. For late-type stars the 'clump' (core-He burning giants) lends itself to this, since the absolute magnitude of the clump is hardly sensitive to age or metallicity effects, lies at $M_V \sim 0.8$, and is a narrow enough feature in M_V to provide a good reddening indicator. Such stars are ideal disk mass density tracers.

It has long been thought that the spiral arm and inter-arm regions of the disk may vary in matter density by as much as 30 per cent or even more. GAIA will measure such variations easily. This will allow extrapolation of the local disk mass measurement to the whole disk and to determine accurately the amount of dark halo matter needed to maintain a flat rotation curve.

Wide binaries The Galactic tidal field causes a cut-off in the distribution of binary separations at semi-major axes of about 1 pc. The distribution with semi-major axes larger than about 0.1 pc is sensitive to gravitational perturbations caused by passing luminous stars and dark masses, i.e., to the total amount of mass near the Sun. The theoretical framework is available to infer properties of the perturbers from the distribution of binary separations (e.g., Wasserman & Weinberg 1991). Useful constraints require accurate data for thousands of wide binaries covering a range of separations. These can be identified to 100 pc if astrometry is available for all stars to $V \gtrsim 16$ mag. GAIA will provide this, and will allow similar analysis to be carried out in larger volumes, and for halo binaries only, which will constrain the vertical distribution of dark matter.

The best determinations to date are based on the ground-based proper motions of ∼1000 relatively bright stars in the nearest 25 pc, and find less than ten wide pairs (e.g., Wasserman & Weinberg 1991). The HST imaging survey of Gould et al. (1995) to $V \sim 21$ covered only 0.15 square degrees, and revealed a similar small number of possible wide pairs, in the nearby halo. GAIA will reach nearly as faint, and the astrometry, radial velocities and photometry will allow reliable identification of physical wide binaries over the entire sky, and will enlarge the available wide binary samples by 3–4 orders of magnitude.

The Local Escape Velocity The local Galactic escape velocity is a key constraint on the Galactic potential, but is not known very accurately (e.g., Carney et al. 1988; Kochanek 1996). Observationally, one tries to find the fastest moving halo stars, i.e., to determine the population of the high-velocity wing of the Galactic phase-space distribution function. Studies to date are restricted to relatively bright objects ($V \sim 14$ mag), and suffer from small-number statistics. GAIA's complete sampling of halo objects to faint magnitudes will provide an accurate distribution function, allowing an accurate measurement of the local Galactic escape velocity.

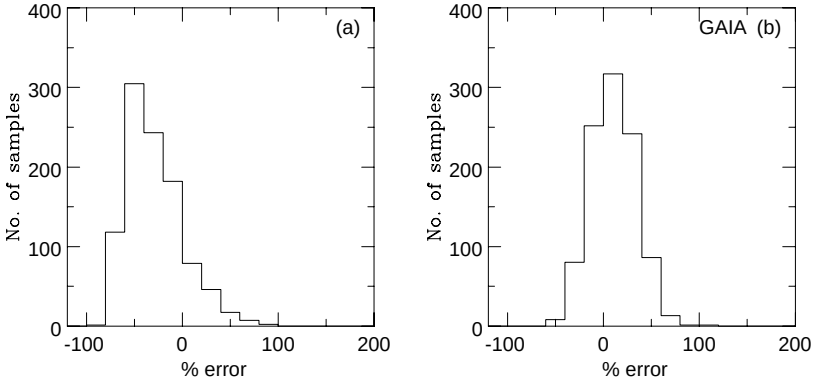


Figure 1.11: Histograms showing how many out of 1000 data sets yielded a given percentage error in the mass of the Milky Way. The left panel shows the present-day situation with 30 data points with radial velocities only, the right panel shows the situation after GAIA with 30 data points with radial and proper motions. Notice that after GAIA the systematic tendency for underestimates is removed, and the histogram is narrower indicating a smaller spread in the uncertainty.

Space Motions of Distant Globular Clusters and Satellites The total mass of the Milky Way is presently fixed by a dataset of 27 objects beyond 20 kpc. The distances and radial velocities of all the members of the sample are known, and additionally 6 also possess measured proper motions. This is evidently a scanty dataset on which to base measurements of one of the most fundamental Galactic parameters. Wilkinson & Evans (1999) provide a detailed analysis of the errors in our current estimate of the mass of the Milky Way. The most serious uncertainties come from the size of the dataset, which may cause a systematic underestimate by a factor of two, and the measurement errors, which cause a scatter in the mass of the order of a factor of two. Wilkinson & Evans (1999) concluded that the total mass of the Milky Way galaxy is $\sim 2.3^{+3.9}_{-1.6} \times 10^{12} M_{\odot}$, while the mass within 50 kpc is $\sim 5.5^{+0.1}_{-1.1} \times 10^{11} M_{\odot}$. So, the mass (or equivalently the extent) of the Milky Way halo is amongst the most poorly known of all Galactic parameters. It is much more uncertain than the distance to the Galactic Centre or the Oort constants, for example. There is a pressing need for more data and here GAIA will provide valuable help. As the sample of satellite galaxies is nearly complete, the dataset can be extended only by measurements of the proper motions. GAIA will (i) remove the bias to lower masses if only radial velocities are available, and (ii) reduce the spread in mass estimates from the measurement errors. This will provide the mass of the Milky Way halo to within ∼20 per cent.

For GAIA, the target is 10 μas yr^{-1} in proper motion accuracy at $V = 15$ mag and 100–200 μas yr^{-1} at $V = 20$ mag. The colour-magnitude diagram of even the distant Leo I (230 kpc) shows that the tip of the giant branch is still visible at $V = 20$ mag (Cabrera et al. 1998; Hernandez et al. 1999b). The individual proper motions of bright stars at the distance of Leo I are only accurate to ~ 240 km s^{-1}. The Cabrera et al. (1998) colour-magnitude diagram is derived from three Wide Field Planetary Camera (WFPC2) frames with the aperture centred on Leo I. The field of view is ~ 1.7 arcmin^2. There are ~ 50 stars brighter than $V = 20$ mag visible on the colour-magnitude diagram. Leo I subtends perhaps ~ 10 arcmin^2 on the sky, using the exponential radius given in Mateo (1998). In total, therefore, Leo I has perhaps ~ 290 stars brighter than $V = 20$ mag, and so the error on the proper motion of the ensemble is less by a factor of ~ 17. In other words, the components of the space motion of Leo I are obtainable to an accuracy of ~ 14 km s^{-1} with GAIA. For closer satellites like Draco and Ursa Minor, the situation is even more favourable. Hernandez et al. (1999b) provide a colour-magnitude diagram for Ursa Minor which has ~ 17 stars brighter than $V = 20$ mag and is derived from single chip WFPC2 observations. Each chip represents a field of view of 0.6 arcmin^2. Taking the exponential scale-length as 8.0 arcmin (Mateo 1998), then the number of stars in Ursa Minor brighter than $V = 20$ mag is ~ 5700. So, GAIA can provide the components of the space motion of Ursa Minor to an accuracy of ~ 1 km s^{-1}. It is thus reasonable to assume that it can provide the space motions to better than 10 per cent.

In order to investigate both the likely error caused by the small number of datapoints available, as well as future prospects of GAIA, considere 1000 simulated datasets with 30 data points representing the satellite galaxies and globular clusters, using the self-consistent halo model described in Wilkinson & Evans (1999). Two cases are examined — in the first, knowledge of only radial velocities is assumed and in the second, the full space velocities are presumed to be measured by GAIA to at least 10 per cent. The results are shown in Figure 1.11. Panel (a) shows

that when the number of data points is ~ 30, the probability of obtaining an estimate of M which differs from the true value by more than a factor of two is about 30 per cent. There is also evidence for a systematic underestimate in the mass when only radial velocities are used. When the proper motions are included, this systematic effect is removed and the probability of obtaining a mass estimate more than a factor of two different from the true value is reduced to just 2 per cent. The average absolute deviation about the mean due to the lack of data is $\sim$20 per cent.

Using Tidal Streams to Determine the Mass Profile Stars in a tidal stream trace a narrow set of orbits in the vicinity of that of the centre of mass, and are correlated in orbital phase: they can all be traced back to a small volume (e.g., near the pericentre of the satellite orbit) where they were once bound to the satellite. Hence individual proper motions of a few hundred stars along a stream provide a tight constraint on parameters of the Galactic potential and the initial condition of the center-of-mass of the satellite. The results of Johnston et al. (1999b) show that 5–10 per cent precision in the rotation curve, flattening and triaxiality of the halo is reachable by mapping out the proper motions and radial velocities along a tidal stream > 20 kpc from the Sun. The fairly large error in distance measurements to outer halo stars presents no serious problem since one can predict distances theoretically using the known narrow distribution of the angular momentum or energy along the tails associated with a particular Galactic satellite.

Tidal streams are excellent tracers of the Galactic potential as long as a stream maintains a cold spaghetti-like structure. The results for static Galactic potentials are likely to be largely generalizable to moderately time-evolving potentials. Perhaps the most exciting implication is that by mapping the proper motions along the debris GAIA can set limits on the rate of evolution of the Galactic potential, and distinguish among scenarios of Galaxy formation.

1.2.9 The Stellar Radiation Field and the Interstellar Medium

The interstellar medium (ISM) is the galactic atmosphere which fills the space between stars. The ISM plays a vital role in galactic evolution through a complex feedback process. When clouds within the ISM collapse, stars are born; when stars die, they return their matter to the surrounding gas. GAIA will have a key role to play in future studies of the ISM for gas components within 10 kpc of the Sun. Once a large fraction of the intrinsically luminous stars have been accurately pinpointed within the ISM, it will be possible to examine in detail how they interact with their environment. Future and current optical, radio and millimeter interferometers will produce detailed kinematic maps $V_r(\ell, b)$ of the warm, neutral and molecular components of the ISM. With the aid of a self-consistent dynamical model for the Galaxy derived from the GAIA data, the spectral information can be inverted to produce accurate density distributions for the different phases in Galactic coordinates, $n(r, \theta, \varpi)$.

The ISM is heated by both extended sources (e.g. coronal gas) and by compact sources (e.g. stars). The former have already been mapped by far-UV, x-ray and gamma-ray satellites. However, unlike stars, the extended sources cannot be inverted directly to provide a 3-d emissivity distribution.

Stars are the principle source of energy for the interstellar medium. Starlight photons produce photoelectric emission from dust grains which heat the neutral gas. Ultraviolet photons from the youngest stars ionize atoms and dissociate molecules. Supernovae provide the main source of kinetic energy: these drive shock waves into the surrounding ISM and are largely responsible for its complexity. Stellar winds and supernovae enrich the gas with heavy elements over the course of billions of years.

ISM Overview The interstellar medium includes starlight, gas, dust, planets, comets, asteroids, cosmic rays and magnetic fields. By number of nuclei, about 90 per cent of interstellar matter is hydrogen, 10 per cent is helium. All of the elements heavier than helium constitute about 0.1 per cent of the interstellar nuclei, or about 2 per cent by mass. Roughly half of the heavier elements are in the form of dust about a micron or less in size. Most of the refractory elements (Si, Ca, Fe) are depleted from the gas phase, and are locked up in small dust grains mixed in with the gas.

Table 1.4: The component properties of the interstellar medium.

component	temperature (K)	midplane density (cm^{-3})	filling fraction (%)	average height (pc)
Clouds:				
H_2	15	200	0.1	75
HI	120	25	2	100
Inter-cloud:				
Warm HI	8000	0.3*	35*	500
Warm HII	8000	0.15	20	1000
Hot HII	$\sim 10^6$	0.002	43*	3000*

* Value uncertain by at least a factor of 2.

The gas can be further divided into hot, warm and cold components, each of which appear to exist over a range of densities, and therefore pressures. Clouds account for only half the mass and 2 per cent of the interstellar volume; the remainder is in the form of an inter-cloud component. Remarkably, the diverse gas components, cosmic rays, magnetic fields and starlight all have very roughly the same energy density of about 1 eV cm^{-3}. All the major constituents (or phases) of the interstellar medium appear to be identified now (Table 1.4), although complete multi-phase studies have been extremely difficult beyond a few kpc from the Sun.

Clouds Half of the neutral atomic hydrogen and all of the molecular hydrogen in the ISM is concentrated into relatively high density and low temperature clouds. H I clouds have complex shapes resembling thin extended sheets or filaments with embedded small clumps. Cloud properties are determined primarily from the hyperfine ground state transitions of hydrogen at 1420 MHz (21 cm) although interstellar absorption lines of trace elements (e.g. Ca^+) also continue to play a key role. For the most common clouds, where the 21-cm radiation escapes freely, the brightness of the emission provides a direct measurement of the H I column density $N_H = \int n_H \, ds$ where n_H is the atomic hydrogen density. When an H I cloud lies in front of a bright source of radio continuum emission, the decrease in the brightness of the background source at 21 cm is proportional to $N_H = \int n_H / T_H \, ds$, where T_H is the temperature of the H I cloud. Table 1.4 summarises basic properties of the cold and warm neutral media.

Molecular hydrogen is confined to the interiors of the densest and most massive clouds where starlight capable of dissociating molecules cannot penetrate. These clouds constitute the active star-forming component of the interstellar medium. Because H_2 has no electric dipole moment, radiative transitions of H_2 are greatly suppressed. Therefore, most of the structural information about molecular clouds in the ISM is obtained through observations of the rotational transitions of the trace molecule CO at 115 GHz (2.6 mm). In addition, a wide variety of other molecules, including complex hydrocarbon chains, have been detected within the H_2 clouds. Molecular clouds are small (40 pc), dense (200 cm^{-3}), with structure on scales of less than 0.1 pc (see Table 1.4). Some of the small condensations can have densities as high as 10^5 cm^{-3}. In disk galaxies like our own, the cold neutral and molecular gas are confined to a disk which is much thinner than the stellar disk.

Inter-Cloud Medium Spitzer (1956) initially speculated that a rarefied ($\sim 10^{-3}$ cm^{-3}), hot gas ($\sim 10^6$ K), extending a kpc or more above the galactic plane, would confine the diffuse gas clouds observed far above the plane and would prevent their expansion and dissipation. By the late 1960s, the existence of widespread hot gas in the disk of the Galaxy was established from observations of O^{+5} ions at ultraviolet wavelengths, and the direct detection of soft x-ray emission. A clear demonstration of coronal halo gas far from the disk has been harder to come by. A wide range of ions is observed (Si^{+3}, C^{+3}, N^{+4}) towards the halo. Current models suggest that ultraviolet light from disk stars can only account for some of the ionization. The N V absorption lines appear to require collisional heating from a pervasive hot corona (Sembach & Savage 1992).

In addition to hot gas, there are two additional components of the intercloud medium. By mass, most of the intercloud medium is in the form a 'warm neutral' or a 'warm ionized' medium. These

phases extend far beyond the thin disk of cold gas (Table 1.4). The warm ionized medium (Reynolds layer) was firmly established by 1973 from three independent observations: low frequency radio observations, time delays in radio pulses from pulsars, and through direct observations of H^+ recombination. The dominant ionization source appears to be dilute flux from young disk stars.

Roughly half of the interstellar HI appears to be located in the 'warm neutral' component of the intercloud medium. This intercloud HI was first identified in 1965 as the source of the ubiquitous, relatively broad (velocity dispersion ~ 9 km s^{-1}) 21-cm emission features that had no corresponding absorption when viewed against bright background radio sources. The large velocity dispersions and the absence of absorption imply temperatures of 5000–10 000 K. Observations of the Lyα absorption line of HI toward bright stars show that this gas has a mean extent from the midplane of 500 pc, i.e., much thicker than the cold neutral disk. If the warm neutral medium is in pressure equilibrium with the cold component, then it would be clumped into regions occupying 35 per cent of the intercloud volume with a density of 0.3 cm^{-3} at the midplane (Table 1.4), although these numbers are highly uncertain.

The Stellar Radiation Field A major uncertainty of ISM studies has been an accurate determination of the radiation field along an arbitrary sight line through the Galaxy. Besides kinematic information, spectral lines provide information on densities, temperatures and elemental abundances. This interpretation is routinely hampered by the uncertain external radiation field impinging on the gas. In published work, fundamental physical quantities are usually quoted for a range of external fields, through the use of sophisticated photoionization codes, with the expectation that the external field will ultimately be tied down.

With an accurate knowledge of the stellar and gas distributions, it will be possible to derive the spectrum of the local radiation field everywhere within a 20 kpc volume. Some regions may be dominated by coronal emission but these will be easy to identify since the spectrum will be much harder than produced in stellar atmospheres. Regions that cannot be understood from the stellar radiation field or coronal emission alone will be interesting in their own right. Possible explanations could include (i) a flawed model atmosphere for a given stellar type, or (ii) cosmic rays or hard radiation from variable or impulsive sources like x-ray binaries, soft gamma-ray repeaters or supernovae (perhaps evident from follow-up radio or high-energy observations).

The dominant source of ionizing photons is hot, young stars in the disk. These are thought to account for a major fraction of the Reynolds layer ionization although it is unknown just how half the UV photons manage to escape the dense star-forming regions. It is possible that HII regions are highly porous or that some (runaway) O stars are ejected from the stellar nurseries. The observed Hα flux along the Magellanic stream suggests that $5-10$ per cent of the UV field escapes into the Galactic halo (Bland-Hawthorn & Maloney 1999).

Cloud Distances and Associations A comprehensive understanding of ISM physics is hampered by the difficulty in assigning distances to clouds. As a result, fundamental physical quantities such as size ($\propto$ distance) and mass ($\propto$ distance^2) are unknown for all but a few clouds. In the case of high-velocity clouds, distances are so poorly known in most cases that the debate over whether the clouds are galactic or extragalactic entities continues nearly 40 years after their discovery in 1963. Distances to neutral and ionized clouds as well as dust clouds can be bracketed by stars observed along the line of sight. This technique has only been tried for a handful of clouds within a few kpc. The method depends on the availability of stars with known distances where some fall in front of the cloud and some lie beyond the cloud and are hence seen in absorption. Knude & Høg (1998) have combined Hipparcos parallaxes, Tycho colour indices, and spectral classification data to estimate distances of molecular clouds. In Knude & Høg (1999) 21 cm emission data were also taken into account resulting in distances of intermediate- and high-velocity clouds. GAIA will provide distances within 10 per cent to hundreds of clouds in the halo and disk.

The disk UV field provides the basis for an important distance indicator for several different classes of sources: pulsars (Taylor & Cordes 1993), high-velocity HI clouds and H^+ clouds in the Galactic halo (Bland-Hawthorn et al. 1998; Sembach et al. 1999). The radio signal from pulsars is broadened by interstellar plasma scintillation. The distance is established in part by interpreting the inferred free electron column in the context of a spiral arm model. High-velocity cloud distances are derived from the Hα recombination flux. If the HI is opaque, the Hα flux is directly related to the external field strength. If the distribution of O stars in the Galaxy could be established, the pulsar and high-velocity cloud distance uncertainties would be reduced by more than order of magnitude.

There are many instances where galactic emission nebulae or dust clouds are being heated by known stellar sources often in close association with gas or dust. Particular examples include H^+ regions due to O stars, B star nebulae, potassium shells around K stars, planetary nebulae, supernova remnants, Wolf-Rayet nebulae, and photodissociation regions. A full stellar inventory within a large volume will allow a far wider association to be made between star groups and structures in the ISM. It will be possible to derive quantitative heating estimates of the warm neutral medium through the known starlight impinging on PAH molecules and dust grains. Furthermore, all stars generate winds with known mechanical luminosities (Abbott 1982) and it will be possible to include their contribution through collisional heating. The ISM is replete with complex structures like wind-blown bubbles, chimneys, loops, arches, and holes, particularly in the vicinity of spiral arms. A large body of theoretical work has shown how these cavities are created by stellar winds or by young stellar associations. Through their common association, stellar positions from GAIA will allow accurate distances to many of these structures.

1.3 The Star Formation History of the Milky Way

A primary scientific goal of the GAIA mission is the determination of the star formation histories, as described by the temporal evolution of the star formation rate, and the cumulative numbers of stars formed, of the bulge, inner disk, Solar neighbourhood, outer disk and halo of the Milky Way. Given such information, together with the kinematic information from GAIA, and complementary chemical abundance information, again primarily from GAIA, the full evolutionary history of the Galaxy is determinable.

The star formation history defines the luminosity evolution of the Galaxy directly. In combination with the relevant chemical abundance distributions, the accretion history of gas may be derived. Together with kinematics, the merger history of smaller stellar systems can be defined. The sum of these three processes forms what is loosely known as 'galaxy formation'. Analysis of the GAIA results will provide the first quantitative determination of the formation history of our Galaxy.

The determination of the relative rates of formation and/or accumulation of the stellar populations in a large spiral, typical of those galaxies which dominate the luminosity in the Universe, will provide for the first time an ability to test galaxy formation models in a quantitative manner. Do large galaxies form from accumulation of many smaller systems which have already initiated star formation? Does star formation begin in a gravitational potential well in which much of the gas is already accumulated? Does the bulge pre-date, postdate, or is it contemporaneous with, the halo and inner disk? Is the thick disk a mix of the early disk and a later major merger? Is there a radial age gradient in the older stars? Is the history of star formation relatively smooth, or highly episodic? In addition to their immediate and direct importance, answers to such questions will provide uniquely a template for analysis of data on unresolved stellar systems, where similar data can never be obtained.

1.3.1 Method

Determination of the evolutionary history of the Galaxy is a sufficiently major project that one must consider both the acquisition of the relevant data, and one's ability to analyse such data. In general, degeneracy in a star's observational parameters between age, metallicity and extinction, convolved with observational errors and uncertain distances, have made determination of the star formation history of a mixture of stellar populations unreliable and non-unique. The best available analyses involve comparison of an observed colour-magnitude diagram with a model population. While such analyses are powerful, they can never be proven unique. The GAIA data combined with new analytical tools, specifically developed for GAIA, will resolve this ambiguity.

The increasing application of HST studies which resolve the stellar populations of nearby systems has initiated quantitative investigation of star formation histories in these systems through comparison of the observed H-R diagram with synthetic ones (e.g. Chiosi et al. 1989; Aparicio et al. 1990 and Mould et al. 1997 for Magellanic and local clusters, and Mighell & Butcher 1992; Smecker-Hane et al. 1994; Tolstoy 1995; Aparicio & Gallart 1995 and Mighell 1997 for dSph companions to the Milky Way). This is customarily done by constructing a statistical estimator of how closely a synthetic H-R diagram constructed from an assumed star formation history resembles an observed one, and then selecting that star formation history from amongst the set of calculated models which maximizes the value of this estimator (e.g. Tolstoy & Saha 1996). The robustness of the approach is undermined by the degree of subjectivity associated with defining the set of plausible models one considers.

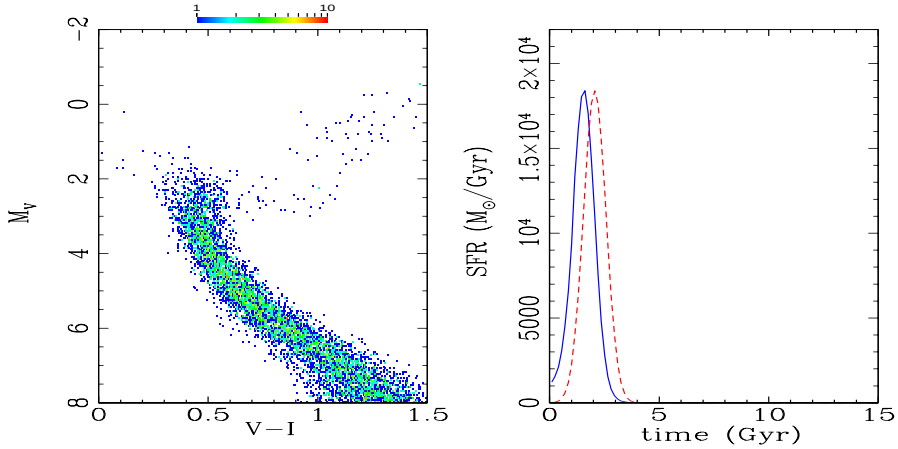


Figure 1.12: Left: synthetic H-R diagram for an 'arm' component, using isochrones of Solar metallicity, and including a 10 per cent observational error. Right: inferred star formation rate versus time for the 'arm' component, assuming a metallicity slightly off, of [Fe/H] = 0.18, solid curve. The dotted curve shows the input star formation rate. Comparing these two curves shows the extent of the age-metallicity degeneracy in this case, i.e., fine details in the build up history of the disk could be accurately recovered, if metallicities of the relevant stars were known to ∼0.2 dex accuracy, and distances to 10 per cent.

In the GAIA study of the Galaxy, it is the precise form of the star formation history which serves as a constraint on a theory (e.g., a collection of randomly located bursts as fragments accrete or a more uniform function as gas cools, for the build-up of the Galaxy). Thus one must avoid *a priori* assumptions about the star formation history. This is possible only by use of direct inversion. A maximum likelihood variational calculus has therefore been developed for the GAIA analysis. Hernandez et al. (1999a) give a full description of the method and its applicability. The model has been used to analyse a realistic simulation of GAIA observations of an old metal-poor stellar population, a young metal rich population, and a mixed population. These simulations show that GAIA data can indeed meet the scientific goal of determination of the star formation history of the Galaxy's stellar populations.

The part of the H-R diagram which is maximally sensitive to star formation histories is near the main-sequence turnoff, the region occupied by F and G stars. All analyses require data down to this luminosity limit. Given the degeneracy of isochrones of different ages in the main sequence region, lower-luminosity stars are not needed for this analysis. The age-metallicity degeneracy of stellar isochrones however necessitates star by star abundance data. Thus, there are three distance regimes to consider: (i) where high-precision data are available at the turnoff: this corresponds to a distance limit of ~ 4 kpc for an old population; (ii) where lower quality data are available at the turnoff: this extends the analysis volume to the Galactic bulge; (iii) where only higher mass stars, and recent star formation, can be studied: this is possible out to the edge of the Galactic disk.

These regimes have been quantified through detailed simulation. Analysis of the type of results shown in Figure 1.12 shows that ages may be recovered reliably with an accuracy of 10 per cent given luminosity and temperature data to 10 per cent accuracy, and metallicities with an accuracy of ~ 0.2 dex. These constraints are met by GAIA for old turnoff stars with $V < 17$ mag, distance within ~ 4 kpc. For such stars GAIA will have complete spectroscopic data (from the radial velocity spectrometer) as well as high-precision multi-colour photometry, proper motions and distances. For low-reddening stars within this distance, Figure 1.13 shows that GAIA photometric, spectroscopic and parallax data allow determination of the star formation histories of the stellar populations to an accuracy of better than one-tenth of the age of the Galaxy. The right panel in Figure 1.13 shows the result of the inversion procedure, in cases where the GAIA data are of high quality, i.e., for all stars brighter than $\sim 15 - 16$ mag. In such cases, even the few stars in the rather rare Population II oldest component can be separated from the younger main sequence to accurately recover the shape of the star formation history throughout the entire lifetime of the Galaxy.

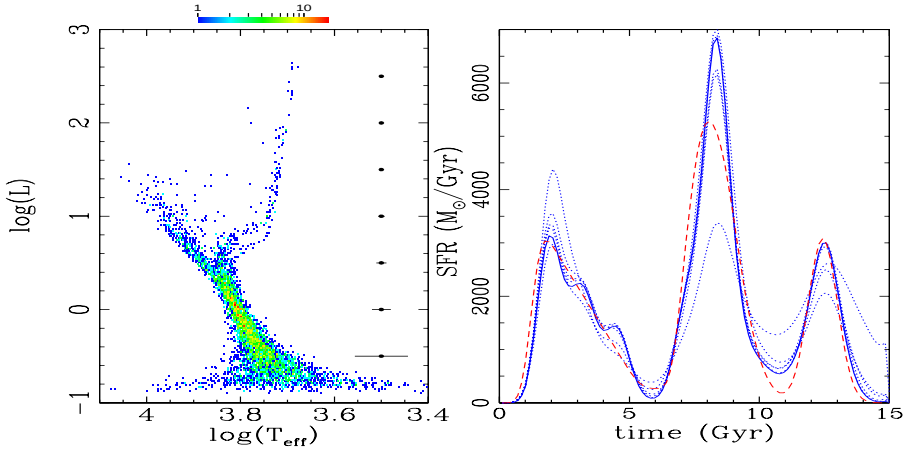


Figure 1.13: Left: synthetic H-R diagram appropriate for good quality GAIA data. The length of the bars to the right corresponds to 2σ in the log (L) error to each side of the dots. Right: the derived star formation history following inversion of the data in the left panel. The true input star formation history is shown by the long dashes. The dotted lines are the successive 2, 4, 6, 8 and 9 iterations of the inversion method. The 10th iteration is given by the solid curve, showing rapid convergence and a good recovery of the input star formation history, when GAIA-quality data are available.

1.3.2 Disk, Bulge and Halo

Given the very considerable radial range of stellar orbits in the Galactic potential, the volume of the Galaxy within 4 kpc of the Sun is sufficiently large to provide a fair sample of the entire stellar population structure of the Milky Way, with two exceptions. One is the very outer Galactic disk, together with the speculative possibility that there is a significant high angular momentum population in the outer halo. Such (rare) stars would have to be on orbits such that they never come within ~ 15 kpc of the Galactic centre. The second case involves stars on low-energy orbits, which are always interior to the sampled volume. This includes the important cases of the inner disk, and the Galactic bulge. Both these situations can be studied by GAIA. Since the limiting magnitude is $V = 20$ mag, old turnoff stars at the Galactic bulge and to the disk limits will be observed by GAIA, while any anomalous outer halo population will be discovered, if present, and analysed from the GAIA observations of K-giants, which are complete within the entire Galaxy. GAIA data at these faint limits is however less precise, and no spectroscopy will exist.

In the outer disk the situation is well defined. Disk-halo discrimination from GAIA proper motions is possible. The metallicity range is determinable from observations of the brighter K giants, as is the spatial density distribution. This information can be included in the formal inversion process, as an additional constraint function. The GAIA data may need to be supplemented by ground-based abundance data, if the abundance range is wide. GAIA will however provide a complete and reliable selection of the appropriate sample for observation. Derivation of the star formation history of the outer disk will thus be feasible, even for old stars, given feasible further development of the analysis algorithms. The situation in the inner Galaxy is more complex, and arguably more important. A line of sight towards the bulge includes the foreground disk and spiral arms, and a complex extinction distribution. Figure 1.14 illustrates the present observational situation.

A substantial observational complication in this case is that the non-uniform extinction distribution tilts the disk zero-age main sequence such that it overlaps with any young bulge stars. An additional complication is that the metal abundance distribution in the bulge proper is known to be extremely broad (see Figure 2 of Wyse et al. 1997), so that even error-free photometric data near the turnoff will show a wide colour range, which must be distinguished from a range of ages and reddenings. Recent analyses of COBE data suggest a scale length for the bulge of ~ 300 pc, so that the line of sight distance range is small by comparison. The analysis is not limited by parallax errors, but population discrimination. Determination of the true bulge age distribution is thus impossible without further data. GAIA uniquely will provide such data.

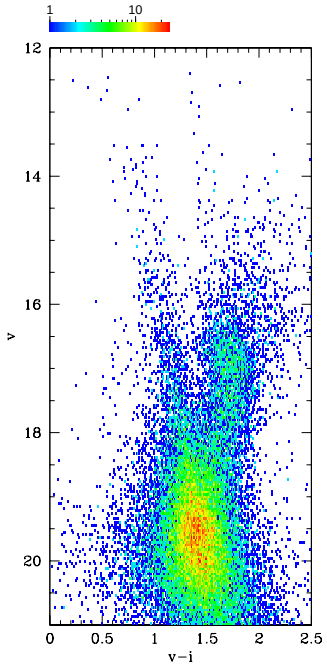


Figure 1.14: An observed colour-magnitude diagram for a field in Baade's Window, with similar completeness and photometric accuracy to the GAIA data. The bulge turnoff appears at V∼19.5 mag. The stars in the blue 'plume' from (V,V-I) = (12,0.5) to (19,1.5) are a mixture of foreground young disk and any genuine young bulge stars. GAIA proper motion data will allow this distinction to be made.

Distinction between disk and bulge stars is possible using proper motion data. The (old) bulge stellar rotation curve is roughly linear, at 25 km s^{-1} kpc^{-1} (Ibata & Gilmore 1995). The inner disk rotation curve, at the distances from the centre appropriate for foreground stars, is close to 200 km s^{-1}, as the asymmetric drift is always small in this radial range. The corresponding mean proper motions of disk and bulge differ by 20σ for GAIA accuracy at these magnitudes. Thus GAIA will efficiently identify any young bulge stars, and allow study of their kinematics. Individual metallicities can be derived for any such stars from the ground, defining directly the recent star formation history of the inner Galaxy, and its spatial and kinematic properties. For older stars, GAIA will again define a fair sample of turnoff stars, ideally matched to multi-slit observations on VLT and Gemini-S for abundance determinations. Such a sample, combined with GAIA photometry and kinematic data, will allow a robust determination of the age distribution of the Galactic bulge stellar populations as a function of metallicity and kinematics. A realistic simulation is presented in Figure 1.15.

The main-sequence turnoff of the oldest stellar populations corresponds to apparent magnitude $\sim$ 20 mag at the Galactic centre, and near the apparent outer edge of the disk. GAIA can therefore determine the full star formation history of the near half of the Milky Way for all ages. This provides a sufficiently large sample that the global star formation history of our Galaxy can indeed be determined by GAIA. The star formation history at more recent times can also be determined for larger distances. The implications for GAIA include: (i) the GAIA data set can be analysed to determine the star formation history of the Milky Way Galaxy, one of the primary science goals; (ii) this determination requires a well-known selection function from the central bulge to the outer Galactic disk, at the GAIA limiting magnitude; (iii) individual stellar metallicities are required for a large, well-defined, although not necessarily complete, sample of stars, with an accuracy ideally as good as 0.25 dex; (iv) effective temperatures must be determinable for these same stars with an accuracy of about 10 per cent; (v) this precision shares the final uncertainty equally between stellar temperature, stellar metallicity and distance uncertainty.

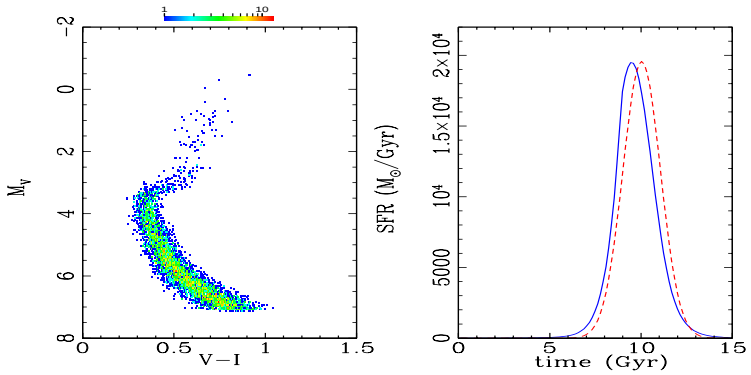


Figure 1.15: Left: synthetic H-R diagram for an 'old bulge' component, using isochrones of [Fe/H] = –1.5, and including a 10 per cent observational error. Right: the inferred star formation history, assuming a correct metallicity of [Fe/H] = –1.5, solid curve. The dotted curve shows the model history. Their comparison shows the accuracy with which the age structure of an old population could be recovered using current statistical methods, if errors of ≤10 per cent in luminosity and temperature, and metallicity measurements, were available.

1.3.3 Recent Birth-Rate of Stars in the Solar Neigbourhood

Very young (pre-main sequence) stars are easily identified in X-rays (e.g., Walter et al. 1988). Many objects as young as the classical T Tauri stars (CTTS) can be identified by their soft X-ray emission (while they lack the characteristic emission line strength, chromospheric activity and large infrared excess of classical T Tauri stars). Such 'Weak-lined T Tauri stars' have been shown to be coeval with the CTTS. ROSAT found these objects in large numbers, and many more will be discovered by XMM. Flux-limited X-ray and EUV surveys have also identified a population of older ($t \lesssim 10^9$ yr) active main-sequence Solar-type stars within 1 kpc from the Sun, with photospheres indistinguishable from older and quieter stars (except for a higher Li abundance). This population is concentrated near the Galactic Plane (Favata et al. 1988; Favata et al. 1993; Jeffries 1995; Micela et al. 1993; Sciortino et al. 1995; Tagliaferri et al. 1994), and is X-ray bright due to the higher dynamo efficiency in young stars, which have not undergone significant magnetic braking.

Such young, activity-selected samples can be used to determine the stellar birth-rate in the last 10^9 years, by using Ca II as an activity indicator (Soderblom et al. 1991), and comparing the observed number counts with predictions of a Galaxy model (Micela et al. 1993; Guillout et al. 1996). X-ray observations by themselves are insufficient for this purpose: they do not constrain the stellar birthrate, nor the detailed spatial distribution of the young stellar population. The GAIA measurements will provide accurate individual distances and motions of activity-selected stars out to a distance of $\sim$ 1 kpc. This will allow a detailed reconstruction of the last 10^9 years of star formation history in the Solar neighbourhood.

GAIA will also measure accurate distances larger than 1 kpc. Fill-in (or emission) of the Ca II lines will be a quantitative indicator of activity (and therefore youth) for late-type stars, and will be available with sufficient S/N down to $V \sim 16$ mag, thus extending significantly the volume within which active stars can be identified and studied.

1.3.4 Local Luminosity Function

The local stellar luminosity function, the absolute number of stars in the Solar neighbourhood as a function of intrinsic luminosity, is the outcome of the history of star formation, chemical evolution, dynamical evolution, stellar evolution, and any temporal evolution of the stellar initial mass function, throughout the history of the Milky Way. A huge part of modern astrophysics is critically dependent on a reliable determination of the luminosity function. The primary requirement is that this determination be complete in a sufficiently large volume as to be representative, and of high statistical weight, and that it be complete to the hydrogen burning limit.

Among the most important consequent analyses of wide significance, one notes explicitly the following:

- the luminosity function, through the mass-luminosity relation, determines the initial mass function, a key concept in understanding the physics of star formation (Section 1.3.5);
- any structure in the luminosity function near the main-sequence turn-off contains the history of star formation in the local Galaxy;
- the total mass in stars and their remnants, when combined with measurements of the local Galactic gravitational potential, determines the distribution of any dark matter in the Galactic disk (Section 1.2.8);
- spatial, temporal or abundance-dependent gradients in the luminosity function determine any systematic changes in the initial mass function;
- the luminosity function for stars evolved past the main-sequence turnoff is the test of stellar evolutionary models, through the relative numbers of stars in different evolutionary phases;
- the white dwarf luminosity function is a test of past high-mass star formation histories, and is an independent chronometer of stellar isochrones;
- the binary fraction as a function of primary luminosity (mass) is a test of angular momentum transport in star formation;
- the distribution function of wide binaries is a measure of the Galactic disk tidal field.

The stellar luminosity function is definable only in one way, by completion of an accurate census of all the stars in some volume of space. Accurate distances and apparent luminosities for large samples of apparently faint stars are crucial: their absence has bedevilled luminosity function determinations until now. Very large volumes are crucial for intrinsically rare or short-lived classes of object. Faint limiting magnitudes are essential for intrinsically low-luminosity objects.

The absolute magnitude of the hydrogen-burning limit is uncertain by a factor of two in luminosity with current data and models, but is close to $M_V = 15$, for both Solar abundance and very metal-poor stars. Coincidentally, a rather similar absolute magnitude corresponds to white dwarfs of the Hubble age. Thus, the GAIA survey limit, near $V = 20$ mag, corresponds to a distance of approximately 100 pc for all these classes of stars. The corresponding parallax is ~ 10 mas, so that GAIA will provide a distance accuracy of $\sim$3 per cent for each star at the sample limit. The number of stars in the GAIA luminosity function cannot be predicted reliably, simply because the luminosity function is at present very poorly known at low luminosities. The current range of estimates corresponds to predictions between 1000 and 40 000 stars per magnitude at the hydrogen burning limit for disk stars, a few to a few hundred metal-poor halo stars, between 100 and 5000 thick disk stars, and in total a few hundred old disk white dwarfs, and some 1000 halo white dwarfs (cf. Section 1.4.7).

The current situation is summarised in Figure 1.16. This shows the only complete (if not to 20 mag) sample for low-luminosity stars, which is based on a few dozen stars within 5.2 pc of the Sun. For comparison, the formally most exact determinations of the luminosity function from discrete star counts, using modeling of photometric parallax, are also shown. In addition to illustrating the poor statistics, and poor agreement between determinations, this figure also illustrates the very large effect of unresolved binaries. The solid model curve and the lower dotted model curve represent the same mass function: the differences between them are a manifestation of spatial resolution, and the very large Malmquist corrections necessary to convert pencil beam photometric parallax data to a luminosity function (see also Section 1.7.4).

1.3.5 Initial Mass Function in Clusters and Associations

GAIA will provide a complete and homogeneous census of the stellar content of a large number of clusters and associations or moving groups (Section 1.2.4). This will enable a statistically significant study of the initial mass function within each group separately and a meaningful intercomparison of the results for different groups. Establishing the initial mass function for groups of all ages over a large volume will greatly advance the understanding in detail of the origin of the field-star population and its mass function.

Reliable and precise determinations of the IMF still suffer from a number of major problems. In the case of studies of field stars in the Galactic disk large and often uncertain corrections have to be made for the effects of sample selection, stellar evolution, variations of disk scale-height with stellar age and the star formation history of the Galaxy. Therefore open clusters and OB associations, where these problems are absent or much less severe, are preferred sites for IMF studies. However, in this case the drawbacks are the contamination of cluster membership lists by field stars, the incompleteness at small stellar masses, the presence of mass-segregation (even in the very young systems) and the small sample size (Scalo 1998). Finally an often overlooked but nonetheless very important

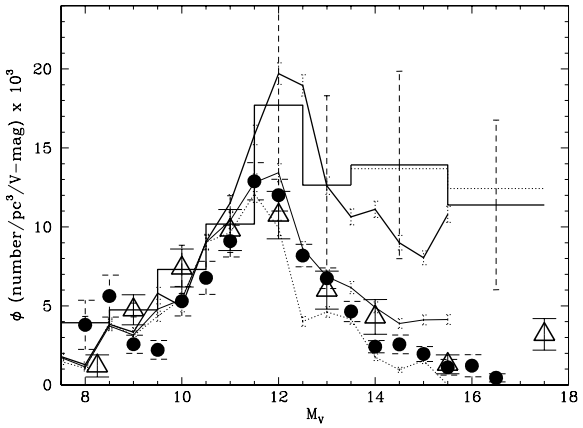


Figure 1.16: The present status of knowledge of the Solar neighbourhood stellar luminosity function. The various curves are as follows: solid histogram: derived from stars within 5.2 pc, contained in the Gliese catalogue (Kroupa 1995); dotted histogram, as for the solid histogram, but illustrating the effect of adding one newly-discovered star within 5.2 pc (Gl 866C; Kroupa 1998); triangles, derived from modeling deep HST star counts, containing 47 dM stars (Gould et al. 1997); solid dots, combined photometric pre-HST luminosity function (Kroupa 1995). The star count data are overlaid with best-fitting models assuming single stars (thick solid line), a 50 per cent binary fraction (thin solid line), and a 100 per cent binary fraction (thin dotted line). The models are based on the Kroupa et al. (1993) mass function.

problem is that of the uncertainties in our knowledge of the mass-luminosity relation. The derivative of this relation enters into many IMF determinations, making them very sensitive to any features in the mass-luminosity relation, such as inflections or slope changes (see D'Antona 1998).

Based on the recent literature on the IMF, Scalo (1998) concludes that either: (i) the uncertainties are so large that very little can be said about an average IMF or IMF variations; or (ii) if the observations are taken at face value, there are strong indications of IMF variations. The large uncertainties are essentially due to the difficulties in converting the observed luminosity function into mass functions and to the generally small sample sizes from which the IMF of a stellar cluster is determined. Establishing whether or not real variations exists in the IMF and how these correlate with physical conditions forms a crucial test of star formation theories (e.g., Elmegreen 1999).

GAIA will be uniquely capable of tackling the problems associated with the determination of an accurate initial mass function:

- GAIA will determine high-quality orbits for thousands of nearby visual binaries (Section 1.5.4). This will lead to a much better empirical determination of the mass-luminosity relation. This relation will be established within the photometric system of GAIA which is also the system used for the observation of the clusters;
- accurate astrometry is a very powerful tool for obtaining a stellar census of open clusters and associations, reducing the number of candidate members in stellar clusters by very large factors (de Zeeuw et al. 1999). Moreover, GAIA will also provide radial velocities, photometry and spectroscopic information, which will lead to a further refinement of the membership lists of stellar clusters and hence very clean raw data with which to determine the initial mass function;
- once an accurate membership list of the clusters has been established it becomes feasible to carry out follow-up high-resolution spectroscopic observations from the ground, necessary to determine masses at both the high (> 10 M$_\odot$) and low-mass (< 1 M$_\odot$) stars;
- in the Pleiades cluster, which is ∼125 Myr old, the low end of the main sequence (i.e., the location of masses $M \simeq 0.075\,M_\odot$ which are at the lithium depletion boundary) is at $I \leq 17.5$ mag (Stauffer et al. 1998b). Hence, as every object brighter than $I = 20$ mag (including the Pleiades brown dwarfs with $I \sim 19$ mag) will be observed by GAIA, a complete stellar census of hydrogen burning stars can be obtained for all clusters and associations younger than about 100 Myr within ∼1000 pc from the Sun;
- because in all cases the entire cluster is surveyed the problems caused by mass-segregation in clusters are circumvented. Although for older clusters the dynamical evolution will still adversely affect the IMF determinations, one can then undertake a systematic study of this phenomenon for large samples of clusters;
- any determination of the IMF must account for the presence of binaries, triples, etc. This will be greatly facilitated by the multiplicity survey that GAIA will provide (Section 1.5).

1.4 Stellar Astrophysics

GAIA will provide distances of astonishing accuracy for all types of stars of all stellar populations, even the brightest, or those in the most rapid evolutionary phases which are very sparsely represented in the Solar neighbourhood. With the parallel determination of extinction/reddening and metallicities by the use of multi-band photometry and spectroscopy, this huge amount of basic data will provide an extended basis for reading *in situ* stellar and galactic evolution. All parts of the Hertzsprung–Russell diagram will be comprehensively calibrated, including all phases of stellar evolution, from pre-main sequence stars to white dwarfs and all existing transient phases; all possible masses, from brown dwarfs to the most massive O stars; all types of variable stars; all possible types of binary systems down to brown dwarf and planetary systems; all standard distance indicators (pulsating stars, cluster sequences, supergiants, central stars of planetary nebulae, etc.). This extensive amount of data of extreme accuracy will stimulate a revolution in the exploration of stellar and Galactic formation and evolution, and the determination of the cosmic distance scale.

1.4.1 Stellar Structure and Evolution

One of the triumphs of stellar evolution theory is a detailed understanding of the preferred location of stars in the physical Hertzsprung–Russell diagram, which plots luminosity versus temperature (Figure 1.17). There are a number of uncertainties associated with stellar evolution models, and hence age estimates based on the models.

Probably the least understood aspect of stellar modeling is the transport process of matter, angular momentum and magnetic field at macroscopic and microscopic levels, including in particular the process of convection. Numerical simulations hold promise for the future, but at present one must view properties of stellar models which depend on the treatment of convection to be uncertain, and subject to possibly large systematic errors. Main sequence stars and red giants have surface convection zones. Hence, the surface properties of the stellar models (such as its effective temperature, or colour) are rather uncertain. Horizontal branch stars have convective cores, so the predicted luminosities and lifetimes of these stars are subject to possible systematic errors. Other domains such as the statistical physics at high density and/or low temperature or the nuclear reaction rates of heavy nuclei also require improvement.

This lack of knowledge has consequences on topics as fundamental as the chemical evolution of the Universe, the rate of formation of heavy elements and of dust in the interstellar medium, and on the measurement of the age of the Universe. Understanding the dynamics of stellar interiors remains a key challenge for astronomy.

A stellar model is constructed by solving the four basic equations of stellar structure: (1) conservation of mass; (2) conservation of energy; (3) hydrostatic equilibrium and (4) energy transport via radiation, convection and/or conduction. These four, coupled differential equations represent a two point boundary value problem. Two of the boundary conditions are specified at the centre of the star (mass and luminosity are zero), and two at the surface. In order to solve these equations, supplementary information is required. The surface boundary conditions are based on stellar atmosphere calculations. The equation of state, opacities and nuclear reaction rates must be known. The mass and initial composition of the star need to be specified. Finally, as convection can be important in a star, one must have a theory of convection which determines when a region of a star is unstable to convective motions, and if so, the efficiency of the resulting heat transport. Once all of the above information has been determined a stellar model may be constructed. The evolution of a star may be followed by computing a static stellar structure model, updating the composition profile to reflect the changes due to nuclear reactions and/or mixing due to convection, and then re-computing the stellar structure model.

The agreement between predicted and observed properties of stars has remained qualitative due to the modest accuracy and relative scarcity of the relevant observed quantities. The development of accurate astrometry with Hipparcos, and high-resolution, high signal-to-noise ratio spectroscopy has allowed major progress in this area, but measurement of the global stellar parameters is often insufficient for testing the internal regions of the stars, where the evolution proceeds. To date, direct information on these regions is available only for the Sun, but even there theory is unable to reconcile the observed neutrino flux with the helioseismology data.

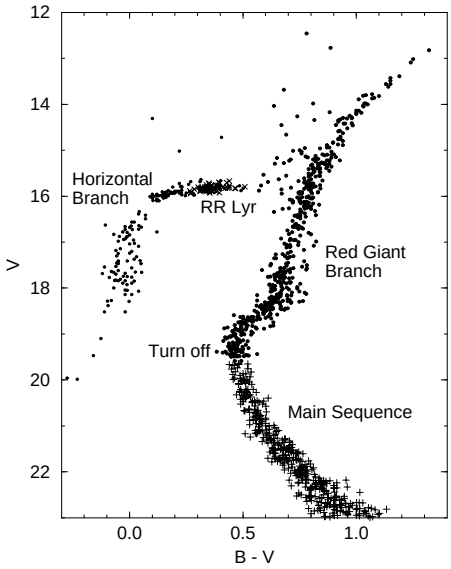


Figure 1.17: Colour-magnitude diagram of the globular cluster M15 (Durrell & Harris 1993).

The next decade will see the opening of the field of asteroseismology from space. It will provide for the first time direct indicators of the physical status of stellar interiors. The corresponding data, combined with measurement of global parameters, as provided by GAIA will allow a major step improvement in our understanding of stellar evolution.

Stellar oscillation frequencies can be used to constrain stellar evolution through both a direct as well as an inverse approach. In the former, given initial approximate stellar parameters and a set of stellar models, the frequency information can be used to derive the fundamental stellar parameters, i.e., the mass and radius of the oscillating stars, thus placing it on the H-R diagram and determining its evolutionary status (Christensen-Dalsgaard et al. 1995; Guenther & Demarque 1996; Petersen & Christensen-Dalsgaard 1996; North et al. 1997; Popper 1997). In the inverse approach, the observed fundamental stellar parameters are taken as a starting point, i.e., the mass M, the stellar radius R, the effective temperature $T_{\rm eff}$, and the chemical composition. With the exception of the double-lined eclipsing binary systems, the accurate determination of fundamental stellar parameters requires knowledge of the distance. Given a stellar structure model, the oscilation frequencies are then predicted, and compared with the measurements, thus testing the underlying physics of the assumed stellar structure.

In the few cases for which asteroseismology has been used to derive accurate stellar parameters, i.e. for the nearby G sub-giant η Boo (Bedding et al. 1998) and for the two δ Scu stars SX Phe and AI Vel (Høg & Petersen 1997), the resulting uncertainties on the stellar parameters, and in particular on the absolute luminosity, match very well the uncertainty deriving from the Hipparcos parallaxes. With the precision offered by GAIA parallaxes, the challenge for the models (and for asteroseismology) to match the accurately measured luminosity will be much greater.

As described in Section 2.3, the global stellar parameters used in the field of stellar structure and evolution need to be obtained by GAIA itself. The key to success is the building of a complete and homogeneous sample covering a large variety of independent parameters. The stellar absolute luminosity is derived from the parallax and the apparent magnitude, corrected for extinction, which can be deduced from the GAIA photometric and spectroscopic data. These also provide the effective temperature $T_{\rm eff}$, metallicity indicators, and the projected rotational velocity of the stars, $v \sin i$. Ages can then be inferred from the location of stars in the Hertzsprung–Russell diagram.

Unfortunately, masses can be measured directly only in special cases, where the gravitational interactions with other bodies are easily measurable. This is the case for binaries, discussed in Section 1.5. An additional possibility comes from astrometric microlensing (Section 1.4.10). The large number of systems for which the mass will be measured by GAIA will be used to validate the modeling of stars for which mass is known, providing in turn indirect estimates of the mass of other stars, for example through the mass-luminosity relation.

Single stars, which will most likely constitute the majority of targets for future asteroseismology space missions such as COROT, MOST and MONS are, from the point of view of *a priori* accurate stellar parameters, the worst case. Dynamical mass determinations cannot be performed on individual stars, for which the only possible mass determination will be through models. Accurate luminosities can be derived from accurate parallaxes and photometry, which, given $T_{\rm eff}$, will yield stellar radii.

1.4.2 Luminosities Through Distances

Luminosity estimates are based on determinations of stellar distances and interstellar absorption (the former can be determined directly only by measurement of the trigonometric parallax, while the latter will be estimated from the medium-band photometry). GAIA will provide distances to an unprecedented 0.1 per cent for 7×10^5 stars out to a few hundred pc, and to 1 per cent accuracy for a staggering 2.1×10^7 stars up to a few kpc (see Tables 1.5 and 1.6). Distances to 10 per cent will reach beyond 10 kpc, and will cover a significant fraction of our Galaxy, including the Galactic centre, spiral arms, the halo, and the bulge, and—for the brightest stars—to the nearest satellites. The faint limiting magnitude allows investigation of white dwarfs (Section 1.4.7), as well as the bottom of the main sequence down to brown dwarfs (Section 1.4.6). For the first time, this will provide an extensive network of accurate distance measurements for all stellar types.

With Hipparcos, direct distance measurements reached 100 pc with 10 per cent precision (and perhaps as far as 200 pc in the best cases), but this distance horizon was insufficient to include rare but astrophysically important categories of stars such as O stars, Cepheids, and RR Lyrae variables. Once the parallax error drops significantly below the ~ 0.1 per cent level (something that for GAIA will be true for stars out to ~ 100 pc), the uncertainty in the distance ceases to be the true limiting factor in the accuracy with which a star can be positioned on the H-R diagram (Lebreton 2000). At that level, the uncertainty in the determination of effective temperatures, bolometric corrections, as well as photospheric metallicities, dominates over any distance-related uncertainty. Effects related to differences between the surface and interior chemical compositions (due, for example, to diffusion) are also likely to play a role. Another limiting factor in the accuracy with which absolute magnitudes can be determined will be intrinsic photometric variability, with Hipparcos results showing that most stars are variable to ~ 0.1 per cent. The uncertainties in temperature and metallicity determinations result from errors in the measurements as well as from the models used to interpret stellar spectra. In particular, the limitations of the relatively simple atmospheric models used to interpret stellar spectra, together with the errors in the atomic rate coefficients, will be apparent.

As importantly, the volume over which 'good accuracy' parallaxes (i.e., ≤ 1 per cent) will be available will increase: for stars with $M_V < 5$ it will extend out to ~ 1 kpc, and even for M dwarfs it will reach some tens of pc. This means that essentially every stellar type, even those in fast stages of stellar evolution, will be sampled in large numbers. This will allow to accurately measure the luminosity of every type of star in the H-R diagram, providing formidable constraints to stellar structure models and evolution theory. With large samples of 'identical' stars one can separate apparent 'twins' and focus on properties presently neglected (such as the magnetic configuration).

For stars brighter than $M_V = -5$, the limiting factor is the precision in the parallax. At the other end, for stars fainter than $M_V = 14.5$, the limiting factor is the apparent magnitude ($V \leq 20$ mag): for all these stars the relative parallax error is always smaller (or even much smaller) than 10 per cent. For all other stars the limit comes from a combination of the distance and the apparent magnitude. Extremely accurate absolute luminosities will be obtained for the entire stellar content of the 50 pc sphere radius around the Sun, brown dwarfs included.

Hipparcos showed that many distances were previously underestimated (e.g., Binney et al. 1997a; Turon 1999), and that even stars within 25 pc from the Sun were not well known: some 40 per cent of CNS3 stars (Catalogue of Nearby Stars, Gliese & Jahreiß 1991) are further than 25 pc from Hipparcos observations (Turon & Perryman 1999). GAIA will provide the accuracy given by Hipparcos within 25 pc up to distances of more than 6000 pc for stars brighter than $V = 15$ mag, up to 2500 pc for stars brighter than $V = 18$ mag, and even up to 125 pc for stars brighter than $V = 20$ mag. Hipparcos also showed that the definition of luminosity classes needs revision: most luminosity classes are spread over one or two magnitudes, and there is no clear separation between many of them in some zones of spectral types (Gómez et al. 1997; Jaschek & Gómez 1998; Paunzen 1999).

1.4.3 Variability

In addition to providing the chromaticity correction needed for the astrometry (Section 3.2.4), and allowing the determination of basic astrophysical parameters, the GAIA large-scale photometric survey will have significant intrinsic scientific value (see also Paczyński 1997). The high photometric accuracy (see Section 8.2), the multi-colour simultaneous coverage and the 100–150 observations per target spread over five years, will provide massive samples of variable stars of nearly all types, including detached eclipsing binaries (Section 1.5.4), contact or semi-contact binaries (Section 1.5.5), and pulsating stars. Phenomena occurring on significantly shorter (e.g. stellar flares) or longer (e.g. Solar-like cycles) time scales fall outside the GAIA variability domain.

Table 1.5: Distances for given relative parallax errors, as a function of M_V (and corresponding V). N(HIP) is the approximate number of Hipparcos stars with the given relative accuracy, while N(GAIA) is the number expected for GAIA according to the Galaxy model described in Section 6.4.4.

σ_π/π	N(HIP)	N(GAIA)	M_V	V	d (pc)
< 0.1 per cent	0	0.7×10^6	0	7	270
			5	12	270
			10	15	100
			15	17	25
< 1 per cent	188	21×10^6	−5	7	2 700
			0	12	2 700
			5	15	1000
			10	17	270
			15	20	100
< 10 per cent	22396	220×10^6	−5	12	27 000
			0	15	10 000
			5	17	2 700
			10	20	1000

The pulsating stars include key distance calibrators such as Cepheids and RR Lyrae stars (Section 1.4.9) and long-period variables (Section 1.4.8). The existing samples are incomplete already at magnitudes as bright as $V \sim 10$ mag. A complete sample of objects will allow determination of the frequency of peculiar objects, and will accurately calibrate period-luminosity relationships across a wide range of stellar parameters (i.e., metallicity). Expected numbers of objects are highly uncertain, but they are very large, with estimations by Eyer & Cuypers (2000) leading to some 18 million variable stars in total, including some 5 million 'classic' periodic variables, 2–3 million eclipsing binaries, $\sim 3 \times 10^5$ with rotation-induced variability, 2000–8000 Cepheids, 60 000–240 000 δ Scuti variables, 70 000 RR Lyrae of which some 15 000–40 000 will be in the bulge, and 140 000–170 000 Miras (see also Section 6).

A systematic variability search will also allow identification of stars in short-lived but key stages of stellar evolution, such as the Helium core flash and the helium shell thermal pulses and flashes. An example of this is FG Sge, the nucleus of a planetary nebula, currently undergoing a He shell flash. Prompt processing of the GAIA photomety will identify many candidate objects of this kind, which will constitute ideal targets for detailed follow-up ground-based studies of these rare but important objects. GAIA observations of ∼50 000 Type Ia extragalactic supernovae and of variability in quasars and active nuclei are discussed in Sections 1.8.7, 1.8.8 and 1.8.9. Extra-solar planetary transits are considered in Section 1.6.

1.4.4 Physics of Stellar Interiors

Lebreton (2000) has reviewed recent developments in stellar structure and evolution in the light of Hipparcos results. The accurate and homogeneous astrometric and photometric data has resulted in more precise characteristics of individual stars and open clusters, and the consequent confirmation of certain aspects of internal structure theory. Further progress on atmospheric modelling is required, since it has implications on a number of observational parameters (effective temperatures, gravities, abundances, and bolometric corrections), theoretical models including outer boundary conditions, and colour calibrations. A better description of the transport processes of convection and diffusion, and related effects of rotation and magnetic fields, is required to improve theoretical stellar models, as well as further improvements or refinements in microscopic physic effects such as low-temperature opacities and nuclear reaction rates in advanced evolutionary stages. Lebreton

Table 1.6: GAIA observations: limiting apparent magnitude and distance for stars with a relative parallax error smaller than 10 percent, for zones with no extinction. The last column indicates the limiting factor: the G magnitude (G) or the distance (d).

Mv [mag]	Stellar type	$\langle V-I \rangle$ [mag]	$V_{\rm lim}$ [mag]	$V_{\rm lim}$ [mag]	$d_{\rm lim}$ [pc]	Limiting factor
	O V	–0.3	12.2	12.2		
–5	B0-G0 Ib	to			27 000	d
	all Ia and Ia0	4.7	12.2	8.5		
0	A0 V	0.01	15.0	15.0	10000	d
	K3 III	1.19	15.2	14.7	11000	d
5	G5 V	0.8	17.6	17.3	3300	d
10	M2 V	2.0	20.3	19.2	1150	d
	DB	0.0	19.7	19.7	870	d
15	M7 V	3.0	22.5	20.6	320	d
	DG	0.8	21.3	21.0	180	G
17	M8 V	3.2	23.1	21.0	170	G
20	brown dwarfs	4.5	24.5	21.0	80	G

(2000) concludes that what is needed from the observational side is more numerous samples of rare objects, including therefore distant objects as well as objects undergoing rapid evolutionary effects, an increased number of more common objects with extremely accurate data including masses, and a census over all stellar populations. The following effects are amongst those that would be probed in detail with the GAIA data.

Combining GAIA and Asteroseismology: the Size of Convective Cores Asteroseismology will be a considerable tool in the field of stellar structure. A key example illustrating the power of seismology when combined with accurate estimates of global parameters is the determination of the size of the convective cores, which plays an important role on the evolution of intermediate mass stars and defines the amount of nuclear material available to sustain the luminosity. A combination of seismology measurements and of accurate determinations of global parameters can probe the convective core, as in intermediate mass stars low-order modes at quite low frequencies are very sensitive to the structure of the inner regions (e.g., Michel et al. 1992; Dziembowski & Pamyatnykh 1991). Lebreton et al. (1995) proposed to use this property to determine the amount of overshooting. Due to the extreme accuracy on frequencies, if the luminosity is known to 2 per cent, then the overshooting parameter can be determined with an accuracy of the order of 0.03 H_p, where H_p is the pressure scale height. It then reduces the contribution of this process to the error on age to a few per cent.

It is likely that these desired eigenmodes will not be excited and measurable in all objects for which one would like to have a primary determination of age. However, one can reasonably expect that in the region of the H-R diagram which includes both the instability strip and the domain of maximum predicted amplitudes for pressure modes stochastically excited by the convective zone, enough objects will oscillate in these modes. These stars will then serve to calibrate the value of the overshooting parameter, and will indicate whether it can be used as a 'universal' value, applicable also to non-oscillating stars in a wider mass range. This project implies a large seismology programme, which has to be carried out from space. COROT, MOST and MONS will be its first steps.

Internal Diffusion of Chemical Elements Diffusion of chemical elements in stellar radiative zones may have important consequences for stellar evolution, in particular for stellar ages when fresh helium is brought to the stellar cores. Diffusion may also modify the composition at the surface of stars during their life implying difficulties to link the abundances of the elements presently observed to the initial abundances of the protostellar cloud. It may occur through various very different processes: microscopic diffusion due to gravitational settling and radiative forces or turbulent diffusion resulting for example from hydrodynamical instabilities in rotating stars. The

high-precision positions in the H-R diagram of stars of known surface abundances, provided by Hipparcos and by high-resolution spectroscopy, has already shown that there exists a discrepancy between the observed positions and the positions predicted by the 'standard' stellar models (Lebreton et al. 1999). This discrepancy remains even after a careful discussion of all sources of error, in particular the atmospheric treatment of abundance determinations. It seems to be the signature of microscopic diffusion as it shares its major characteristics: exponential increase on a time scale of a few Gyr, and dependence on metallicity (Morel & Baglin 1999). This result is based on a few dozen stars for which measurements of sufficient accuracy are available. GAIA will provide a much larger homogeneous sample of stars with at least the same accuracy. The knowledge of the masses of at least a fraction of the stars will bring additional information, in particular on the helium abundance, and therefore stronger constraints on diffusion.

Outer Convective Zones Most stellar models are still built by treating convection according to the classical parametric mixing-length theory. This procedure is useful for stars on the lower main sequence, as the radius of these objects is not very sensitive to the details of the convective model, and a Solar-like tuning of the mixing-length parameter gives reasonable agreement with observations (Fernandes et al. 1998; Perryman et al. 1998; Lebreton et al. 1999). By contrast, the mixing-length theory has little predictive power for stars on the upper main sequence, and in other evolutionary phases: observational comparisons show that the tuning of the mixing-length parameter depends both on mass (Stothers & Chin 1995) and metallicity (Chieffi et al. 1995).

The convective model proposed by Canuto & Mazzitelli (1991) has more predictive power than the mixing-length theory, from pre-main sequence (Luhman & Rieke 1998) to white dwarfs (Althaus & Benvenuto 1998), but is still a local model. A more realistic treatment of convection will have to predict both stellar radii and the amount of overshooting, and as such, will have to be non-local. At present, both large-eddy numerical simulations (Nordlund & Stein 1999) and analytic models (Canuto 1992; Kupka 1999) are providing first insights on the problem, even if the former are still confined to thin and relatively viscid layers, and the latter to polytropic stellar models.

The careful calibration of the H-R diagram allowed by GAIA as a function of metallicity and age, combined with the astroseismological results for overshooting, will greatly enhance our capabilities of dealing with non-local convective models for stellar interiors.

1.4.5 Stellar Ages, Galactic Evolution, and Age of the Universe

Precise stellar age determinations are required for a variety of topics related to dynamical studies, Galactic structure and evolution, and cosmological time scales. The 'primary' determination of ages relies on comparisons of stellar models or isochrones with the best-available data, in particular luminosity, effective temperature and abundances, on individual stars or stellar groups. Although the principle of the method is general, its application to different types of stars requires specific considerations. The uncertainty on age depends on the status of the object, individual or member of a group, and is strongly related to its intrinsic properties, mass and age.

A/F Stars and Galactic Evolution The determination of the age of relatively young objects, with ages ranging from several 10^6 to a few times 10^9 yr, are relevant for Galactic evolution studies (see Section 1.3). The determination of the age of the oldest objects in our Galaxy provides a lower limit to the age of the Universe. This estimate can be used to constrain cosmological models, as the expansion age of the Universe is a simple function of the Hubble constant, the average density of the Universe, and the cosmological constant. The oldest objects in the Milky Way are the metal-poor stars located in the halo, field Population II stars and globular cluster stars.

In a detailed analysis of the uncertainties involved in matching observations of main sequence A-F type stars to theoretical stellar models, Lebreton et al. (1995) considered the uncertainties in the physical prescriptions of the models (reaction rates, opacities, mixing-length parameter for envelope convection and overshooting of the convective core), chemical composition, effective temperatures and bolometric corrections. They showed that with combined efforts on atmospheric modeling (to improve the global parameters), on asteroseismology (to improve the physical description of the stellar interior) and on distance determinations (to improve luminosities) an accuracy on age determination of about 10 per cent is foreseeable. The luminosity uncertainty remains one of the major sources of error on age determinations, and improvements of a factor 20 compared with Hipparcos are required to provide distance calibrations below the projected contribution of atmospheric modeling.

Star Clusters H-R diagrams of open and globular clusters have been exploited for decades to constrain internal structure models and stellar evolution theories. Although millions of field stars will be placed very accurately on the H-R diagram, open and globular clusters will remain unique tools because cluster members share the same age and chemical composition with masses spread all along the mass spectrum. Most clusters will contain hundreds to thousands of members observed by GAIA, and very clean H-R diagrams will be obtained by using the GAIA astrometry, photometry and radial velocities to discard non-members. The mean value of cluster metallicity will be determined with a much higher accuracy than those of individual members.

H–R diagrams of the few hundred clusters closer than 3 kpc, with typical precision of the mean cluster distance modulus between 0.003 and 0.03 mag, will allow unprecedented high-precision comparisons with theoretical tracks of several ages and metallicity. These comparisons will be strongly reinforced by the direct determination of individual masses for about 10 SB2 or visual binaries per nearby open cluster, from GAIA epoch astrometry combined with radial velocity data. Conversely, clusters will also give strong constraints on the zero points of [Fe/H] calibrations of the wide-band and medium-band photometric systems of GAIA for stars of different spectral types.

The determination of ages of young open clusters by means of theoretical isochrones from the main sequence turnoff suffers from many uncertainties, first of which is the lack of a sufficient number of stars to define the turn-off accurately. This is due to the limited number of massive stars in clusters, and cannot be helped by GAIA. Further uncertainty in the position of the turn-off may result from inaccurate global parameters but also from stellar rotation and binarity. The age obtained through the comparison with theoretical isochrones suffers from uncertainties in the chemical composition of the stars, in the amount of overshooting of the convective cores, and in the difficulties to transform theoretical isochrones to the observed H-R diagram. For clusters up to the age of the Pleiades, one can improve the dating by using the pre-main sequence turn-on, which GAIA will help to define accurately for a number of open clusters with various chemical compositions, assessing the membership of enlarged numbers of low luminosity stars. This will certainly provide stronger constraints for stellar models as well as new insight in the understanding of the various components of the galactic disk.

Hipparcos provided accurate individual distances to about 100 members of the Hyades. These data complemented with ground-based high-resolution spectroscopic observations were used by Lebreton et al. (1997), Perryman et al. (1998), and de Bruijne & Hoogerwerf (2000) to estimate the helium content and improve the cluster age determination. GAIA will allow similar studies for a large number of open clusters with various chemical compositions.

Helium Abundance and Chemical Evolution of the Galaxy The position of the zero age main sequence in the H-R diagram depends critically on the chemical composition of stars. A large sample of non-evolved low-mass stars with determined metallicities and accurate positions in the H-R diagram (including many binaries with determined masses) will be a key tool to discuss the helium abundance of these stars, not accessible from the spectrum, and of the possible relation between the helium abundance and metallicity. This requires surveying down to masses below 0.7–0.8 $M_{\odot}$ in order to eliminate evolutionary effects. Hipparcos measured some suitable low-mass stars, although few yet have sufficiently accurate metallicity determinations from ground-based high-resolution spectroscopy. GAIA will allow progress because it will give access to a large number of stars with spectral types K–M.

Table 1.7: Age estimates of the oldest globular clusters, from Chaboyer (1999). According to this reference, the higher age estimate from Pont et al. (1998) arises from the effect of binary systems.

Age (Gyr)	Distance determination	Reference
11.5 ± 1.3	five independent techniques	Chaboyer et al. (1998)
12 ± 1	main sequence fitting (Hipparcos)	Reid (1997)
11.8 ± 1.2	main sequence fitting (Hipparcos)	Gratton et al. (1997)
14.0 ± 1.2	main sequence fitting (Hipparcos) including binaries	Pont et al. (1998)
12 ± 1	theoretical HB and main sequence fitting	D'Antona et al. (1997)
12.2 ± 1.8	theoretical HB	Salaris et al. (1997)

The Oldest Stars and the Age of the Universe A minimum age of the Universe can be estimated directly by determining the age of the oldest objects in our Galaxy. This estimate can be used to constrain cosmological models, as the expansion age of the Universe is a simple function of the Hubble constant, the average density of the Universe, and the cosmological constant. Currently the best estimate for the age of the oldest stars is based on the absolute magnitude of the main-sequence turn-off in globular clusters, requiring that the distance to the globular cluster be known (Chaboyer 1999). Field subdwarfs observed by Hipparcos can be used to estimate the distance of globular clusters through main-sequence fitting. This yields ages of the oldest clusters of around 11.5 ± 1.3 Gyr (Table 1.7), implying a minimum age of the Universe of ≥ 9.5 Gyr (95 per cent confidence level). This is about 3 Gyr younger than previous determinations, partly due to improved input physics used in the models, and partly due to the longer distance scale used for the globular clusters.

Hipparcos provided an opportunity to estimate the age of the local halo, yielding a lower limit of the age of the Universe independent of globular clusters. Cayrel et al. (1997) selected a few Population II field stars (subdwarfs and subgiants), with the most precise parallaxes determined by Hipparcos, and metallicities in a narrow range near the most frequent metallicity in the halo. These stars delineate an isochrone-like curve in the H-R diagram with a turn-off corresponding to an age of 14 ± 2 Gyr. Subgiants are about one hundred times rarer than subdwarfs in the halo and only two of them have parallaxes determined by Hipparcos with an accuracy better than 12.5 per cent (Reid 1997). A few others have good Hipparcos parallaxes, and metallicity from Coravel, but the sample is very small, and this limits the precision on the location of the subgiant branch and implies a poor accuracy on age.

GAIA will improve the age estimate of the oldest stars for at least two reasons. The number of subdwarfs with accurate distances will considerably increase in each metallicity interval allowing to apply the main-sequence fitting technique to derive the distance of an increased number of globular clusters of various chemical compositions. Furthermore, distances of a substantial number of field subgiants will be measured, improving the age determination of the field halo stars.

Independent ages of old objects can be obtained from nucleochronology (which GAIA does not provide) and from comparison of white dwarf cooling sequences with lower luminosity limits of white dwarf samples. The age of the old disk and of the halo of the Galaxy have been estimated by considering white dwarfs observed in the Solar vicinity, in old open clusters, and in globular clusters (Winget et al. 1987; Richer et al. 1997; Richer et al. 1998). This method needs the distance of the objects to be known and will benefit from any improvement on distances.

It has long been realized that diffusion (the settling of helium relative to hydrogen) could shorten the predicted main sequence lifetimes of stars. However, it was not clear if diffusion actually occurred in stars, so this process had been ignored in most calculations. Recent helioseismic studies of the Sun have shown that diffusion occurs in the Sun (Dale et al. 1993; Guenther et al. 1996). The Sun is a typical main sequence star, whose structure (convective envelope, radiative interior) is quite similar to main sequence globular cluster stars. Thus, as diffusion occurs in the Sun, it appears likely that diffusion also occurs in main sequence globular cluster stars. Modern calculations find that the inclusion of diffusion lowers the age of globular clusters by 7 per cent (Chaboyer et al. 1996). The recent use of an improved equation of state has led to a further 7 per cent reduction in the derived globular cluster ages (Chaboyer & Kim 1995). The equation of state now includes the effect of Coulomb interactions (Rogers 1994).

1.4.6 Isolated Brown Dwarfs

GAIA will detect a large number of isolated brown dwarfs (brown dwarfs in binary systems are considered in Section 1.6), including most of the brown dwarfs liable to be found by currently ongoing near infrared surveys (DeNIS, 2MASS). Figure 1.18a shows the bottom of the H-R diagram with evolutionary tracks of brown dwarfs of different masses, and with 0.1 and 1 Gyr isochrones (thick lines). The isochrone demonstrates that the luminosities of brown dwarfs fade rapidly to very faint absolute magnitudes, the lighter brown dwarfs being more sensitive to this effect. Brown dwarfs detected at $V < 20$ mag will be strongly biased towards very young objects and those in the upper mass interval. Young brown dwarfs are however visible at relatively large distances. Figure 1.18b shows that, for objects at the lithium burning boundary as reference, GAIA will see Pleiades-age brown dwarfs out to 400 pc, and younger 'Gould Belt' brown dwarfs (Section 1.2.4) could be visible out to about 1 kpc. Old brown dwarfs will only be visible if they are very nearby. Figure 1.18a also implies that the youngest brown dwarfs (up to a few 0.1 Gyr) have absolute magnitudes comparable to M dwarfs, with $M_V \sim 10 - 16$. It is reasonable to expect that brown

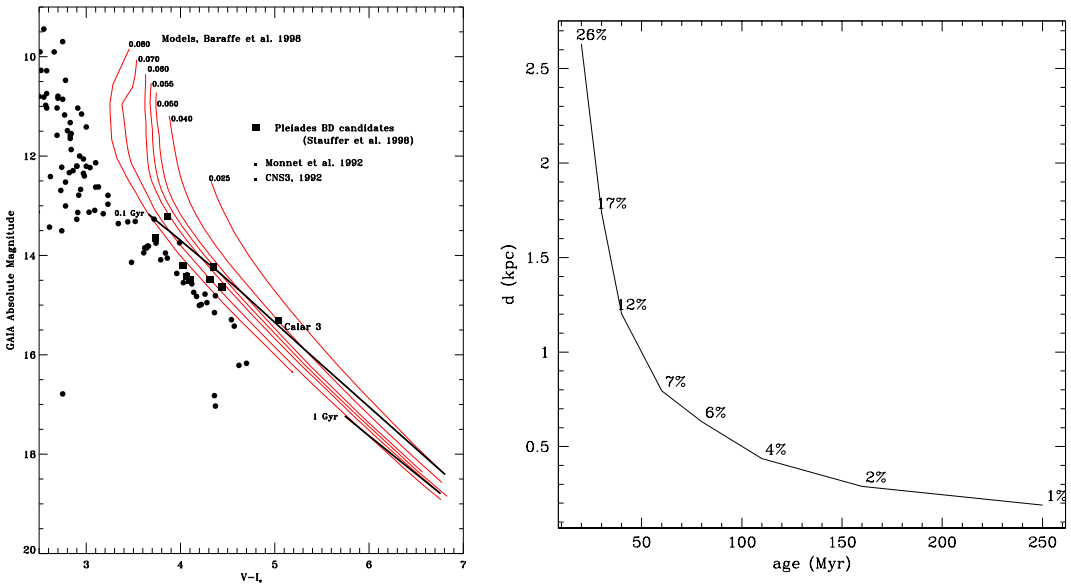


Figure 1.18: (a, left) The bottom of the H-R diagram. Evolutionary sequences of brown dwarfs from Baraffe et al. (1998), labeled with their masses ($M_\odot$), are plotted together with data for M dwarfs (dots), brown dwarf candidates in the Pleiades (squares) from Stauffer et al. (1998a) and Calar 3, a confirmed brown dwarf in the same cluster (Rebolo et al. 1996). Thick lines are brown dwarf isochrones of 0.1 and 1 Gyr. (b, right) The distance at which brown dwarfs at the lithium burning limit can be seen by GAIA, as a function of age; even at 20 mag, GAIA will measure proper motions accurate to ∼0.1 mas yr^{-1} and parallaxes accurate to ∼0.1 mas. Labels give the percentage accuracy $\Delta\pi/\pi$ for I∼20 mag stars at the given distance.

dwarfs moderately affected by interstellar extinction could be seen by GAIA to distances of a few hundred parsec. In particular, all Pleiades brown dwarfs down to approximately 0.01 $M_\odot$ should be sampled.

The number of expected brown dwarfs is difficult to quantify, because of the poor statistics of known objects, and our ignorance of their formation processes. It is possible to make a crude projection, by extrapolating current estimates of the population of young M dwarfs. Methods to select young M dwarfs are based on their chromospheric and/or coronal activities. Chromospheric activity seems to be an ambiguous diagnostic of youth (Reid et al. 1999), but coronal X-ray emission clearly traces young populations (Section 1.2.4). In the immediate Solar neighbourhood (within 25 pc), X-ray active stars represent approximately 18 per cent (or 0.003 $M_\odot$ pc^{-3}, see Fleming 1998) of the M dwarf star population. Although the level of activity of these stars indicates they are young, these objects are not associated with star-forming regions (the local bubble, < 80 pc, is notably devoid of star formation activity), although it is clear that a significant population of young main sequence field M dwarfs exists. For an initial mass function continuous over the stellar/sub-stellar limit — consistent with results for the Pleiades (Bouvier et al. 1998) and preliminary analysis of near infrared surveys (Reid et al. 1999) — and a constant formation rate for brown dwarfs, one expects $> 20\,000$ brown dwarfs older than 0.5 Gyr in a standard Galaxy model.

Although most brown dwarfs will be at the faint end of the GAIA survey, their proximity ensures that the relative precision on parallaxes will typically be better than a few per cent. Therefore, it is expected that the positioning of these objects in the H-R diagram will be excellent, allowing the determination of ages and masses by sequence fitting. This will give an accurate picture of the recent brown dwarf formation history, including their formation rate and mass function.

1.4.7 White Dwarfs

White dwarfs are well-studied objects, and the physical processes that control their evolution are relatively well understood. Most phases of white dwarf evolution can be succesfully characterized as a cooling process. That is, white dwarfs slowly radiate at the expense of the residual thermal

energy of their ions. The release of thermal energy extends over very long time scales (of the order of the age of the Galactic disk, $\sim 10^{10}$ yr). While their detailed energy budget is still the subject of some debate, their mechanical structures, which are largely supported by the pressure of the gas of degenerate electrons, are very well modelled, except for the outer layers. These layers control the output of energy and a correct modelling is necessary to understand the evolution of white dwarfs. This, in turn, demands precise spectrophotometric data. Accurate parallaxes — like those that GAIA can provide — will provide very tight new constraints on the models.

White Dwarf Luminosity Function of the Disk The luminosity function of white dwarfs at low luminosities, and especially the position of its cut-off, provides important information about the age of the Galactic disk. The main sources of observational uncertainty are (Knox et al. 1999): the distance to the lowest luminosity white dwarfs, the bolometric corrections, and the chemical composition of the outer layers (i.e., DA if hydrogen is present, or non-DA if hydrogen is absent). The contribution of the current observational uncertainties to the total error budget of the galactic age is as large as 2 Gyr. Of this, 1 Gyr comes directly from the binning and sampling procedure and the statistical noise of the low-luminosity bins. This uncertainty will be reduced by a factor of roughly 4–5 (to ~ 0.2 Gyr) because of the large number of white dwarfs (of the order of 200 000) that GAIA will observe (Figueras et al. 1999). Follow-up spectroscopic observations with ground-based telescopes of the white dwarfs in the lowest luminosity bin ($\sim 50 - 60$ objects at $M_v \sim 16$ mag) with 1 per cent accuracy in distance provided by GAIA, together with improved theoretical models, will result in a factor of five improvement in the present 1 Gyr atmospheric uncertainty. Therefore, after GAIA, the age of the disk will be known to ~ 0.5 Gyr.

An accurate luminosity function not only provides a tight constraint on the galactic age but also has the bonus of providing important information about the temporal variation of the star formation rate (Isern et al. 1999). GAIA will improve the present status and will also be able to distinguish among the white dwarf luminosity function of the thin and the thick disk from the kinematic properties (García-Berro et al. 1999), and provide in this way an unprecedentedly deep insight into Galactic history.

White Dwarf Luminosity Function of the Halo The scarcity of known bright halo white dwarfs, and the lack of good kinematical data necessary to distinguish halo white dwarfs from those in the disk, so far prevent the construction of a good luminosity function for the halo (Torres et al. 1999). GAIA's high-quality parallaxes and proper motions will result in accurate tangential velocities, allowing discrimination of these two populations. GAIA will observe hundreds of halo white dwarfs with errors in tangential velocities smaller than 5 km s^{-1} (Figueras et al. 1999). A robust determination of the bright part of the halo luminosity function will narrow the range of allowed IMFs for the halo (Isern et al. 1998), and will constrain the nature of the microlensing events observed in the direction of the Magellanic Clouds (e.g., Alcock et al. 1997; de Zeeuw 1999).

Theoretical models indicate that if the halo is not too old, about 12 Gyr, there is a reasonable chance to detect the corresponding cut-off ($M_V \sim 16$) with surveys as deep as $V \sim 20$ mag, provided that the DA population were dominant. By contrast, if non-DA white dwarfs turn out to be dominant, the cut-off would be placed at absolute magnitudes as large as 20, and there would be no chance to detect it and, thus to constrain in this way the age of the halo (Isern et al. 2000). Indirect information about the halo (such as its age and duration of the main star formation burst) will come from the comparison of the absolute numbers of red halo dwarfs and white dwarfs in a complete volume-limited sample.

White Dwarfs as Laboratories for Fundamental Physics White dwarfs are well suited to test any departure from standard physics, since even small changes in physical constants can result in prominent effects when the relevant time scales of white dwarf cooling are taken into account. Such is the case, for example, of a hypothetical change in the gravitational constant, G. Using the white dwarf luminosity function García-Berro et al. (1995) derived an upper bound of $\dot{G}/G \leq -(1 \pm 1) \times 10^{-11}$ yr^{-1}, which is comparable to bounds derived from the binary pulsar PSR 1913+16 (Damour et al. 1988). Since this is a statistical upper limit, any improvement in our knowledge of the white dwarf luminosity function of the Galactic disk will translate into a more stringent upper bound for $\dot{G}/G$ (see Section 1.9.3 for further details). This method is very

powerful but demands error bars as small as possible and this can only be achieved through a deep ($V \sim 20$ mag) and all sky survey.

Pulsating White Dwarfs The rate of change of the period of pulsation of several white dwarf variables (DAVs, DBVs and DOVs) can also be used to place stringent upper limits on the properties of various weak interacting particles such as axions (Isern et al. 1992) and neutrinos (Blinnikov & Dunina-Barkovskaya 1994). Theoretical models predict that the rate of change of period is directly related to the rate of cooling. Therefore, any additional source or sink of energy directly modifies the rate of cooling and, hence, the rate of change of the period. Since this can now be measured with high precision ($\simeq 10^{-15}$ s s^{-1}) variable white dwarfs provide us with a unique tool to measure or at least place constraints on the properties of these exotic particles. This technique can be extended to other particles as well. However, the major drawback of this technique is the poor knowledge of the parallax of these pulsators. The many white dwarfs that GAIA will observe will include a handful of these stars and will thus provide an important tool to test the physics of these particles.

Physics of Dense Matter Another field in which these non-radial pulsators have been used and in which GAIA could shed some light is the direct determination of the crystallized mass of white dwarfs. More precisely, the size of the crystallized core strongly affects the power spectrum of the non-radial pulsations by introducing a boundary condition: the position of the crystallization front. This front moves with velocities of a few cm/yr. The accurate determination of the position of the crystallization front can be only done today for BPM 37093, a very massive white dwarf (Winget et al. 1997), but it places very precise requirements to be met by the crystallization theory. The study of a large enough sample of very massive non-radial pulsators with very well-known fundamental properties accessible to GAIA will significantly improve our knowledge of the behaviour of matter at very high densities.

1.4.8 Other Specific and Rare Stellar Types

Perhaps the most dramatic effect of GAIA's contribution to stellar astrophysics will be visible for the rarer stages in stellar evolution, for which Hipparcos has not been able to supply strong constraints on the luminosity, given their large average distances. Examples include Tc-rich S stars, the central stars of planetary nebulae, and Population II stars, as well as many others. For all of these stellar types the few available Hipparcos parallaxes show important discrepancies with existing theoretical models or earlier calibrations. The classes of objects in use as extragalactic distance indicators are discussed in Section 1.4.9.

Massive Stars Although only a small fraction of stars in the Galaxy are more massive than $20M_{\odot}$, such stars, which spend most of their short lives as H-burning O-type stars, play an important role in Galactic structure and evolution. Thus, accurate knowledge of the luminosity of these stars is important for comparing masses derived from stellar evolutionary models with those derived from stellar atmosphere models, for determining initial mass functions, and for studying stellar evolution in the high luminosity/high mass region of the Hertzsprung–Russell diagram. The absolute magnitudes of O stars are presently poorly determined (no O star is sufficiently close to the Sun to have a trigonometric parallax accurately measured), the absolute visual magnitudes coming primarily from O stars in clusters and OB associations whose distances are themselves uncertain, but are typically around 1–2 kpc. Typical apparent magnitudes are $V = 4 - 6$ mag.

Wolf-Rayet stars These are a stage in the evolution of stars more massive than about $30M_{\odot}$, which have been stripped of their H-rich envelopes by winds in previous evolutionary phases. There are two types of WR stars: the C-rich WC stars and the N-rich WN stars. Because of their short total lifetime ($\lesssim 4\ 10^6$ yr), they are indicators of recent star formation, or of young clusters. The stars are very luminous, $M_V < 3$, with the late WN stars as bright as $M_V \simeq -6$, and can be observed up to very large distances with GAIA. Because of their remarkable optical spectra, dominated by strong emission lines, they can be distinguished easily with low-resolution spectroscopy or with narrow-band filter photometry. Accurate distances measured by GAIA are important for several reasons: (i) the distance determinations of very young clusters by means of WR stars makes it possible to trace the spiral arms across the Galaxy; (ii) evolutionary calculations show that the ratio of WC to WN stars increases with increasing metallicity. This is because the mass-loss rates in the previous phases increase with increasing metallicity, and the WC stars have lost more mass and are stripped further than the WN stars. So by measuring the distances of the WR stars and their WC/WN ratio as a function of location in the Galaxy, the metallicity can

be traced as a function of Galactic location; (iii) because of their intrinsic brightness and their remarkable spectra, late-WN stars can be observed in distant Galaxies up to about 60 Mpc (for a visual magnitude limit of $V < 28$ mag and low extinction). This makes these stars excellent candidates for measuring distances significantly beyond the Cepheid limit. This requires a careful calibration of the M_V–subtype relation for WR stars. This is useful also for the interpretation of spectra of starburst galaxies. GAIA's parallaxes are crucial here.

Hipparcos observed a few WR stars, but the parallax accuracy was insufficient for useful luminosity calibration except for γ Vel and EZ CMa. These stars were found to be members of nearby young stellar groups, allowing an accurate distance measurement (de Zeeuw et al. 1999). The inferred absolute luminosities are an order of magnitude smaller than assumed previously, demonstrating that a re-calibration of WR distances is urgently needed.

Tc-Rich S Stars These are interesting for understanding stellar structure because they accurately time the occurrence of a dredge-up of core material to the surface (van Eck et al. 1998). The Hipparcos data show that their status is ill-understood: the derived core masses (0.52–0.56 $M_\odot$) are smaller than the theoretical lower limit (0.58 $M_\odot$) for third dredge-up (Groenewegen et al. 1997). However, the errors on the Hipparcos parallax (given the large distance to these systems) cause a large uncertainty in the derived masses. GAIA parallaxes will resolve this issue.

Long-Period Variables These red giants (covering Miras, SRa/b, and OH/IR stars) are on a critical, short-lived stage of the evolution of intermediate mass stars (the Asymptotic Giant Branch), leading to planetary nebulae and white dwarfs. They also strongly contribute to the chemical evolution of the Galaxy (mass-loss) and, due to their strong infrared flux, are a useful probe of galactic structure (e.g., Habing 1996). They are promising distance indicators, as a complement to the Cepheids, which are some five times less numerous at the same luminosity. Hipparcos has provided only a few parallaxes, usually with poor precision. Moreover, luminosity calibrations have shown that the period-luminosity relation observed in the LMC cannot be trivially transposed to other galaxies, because of its dependence on metallicity, the respective proportions of Miras and SRb stars, and the star formation history (Barthès & Luri 1999; Barthès et al. 1999). The pulsation models (a tool for determining present masses and mass-loss as a function of period, luminosity and metallicity) still suffer from significant uncertainties, such as non-linear effects, time-dependent convection coupled to pulsation, and coupling of the stellar and circumstellar envelopes (Ya'ari & Tuchman 1996; Ya'ari & Tuchman 1999).

For a large sample of long-period variables in our Galaxy (probably in excess of 10 000 objects, with nearly no bias except that due to interstellar extinction), GAIA will provide parallaxes and proper motions, along with magnitudes in bands where a period-luminosity relation exists both at mean and maximum brightness. It will also provide useful colour indices (such as $V - I$), radial velocities and rough metallicity estimates. In the Magellanic Clouds most individual parallaxes will be insignificant, but the other information will be exploitable. All these data will make possible a precise luminosity calibration, together with a reliable separation of the different populations and variability types, both in our Galaxy and in the Magellanic Clouds. This will provide us with the 'period-luminosity-other' relations necessary to estimate extragalactic distances. These abundant luminosity and metallicity estimates, together with the masses of a few binaries (none were obtained from the Hipparcos mission) and the (imprecise) photospheric velocity amplitude of some stars, will provide strong constraints allowing the test and calibration of pulsation models and their underlying physics, as well as the theoretical and empirical mass-loss models. Important consequences are to be expected concerning the initial and final mass relation, and the ages and the history of star formation, both in our Galaxy and in the Magellanic Clouds.

Physics of Cepheids Accurate distances of Cepheids throughout the Galaxy will allow derivation of period-luminosity relations at different metallicities. In the H-R diagram, the instability strip will be delineated precisely, which will constrain the structure of the outer convective zones, and the properties of turbulent convection. The presently available data is not large enough to allow statistical studies relevant for pulsation-convection coupling and time-dependent convection.

Novae and Nova-Like Variables Distance determinations to novae are required to interpret the energetics of the outburst, and to place these objects more securely within the context of evolutionary models. Distance estimates can be made through modeling of the shell expansion velocity, but such applications are restricted to particular periods after outburst, and also suffer from modeling uncertainties. Most Galactic novae are brighter than $V = 12$ mag at maximum, although measurements to $V = 16$ mag or fainter will also allow the determination of distances to Galactic novae observed over the last few decades. Related objects, such as dwarf novae, AM Her stars, symbiotic stars, and cataclysmic binaries could be studied, providing accurate luminosities needed to distinguish among alternate possible energy generation mechanisms. Many such nova-like variables would lie within the distance horizon and the magnitude limit (say, brighter than $V = 16$ mag) necessary to provide distances to better than 5 per cent.

T Tauri Stars Low-mass pre-main sequence stars (i.e., T Tauri stars, or TTS) are of particular interest to stellar evolution, among other things because they are, if sufficiently young, in a fully convective stage of evolution. The GAIA parallaxes will completely remove the distance uncertainty in the determination of the fundamental stellar parameters for pre-main sequence stars. GAIA is the only tool to help observationally in exploring these fundamental problems of stellar evolution. Accurate distances will be the key in pinning down the ages of the rich pre-main sequence population available in the nearest kpc (cf. Section 1.3.3).

One of the brightest and best studied weak-lined T Tauri stars is HD 283572 in the Taurus cloud, one of the nearest active star-forming region. Its evolutionary status on the basis of the Hipparcos parallax has been studied by Favata et al. (1998). The parallax for HD 283572 is $\pi = 7.81 \pm 1.30$ mas, corresponding to a distance of 128 pc (implying $M_V = 3.48$), with a 1σ range of 110–154 pc, and a range in absolute magnitude of 3.81–3.08, or a luminosity range of a factor of 2, and correspondingly large uncertainties in the model age and mass of the system. For other T Tauri stars in Taurus the situation is even less favorable, with most of them having, in the Hipparcos data, $\sigma_\pi \sim 2$ mas. Indeed, an accurate determination of their evolutionary status is not possible, as the parallax error still dominates the other sources of uncertainty.

X-ray selection of active stars has shown that Gould Belt stars outside the star formation regions are older than those in star formation regions, and are likely to populate the whole range of ages between the T Tauri phase (up to a few million years old) to hydrogen ignition (several tens of million years, corresponding to the age of the youngest studied open clusters). There is little direct information about this evolutionary phase, although it is the fundamental stage in the evolution of low mass stars, during which some 'hidden' parameters as rotation, angular momentum evolution, and magnetic fields play their most important role. During this phase the following events take place: (i) the main rotational evolution of the star: the objects accelerate (or not) from the T Tauri relatively slow rotation rate to the possibly high rotation rates of the stars in young clusters like α Per. The theory of angular momentum transfer through the star and the role of the protostellar disk (e.g., Bouvier 1994) needs to be constrained observationally; (ii) planet formation and disk evolution: information is needed on the lifetime of disks and on the possible role of planets in the rotational history; (iii) pre-main sequence lithium depletion: 35 years after the recognition that pre-main sequence nuclear burning of lithium occurs in the low-mass stars, we still do not know how much lithium is depleted during the early evolution of the Sun (D'Antona & Mazzitelli 1997; Martín 1997). This problem reflects on many others, from the understanding of stellar structure to that of constraining the Big Bang model by observing the primordial abundances of light elements. Monitoring directly the phases of lithium depletion in a large number of stars, with different rotation periods, masses and disks, will finally shed light on this issue; (iv) the transition from the Hayashi convective track to the 'radiative' track which approaches the main sequence: how does the magnetic structure (a dynamo associated with the deep convective zones is likely to be responsible for the observed X-ray emission) change, and how is this reflected in the angular momentum evolution?

Pulsars GAIA, like SIM, will be able to measure the parallax and the proper motion for the Crab pulsar. No other known optical counterparts to radio pulsars are bright enough for such a study. But GAIA offers the unique opportunity to discover fast-moving objects that may be radio-quiet pulsars (such as Geminga) should such objects exist and have escaped detection so far at other wavelengths. The determination of the properties of the interstellar medium by establishing distance indicators (see Section 1.2.9) will indirectly lead to significant improvement in the estimates of pulsar distances.

Table 1.8: GAIA observations of stellar distance indicators.

Distance indicator	GAIA observations
Open clusters	complete membership census 3-D observations up to $\sim$ 1000 pc all mean distances to better than 1 % many new clusters to be discovered
Globular clusters 140 known clusters	complete membership census except on some central zones $\sim$ 20 with $\sigma_\pi/\pi < 10$ % per star $\sim$ 40 with $\sigma_\pi/\pi < 20$ % per star mean distance to < 1 % for 110 clusters mean distance to < 5 % for all clusters
Cepheids	$\sigma_\pi/\pi < 1$ % up to 3000 pc $\sigma_\pi/\pi < 4$ % for all galactic Cepheids membership to clusters all over the Galaxy
RR Lyrae	$\sigma_\pi/\pi < 1$ % up to 3000 pc $\sigma_\pi/\pi < 10$ % for most galactic RR Lyrae
Mirae	$\sigma_\pi/\pi < 1$ % up to 3000 pc $\sigma_\pi/\pi < 6$ % for all galactic Mirae

1.4.9 Cosmic Distance Scale

Numerous questions related to the cosmic distance scale are dealt with comprehensively in a number of recent reviews (see, for example, Egret & Heck 1999; Reid 1999). The present debate over the conflicting information provided by various distance indicators to, for example, the Large Magellanic Cloud (Walker 1998), are presumably saying as much about the physics of these objects as they are about the issue of the distance scale itself. For the foreseeable future, attempts to establish unambiguous distance estimates to the Local Group galaxies will rest on the critical discussion of data from many different classes of objects, including those derived from orbital gas motions, as reported for NGC 4258 (Hernstein et al. 1999). While elements of the distance scale determination will be excellent targets for SIM (which will measure pre-selected samples of each distance scale calibrators), GAIA will measure them all, and in an unbiased way.

The impact of Hipparcos on the cosmic distance scale was reviewed by Turon & Perryman (1999). The only very firm result is the complete three-dimensional study of the Hyades open cluster for which more than 200 members were individually observed with good accuracy (Perryman et al. 1998). The number of other stellar distance indicators observed with sufficient accuracy was not large enough to reliably explore, for example, the effect of metallicity on the position of open cluster main sequences (van Leeuwen 1999; Robichon et al. 1999; Pinsonneault et al. 1998) or on period-luminosity relations for pulsating stars (Feast & Catchpole 1997; Oudmaijer et al. 1998; Luri et al. 1998; Fernley et al. 1998; Tsujimoto et al. 1998; van Leeuwen et al. 1997; Whitelock et al. 1997; Bergeat et al. 1998). Similarly, Reid (1999) has carried out a detailed review of the luminosity calibration of primary distance indicators and of the Galactic distance scale, and concluded that *'Many of these issues remain to be resolved'*.

GAIA will provide accurate distances (and proper motions) for such huge numbers of each category of stellar distance indicators that, again in this domain, the analysis methods can be drastically changed. Some illustrative numbers are given in Table 1.8. The sampling of open and globular clusters in age, metal, oxygen or helium content will be complete all over the Galaxy. Parallel improvement in the transformation between the observational and the theoretical H-R diagram will be required to take full benefit of these accuracies in terms of stellar evolution and age determination: photometric and/or spectroscopic data should allow the determination of the bolometric magnitude and of the effective temperature from the observed magnitudes and colours.

For pulsating variables, the sampling versus period, populations, colours, and metal content will be as good as possible as excellent distance determinations will be obtained for all observable galactic stars, and a first reliable estimation of the intrinsic dispersion of the period-luminosity relations will be possible. Moreover, a first check of the universality of these relations (not only the slopes, but also the zero-points) will be possible, directly for LMC Cepheids, or using GAIA mean distances for the closest galaxies of the Local Group, at least LMC, SMC and Sagittarius (see Section 1.4.8).

Table 1.9: Comparison between different methods to determine RR Lyrae absolute magnitudes (from Popowski & Gould 1999).

Method	Absolute magnitude at [Fe/H] = −1.6	Grade	Main problems	Potential future usefulness
Statistical parallax	0.77 ± 0.13	A	—	A−
Trigonometric parallax	0.71 ± 0.15	A−	done for non-RR Lyrae stars small number statistics	A+
Cluster kinematics	0.67 ± 0.10	B+	modeling of rotation uncertain density profile proper motions	A−
Baade-Wesselink	0.45–0.70	B	temperature scale p factor, bolometric corrections	B
Theoretical models	0.45–0.65	B−	input physics	A
Main sequence fitting	0.45 ± 0.04	C	metallicity scale reddening uncertainties small number of calibrators	B
White dwarf fitting	0.67 ± 0.13	C	WD masses from theory reddening uncertainties	A

Cepheids and RR Lyrae Stars In addition to the importance of these stars to models of stellar structure and evolution (see Section 1.4.8), Cepheids and RR Lyrae stars form the cornerstone of the extragalactic distance scale. With the use of the Hubble Space Telescope and of large ground-based telescopes, these stars are now observed in many nearby galaxies, out to nearly 25 Mpc for Cepheids, and thus reaching the Fornax and Virgo clusters. Cepheids are the main Population I bridge between our Galaxy and the LMC to spiral and irregular galaxies, RR Lyrae to spirals and ellipticals. Some 55 Cepheids and 26 RR Lyrae stars are known to lie within about 1 kpc, and were already contained with the Hipparcos observing programme, but most of these lie beyond about 300–400 pc. GAIA parallaxes will allow a definitive resolution of the controversy about the zero points of the period-luminosity-colour relationships by providing distance estimates to better 1 per cent for most galactic Cepheids and for RR Lyrae up to about 3 kpc, better than 10 per cent for all galactic RR Lyrae and for Cepheids in the Sagittarius galaxy, and still between 10–30 per cent for Cepheids in the Magellanic Clouds. A census of Cepheids and RR Lyrae with well-determined light curves will be obtained from the GAIA multi-epoch photometry in our Galaxy and its nearest neighbours. As a result, the details of the period-luminosity-colour relationship for galaxies with different metallicities will be significantly improved.

Cepheids have now been observed by HST in 25 galaxies with the aim of providing a firm basis for the calibration of secondary distance indicators (Tanvir 1999; Mould et al. 2000). The distances of these galaxies have been obtained with the assumption of a true distance modulus and an average line-of-sight reddening to the Large Magellanic Cloud of 18.50 ± 0.13 and E(B-V) = 0.10, and neglecting the metallicity dependance of Cepheids properties. In addition to the zero-point calibration, another controversy in the Cepheid luminosity calibration is the issue of metallicity dependance (Tanvir 1999; Sasselov et al. 1997; Fry & Carney 1997). The large sample of Cepheids with accurate distances provided by GAIA will provide a strong basis for further analysis of the this effect.

The determination of absolute magnitudes and distances for RR Lyrae stars has been central in establishing the cosmological distance scale. Table 1.9, from Popowski & Gould (1999), summarizes the results reported in that review. The errors quoted are statistical only. Methods were ranked and graded (A: very good; B: good; C: acceptable) according to their judgement of their susceptibility to systematic effects. The table also lists Popowski & Gould's grade of each method's potential reliability with forseeable improvement in the data. The methods of statistical parallax, trigonometric parallax and main sequence fitting have all been affected by Hipparcos.

Planetary Nebulae The central stars of planetary nebulae have masses in a very narrow range, and thus provide the possibility of being good distance indicators. However, because of their rarity, and therefore their typical distances, and their nebulosity, no satisfactory method yet exists for their distance estimation. Parallax measurements of the central stars would lead to significant advance in the understanding of the formation and evolution of the shells, the status of the central stars, and

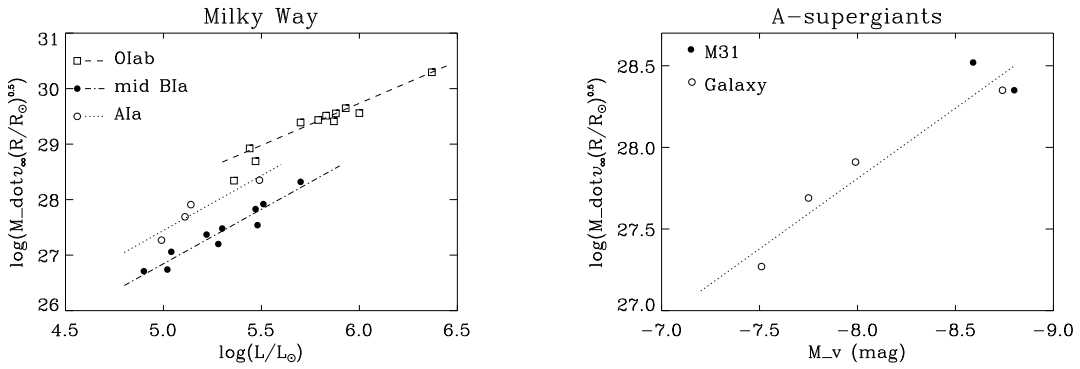


Figure 1.19: (a) The observed wind-momentum luminosity relationship of the brightest galactic supergiants of spectral type O, mid B and early A. Data from Puls et al. (1996) and Kudritzki et al. (1999). Wind momenta are given in cgs-units. (b) Wind momenta of two A-supergiants in M31 and galactic A-supergiants as function of absolute visual magnitude. Results from McCarthy et al. (1997) and Kudritzki et al. (1999).

the role of these objects as standard distance indicators. Many tens of planetary nebulae would be measurable down to $V = 16$ mag. The central stars appear to be, from their Hipparcos parallaxes (Acker et al. 1998), fainter than expected from stellar structure models. Accurate trigonometric parallaxes for many hundreds of planetary nebulae are required to calibrate the planetary nebula luminosity function, which is used extensively as an extragalactic distance indicator.

Blue Supergiants Recent advances in the theory of stellar atmospheres and winds (Puls et al. 1996; Gabler et al. 1989; Santolaya-Rey et al. 1997) have revealed a new way to determine luminosities of blue supergiants in other galaxies by means of the wind-momentum luminosity relationship with an accuracy rivalling that of the period luminosity relationship for Cepheids. The basis for the successful application of this method is its galactic calibration using blue supergiants with accurate distances in the Solar neighbourhood. At the present stage, this is the weakest point in the development of the new method. Accurate trigonometric parallaxes for blue supergiants of extreme brightness are not available so far. With distances larger than 400 pc (Ori OB1) they proved to be out of the reach of Hipparcos. GAIA, on the other hand, will allow measurement of trigonometric parallaxes for dozens of supergiants of extreme absolute magnitude. This will yield a solid foundation of the wind-momentum luminosity relationship method as a new and superior way to determine extragalactic distances.

The method relies upon the fact that luminous hot stars have winds driven by photospheric photon momentum absorption through metal lines. The theory predicts that the 'modified' wind momentum, defined as the product $\dot{M}v_\infty R^{0.5}$ of mass-loss rate, terminal velocity and square root of stellar radius, is proportional to a power of the stellar luminosity L (Castor et al. 1975; Pauldrach et al. 1986; Kudritzki et al. 1989; Kudritzki 1998), i.e., $\dot{M}v_\infty R_*^{0.5} \propto L^{1/\alpha}$, where α is a well-determined dimensionless number of the order of 2/3. It represents the power-law exponent of the distribution function of the line strengths of the thousands of spectral lines driving the wind. Both α and the proportionality constant depend on spectral type, since the dominating ions driving the wind through their line absorption change with temperature. This prediction has been confirmed by observations of winds of galactic supergiants of spectral type O, B, A, where the mass-loss rates have been determined from stellar H_α line profiles with high precision and the terminal velocities are obtained from UV P-Cygni profiles in the case of O,B-supergiants and from H_α in the case of A-supergiants (see Figure 1.19a; also Puls et al. 1996; Kudritzki et al. 1999). As is also predicted by the theory, and confirmed by observations of blue supergiants in the SMC and LMC (Kudritzki et al. 1996; Puls et al. 1996), the wind-momentum luminosity relationship also depends on metallicity. The calibration of the metallicity dependence using HST and ground-based observations of a larger sample of B- and A- supergiants in LMC and SMC is presently under way. The best spectroscopic targets for extragalactic distance determinations are A-supergiants. Since massive stars evolve at almost constant luminosity towards the red, A-supergiants are the optically brightest 'normal' stellar objects (with absolute magnitudes up to $M_V = -9$ mag) because of the effects of Wien's law on the bolometric correction. In addition, for these objects the wind momentum can be determined solely by optical spectroscopy at H_α (without need of the UV), so that 8m class ground-based telescopes can be used rather than HST (which is then needed for accurate photometry only).

Figure 1.19b gives an example from the recent work by McCarthy et al. (1997) and Kudritzki et al. (1999).

A-supergiants have V between 20 and 21 in galaxies 6 Mpc away, certainly not a problem for medium (0.22 nm) resolution spectroscopy with 8 m class telescopes. Even in a galaxy like M100 at a distance of 16 Mpc (Freedman et al. 1994) these objects will still be accessible at V around 22.5. Indeed, the HST colour-magnitude diagram published by Freedman et al. (1994) may show the presence of such objects in M100. McCarthy et al. (1997) estimate that with 10 to 20 objects per galaxy it will be possible to obtain independent distance moduli with an accuracy of 0.1 mag out to the Virgo and Fornax clusters of galaxies. The fundamental advantage of this new method relies on the fact that it is based solely on spectroscopy which — simultaneously with the stellar wind momentum — will always yield the intrinsic stellar parameters including metallicity and also the interstellar extinction.

Globular Clusters Mean distances to better than 1 per cent will be obtained for about 110 of the 140 globular clusters of our Galaxy (see Section 6.4.6 for an assessment of the observational limitations for globular clusters). Combined with their space motions, this will allow derivation of their orbits, which in turn constrain the mass of the Milky Way (Section 1.2.8). The GAIA mean distances also calibrate — as a function of metallicity — the distances of external globular clusters derived via main-sequence fitting of, e.g., HST colour-magnitude diagrams (see also Section 1.4.5). And the luminosity function of the ensemble of Galactic globular clusters calibrates the method that uses the entire globular cluster luminosity function of a galaxy as a distance indicator.

GAIA will provide a complete census of member stars in the non-central parts of all Galactic globular clusters (see Sections 1.2.7 and 1.10). This means between 100 and 100 000 stars per globular cluster. In half of the clusters, more than 5000 stars will be observable; more than 1000 stars for 100 of the 140 clusters. 83 of the 140 clusters are nearer than about 10 kpc, while 108 are nearer than 15 kpc. The blue horizontal branches range from $V = 12.5$ mag for the closest clusters to $V = 19$ mag for clusters in the LMC, with a typical magnitude between 15–17 mag. The brightest turn-off will be at $V \approx 16$ mag, most of them between 18–20 mag.

Nearby Galaxies Direct mean distances to the closest galaxies of the Local Group will be within reach and, for example, the controversy between the LMC distances determined from Cepheid (mainly located in the bar) and RR Lyrae (mainly in globular clusters) will be resolved. A sufficiently large number of Cepheids and RR Lyrae will be observed by GAIA to obtain a mean distance without the use of intermediate objects or indirect methods, and these can be compared with independent estimates, for example, from orbital motions in the nuclear gas disk in the case of NGC 4258 (Hernstein et al. 1999).

For example, by using the restricted sample of Cepheids with V photometry from Tanvir (1997), and the GAIA accuracy as a function of apparent magnitude, a mean distance with a relative accuracy of 3 per cent is obtained. Adding the nearly 1500 new Cepheids discovered by microlensing surveys (Beaulieu et al. 1995; Bersier et al. 1998) — and GAIA will certainly discover many others — a relative accuracy better than 1 per cent is obtained. The use of the nearly 8000 RR Lyrae discovered by microlensing surveys (Alcock et al. 1996), with mean V magnitudes between 18.7 and 19.7, lead to a mean distance accuracy better than 5 per cent. GAIA epoch sampling will lead to the discovery of many new RR Lyrae.

1.4.10 Masses from Microlensing

Gravitational lensing allows measurement of masses of astronomical objects with an accuracy of a few per cent. Conventional ground-based microlensing surveys (such as MACHO, EROS and OGLE, Alcock et al. 1993; Aubourg et al. 1993) measure the temporary amplification of a source star when lens and source are aligned. The mass of the lens is not determined unambiguously, but folded in with the distances and the proper motions of lens and source. A microlensed source has two (generally unresolvable) images. The centroid of the two images makes a small excursion (of the order of a fraction of a mas) around the trajectory of the source as a result of varying magnification and image positions during lensing (e.g., Høg et al. 1995; Boden et al. 1998), and it is this centroid motion that GAIA can detect.

Astrometric microlensing has two advantages over photometric microlensing. First, the astrometric cross-section is substantially larger than the photometric, and second, the degeneracy with regard to the mass of the lens is removed. GAIA can probe this very different régime. The photometric optical depth $\tau_{\rm ph}$ towards the Galactic Centre is $\sim 10^{-6}$ and it is $\sim 10^{-7}$ towards the Large Magellanic Clouds. GAIA will monitor $\sim 10^9$ stars to $V \sim 20$ mag. Over the course of the five

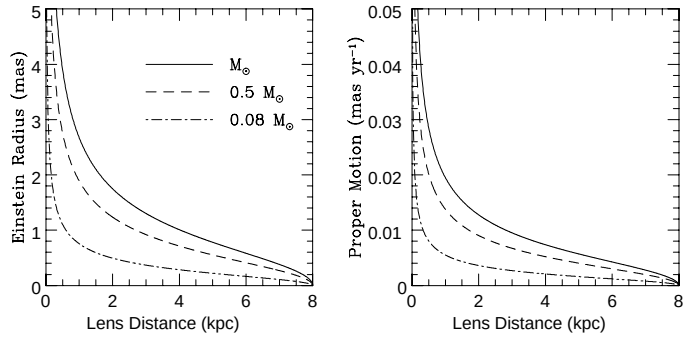


Figure 1.20: The left panel shows the Einstein ring radius and the right panel the relative proper motion between the lens and the source as a function of the lens distance. The source star is in the bulge and the different curves correspond to different lenses with masses $1M_\odot$, $0.5M_\odot$ and $0.08M_\odot$ respectively. The time scale of the event is taken as 50 days. The Einstein radius is related to the maximum angular deflection and so the graphs show the typical sizes relevant for astrometric microlensing towards the bulge.

years of elapsed mission time, GAIA samples each star typically 100–150 times. With this sampling, the efficiency of a microlensing experiment is $\lesssim 10$ per cent (e.g., Dimeo et al. 1997); by contrast, MACHO samples light curves at least once a night and the efficiency is ∼30 per cent. So, the expected number of photometric microlensing events detected by GAIA is < 100. The astrometric optical depth $\tau_{\rm as}$ is related to the photometric optical depth by $\tau_{\rm as} \sim \tau_{\rm ph}\theta_{\rm E}^2/\theta_{\rm min}^2$ (Miralda-Escude 1996; Paczyński 1998), where $\theta_{\rm E}$ is the angular Einstein radius, and an event is detected if the maximum angular deflection is greater than $\theta_{\rm min}$. For $V = 15$ mag, GAIA's astrometric accuracy for an individual measurement is around 60–100 μas, whereas for $V = 20$ mag, the accuracy is a little better than 1000 μas. Adopting the more pessimistic estimate, the astrometric optical depth $\tau_{\rm as}$ towards the Galactic Centre is $\sim 10^{-6}$ (inferred from numerical integrations in models of the Milky Way). Taking into consideration also the time variation of the centroid shift, which is necessary for the effect to be observed, the expected number of astrometric microlensing events detected by GAIA will be a few hundred.

Most of the lenses whose masses GAIA can measure are disk and bulge stars. To gain insight into typical values, consider a microlensing event with a time scale ~ 50 days and maximum magnification 2.64, which corresponds to an impact parameter $u = 0.4$. Suppose the source is a bulge star at a distance $D_{\rm s} \sim 8$ kpc. For three lens masses of $1.0M_\odot$ (full lines), $0.5M_\odot$ (broken lines) and $0.08M_\odot$ (dot-dashed lines), Figure 1.20 shows the angular Einstein radius $\theta_{\rm E}$ and the relative proper motion as a function of the lens distance. The maximum astrometric deflection is proportional to the Einstein radius, and so this sets the characteristic scale of the phenomenon. The goal of astrometric microlensing is to extract information about the lens mass by fitting the motion of the light centroid. Figure 1.21a shows two simulated events towards the bulge — the left panel shows a lens of mass $1M_\odot$ at 4 kpc, while the right panel a brown dwarf lens of mass of $0.08M_\odot$ also at 4 kpc. The astrometric deflection is shown for a barycentric observer (dashed lines) and a terrestrial observer (full lines). There are at least six parameters — namely the Einstein angular radius, as well as the zero point, the proper and parallactic motion of source and lens (e.g., Boden et al. 1998). If the lens is luminous, then the blending fraction is also unknown. It takes at least six independent measurements to construct a centroid trajectory. The marks on the full curves seen by a terrestrial observer in Figure 1.21a correspond to GAIA's typical sampling rate. The construction of the centroid trajectory and the extraction of the lens mass is straightforward (e.g., by least square fitting or by downhill simplex). The astrometric microlensing signal differs from the time-harmonic excursion if the source has a gravitational companion. The latter can be detected and removed if present — for example, by model fitting or by finite differencing.

Figure 1.21b shows the typical error in GAIA's mass determination for microlensing towards the bulge. The expected error depends on the mass of the lens and its distance — as these control the size of the Einstein radius and the magnitude of the parallactic effects. The astrometric accuracy of each individual measurement is taken as 100 μas (the nominal individual accuracies are about 60 μas for $V = 15$ mag, with the final mean astrometric accuracies a factor of 6 or so better). Here, the lens mass can be recovered very comfortably, even in the brown dwarf régime. This suggests that GAIA will be a powerful probe of the present-day mass function towards the Galactic Centre. The degrading of the astrometry has a deleterious effect on the accuracy with which the lens mass can be recovered. This may be unduly pessimistic as it is assumed only the astrometry is used. For at least some (perhaps most) of the events, additional photometric data will be available from ground-based monitoring campaigns.

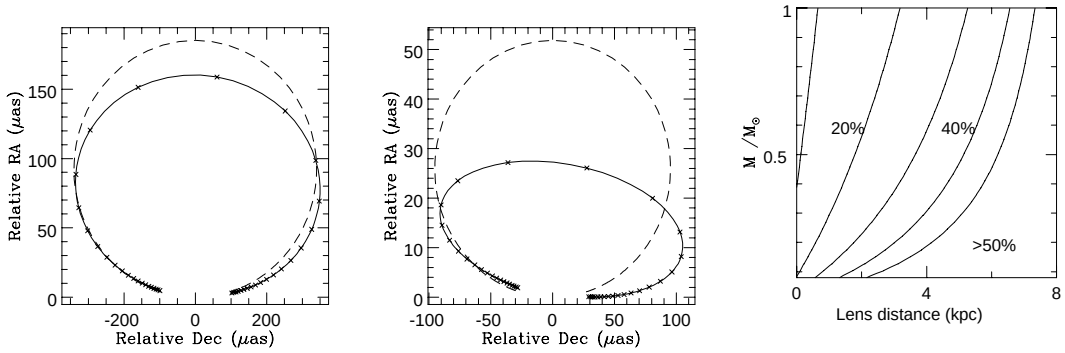


Figure 1.21: (a) Two simulated events for microlensing towards the bulge. The left panel shows a lens of mass $1M_\odot$ at 4 kpc, while the right panel a brown dwarf lens of mass of $0.08M_\odot$. The astrometric deflection is shown for a barycentric observer (dashed lines) and a terrestrial observer (full lines). The marks on the full curves correspond to GAIA's typical sampling rate. (b) Contours of expected error in GAIA's mass determination for microlensing towards the bulge. The percentage error in the mass estimate is contoured in the plane of lens mass and lens distance, assuming the source lies in the bulge. The assumed astrometric accuracy of a single measurement is 100 μas, corresponding to $V = 15$ mag.

The main microlensing objective of GAIA is to determine the present-day mass function of the disk and bulge. The faint end of the mass function is very poorly determined and GAIA can hope to make a very real contribution here.

There are 60 000 high proper motion stars known (Luyten 1976) with proper motion > 0.2 arcsec yr^{-1} — of course, this is a lower estimate and an all-sky catalogue of high proper motion stars complete to $V = 20$ mag will only be available after GAIA. For a given astrometric displacement, the minimum impact parameter and the time scale for the lensing event can be calculated for typical stellar masses. The total area swept by the 60 000 stars in the 5-year mission time is $\sim 10^{-7}$ of the whole sky. There are ~ 100 stars brighter than $V = 20$ mag in the total area swept by these stars. The lensing events have typical time scales between a few months and several years. For each of these high proper motion stars, a first assessment suggests that the mass can be determined to $\sim$1 per cent.

The faintest of the lenses will be the most astronomically interesting, being the brown dwarfs and the old, degenerate dwarfs. Planet-searching by photometric and astrometric microlensing has also been suggested (e.g., Mao & Paczyński 1991; Safizadeh et al. 1999). A planet orbiting a lens can perturb the images and distort the astrometric signal (see Figure 1.22 for some ideal cases from Safizadeh et al. 1999). The planet's effect is short, unlike parallactic and blending effects which are important over the entire duration of the event. However, the magnitude of the planetary perturbation can still be large for Jovian-like planets. On average, the time spent above 10 μas is a few days. This suggests that the efficiency of planetary detection via microlensing will be very low. More relevant for planet research will be the astrometric follow-up of events detected photometrically, a project ideally suited to SIM rather than GAIA.

Forward Microlensing Although this section has focussed on the effects of the 'inverse' microlensing problem in which (as in the MACHO-type searches) many sources are regularly examined for lensing from unknown objects, the 'forward' microlensing problem is also relevant. In this case the lensing effects of known objects are predicted and (hopefully) observed. Every star will cause a displacement of background objects, which can be predicted for stars with known positions, proper motions, and parallaxes. As the star moves, this displacement will change, and can thus be detected. From such measurements, the mass of the star can be directly determined. The area swept out by each star depends on its mass, distance, and proper motion. Some simple calculations were carried out by (Paczyński 1998) which showed that, using the 12 204 Hipparcos Catalogue stars with distances < 100 pc and proper motions > 100 mas yr^{-1}, about 10^{-7} of the whole sky would be subject to displacements > 100 μas from these stars in a 3-year mission. Since GAIA will produce a catalogue of $\sim 10^8$ objects to about 15–16 mag, there should be of order 10 observable microlensing events from Paczyński's Hipparcos Catalogue stars alone.

Reconciling what seems to be a higher probability of detecting the forward events, compared with the of order 100 GAIA standard microlensing events, will require further study. The Hipparcos subset already represents a large fraction of the total number of high proper motion stars, so that the predicted events will not scale simply with catalogue size. The details of the individual measurement errors as a function of magnitude (the faintest GAIA stars

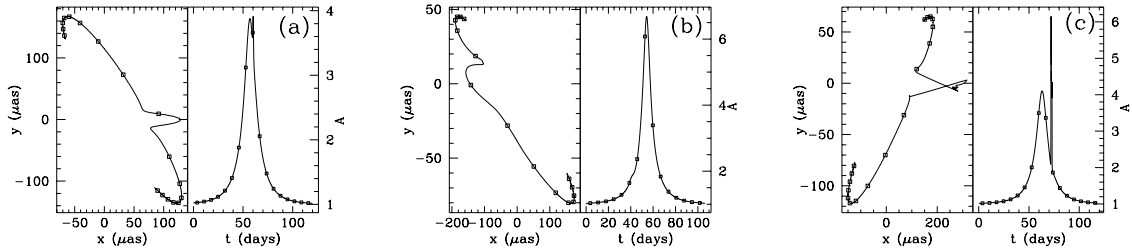


Figure 1.22: Planet lensing examples from Safizadeh et al. (1999). All examples assume $q=10^{-3}$ (where q is the planet to lens mass ratio), with a primary lens Einstein radius of 550 μas, corresponding to a Saturn mass planet. Squares are plotted one per week: (a) x_p = 1.3; (b) x_p =0.7; (c) a caustic crossing event with x_p = 1.3; x_p is the projected planet-lens separation in units of the Einstein radius (courtesy Neda Safizadeh).

will be less suitable targets) also require assessment. On the other hand, since the lensing events in the forward case can be predicted, it will be possible to apply additional resources, such as SIM, to interesting events. This implies that the entire GAIA Catalogue should be examined for potentially observable microlensing events of one object on another, and also for the possibility of microlensing by stars of objects in the Radio Reference Frame (where retrodictions from 1980 to the present will also be useful). This might also serve as a different means of doing a planet search, where slow microlensing events from distant objects with small proper motions will be most useful.

1.5 Binaries and Multiple Stars

In the coming decades, large telescopes and interferometry from both the ground and from space will revolutionize the study of most classes of objects, including binaries and multiple stars. As at the present time, however, the selection of observed objects is likely to be rather random, and one of GAIA's unique features is the well-defined sampling.

1.5.1 Census of Binaries

GAIA will not resolve binaries with separations below ∼20 mas which today is routinely achieved from the ground by speckle interferometry. At this angular separation, only systems with nearly equal component magnitudes can be measured, but at a few arcsec, even very faint secondaries can be detected. For a distance-limited sample, the detection-efficiency for resolved binaries will thus be low at periods corresponding to 20–50 mas separation, but it will increase significantly for wider systems. The resolved binaries will be studied in different populations of stars, including nearby star forming regions, open clusters and associations as well as field binaries of different ages. The high-precision parallaxes and proper motions will also allow identifying (among individually observed stars) the very widest binaries and studying their relation to 'common proper motion' pairs and associations. They also constrain the gravitational potential in the Galactic disk (Section 1.2.8).

GAIA is extremely sensitive to non-linear proper motions. A large fraction of all astrometric binaries with periods from 0.03–30 years will be immediately recognized by their poor fit to a standard single-star model. Most will be unresolved, with very unequal mass-ratios and/or magnitudes, but in many cases a photocentre orbit can be determined. For this period range, the absolute and relative binary frequency can be established, with the exciting possibility of exploring variations with age and place of formation in the Galaxy. Figure 1.23 shows the detection efficiency as a function of period, for two different distance-limited samples and for four ranges of apparent magnitude. The left peak contains the astrometric binaries, while the right peak contains resolved binaries. The gap between the two becomes more pronounced as the mean distance increases.

There is some non-intuitive fine-structure, mainly arising from the varying slope of the mass-luminosity relation. The simulations assume a mass-independent distribution of mass-ratios that gives a much broader Δm-distribution for more massive (brighter) than for fainter pairs, and the fraction of resolved binaries will thus seem abnormally low for the *a priori* easier bright systems.

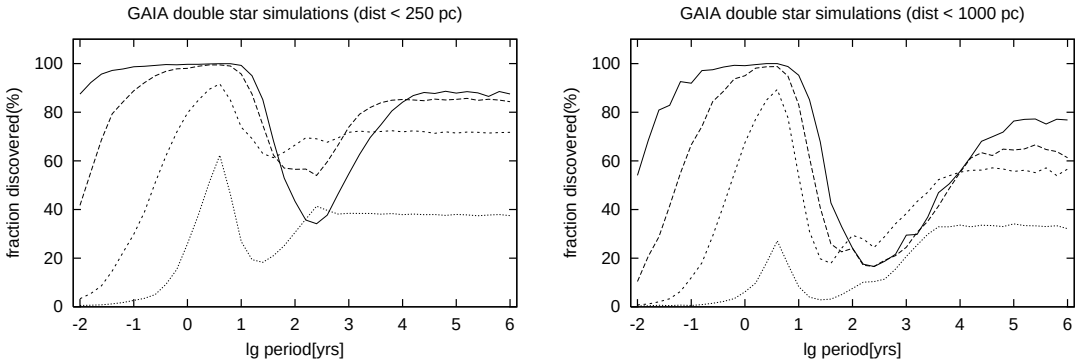


Figure 1.23: Total fraction of the double stars that are observed as non-single by GAIA, as a function of the binary period. The four curves are for the successive magnitude ranges 10–12.5, 12.5–15, 15–17.5 and 17.5–20 mag. (a) stars nearer than 250 pc. (b) stars nearer than 1000 pc.

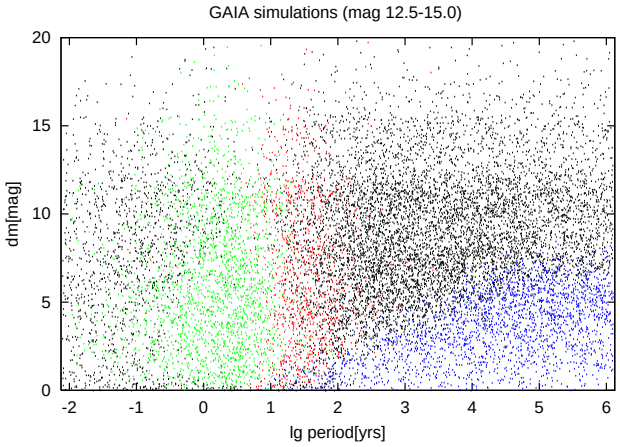


Figure 1.24: Simulated bright binaries (black) in a log P versus Δm diagram. The other colours denote systems detected as non-single by GAIA: blue for resolved system, red for 'quadratic', and green for 'stochastic' proper motion deviations.

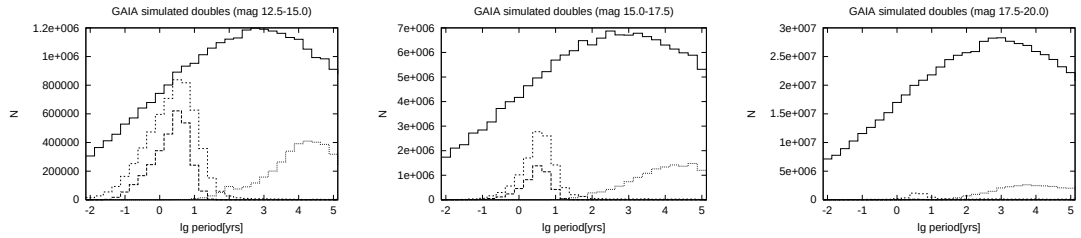


Figure 1.25: Total numbers of binaries per 0.25 dex in log period (solid curves) and the numbers resolved (dotted) or detected (lighter dashed) as astrometric binaries. The lower dashed curve gives the number of successful astrometric orbit determinations. Apparent magnitude intervals: (a) 12.5–15.0; (b) 15.0–17.5; (c) 17.5–20.0 mag.

Table 1.10: Total numbers (millions) and fraction detected of the binaries within different distance-limits and magnitude-intervals.

mag	parameter	62.5 pc	125 pc	250 pc	500 pc	1000 pc
10.0–12.5	N(bin)	0.035	0.17	0.80	2.6	4.8
	N(det)	0.034	0.16	0.69	2.0	3.4
	fraction (%)	96	95	87	78	71
12.5–15.0	N(bin)	0.077	0.37	1.62	6.5	22
	N(det)	0.067	0.33	1.37	4.8	13
	fraction (%)	86	89	84	73	60
15.0–17.5	N(bin)	0.079	0.69	3.58	14.2	47
	N(det)	0.051	0.48	2.52	9.1	24
	fraction (%)	64	69	70	64	50
17.5–20.0	N(bin)	0.004	0.34	4.81	29.8	110
	N(det)	0.003	0.14	1.75	8.9	24
	fraction (%)	70	41	36	30	22
All	N(bin)	0.196	1.58	10.8	53	183
	N(det)	0.152	1.11	6.3	25	64
	fraction (%)	77	70	59	47	35

Table 1.10 summarizes the results of the simulations for five different distances. It shows that GAIA will detect a majority (59 per cent) of maybe 10 million binaries closer than 250 pc to the Sun. While this fraction drops to 35 per cent out to 1000 pc, this represents key information on 64 million binaries. This huge sample can be subdivided in, e.g., age-groups, and the variability and possible evolutionary trends of the binary frequency at different orbital periods can be studied.

Figure 1.24 shows the distribution of the detected binaries in the $(\log P, \Delta m)$ plane. They can be divided into two main categories. The resolved binaries to the right are generally main-sequence pairs with small Δm, while the astrometric binaries on the left include also large Δm's from systems with white dwarf components. With increasing distance, the resolved binaries will have longer periods, and there will be a lack of detections of the otherwise very common 10–1000 year pairs. The expected all-sky number counts are illustrated in Figure 1.25, which demonstrates that the faint majority of GAIA binaries will mostly go undetected. But they are present, and the statistical results from the nearby samples allow making realistic corrections for the duplicity.

The number of detected astrometric binaries will be larger than the number of actual orbit determinations. The success rate has been checked by simulations, and the number of 'good' orbits is given by the lower curve to the left in Figure 1.25. Although the relative number of astrometric orbits declines fast towards fainter magnitudes, the absolute numbers (again some 10 million orbits) are seen to be dominated by the 15–17.5 mag interval. To determine starting elements for all these unknown orbits (enabling a subsequent least-squares refinement as used in the simulations) will be a formidable task, which will require significant experimentation and development of algorithms. Figure 1.26 is an example of an astrometric binary discovered with Hipparcos, classified only as a 'stochastic solution', but with a full orbital solution subsequently determined.

A key issue regarding double star formation is the distribution of mass-ratios q. For wide pairs (> 0.5 arcsec) observed as independent GAIA targets, this is indirectly given through the distribution of magnitude differences, and one may strive for observations to maybe $\Delta m = 10$ mag at 4–5 arcsec separation. At smaller separations, the wings of the point spread function from the primary will interfere, while at larger separations there will be an impractical number of optical companions. With this dynamical range, one will have at least a photometric determination of the q-distribution down to $q \sim 0.1$, covering the expected maximum around $q \sim 0.2$. Furthermore, the large numbers of ('5-year') astrometric orbits, will allow derivation of the interesting statistics of the very smallest (brown dwarf) masses (Section 1.5.6), as well as the detailed distribution of orbital eccentricities.

The large-scale GAIA observations will also clarify the statistics of multiple systems. For stability reasons, the period-ratio in hierarchical systems has to be above about 10:1, but the observed distribution goes from 10 to at least 10^6, with 1000 as a typical value. The '5-year' astrometric pairs can thus be expected to contain many

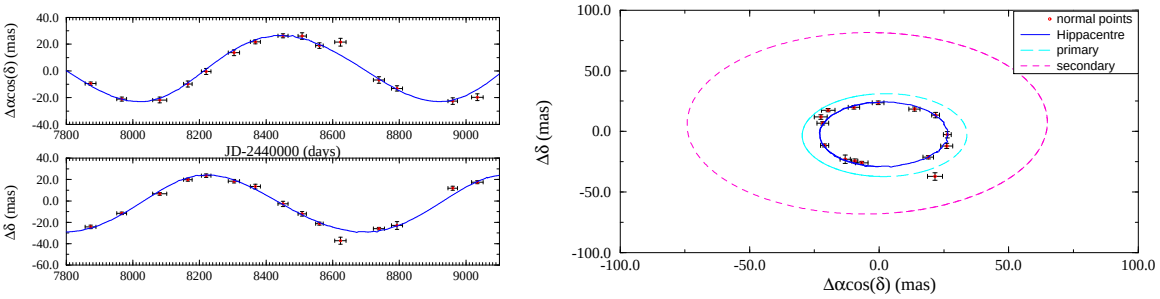


Figure 1.26: HIP 39903: an example of a new astrometric binary from Hipparcos data. The left panels show the fits in right ascension and declination, and the right panel shows the reconstructed orbit in the $\Delta\alpha, \Delta\delta$ plane. Combining the astrometric data with the spectroscopic data from Murdoch & Hearnshaw (1993), and assuming a mass of 1.27 $M_\odot$ for the primary gives a mass 0.58 $M_\odot$ for the secondary and a 2.8 mag difference between components (probably a F5V+K6V system). The period is $P = 899.28$ days, $a_0 = 26.47$ mas, $e = 0.112$, $i = 21°.8$. For previously unknown orbital systems like this, the GAIA radial velocity resolution will allow determination of the orbital period, and will thus play a key role in the determination of the orbital solution based on the astrometric data (courtesy F. Arenou).

spectroscopic and/or eclipsing subsystems, and GAIA will be able to provide reliable triple-star statistics. Although one will seldom be able to specify completely the geometry in a specific system, the fraction of eclipsing components is expected to vary with the astrometric inclination, shedding some light on the distribution of relative inclination between the two orbits. This is an important parameter in triple-star orbital dynamics, and it may also help discriminate between different theories of (multiple) star formation.

1.5.2 Proper Motion Bias from Undetected Binaries

A key issue for GAIA is the extent of the 'noise' in the proper motions due to undiscovered binaries. In a Galaxy model with 100 per cent binaries (certainly more realistic than 0 per cent) and no galactic-scale motions, the simulated observed proper motions come from the magnitude-dependent observational uncertainty, but with a contribution from orbital motion in unresolved binaries. Figure 1.27 shows such simulated proper motions (converted to velocity via the well-observed parallaxes) for 'apparently single' stars, as a function of their true orbital periods.

This is mostly a problem for the bright stars. In Figure 1.27a, the typical observation error (horizontal line) is small, and the plot shows the slow $P^{(-1/3)}$ Keplerian taper (with a large spread due to varying orbital parameters). The many systems with about 30–1000 year periods add significant noise to the observed galactic motions, but because of the random orbital parameters, any systematic effects will be small. The 1–30 year binaries are detected with high efficiency by their curved proper motions, causing the gap in the period-distribution. At even shorter periods, the means over 5 years will correspond closely to center-of-mass velocities. For the faint (majority) of stars observed by GAIA (Figure 1.27b), the observational errors are mostly as large or larger than the orbital effects, and again the duplicity will not have major consequences for galactic studies. It is nevertheless important to estimate the orbital contribution to the velocity spread, which needs careful extrapolation from the bright-star results.

1.5.3 Masses from Visual Binaries

GAIA's precision astrometry will be exploited for visual binary orbit- (and hence mass-) determinations of unusual accuracy. Although the number of resolved systems will be relatively modest, and although a fraction of these will have short enough periods for the orbits to be well-defined, the absolute numbers of new high-precision masses will be very impressive. GAIA observations (astrometry and radial velocities) for thousands of close (20–100 mas) but resolved binaries have

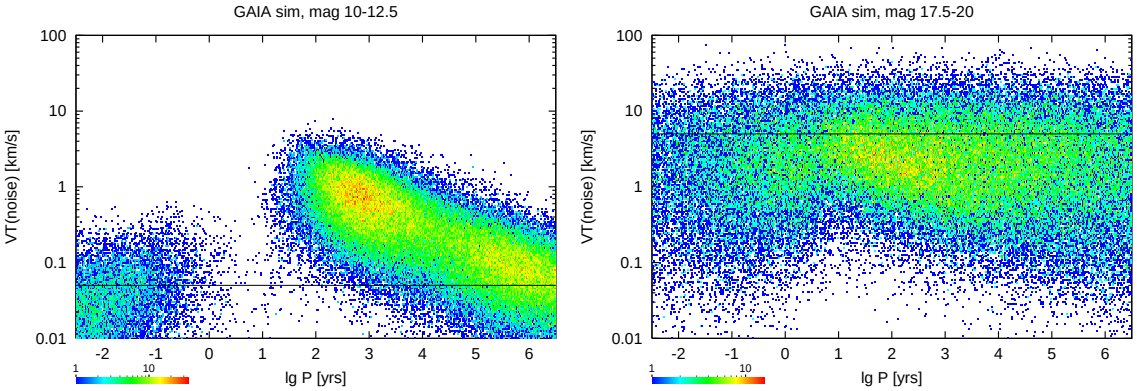


Figure 1.27: Simulated proper motion errors (converted to velocity via the observed parallax) for the photocentres of undiscovered binaries with different periods: (a) in the range 10–12.5 mag; (b) in the range 17.5–20 mag. The horizontal lines indicate a median mean error in the absence of orbital motions.

been simulated, based on preliminary assumptions about the per scan observational errors. Least squares fitting of orbits to these measurements shows that the full set of orbital elements can be recovered for periods in the 4–40 year range.

The mass-ratios are surprisingly well-determined even for 50-year periods. A value for the total mass can thus be divided between the two component stars with almost no extra error. Secondly, the mass uncertainties due to errors in the parallaxes are almost always negligible compared to the uncertainties due to errors in the orbit determination. And thirdly, although radial velocity observations give limited quantitative information for the mass determinations, they play a key role in helping the orbit determinations converge. In the simulations, the least-squares solutions are started from a 'known' input. In reality, one will face a much more difficult problem with no *a priori* knowledge of even the binary periods. By not being 'hidden' in the scanning model, the radial velocities will be crucial for locating plausible periods.

Table 1.11 summarizes the estimated numbers of orbits (over the whole sky) with mass-determinations of different precision. The smallest errors are obtained for the brighter objects, but in the (still useful) $3-10$ per cent bin, at least a third of the stars are fainter than $V \sim 15$ mag. As to the absolute numbers, one may obtain more than 10^4 masses accurate to 1 per cent. This represents a very important astrophysical result, allowing detailed checks of the stellar evolution models. Comparably accurate results are otherwise only available from short-period eclipsing binaries, where interaction effects cannot be excluded *a priori*. The GAIA masses will be dominated by relatively bright ($V < 15$ mag) main-sequence G-K dwarfs, but one will also have a fair share of fainter lower-mass objects. The mass-ratios are well-determined, but as for most resolved systems, they will not be far from unity. Figure 1.28 shows results from Söderhjelm (1999).

1.5.4 Fundamental Data from Spectroscopic and Eclipsing Binaries

GAIA radial velocity observations of the brighter astrometric binaries will allow spectroscopic orbit determinations, giving single-spectrum mass-functions $f_s(m) = m_2^3/(m_1+m_2)^2 = m_1 q^3/(1+q)^2$. The usually unknown inclinations follow from the astrometric orbits, and once a primary mass is estimated, the mass-ratio follows. The (so far uncertain and controversial) distribution of mass-ratios will be obtained with high accuracy for large numbers of main-sequence systems.

GAIA radial velocities for early-type stars will have fairly large errors (Section 8.3), so that high-quality (1 per cent) masses will be obtained only for pairs with at most a few days period. Because the probability of eclipses increases steeply towards shorter periods, this may still be about half of the systems. Ground-based spectroscopy will help here. For late spectral types, the radial velocity errors are smaller, but so are the velocity amplitudes, and the probability of eclipses. The majority of mass-determinations will again be at periods of a few days.

Table 1.11: Total numbers of systems with different estimated mass uncertainties (per cent) for three different magnitude intervals.

V	$< 0.3\%$	0.3–1%	1–3%	3–10%
10.0–12.5	2 000	3 000	4 000	8 000
12.5–15.0	2 000	9 000	9 000	12 000
15.0–17.5	0	1 000	5 000	10 000
Total	4 000	13 000	18 000	30 000

The GAIA radial velocities will cease to be useful for orbit determinations at rather bright magnitudes, around 11 mag for early-type stars and 14 mag for late-type. The multi-epoch photometry will however provide a sample of several million eclipsing binary light-curves, and among these, many will be members of clusters or associations with well-defined ages and metallicities. These will be prime targets for ground-based follow-up spectroscopy (and sometimes also photometry), giving in the end masses, radii and temperatures for stars of known age and composition (Section 1.5.6).

The best available mass-luminosity and mass-radius relations (errors ≤ 1 per cent) are based on less than 60 eclipsing binaries. While these systems are relatively common (about one star in 1000, Kaluzny et al. 1995), none are known for many important but short-lived stages of stellar evolution. GAIA will have a dramatic impact by identifying large numbers of such systems, and determining masses and radii for them. A typical application might be the determination of the convective core overshooting parameter as a function of mass (Rosvick & VandenBergh 1998), where the comparison with theoretical models can not yet be done for lack of suitable test binaries. The mass determination of a single pre-main sequence eclipsing binary was a very important step in calibrating the pre-main sequence evolutionary tracks. Even bright systems are not well-surveyed, as demonstrated by the large number of new eclipsing (as well as contact) binaries discovered by Hipparcos at magnitudes as bright as 7 mag.

GAIA will observe large numbers of double-lined spectroscopic binaries. All of them give directly a mass-ratio from the ratio of velocity-amplitudes, and the resulting q-distribution (near $q = 1$) can be combined with the lower-q distribution from the single-lined systems. Even more importantly, a sizeable number of the shorter-period systems will be found to be eclipsing. These double-lined eclipsing binaries will then give the masses and radii that are so crucial for checking and calibrating the theoretical models for stellar interiors (e.g., Schwarzschild 1958). The effective temperatures follow from the radii and absolute luminosities in principle from the relation $L = 4\pi R^2 \sigma T_e^4$, where the largest uncertainties are in the bolometric corrections. With many systems to work with, different stellar models can be checked, and a self-consistent set of calibrations derived.

Considering eclipsing binaries that are members of physically bound multiple systems also constrains stellar evolutionary models. All the information that can be extracted for the companion(s), i.e. effective temperature, magnitude difference, mass, etc., should also be fitted by the same isochrone as the eclipsing binary pair. Few such systems are known at present, but GAIA will discover numerous examples.

For the millions of faint eclipsing binaries with only light-curves available, GAIA will uniquely provide also the absolute luminosities and temperatures (from the parallaxes and colours). It is then possible to estimate the absolute stellar radii, and via the light-curves the sizes of the orbit follow, and thus the masses of the systems. The accuracy is moderate, but this will be a new way to look for large deviations from the 'normal' mass-radius-luminosity relations. Finally, the period-statistics gives the short-period continuation of the distributions obtained from the astrometric pairs. Due to the very large numbers of systems, it will be possible to study and correct for the selection effects and to find evolutionary effects by looking at different age-groups.

1.5.5 Interacting Binaries, Accretion Disks, and Black Holes

Certain types of variables, while common, are poorly understood: W UMa-type contact binaries (with two Solar-type components, periods of 0.2–1.0 day and variability of ~ 0.6 mag), represent ~ 1 per cent of the total number of Solar-type stars (Ruciński 1994), and thus a relatively normal (and long-lived) stage of the evolution of Solar-type stars, yet their origin, structure and evolution is poorly understood. The distribution of currently known contact binaries (~ 500) already shows

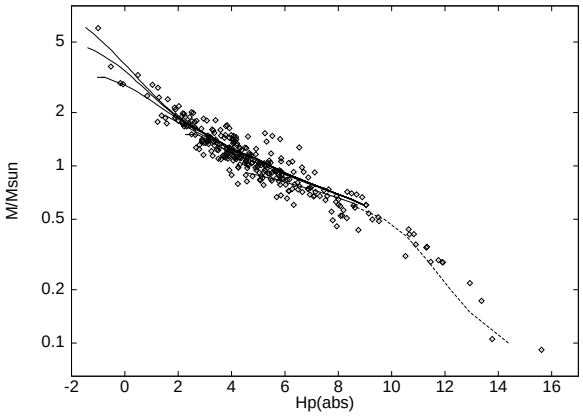


Figure 1.28: The mass-luminosity relation for 276 main sequence stars with mass uncertainties below 30 per cent, as derived from the Hipparcos data by Söderhjelm (1999). Theoretical isochrones are shown for log(age) = 7.5, 8.0, 8.5, 9.0, 9.4, and 10.2.

severe incompleteness at $V \sim 9$ mag. GAIA will discover many more, and may well identify some objects in the critical, short-lived stages leading to their formation. The sample will be unbiased, and thus allow the spatial distribution and evolutionary state as a population to be studied. This is not possible with current (or foreseen) ground-based samples, which do not reach sufficiently faint magnitudes and show a high degree of patchiness in their spatial distribution.

Similar considerations apply to Algol-type binaries (eclipsing binaries in which one of the components fills its Roche lobe accreting on the other component): they are fairly common among bright stars, so that accretion must play a role in the evolution of a significant fraction of stars. The existing sample of Algols is already severely incomplete at $V \sim 9$ mag. In many cases, radial velocities (GAIA or ground-based) will allow mass/radius-determinations which may be compared with theory. Answering fundamental questions about the mass-transfer (time-scales, systemic mass-loss, orbital evolution) requires these complete samples of systems with well-determined parameters.

RS CVn-type binaries (two cool, partly-evolved stars with orbital periods of a few days, and rotational period locked to the orbit) are also common, and prominent in the X-ray sky. Establishing their space density and distribution is essential for understanding their contribution to the Galactic component of the X-ray background. The GAIA sampling of their light-curves will also help elucidate their complex rotation/activity/age-relations. Because their components are often similar, radial velocity observations will often give accurate masses and dimensions.

A rich variety of astrophysical problems related to interacting binary systems are accessible with parallaxes in the 10 μas range. The evolutionary history of interacting binary systems, and the origin of Type I supernovae, millisecond pulsars, low mass X-ray binaries, and globular cluster X-ray sources is intimately bound up with the behavior of compact binaries with mass transfer and mass loss. Accurate stellar parameters in such systems can be derived from the many eclipsing pairs discovered by GAIA, supplemented by additional radial velocity observations (Section 1.5.4). In some favourable cases, astrometric orbits are measurable, constraining the masses even without eclipses. Many specific questions about accretion rates, precursors, mass distributions, and kinematic behaviour can be addressed with these data, including studies of the black hole candidates.

Galactic black hole candidates have bright secondaries (9 mag in the case of Cyg X-1, and 12 mag or fainter in the case of V404 Cyg) and wide orbits (with orbital periods of about 6 days), which should yield definitive black hole masses by determining orbital separation and inclination.

Be star X-ray binaries are believed to consist of a recently formed neutron star and a Be star companion. The orbit has not yet circularized, and the eccentric motion produces periodic eruptions at periastron as the compact star passes through the mass outflow from the Be star. Measurement of the orbital parameters will constrain the anisotropy of the supernova mass ejection mechanism, and so shed light on the kinematics of isolated pulsars.

To understand the core collapse of massive stars, the supernova phenomenon, and the existence and properties of neutron stars, an accurate equation of state for matter at supra-nuclear density is essential. Models tested on Earth using heavy-nuclei collision experiments do not reach sufficiently high densities. Hence, it is necessary to compare model predictions with observed neutron-star properties. The Be/X-ray and Be/pulsar binaries, in which a neutron

star is an X-ray or radio pulsar in a wide orbit with a Be star, offer good prospects for the determination of neutron-star masses, by combining timing orbits of the neutron star with astrometric orbits of the massive companion. For typical systems, the neutron-star mass estimates will be accurate to $\lesssim 0.1\,M_\odot$. In these systems, the neutron star cannot yet have accreted much material, so the range of masses found in different systems will reflect directly the supernova process. A constraint on the equation of state will come from the largest measured mass, as the maximum mass a neutron star can have before collapsing into a black hole depends on the equation of state.

1.5.6 Low-Mass Stars

The stellar mass function at low masses remains poorly known, in spite of massive observational and theoretical efforts over many years. The stellar luminosity function has been determined to improving accuracy recently, especially in galactic and globular clusters, using HST, but also in extensive field star surveys. Hipparcos results provide an accurate luminosity function for luminous stars near the Sun. The improvements GAIA will provide were discussed in Section 1.3.4.

Conversion to a mass function remains problematic. Alternative analyses of the available data conclude that the initial mass function is decreasing (De Marchi & Paresce 1995, for the dynamically evolved globular clusters), flat (Piotto et al. 1997; von Hippel et al. 1996), slightly rising (Kroupa et al. 1993 for the disk; King et al. 1998; Gould et al. 1997; Kirkpatrick et al. 1994), or steeply rising (Mera et al. 1996 for the halo). The problem arises since theoretical mass-luminosity relations do not reproduce the observed structure in the luminosity functions, particularly failing to describe the precise Hyades data from Hipparcos. Empirical calibrations at low masses are limited in quality and number to only two eclipsing dM binaries, one of which is probably a halo pair, and a small set of field binaries studied by speckle methods, none of which has accurately known age or metallicity, and few of which have very good distances. Age and metallicity have large effects on the stellar models, while distance and a reliable extinction measure are necessary to derive luminosity (e.g. Chabrier & Baraffe 1995; Baraffe et al. 1995). None of these parameters can be derived reliably from available model atmosphere analyses of very cool stars.

The mass range of greatest interest is $0.5 \geq M/M_\odot \geq 0.1$, in which there is substantial curvature, and possibly higher-frequency structure, in the mass-luminosity relation. Since the conversion of luminosity to mass depends on the derivative of this relation, it is this poorly-quantified structure which causes the uncertainty. The current state-of-the-art and its limitations is illustrated in Figure 2 of Henry & McCarthy (1993). GAIA will greatly improve the empirical mass-luminosity relation in this mass interval. On the one hand, high-quality mass determinations will be made from several hundred visual binary orbits (see Section 1.5.3). The drawback is that one still has hard-to-disentangle effects of age and metallicity for these nearby field binaries. On the other hand, GAIA will discover many new low-mass eclipsing binaries in clusters, where these parameters can be derived with very high precision from studies of the turn-off members of the same clusters (Section 1.2.4). Because the low-mass stars are so faint, their light- and radial velocity curves will however have to be derived from ground-based observations with better accuracy than is possible with GAIA. As an additional benefit of such surveys of the lower part of cluster main sequences, we will get improved knowledge of the luminosity functions and of the binary statistics.

The location of the stellar masses below $\sim 0.5 M_\odot$ (GAIA absolute magnitudes from 9–12) in the H–R diagram depends sensitively on metal abundance (e.g. Chabrier & Baraffe 1997). This dependence can be calibrated observationally through open cluster observations. Subsequently, these low mass stars can be used as tracers of the disk metallicity gradients up to large distances.

The distribution of separations, mass ratios, and orbital eccentricities in the low-mass binary population is unknown. Available studies are restricted to relatively long periods (tens of years) to allow speckle resolution. Radial velocity discoveries however (e.g. Marcy & Butler 1998) have found planetary systems with properties quite like those of the Solar system (see Section 1.6). The only known eclipsing dM binaries have periods of about one day, presumably due to observational selection. We may then adopt, from Kepler's laws, (period)$\propto m/r^3$, so that the observable parameters for stellar systems with total masses of order 0.5 $M_\odot$ can be assumed to follow similar relations to those which apply in the Solar system.

For simplicity we consider only non-interacting binaries, so limit consideration to separations larger than a few stellar radii, and exclude cataclysmic variables. An eclipse will occur if $R_\star \geq R \sin i$, with R the star-star separation, and i the inclination of the orbit to the line of sight, defined so $i = 0$ is the eclipse plane. This implies, for a $0.3 M_\odot$ star, at periods of 1 day (like CM Dra, YY Gem), $i \leq 17°$, period = 100 d (Mercury's orbit) $i \leq 0.34°$, period = 1 yr, $i \leq 0.11°$. The corresponding probabilities of this alignment are 0.1, $2.10^{-3}, 6.10^{-4}$. For a log-normal period distribution function, comparable to that observed in radial velocity surveys of higher mass field stars, one expects equal numbers of systems to exist in each of these intervals, so that of order 10^3 star-observation sets must

be observed to detect of order 1 system. For a conservative estimate, we restrict to stars within 1 kpc, to minimise extinction, and maximise reliable association with a turnoff. In that case, again conservatively, scaling from related HST studies of NGC 2024, a 'representative rich cluster' will contain some 1000 star-observations per crossing of the GAIA field-of-view. Thus one to a few accurate masses will be determinable per cluster, each of which will be for stars with accurately known age and metallicity. This calculation is conservative, in that a rich open cluster such as M67 is known to have some 20 000 members.

Since dM stars are prone to flare, and have star spots which may mimic shallow eclipses during a rotational period, several multi-colour observations per star are optimal for detection, and are naturally provided by GAIA. A strength is that one does not need contiguous data runs. So long as the data sets have some sampling on times shorter than about an hour any random spacing of observations will do, since the orbits must have random phases.

For a system with inclination $i = 0$, one may calculate the eclipse duration. From Kepler's laws $\omega^2 \propto r^3$, while the angle subtended at the primary by the secondary goes as r^{-1}, so there is only a very weak distance dependence of the eclipse duration, (angular size)/(angular speed). The angular speed at 1 AU is close to 1 degree per day, for a 1 Solar mass system. For low-mass stars, stellar radius and mass closely follow $R/R_\odot \approx M/M_\odot$ (Baraffe et al. 1995), so that the angle subtended is typically $10^{-3}(1/d(\mathrm{AU}))(M_\star/0.1)$ and a representative eclipse duration becomes $\sim 1 \leq \tau \leq 5$ hours, well-matched to GAIA's observational capabilities.

1.6 Brown Dwarfs and Planetary Systems

Sub-stellar companions can be divided in two classes, namely planets and brown dwarfs. There exist three major genesis indicators that can help classify sub-stellar objects as either brown dwarfs or planets: mass, shape and alignment of the orbit, and composition and thermal structure of the atmosphere. Despite recent attempts (Oppenheimer & Kulkarni 2000), up to now theory has not been able to establish either a firm lower limit to the mass of a brown dwarf or the upper bound to the mass of a planet. Indeed, giant planets and brown dwarfs may populate a common mass range. Therefore, mass alone cannot be decisive for the classification of a low-mass companion to a star as a planet, unless this mass is close to that of the Earth (Black 1997).

It is thought that stars form from large-scale dynamical instabilities, while planets form by core accretion or dynamical instability of protoplanetary disks. It has recently been shown (Black 1997) that correlations between eccentricity and the logarithm of orbital period for pre-main sequence (and main-sequence) binaries and for objects thought to be the result of accretion in a disk (like the giant planets in the Solar system) are significantly different. The majority of the candidate planets discovered so far by the radial velocity programmes appear to follow the $(e, \log P)$ relation that can be established for pre-main sequence binaries. Low eccentricity alone is not sufficient to classify as planet a newly discovered low-mass companion, as low eccentricity is expected for stellar companions orbiting close to the other star.

Different correlations among orbital parameters (eccentricity, period or semi-major axis) and measurable differences in planetary frequency are likely to be generated by diverse planetary formation scenarios (core accretion and disk instability are the two known to date) and evolution mechanisms, as well as by different formation and evolution processes of the parent star (binarity, spectral type, metallicity, age). Orbital evolution mechanisms through the stellar disk (like gravitational migration) could generate differences in the distribution of orbital parameters (e.g., Jupiter-size planets on large orbits with very high eccentricity) with age. If disk instability is the preferred mode for the formation of giant planets, the gravitational pull induced by them on the parent star should already be measurable in very young pre-main sequence stars (there are several star forming regions closer than 200 pc), and planetary frequency is probably enhanced in binary stars (Boss 1998a; Boss 1998b).

The ability to simultaneously and systematically determine planetary frequency and distribution of orbital parameters for the stellar mix in the Solar neighbourhood, is a fundamental contribution that GAIA will uniquely provide, the only limitations being those intrinsic to the mission, i.e., to the actual sensitivity of the GAIA measurements to planetary perturbations. In the next section we discuss those limits and provide quantitative estimates of sample sizes accessible to GAIA.

Indications on the history of planetary frequency will come from the observations of the local population of old (thick disk and halo) stars. Galaxy models predict $\sim$ 4000 F–G–K dwarfs and sub-dwarfs to 200 pc, brighter than $V = 13$ mag, and with metallicity [Fe/H] < -1.0. By looking at stars in the age range $\sim$1–8 Gyr, i.e., up to values comparable with the age of the Galaxy, it will be possible to evaluate if metal enrichment has made planetary formation more likely.

Ongoing spectroscopic work in conjunction with ground-based astrometric searches will probably have more than 10 years of continuous observations on a number of sub-stellar companions, so that basic classification can start

before the sky is surveyed by GAIA. The range of masses between 1 $M_\oplus$ and a few Earth-masses (Neptune-class planets), will be marginally accessible to GAIA, but might be studied with SIM.

The major ground-based programmes and space missions employ pointed instruments and therefore require input lists of candidate stars. Although these lists can be optimized to yield much more than just the existence of extra-solar planets of different masses, the target selection criteria will depend on current understanding of formation and evolution mechanisms. The effectiveness of these pre-selected lists will depend on actual planetary frequencies (to date largely unknown) which, in turn, are likely to be functions of several other parameters, as discussed above.

Assuming that the portion of the orbital parameter space probed by GAIA is adequate, its magnitude limited survey could identify a larger variety of planetary systems, and thus lead to their classification into several different classes, resulting from a possibly different formation history. Also, GAIA will play an important, although indirect, role in the quest for habitable Earth-size planets. These might be found around Solar-type stars with a Jupiter-like planet orbiting at a distance ≥ 3 AU, a configuration which could conceivably protect terrestrial planets from cometary impacts (Wetherill 1994). GAIA measurements are particularly sensitive to such systems as a result of the mission duration and the favorable magnitude of Solar-type stars up to ~ 200 pc from the Sun (brighter than $V \sim 12$ mag). Therefore, GAIA could provide us with a quantitative estimate of the likelihood that inhabitable Earths exist in the Solar neighbourhood.

1.6.1 Brown Dwarfs in Binaries

An isolated brown dwarf is typically visible only at ages < 1 Gyr (Section 1.4.6) because of the rapidly fading luminosity. The same constraint applies if one wants to observe it as a resolved component of a binary system. However, in a binary system, the mass is conserved, and the gravitational effects on a main-sequence secondary remain observable over much longer intervals. GAIA will have the power to investigate the mass-distribution of brown-dwarf binaries with 1–30 year periods, through analysis of the astrometric orbits. From the size of an astrometric orbit of known period and parallax, we get the 'astrometric' mass-function $f_a(M) = (\alpha/\pi)^3/P^2$ where $\alpha = a(\mu-\beta)$. With Kepler's third law, this yields $f_a(M) = [m_2^3/(m_1+m_2)^2](1-\beta/\mu)^3$, where the mass-fraction $[\mu \equiv m_2/(m_1+m_2)]$ and luminosity fraction $[\beta \equiv L_2/(L_1+L_2)]$ are generally unknown. Maximum $f_a(M)$ is obtained for $\beta = 0$, corresponding to the limit of a very faint secondary, and the observed values are thus lower limits for the true $f(M) = m_1 q^3/(1+q)^2 = m_2^3/(m_1+m_2)^2$. This is mainly interesting for the lowest-mass secondaries. For a normal main-sequence primary, m_1 is estimated fairly well from its colour and luminosity, and the low tail of the $f(M)$-distribution will then trace the q-distribution towards the substellar brown dwarf/planet region.

Unfortunately, a small $f_a(M)$ may be due to a $\beta \approx \mu$ also. With a typically steep $L \sim M^4$ mass-luminosity relation, this translates to a two-valued 'mirror solution': either $q \approx [f_a(M)/m_1]^{1/3}$ or $q \approx 1 - [f_a(M)/m_1]^{1/3}$. The relative number of one case or the other depends sensitively on the q-distribution, and one will have to use the 'resolved' distribution around $q = 1$ in order to infer the low-q distribution. (In these cases, the radial velocity information from GAIA is of little help. Either the radial velocity amplitude is indeed very small, or it is a blend of two equal sine-curves with opposite phase and again unobservable.)

Recent simulations by Quist (2000) have been carried out for brown dwarfs orbiting main-sequence stars: among stars within 100 pc from the Sun, most such systems with periods between 0.003 and 200 yr will be detected.

1.6.2 Planetary Systems

There are a number of techniques which in principle allow the detection of extra-solar planetary systems (e.g. Marcy & Butler 1998): these include pulsar timing, radial velocity measurements, astrometric techniques, transit measurements, microlensing, and direct methods based on high-angular resolution interferometric imaging. Spectroscopy has been successful in finding the first such systems around normal stars. Four years since the first discovery of Mayor & Queloz (1995), some thirty candidate exo-planets are now known, with minimum masses in the range 0.5–11 Jupiter-mass (M_J), including the triple planet system of υ Andromedae (Butler et al. 1999). These recent discoveries have raised new questions in our understanding of the properties of planetary systems.

Several of these candidate planets have characteristics that are hard to explain within the context of current theoretical models for the formation and evolution of planetary systems. The fundamental tenets upon which present theories are based include nearly circular orbits and giant planets formed several AU from the central star, in contrast with the very short orbital periods (Mayor & Queloz 1995) and high eccentricities (Latham et al. 1989; Cochran & Hatzes 1994; Mazeh et al. 1996) found for several candidate planets. Their interpretation as *bona fide* planets largely rests on our understanding of correlations shown by their orbital and physical parameters, as recently discussed by Black (1997) and earlier by Duquennoy & Mayor (1991).

The inadequacy of the current models, when confronted with the first actual observations of extra-solar planets, clearly indicates that our knowledge of the complicated physical phenomena governing the formation and evolution of planetary systems is still partial. The interplay between additional theoretical work and more observational data is instrumental for continued improvement in our theoretical understanding of how planets form and evolve, and where Earth-like planets could eventually be found.

A better understanding of the conditions under which planetary systems form and of their general properties requires sensitivity to less massive planets (down to $\sim 10 M_{\oplus}$), better characterization of known systems (mass, and orbital elements), and complete samples of planets, with useful upper limits on Jupiter-mass planets up to several AU from the central star. Astrometric measurements good to 2–10 μas, and in particular those made from space by the NASA mission SIM (Boden et al. 1997) and by GAIA, will contribute substantially to these tasks, and will complement the ongoing radial velocity measurement programmes. SIM will be able to study in detail targets detected by other methods, including microlensing, and to search for low-mass planets (Danner & Unwin 1999). GAIA's strength will be its discovery potential, following from the astrometric monitoring of all of the several hundred thousand bright stars out to distances of ~ 200 pc.

1.6.3 Planet Detection and Orbit Determination

Astrometric techniques aim to measure the transverse component of the photocentric displacement. Reviews of ground-based work related to planet detection are given by Gatewood (1987) and Colavita & Shao (1994). The displacement can be quantified using the 'astrometric signature', $\alpha = (M_{\rm p}/M_{\rm s})(a_{\rm p}/d)$, where $M_{\rm p}$ and $M_{\rm s}$ are the planet and stellar masses, $a_{\rm p}$ is the planet orbital radius, and d is the distance; α is in arcsec if $a_{\rm p}$ is in AU and d in parsec.

Jupiter orbiting the Sun viewed from a distance of 10 pc would cause an astrometric amplitude of 500 μas, while the effect of the Earth at 10 pc is a one-year period with 0.3 μas amplitude. It follows that milliarcsec level astrometry can contribute only modestly to extra-solar planet detection. Nevertheless, the Hipparcos data were used to place upper limits on the planetary companions to 47 UMa, 70 Vir and 51 Peg (Perryman et al. 1996), while Mazeh et al. (1999) have used the Hipparcos data in combination with spectroscopic orbit elements to establish a semi-major axis of 1.4 ± 0.6 mas for the outermost planet in the υ And system, implying a mass of $10.1^{+4.7}_{-4.6} M_{\rm J}$ for that planet.

GAIA's potential can be assessed by simulating observations of a homogeneous set of extra-solar planetary systems, to establish the expected sensitivity to the presence of planets and the potential for accurate estimation of orbital parameters, as a function of semi-major axis, period, and eccentricity, and the distance from the Sun. A natural choice is to experiment with planet-star combinations grossly resembling our own Solar system, i.e., single Jupiter-mass planets orbiting 1-$M_{\odot}$ stars. In practice, this choice for the mass of the parent star encompasses the spectral class range from ∼F0 to K5-type dwarf stars, whose masses are within a factor of ~ 1.5 that of the Sun. For stars brighter than $V \leq 12$ mag (which will have the most accurate astrometry), this translates into a distance cutoff of ~ 200 pc. To this distance, F0–K5 type dwarfs dominate the star counts at bright magnitudes.

GAIA's scanning law is such that one-dimensional position measurements of any given object on the sky are made at successive epochs throughout the 5-year mission, approximately every month, on average. The sensitivity to the astrometric signature α depends on the error of the epoch measurements, which was fixed to 10 μas in the simulations (Lattanzi et al. 2000, and references therein), a value ∼3 times larger than the expected final astrometric error. It applies to the bright magnitude interval $V \leq 12$ mag (as the photon error becomes negligible compared to the magnitude-independent residual system errors), and is consistent with the adopted mission profile and the level of residual system errors of the baseline design. Therefore, in that magnitude range GAIA will have maximum sensitivity to planetary perturbations acting on normal stars. However, actual planetary formation scenarios could be such that astrometric signatures can also be detected at fainter magnitudes.

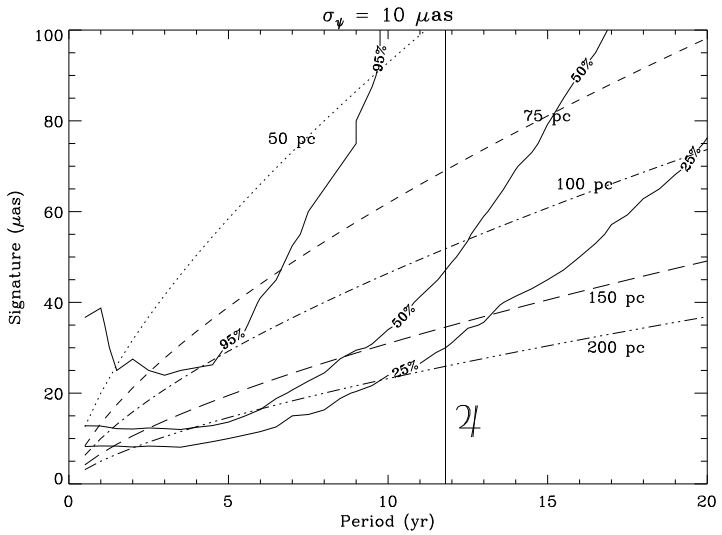


Figure 1.29: Iso-probability contours (solid lines) for 25, 50, and 95 per cent of detection probability, compared with Kepler's third laws (dotted/dashed lines) for systems with Jupiter–Sun masses at D = 50, 75, 100, 150, and 200 pc.

In the same way that the presence of convection currents on the star's surface may limit radial velocity determinations to no better than ± 1 m s^{-1}, even in older less-active stars (Marcy & Butler 1998), the ultimate limit to the detection of Earth-like planets may well be the non-uniformity of illumination over the disk of a star. Woolf & Angel (1998) show that the Earth causes the Sun's centre of mass to move with a semi-amplitude of about 500 km (0.03 per cent of the stellar diameter), while sun spots with up to 1 per cent of the Sun's area will cause the apparent centre of light of the Sun to move by as much as 0.5 per cent (~ 5 μas at 10 pc for a star of radius $1R_\odot$).

Detection of Giant Planets The results of the simulations are summarized in Figure 1.29. The solid lines were derived from the simulated dataset, and express the empirical relationship between the amplitude of the astrometric perturbation and the orbital period for detection probabilities of 25, 50, and 95 per cent. The curves show the behaviour of the astrometric signature α as a function of the orbital period for a Jupiter-mass planet around a Solar-mass star at different distances from the observer, obtained by substituting Kepler's third law in the expression defining α. Essentially all of the existing Jupiter-mass planets (95 per cent detection probability) within 50 pc and with periods between 1.5–9 years will be revealed by GAIA. The range of periods narrows with increasing distance. At 100 pc statistical certainty is possible only for those Jupiters with orbital periods clumped around the mission duration. These results confirm the existence of a good overlap with the range of periods probed by spectroscopy. This is illustrated in Figure 1.30b, which compares the GAIA 50 per cent detection probability curve with known systems.

The detection method used in the simulations applies a χ^2 test to the residuals after fitting the epoch measurements of each planetary system to a single-star model. The test was run on a set of 160 000 planetary systems, uniformly distributed on the sky, as function of period P and α, which are expected to be among the major contributors to detection efficiency. The simulated values ranged from 10 to 100 μas for α, and between 0.5 and 20 years for P. The remaining orbital elements were distributed randomly in the ranges: $0° \le i \le 90°$, $0 \le e < 1$, $0 \le \Omega$ (position angle of the ascending node) $\le 2\pi$, $0 \le \omega$ (argument of periastron) $\le 2\pi$, $0 \le \tau$ (periastron epoch) $\le P$. A value of 50 per cent is taken as the threshold for significant probability of planetary detection. Then, Jupiter-like planets (same orbital period of Jupiter) appear detectable up to a distance of 100 pc. Jupiter-mass planets with shorter periods (roughly between 2.5 and 8 years) can be detected up to 150 pc, and detectability is still effective, although for an even narrower range of periods (4-6 years) around the mission lifetime, at 200 pc.

Table 1.12 summarizes the estimated number of giant planets GAIA will detect, based on a standard Galaxy model and an effective planetary frequency of 1.3 per cent. Estimates by Colorado McEvoy (1999) put the number of detections of Jupiter-mass planets somewhere between 10 000–50 000, depending on details of the detection and orbital distribution hypotheses.

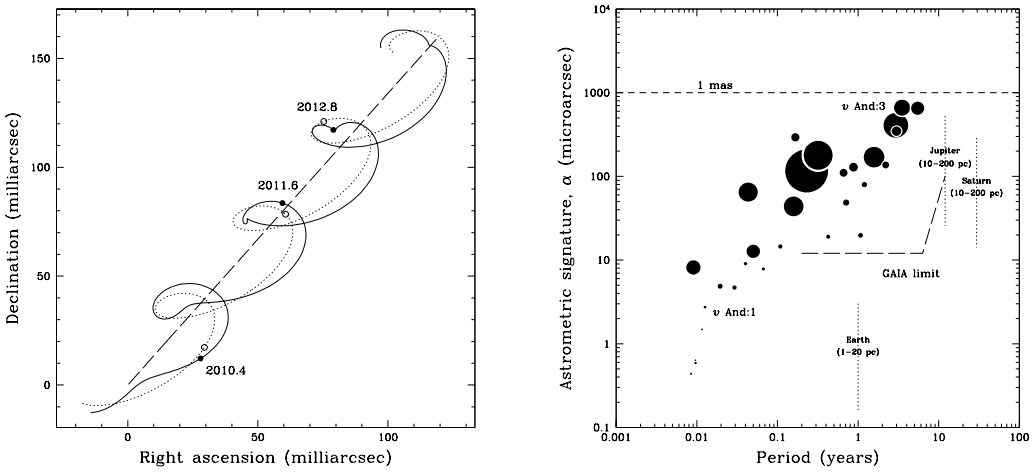


Figure 1.30: Left: the modelled path on the sky of a star at a distance of 50 pc, with a proper motion of 50 mas yr^{-1}, and orbited by a planet of $M_p = 15\,M_{\rm J}$, $e = 0.2$, and $a = 0.6$ AU. The straight dashed line shows the system's barycentric motion viewed from the Solar System barycentre. The dotted line shows the effect of parallax. The solid line shows the apparent motion as a result of the planet, the additional perturbation being magnified by ×30 for visibility. Labels indicate times in years. Right: astrometric signature, induced on the parent star for the known planetary systems as a function of orbital period (adapted from Lattanzi et al. 2000). Circle radii are proportional to $M_p \sin i$. Astrometry at the milliarcsec level has negligible power in detecting these systems, while the situation changes dramatically for microarcsec measurements.

To evaluate the number of potential planetary systems within GAIA's detection horizon one needs to know the stellar content in the Solar neighbourhood and the planetary frequency distribution. Star counts to 200 pc were obtained from simulated catalogues produced with the Besançon model of stellar population synthesis (Robin & Crézé 1986; Bienaymé et al. 1987) constrained to include the $V = 12$ mag limit. According to this Galaxy model, and for spectral types no later than K5, we should expect $\sim$ 66 000 stars within 100 pc from the Sun; the number increases to $\sim$ 223 000 and $\sim$ 529 000 stars, for distances of 150 pc and 200 pc, respectively. The contribution from early-type stars and giants is negligible.

To date, only coarse attempts at establishing planetary frequency have been made, which are based on the radial velocity programmes concluded so far. A recent estimate (Marcy et al. 2000) yields an integral planetary frequency $F_p \sim 4$ per cent ($\sim$300 stars surveyed) for giant planets (defined as having masses in the range 0.5-5 $M_{\rm J}$) orbiting within 3 AU from the parent star. Revising this estimate and extending it to wider ranges of orbital radii is one of the objectives of the current and future surveys. Nonetheless, we can utilize this estimate of F_p to predict the number of giant planets GAIA could reveal. Under the assumption that the actual planetary frequency is uniform with semi-major axis, F_p can be turned into an estimate of the planetary frequency (f_p) simply by dividing it by the limiting orbital distance of 3 AU; this yields the value $f_p \sim 1.3$ per cent per 1 AU bin. The lower limit to the number of detected giant planets N_d at a given distance d (in pc) predicted for GAIA is then obtained from the relation $N_d > 0.5\, f_p\, \Delta a\, N_\star$, where $N_\star$ is the total number of stars within a sphere of radius d centered on the Sun (for given limiting magnitude and spectral type), while the ranges (Δa) of semi-major axis were computed from converting the orbital period intervals generated by the intersections of the 50 per cent iso-probability curve with the Kepler's third law plots at different distances. The factor 0.5 explicitly accounts for the adopted 50 per cent detection probability threshold.

Measuring New Giant Planets Once planets are 'detected', the goal is to reliably determine their orbital characteristics and mass. In the simulations, both the planet orbital elements (P, τ, e , a, i, ω, Ω) and the standard astrometric parameters (λ, β, μ_λ, μ_β, π) are solved for by fitting a non-linear model to the epoch measurements.

In the experiments on orbital solutions (Figures 3–5 of Lattanzi et al. 2000), the distance horizon was fixed at 100 pc, as the 30 per cent convergence fraction for the orbital parameters a and i would quickly drop to values exceedingly low for long periods ($P = 11.8$ years). At 100 pc, the fraction of acceptable solutions reaches its lowest value of $\sim$ 40 per cent for the parameter a. Taking this as the probability value appropriate for the entire set of orbital parameters, a lower

Table 1.12: Number of giant planets that could be revealed by GAIA, as function of increasing distance. A uniform frequency distribution of 1.3 per cent planets per 1 AU bin is assumed.

d (pc)	$N_\star$	Δa (AU)	N_d
<100	∼66 000	1.3 - 5.3	$\gtrsim$ 1700
100–150	∼223 000	1.8 - 3.9	$\gtrsim$ 2100
150–200	∼529 000	2.5 - 3.3	$\gtrsim$ 1600

limit to the number of detected planets for which there will be reliable orbits follows from the data in Table 1.12. More than 680 of the ∼1700 detected planets will have accurate estimates of the orbital parameters. For $P = 11.8$ years the experiments are still useful for a rough calculation of what the number of measured planets would be at 200 pc. The extrapolation suggests that at least 50 per cent of the planets would have orbital parameters good to 30 per cent or better, adding more than 790 planets to the previous count, and a similar number of accurately measured planets should come from the 150 pc distance bin, bringing the total to over 2000 planets. The statistical value of such a sample would be instrumental for critical testing of theories on planet formation and evolution. Finally, one immediately realizes the uniqueness of the GAIA sample of measured planets, as its size is at least comparable to that of the observing lists of the largest ground based surveys, and to the size of the planet-finding programme which is reasonable to expect for SIM.

Once the orbital elements are estimated, planet mass can be derived from the usual expression defining the mass function of binary systems as applied to the case of a planet mass secondary, i.e.: $M_p \sim (a_s/\pi)(M_s^2/P^2)^{1/3}$, where M_p and M_s are in units of $M_\odot$, π in arcsec, P in years, and a_s (the semi-major axis of the parent star) in arcsec. This shows that the mass of a planet cannot be determined directly from the orbital parameters alone, but reliable independent estimates of the mass of the parent star must be provided by some other means (e.g., by spectroscopy).

The orbital solutions provide the means to calculate the convergence probability of each orbital parameter to a well defined accuracy level, i.e., to compute the percentage of stars for which the solution of each parameter converged within a given fraction of its true value. Convergence probabilities as a function of distance from the observer and within the 30 per cent accuracy level for a, P, i, and e were evaluated for systems with Sun-Jupiter masses and given orbital periods (0.5, 5, and 11.8 years). As expected, the 5 year-period case is particularly favourable, due to the optimal sampling of the orbit. The 11.8 year period, although it corresponds to a larger signal (α increases with period), suffers from poorer sampling of the orbit during the mission duration, while in the short-period case (0.5 years) both the smaller signal amplitude and the generally non-optimal timing of the observations contribute to worsen the accuracy of the fit. Also, the statistics of the solutions show that different accuracy levels are to be expected for different orbital parameters. Orbital period and eccentricity come out to be easier to measure, compared to semi-major axis and inclination. Finally, the results obtained allow us to estimate a lower bound, consistent with the assumptions above, to the fraction of giant planets detected by GAIA for which it would be possible to derive orbital elements within well defined accuracy levels.

1.6.4 Formation and Stability of Multi-Planet Systems

Establishing the frequency of multi-planet systems is important, as it is not known at present whether a planetary system like ours, which presents two very different and separated classes of planets (the terrestrial planets and the large planets), is generic or not. The possible discovery of several multi-planets system will also provide an understanding of whether the apparent regularity of the spacing of the planets in the Solar system (usually denoted as Bode's law) is a common feature of planetary systems. This is important, as it seems that the scaling law presented by the planetary distribution can be related to the initial distribution of matter in the protoplanetary disc, and to their dynamical evolution.

The presence of Jupiter-like planets close to the central star is still not fully accounted by current formation models, which invoke a migration of the planets initially formed in a different place (see Ward 1997). A different solution would be systems with many planets, spaced regularly, but with no small inner planets. In this case, a large planet close to the central star could result from dynamical interaction of planetesimals. Due to the improvement of computer facilities, numerical simulations of the formation of planetary systems are evolving rapidly at present, also stimulated by the discovery of the first extra-solar planets. In particular, it will be soon possible to make realistic numerical simulations of the planetary accretion with a number of bodies close to reality, filling the present gap between a purely hydrodynamics approach, valid in the very early stage of the planet formation, and the dynamics of the already formed planets which is now quite well understood since the last ten years. This will certainly continue in the next decades, and the observations of GAIA will provide challenging constraints on these theoretical models.

The knowledge of the parameters of the main planets of a multi-planet system is very important, as they affect dynamically all the remaining bodies by gravitational interactions. When these planets are known, it is possible to conduct stability analysis of the possible orbits for other smaller bodies in the system. In this way, it will be possible to assess the existence of stable zones for Earth-like planets. By contrast, if the habitable zone is largely chaotic due to the presence of the large planets, this will dismiss the possibility of finding a planet similar to the Earth in the corresponding planetary system. GAIA's survey will probably find a number of such multi-planet systems thus identifying possible locations of Earth-like planets. These data would also be extremely useful in the search for terrestrial planets by other means, especially with future transit observations.

Further measurements of υ And has revealed two additional planets in orbit around the star (Marcy et al. 1999) making it the first multiple planetary system (as opposed to single planet) discovered around a main-sequence star other than our Sun. In addition to the $0.6\,M_{\rm J}$ object in a 4.6-day orbit, the two more distant planets have $M_p \sin i$ of 2.0 and $4.1\,M_{\rm J}$, a of 0.82 and 2.5 AU, and large e of 0.23 and 0.36 respectively, as do all known planets in distant orbits. Mazeh et al. (1999) were able to derive a mass estimate for the outer companion of $10.1^{+4.7}_{-4.6}\,M_{\rm J}$ using Hipparcos astrometry, compared to $M_p \sin i = 4.1\,M_{\rm J}$, implying an orbital inclination of 156°. If the three planets all have the same inclination, masses of the inner two planets of 1.8 ± 0.8 and 4.9 ± 2.3 $M_{\rm J}$ would follow. A large difference between the inclinations of the outer two planets appears to be ruled out by dynamical stability arguments (Holman et al. 1997; Krymolowski & Mazeh 1999). A number of numerical studies have been carried out to establish whether the system is stable over a long timescale (Laughlin & Adams 1999; Lissauer 1999; Rivera & Lissauer 2000). All studies indicate that the stability depends strongly on the planetary masses, since the orbits of the second and third planets brings their relative separation close to their Hill radius, a proximity that makes their motion chaotic.

1.6.5 Transits from Photometric Data

The first detection of an extra-solar planetary transit has been reported by Charbonneau et al. (2000) for HD 209458, leading to a determination of the mass, radius, and hence the density of this planet. The transit signal was subsequently found in the Hipparcos epoch photometry data (Robichon & Arenou 2000). Further studies will be required to establish how many such transits will be contained in the GAIA photometric data base, but the number may of order several thousand.

Recent detections of reflected light give a further indication of the type of follow-up observations that will be possible once new candidate planets can be identified in large numbers. The physical interpretation of these extra-solar planets is also now just beginning. For τ Boo Charbonneau et al. (1999) placed an upper limit on reflected light, and further work by Collier Cameron et al. (1999) achieved a possible detection, providing a value for i, and hence an estimate of $M - p \sim 8\,M_{\rm J}$, twice the minimum value for an edge-on orbit. Assuming a Jupiter-like albedo $p = 0.55$ yields $R_p \sim 1.6 - 1.8\,R_{\rm J}$, slightly larger than the structural and evolutionary predictions of Guillot et al. (1997) of $\sim 1.4 - 1.1\,R_{\rm J}$ at ages of 2–3 Gyr.

1.7 Solar System

Solar system objects present a challenge to GAIA because of their significant proper motions, but they promise a rich scientific reward. The minor bodies provide a record of the conditions in the proto-Solar nebula, and their properties therefore shed light on the formation of planetary systems.

1.7.1 Minor Planets

The population of relatively small bodies located in the main asteroid belt between Mars and Jupiter should have experienced limited thermal evolution since the early epochs of planetary accretion. For this reason, minor planets should be representative of the original population of planetesimals that formed in that region of the Solar system. Due to the non-negligible radial extent of the main belt, minor planets provide important information about the gradient of mineralogical composition of the early planetesimals as a function of heliocentric distance. It is therefore important for any study of the origin and evolution of the Solar system to investigate the main physical properties of asteroids including masses, densities, sizes, shapes, and taxonomic classes, all as a function of location in the main belt and in the Trojan clouds.

Masses The possibility of determining asteroid masses relies on the capability of measuring the tiny gravitational perturbations that asteroids experience in case of a mutual close approach. Hipparcos data have been used successfully for such mass determinations (e.g. Bange 1998; Viateau & Rapaport 1998), although at present only about 10 asteroid masses are known, mostly with quite poor accuracy. GAIA will enable a large number of asteroid masses to be determined accurately.

The number of mutual close encounters between asteroids over the years 2004–2015 has been estimated. For each event, the minimum distance and relative velocity have been computed. The predicted deflection has been computed whenever the minimal distance is lower than 0.25 AU, and/or the encounter velocity is smaller than 2 km s^{-1}. It is assumed that a close encounter is effective from the point of view of mass determination when the computed deflection is greater than 10 mas. As a preliminary consideration, a set of 174 perturbing asteroids, having diameters larger than 120 km, has been taken into account. As a set of possible targets (deflected bodies) all the known objects larger than 50 km (729 asteroids) were chosen. The results of the computations are that 136/174 perturbing asteroids are involved in 849 effective encounters. The parameters considered in this analysis were conservative, since the expected S/N ratio were larger than 100. Accordingly, it is expected that GAIA should be able to provide mass determinations for more than 100 asteroids during the mission lifetime.

Sizes and Densities Density can be determined when both mass and size (and shape) of a given object are known. Most asteroids are too small for a direct size measurement from the ground. Alternative techniques based on stellar occultations have provided only a few reliable measurements, due to the intrinsic difficulties in predicting and observing these events. With a resolving power around 20 mas for extended sources, GAIA should be able to directly measure over 1000 asteroid diameters. In addition to main belt objects, 60–90 Jupiter Trojans will also be measured. Measurement of the overall size distribution of the asteroid population is essential in order to understand the general process of collisional evolution of the belt.

These observations will not only constitute the first large catalogue of reliable asteroid sizes, but they also will allow calibration and extension of the major data base on asteroid sizes, namely the IRAS Minor Planet Survey (IMPS), which lists size and albedo data for about 2000 objects with $V \gtrsim 16$ mag, derived by a radiometric technique which uses a poorly constrained thermal model to predict the distribution of temperature on the asteroid surface. Although speckle-interferometry from the ground with Keck/VLT-class telescopes may succeed in giving better results in terms of angular resolution, GAIA will still furnish a data set of exquisite homogeneity and completeness.

Each asteroid will be observed several times at different geometrical configurations with respect to the rotational axis of the target. Due to the expected shape irregularities, this will lead to different measured angular diameters, depending on the apparent area as seen from the satellite. In this way it will be possible to derive also shape estimates. The knowledge of shape, mainly for large asteroids, is important for testing the hypothesis of the existence of triaxial equilibrium bodies as a consequence of fragment re-accumulation after collisional break-up. When combined with independent mass determinations, size determinations will provide reliable values of mean density. This is a crucial parameter, being related to the bulk mineralogic composition, and to the overall physical structure of the bodies. In particular, if asteroids are, at least in some size range, assemblages of different blocks hold together by self gravitation ('piles of rubble') a significant fraction of their volume should consist of empty interstices, leading to fairly small density values.

The GAIA observations will also allow identification of binary objects. This is important both for the possible existence of quasi-contact binaries with a mass ratio of the order of one (equilibrium double ellipsoids), and for quite separated systems (post break-up binaries, like the Ida-Dactyl system).

Albedos and Taxonomy Direct measurements of sizes, combined with simultaneous measurements of the apparent magnitude at visible wavelengths, will provide the surface albedo for ∼1000 objects. Albedo is directly related to the mineralogic characteristics of surface layers. Asteroid albedos range mostly between 0.03 and 0.5. Higher albedo objects are more common in the inner part of the main belt, while lower albedo asteroids tend to be located in the outer regions. This trend continues outwards, and low albedos are exhibited also by Trojans, comet nuclei, and Kuiper-Belt objects.

Albedo is a useful complement to spectrophotometric data for the definition of different taxonomic classes. The GAIA photometry will be much more reliable than most data presently available. The colour indices will provide a taxonomic classification for the whole sample of observed asteroids. The sample will extend down to objects of small sizes, and will provide the distribution of taxonomic classes as a function of size. This is relevant for the origin of ordinary chondrites and for the effects of surface weathering due to Solar wind, cosmic rays and micro-impacts.

It is not clear whether the differences among the known taxonomic classes, which are usually interpreted in terms of differences in mineralogic compositions, correspond also to differences in mean density. This would be the case, if taxonomy really deals with differences in overall composition and if the internal structures, in particular the presence of empty interstices, is not the most important factor in determining density. Of course, if a well-defined relationship between density and taxonomy can be evidenced, this would allow us to make reliable density estimates for objects of a known taxonomic characterization. This would allow a quantitative comparison with densities of meteorites, which would shed light on the genetic relationship between different classes of meteorites and their supposed asteroid sources (a typical example is the debated relationship between S-type asteroids and ordinary chondrites).

Orbits The impressive astrometric capabilities of GAIA will be exploited, both for direct asteroid orbit determinations, and through the derivation of a dense astrometric catalogue of stars, which will be important for many purposes in Solar system studies (e.g., Sections 1.7.3 and 1.9). In particular, this will make it possible to predict stellar occultations by small Solar system bodies with sufficient accuracy for developing efficient observing programmes, which should lead to further measurements of asteroid sizes with very high precision.

Preliminary simulations have been performed in which the covariance matrix of the orbital elements of more than 6000 asteroids were computed using both the whole set of astrometric observations collected from ground-based telescopes since 1895 through 1995, as well as a set of simulated observations carried out by GAIA, computed by considering a 5 year lifetime of the mission. Another set of simulated ground-based observations covering 1996–2015 were also performed, according to the expected results of present activity and planned observational programmes. The results show that for the known asteroids the predicted ephemeris errors based on the GAIA observations alone 100 years after the end of the mission are a factor larger than 30 better than the predicted ephemeris errors corresponding to the whole set of past and future ground-based observations.

New Objects and Near-Earth Asteroids In addition to known asteroids, GAIA will discover a very large number, of the order of 10^5 or 10^6 (depending on the uncertainties on the extrapolations of the known population) new objects. It should be possible to derive precise orbits for many of the newly discovered objects, since each of them will be necessarily observed many times during the mission lifetime. These will include a large number of Near-Earth Asteroids. The combination of on-board detection, faint limiting magnitude, observations at small Sun-aspect angles, high accuracy in the instantaneous angular velocity (0.25 mas/s), and confirmation from successive field transits, means that GAIA will provide a detailed census of Atens, Apollos and Amors, down to diameters of about 260–590 m at 1 AU (Perryman 2000; Zappalà & Cellino 2000).

1.7.2 Trojans in the Inner Solar System

There have been sporadic searches for the Trojans of the Earth and Mars before. With the exception of the two Martian Trojans, 5261 Eureka and 1998 VF31 (e.g., Tabachnik & Evans 1999), these searches have been without success. Existing surveys of the terrestrial Lagrange points rule out only large (100 km) sized objects. GAIA will observe to within 35° of the Sun, so identification of asteroidal companions of Mars, the Earth and even Venus will be possible. This will be the first time that the Venusian Lagrange points will have been scanned.

The Lagrange points of all the terrestrial planets can maintain long-lived orbits at least over timescales of 100 million years (Evans & Tabachnik 2000). Re-simulation of these long-lived orbits for 1 million years and sampling every 2.5 years provides synthetic observations of asteroidal clouds for the terrestrial planets, which allow identification of the likely places of greatest concentration of co-orbiting asteroids, as well as their typical proper motions and magnitude adjustments. Such synthetic observations are presented in the panels of Figure 1.31 for Venus (figures for the other planets are given in Evans & Tabachnik 2000).

GAIA is ideal to look for these objects because of the enormous area of sky that must be searched. Co-orbiting satellites like Trojans librate about the Lagrange points, but the amplitude of libration can be very large. Figure 1.31(c) shows the distribution as a function of proper motions in the plane of, and perpendicular to, the ecliptic. The average velocity in the plane of the ecliptic is 238 arcsec/hr with a full-width half maximum of 28 arcsec/hr. Finally, Figure 1.31(d) shows the distribution of magnitude adjustment versus geocentric longitude. The brightest objects occur

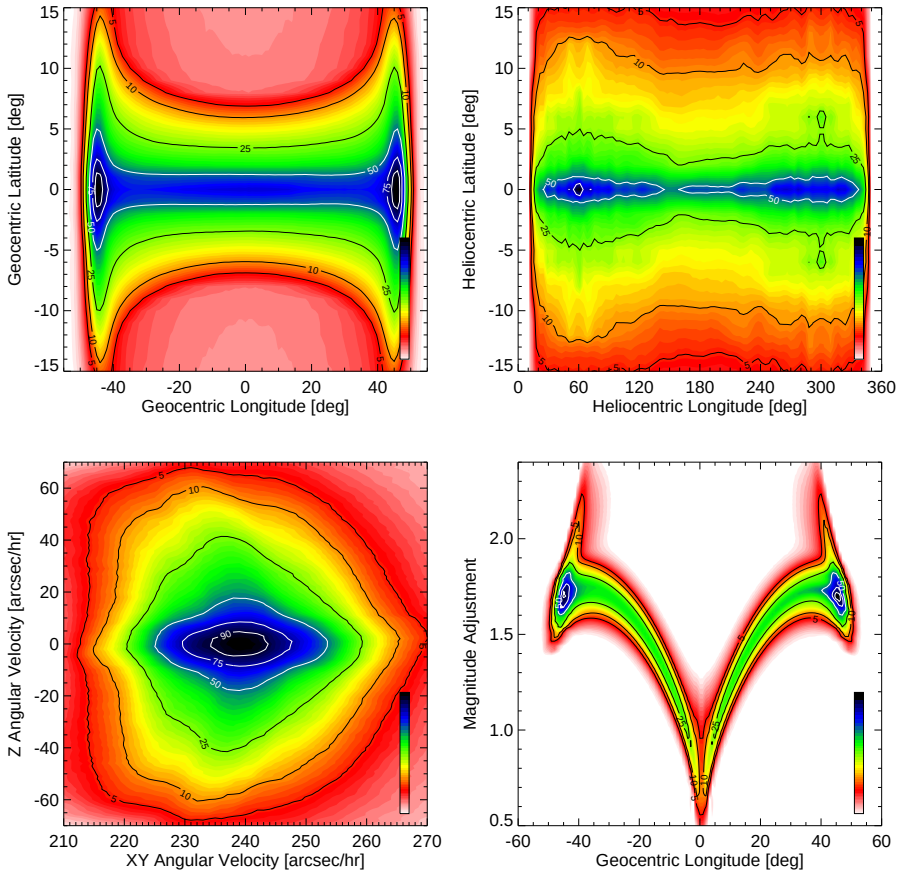


Figure 1.31: Predictions of the expectation of co-orbiting asteroids: (a) in the plane of geocentric ecliptic latitude and longitude; (b) heliocentric ecliptic latitude and longitude; (c) the distribution as a function of proper motions in the plane of, and perpendicular to, the ecliptic; (d) the distribution of magnitude adjustment versus geocentric longitude.

close to the Sun. These are the asteroids on horseshoe orbits at superior conjunction. Even though they are furtherest away from the Earth, this is outweighed by the effects of the almost zero phase angle. The broadest range of magnitude adjustments occurs at the greatest eastern and western elongations $\sim \pm 45°$. Here, the phase angle changes quickly for small changes in the longitude. This portion of the sky is accessible to GAIA. The average value of the magnitude adjustment here is 1.7, i.e., these objects are typically 1.7 mag fainter than their absolute magnitude (at zero phase angle and at unit heliocentric and geocentric distance). For the same distribution of magnitudes, Venusian Trojans are on average brighter than terrestrial or Martian Trojans.

1.7.3 Trans-Neptunian Objects: the Kuiper Belt

The old view of a vast region of empty space extending from Pluto (∼40 AU) to the Oort Cloud (∼10 000 AU) has been conclusively replaced by a picture of a volume richly populated by unexplored new worlds. Ground-based surveys in the past few years have discovered over 100 icy bodies beyond Neptune, members of a population called the 'Kuiper Belt'. Kuiper Belt bodies are related to a wide range of outer Solar system bodies, such as the short-period comets, the Neptunian satellite Triton, and the Pluto-Charon system — indeed, Pluto is now recognized as the largest known member of the Kuiper Belt. The connection between Kuiper Belt objects and so many different Solar system bodies hints at a common origin in the outer Solar system — a

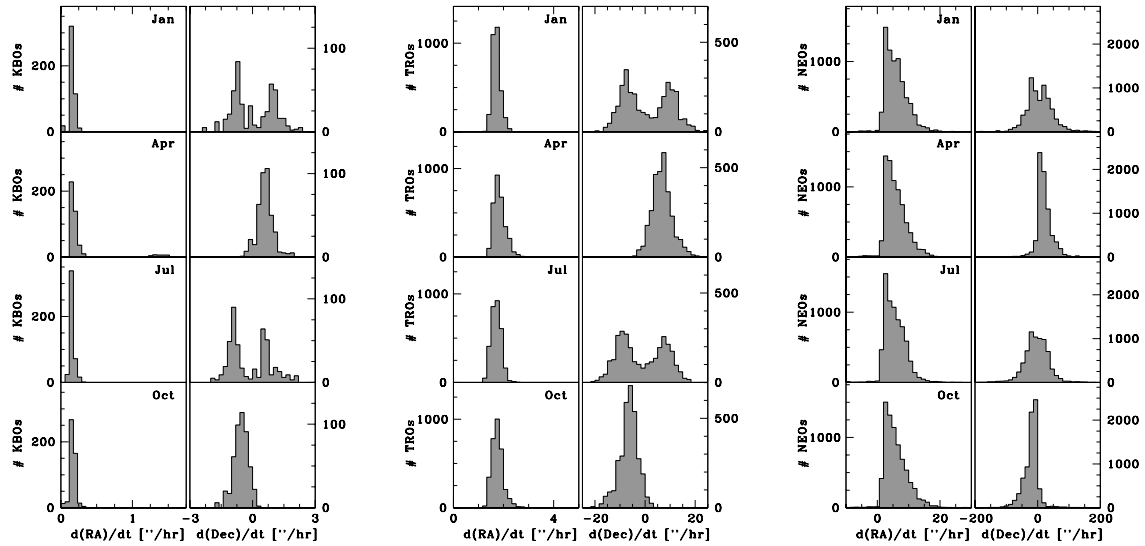


Figure 1.32: Mean angular motions of (a) Kuiper Belt objects, (b) Trojans, and (c) Near–Earth asteroids.

remarkable hypothesis considering the diverse characteristics of these bodies. The Kuiper Belt is also our closest link to the circumstellar disks found around other main sequence stars, and an understanding of the physical processes operative in the Belt (both now and in its early stage) will mark a key step forward in understanding the problem of planetary formation.

The known Kuiper Belt objects possess a highly non-uniform distribution of orbital elements, falling into three distinct dynamical classes: the Resonant KBOs, the 'classical' KBOs, and the 'scattered' KBOs. Each dynamical class provides a different window on the physical processes that shaped the early Solar system.

The Resonant KBOs are objects which lie in or near the mean motion resonances of Neptune. Some have also been found in the 4:3 resonance at 36.4 AU and the 5:3 resonance at 42.3 AU. However, the 3:2 resonance at 39 AU (which counts as its members Pluto and its cousins, the 'Plutinos') is by far the most populated, contributing to $\sim 1/3$ of the known sample. Bias-corrected statistics puts the true Plutino fraction closer to 10–15 per cent.

The Classical KBOs are roughly defined as those with semi-major axes $a \geq 42$ AU and perihelia $q > 35$ AU. Their eccentricities are modest, but their inclinations can be considerable (largest value known $\sim 32°$). Numerical integrations indicate that their orbits are stable for the age of the Solar system, suggesting that they formed *in situ* and have survived ejection from the Belt until now.

The Scattered KBOs are characterized by large, highly eccentric and highly inclined orbits. The only known prototype of this population is the object 1996 TL_{66}, with $a = 85$ AU, eccentricity $e = 0.59$, corresponding to an aphelion at 135 AU. The origin of the scattered objects is unknown but may be related to the swarm of bodies scattered outwards by Neptune at the late stage of its growth. Discovery statistics suggest ~ 6000 scattered KBOs (500 km diameter or larger), although this estimate is highly uncertain due to the sample of 1. These objects are hard to detect because their large, eccentric orbits render them invisible except when near perihelion. In absolute numbers the scattered Kuiper Belt objects may dominate the trans-Neptunian region.

The Kuiper Belt is estimated to contain $\sim 0.1 M_E$ inside the 30–50 AU region, and perhaps a few $\times 0.1 M_E$ exist out to a couple of hundred AUs. Dynamical studies suggest that the Kuiper Belt most likely formed *in situ* in the trans-Neptunian region, with its current spatial distribution resulting from dynamical erosion by the planets. Recent models of accretion in the Belt support this view (e.g., Stern & Colwell 1997; Kenyon & Luu 1998; Kenyon & Luu 1999).

Although the Kuiper Belt is the subject of intensive ground-based observations, there are questions that can be only resolved with an all-sky survey such as GAIA. GAIA will detect a significant number of Kuiper Belt objects during its 5-year mission. The angular motion of a typical object at $\sim 90°$ elongation (where GAIA will be looking) is small: the known KBOs have $d\alpha/dt = 0.02 - 1.0$ arcsec hr^{-1} and $d\delta/dt = 0.002 - 1.2$ arcsec hr^{-1}. The surface density of the Kuiper Belt at $V = 20$ mag is 8×10^{-3} objects per square degree (2×10^{-2} at $V = 21$ mag — see Figure 1.33),

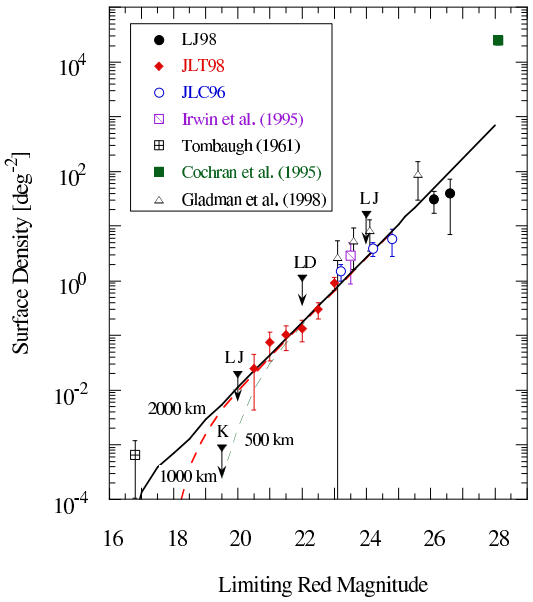


Figure 1.33: Surface density of Kuiper Belt objects as a function of R magnitude, based on the objects discovered before 1999.

implying that GAIA should discover $\sim$ 300 KBOs with $V \leq 20$ ($\sim$ 800 KBOs with $V \leq 21$). However, these numbers are most likely upper limits, since the surface density was calculated from surveys centred at the ecliptic, and may not be representative for the entire sky. It is likely that ground-based surveys would yield a comparable or larger number of Kuiper Belt objects by 2015, but the GAIA-produced sample will be unique because (i) it will represent the complete sample of KBOs to $V = 20$ mag, free of all directional observational bias, and (ii) it will contain all the brightest KBOs, i.e., the best targets for physical studies. Scientific objectives regarding the Kuiper Belt that can be answered only with GAIA are outlined below.

Binaries An exciting prospect offered by GAIA is to search for binaries in the Kuiper Belt. Is Pluto-Charon the only binary in the Kuiper Belt, or are binaries common?

Catastrophic collisions should have been common in the young, more massive Kuiper Belt and may have produced many binaries. Binary KBO systems may manifest themselves as a 'wobble' in the primary KBO, due to the tugging effect of the companion. Depending on its amplitude, this wobble may be detectable by GAIA. For example, GAIA should produce the best measurement yet of the mass ratio of Pluto and Charon. Observations from the ground and from the Hubble Space Telescope have yielded different mass ratios, but GAIA, with its much superior astrometric capability, should provide the definitive answer.

Orbits GAIA will also provide accurate orbits for KBOs, essential for understanding the dynamics of the Kuiper Belt. Ground-based astrometry of Kuiper Belt objects is presently good to $\sim$ 0.5 arcsec, rendering it practically impossible to determine accurate orbits on a few year timescale (one complete orbit at 40 AU takes about 250 yrs).

A good example of the crucial need for good orbital elements is the determination of the populations of the resonances. It has been suggested that the mean motion resonances became populated when the giant planets migrated during the late stages of their growth. As Neptune migrated outward, the resonances must have swept through the protoplanetary disk and thus captured objects. Trapping was particularly efficient in the 3:2 and 2:1 resonances, and the relative populations in these two resonances depend on how fast Neptune migrated (Malhotra 1995). The measurement of the population ratios may therefore allow an estimate of the rate at which Neptune's orbit expanded. This important constraint on early planetary dynamics was previously thought to be out of the realm of direct observation, but can be calculated in the near future with the help of GAIA.

Sizes With an angular size of ~ 0.1 arcsec, even Pluto is barely resolvable by GAIA's resolving power of around 20 mas, and hence most of the known KBOs (the largest angular diameter is 20 mas) will not be resolvable by GAIA. However, GAIA will be a useful tool in determining the sizes of KBOs — from the ground. It will do so by providing (i) very accurate astrometry for detected KBOs, and (ii) a very high quality reference star catalogue.

With the current number of known KBOs, the expected rate of occultation events of stars with magnitude $V \leq 21$ mag that is visible from an 8-m class telescope is only ~ 3 per year. However, at the present time, there is no astrometric star catalogue that approaches $V \leq 20$ mag. GAIA will provide such a star catalogue, as well as increase the number of large KBOs suitable for occultation observations.

Bright (large) KBOs do not lend themselves well to detection in ground-based CCD surveys, since they are rare and even the new large-format CCD arrays have hardly the field-of-view to produce them in sufficient numbers. As a result, KBO statistics at $R \leq 20$ (corresponding roughly to $V \leq 20.5$) are poorly constrained. The only data (upper limits) in this region come from photographic surveys, which are usually less reliable than CCD observations (see Figure 10 of Jewitt et al. 1998). Consequently, the maximum radius of KBOs is far from well known, yet it is a key constraint on any accretion model of the Kuiper Belt. With its all-sky survey, GAIA is in a unique position to unambiguously determine the KBO size distribution at large sizes.

Other Plutos Probably the most exciting question that GAIA can answer regarding the outer Solar system is: have other Pluto-size bodies escaped detection? It is clear that accretion in the outer Solar system is capable of producing 1000 km size bodies such as Pluto and Triton, and there is reason to believe that other undetected Plutos might exist in the outer Solar system. Accretion models of the Kuiper Belt predict the formation of 1–10 Plutos (Kenyon & Luu 1998). Furthermore, the most stringent limit on the mass in the outer Solar system is currently $\sim 1 M_E$ inside 100 AU, or 500 Pluto-masses. These findings do not by themselves prove that other Plutos exist, but nor do they rule out the existence of such bodies. If other Pluto-sized objects exist, GAIA is the best instrument with which to find them.

The most extensive survey of the outer Solar system was completed by Clyde Tombaugh who covered 1530 square degrees, and discovered Pluto (Tombaugh 1961). Although it was a major achievement, Tombaugh's survey had several limitations: (1) its limiting magnitude is only $V = 15$ mag, meaning that Pluto would have escaped detection if it were only slightly further away (e.g. > 42 AU); (2) photographic plates are very difficult to calibrate accurately due to their trailing loss and nonlinearity; and (3) the survey was restricted to within 15° of the ecliptic. It would not have been difficult to hide other 'Plutos' from Tombaugh's survey if they were at higher inclinations or further out (see, e.g., the highly elliptical orbit of 1996 TL_{66}).

1.7.4 Stellar Encounters and Perturbations of the Oort Cloud

Close or even penetrating passages of stars through the Oort Cloud can in principle deflect large numbers of comets into the inner planetary region, initiating Earth-crossing cometary showers and possible Earth impacts. Although the distribution of long-period cometary aphelia is largely isotropic, some non-random clusters of orbits do exist, and it has been suggested that these groupings record the tracks of recent stellar passages, with dynamical models suggesting typical decay times of around 2–3 Myr. GAIA's complete and accurate census of the distribution and space motions of the stars in the Solar neigbourhood (see Section 1.3.4) will allow a determination of the frequency of such close encounters, and will provide, for the first time, sufficiently accurate astrometric data for a large number of stars to carry out a reliable study of the link between comet showers and past impact events and mass extinctions on Earth.

García-Sánchez et al. (1999) used Hipparcos data to investigate close stellar encounters with the Solar system, the consequences for cometary cloud dynamics, and the evolution of the comet population over the history of the Solar System. Effects of individual star passages on comet orbits were studied through dynamical simulations. Algol was the largest perturber in the recent past (although other stars have passed even closer), passing at a distance of about 2.5 pc about 7 Myr ago. Gliese 710 is the most significant known future perturber. At 19 pc from the Sun and approaching at about 14 km s^{-1}, it will pass through the Oort Cloud, at about 69 000 AU from the Sun, in about 1 Myr. But García-Sánchez et al. (1999) concluded that none of the predicted passages could have caused a significant disruption of the Oort Cloud, which supports the hypothesis that the currently observed flux of long-period comets corresponds to a steady-state value.

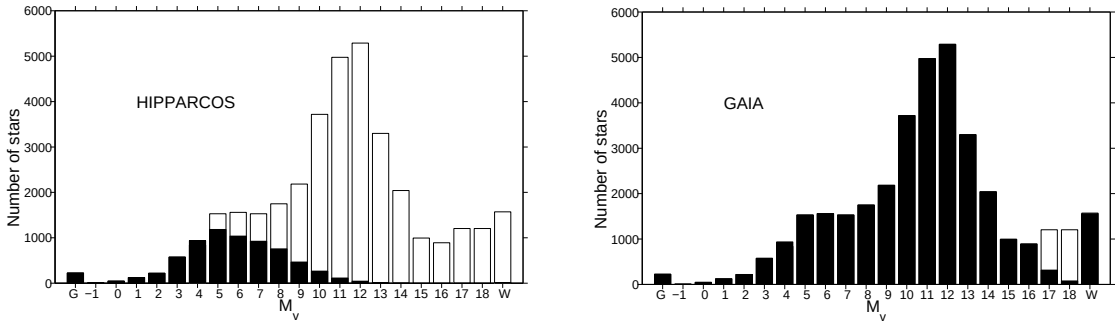


Figure 1.34: Comparison of the completeness from Hipparcos (left) with the expected completeness from GAIA (right). The plots show the number of star systems (individual stars or binary/multiple systems) within 50 pc of the Sun, as a function of absolute magnitude M_v, predicted by a systemic luminosity function based on data from the literature. The bar with the letter 'G' (at left) denotes giant stars, and with the letter 'W' (at right) denotes white dwarfs. The black part of the bars gives the number of star systems detected by Hipparcos or by GAIA (courtesy Joan García Sánchez).

Figure 1.34 shows the number of star systems (individual stars or binary/multiple systems) within 50 pc of the Sun, as a function of the absolute magnitude M_V. The black part of the bars gives the number of star systems detected by Hipparcos (top plot) or expected for GAIA (bottom plot). 'G' denotes giant stars and 'W' indicate white dwarfs. Hipparcos detected about 20 per cent of the nearby star systems, whereas GAIA will detect nearly all of them.

Two explanations for an increased rate of impact events on Earth have been suggested: (i) a collisional breakup of a large asteroid in the asteroid belt that can deliver collision fragments to orbital resonances, resulting in large fragments ejected from the asteroid belt to Earth-crossing orbits; (ii) a comet shower caused by a close stellar passage, increasing significantly the number of comets with Earth-crossing orbits. The reliable determination of a close stellar encounter with the Solar system during the time of the impact events would provide strong support to the cometary origin of such impacts, as opposed to the asteroid hypothesis. In particular, an extinction at the end of the Eocene period, 36 Myr ago, is identified with several large impact craters, multiple iridium layers, and other evidence of a prolonged period of increased cometary flux in the inner planets region. Hipparcos data allowed the study of passages within a few million years. GAIA will enhance this time interval to a geologicaly interesting interval.

The encounters predicted by using GAIA data are expected to establish whether the currently observed comet flux corresponds to an enhanced or a steady-state flux, with implications for the size of the Oort Cloud population. The prediction of future close or penetrating passages through the Oort Cloud may be used to estimate resulting enhancements in the inner Solar system cometary flux.

1.8 Galaxies, Quasars, and the Reference Frame

GAIA will not only provide a representative census of the stars throughout the Milky Way, but it will also make unique contributions to extragalactic astronomy. These include the structure, dynamics and stellar populations in the Magellanic Clouds, M31 and M33, the space motions of Local Group galaxies, and studies of supernovae, galactic nuclei, and quasars.

1.8.1 The Magellanic Clouds

The Magellanic Clouds are substantial galaxies in their own right, which provide the nearest examples of young intermediate-to-low chemical abundance stellar populations for study. The LMC and SMC will provide millions of stars for GAIA analyses, as illustrated in Figure 1.35. The key scientific questions for GAIA involve the dynamics of the LMC-Galaxy and the LMC-SMC interactions, the luminosity calibration of stellar populations, the dynamics of star forming regions, and the dynamical structure of the LMC 'bar'. At the LMC distance of 50 kpc (parallax 20 μas)

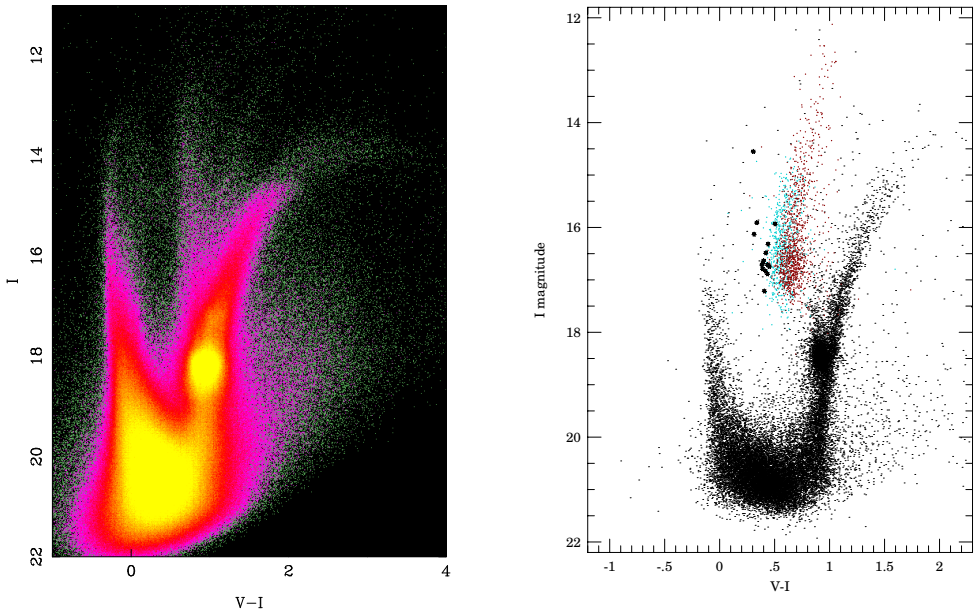


Figure 1.35: Left: A colour-magnitude diagram for the centre of the LMC, from Zaritsky et al. (1997). This is for a field of about 4×4 deg^2, centred 3° north of the bar. It contains 2.4 million stars brighter than $I = 20$ mag (data provided by the Magellanic Cloud Photometric Survey, courtesy of Dennis Zaritsky). Right: Colour-magnitude diagram for an area of 14×57 arcmin^2 in the SMC bar. There are 45 500 stars with I<20 mag. Overplotted are the Cepheids from OGLE, with fundamental, first overtone, and single-mode second overtone indicated separately by colour (from the OGLE consortium, courtesy of Andrzej Udalski).

individual bright stars, with $I = 12 - 16$ mag, will have transverse velocities determined to $\sim 1 - 2$ km s^{-1} (≈ 20 μas yr^{-1}). These stars are primarily OB stars and Cepheids, for which precise radial velocities are difficult to obtain. However, GAIA will allow kinematic mapping and membership analyses of young star forming regions in the LMC and SMC with comparable precision to that presently available in the Milky Way. That is, it will be possible to compare directly the kinematics and structure of star forming regions in a large spiral disk with those in a mid-sized irregular galaxy.

The dynamical evolution of the Solar neighbourhood is dominated by diffusion of stars in velocity space, crudely described as an age-velocity dispersion relation (Section 1.2.3). This process is not well-understood, but presumably involves energy input from spiral arms and molecular clouds. The GAIA kinematics in the LMC and SMC will quantify the age-kinematics relation in a different environment, further constraining the key dynamical processes.

One of the major puzzles in the structure of the LMC and the SMC is their very asymmetric luminosity distribution. While the large-scale, radially-averaged luminosity profiles of both galaxies follow fairly smooth exponentials, both show significant bar-like asymmetries. This is most obvious in the LMC, and in stars of ages less than a few Gyr old. However, the LMC 'bar' is substantially offset from the dynamical centre, and seems unrelated to the stellar dynamical $m = 2$ modes of cold disks. It appears to be sufficiently long-lived to have survived differential rotation for several rotation periods. It is presently unknown what the dynamical status of the bar is, or even if it is in the same plane as the main LMC disk. The possibility that it projects in front of the disk is attractive, as it would provide a natural explanation for the microlensing optical depth from normal stars. GAIA will provide not only three-dimensional dynamics across the whole bar and disk region, quantifying the dynamical relationship between these features, but also excellent relative distances. While an individual parallax to an LMC star will be imprecise (20 per cent error), the very large number of targets will map the spatial structure of the LMC/SMC system with high spatial resolution directly. Additionally, for the LMC disk stars, rotational parallax distance determination will be possible with GAIA (see Section 1.8.3).

The masses of the LMC and SMC are very poorly known. Current analyses involve approximate

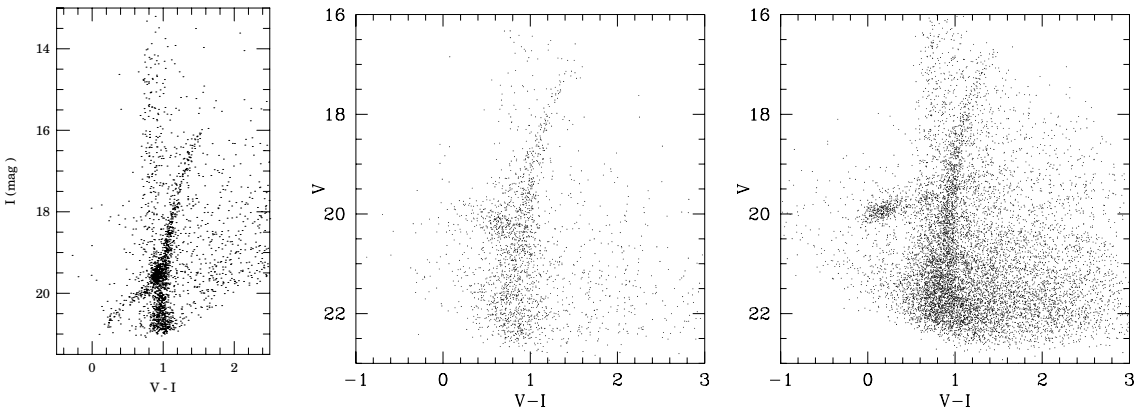


Figure 1.36: Colour-magnitude diagrams for (left) the central regions of the Carina dwarf. This field covers 14×14 arcmin^2, and has 2210 stars brighter than I = 20 mag; (middle) the central 8 arcmin radius of the Draco dSph galaxy; (right) the Ursa Minor dSph galaxy. GAIA will observe a few thousand stars in each system, and will get very accurate data on the brightest.

solutions fitting the poorly known transverse velocity, and assuming simple disk structure, for a small number of test particles. GAIA proper motions will map the membership of the Clouds as far as they extend, including the 'interCloud' regions of young metal-poor star formation, the complex SMC 'wing', and stars associated with the HI Magellanic Stream (Mathewson et al. 1974; Putman et al. 1998; Majewski 1999). This will map the dark halo structures of both the intact LMC and the apparently distorted SMC, determining the extent of their halos, the density of the Milky Way at 50 kpc, and the effects of the LMC-SMC interaction.

The gravitational pull of the LMC distorts the halo of our own Galaxy, producing a coherent dipole and quadrupole signature. N-body simulations show that the peak distortion occurs at a galactocentric radius of about 25 kpc, and produces a coherent mean flow field with amplitude of $10 - 30$ km s^{-1} (Weinberg 2000). If a sufficient number of stellar tracers can be found at these large distances (see Section 1.2.6), GAIA can measure these distortions.

The LMC also induces a significant dipole (i.e., lopsided) distortion of the outer disk (Weinberg 1999). The effect may be at the level of $3 - 5$ km s^{-1}, which will be hard to measure, but might be possible using distant carbon stars. These low-order distortions are ubiquitous in other galaxies, and seem to be generic for halo-disk interaction. Confirming or refuting the effect in our own Galaxy by a direct measurement would be very interesting.

1.8.2 Internal Dynamics of the Dwarf Satellites of the Milky Way

There are eight known dwarf satellite galaxies beyond the Magellanic Clouds, at distances up to 230 kpc. These provide key dynamical tracers of the outer mass distribution of the Milky Way, at larger distances than any other available tracer (Section 1.2.8). For the nearer dwarfs, especially Ursa Minor, GAIA will allow internal dynamical studies. The internal velocity dispersion in UMi is $\sim$ 10 km s^{-1} (Hargreaves et al. 1994), 30 μas yr^{-1}, or twice the single-point error for the brighter stars. UMi is unique among the dSph galaxies in showing marginal evidence for minor axis rotation (Hargreaves et al. 1994), an indicator of possible triaxiality, tidal perturbation by the Milky Way, or non-isothermality in the dark matter. An accurate two-dimensional transverse velocity map of UMi is feasible with GAIA, retaining adequate spatial resolution while attaining precision by spatial binning. Similar studies at lower spatial resolution are possible in the other satellites (see Figure 1.36), including Sgr (Section 1.2.6). Of course, the GAIA proper motions will provide excellent discrimination between field stars, and provide a clean test of the expectation that all these dwarf galaxies are parts of extended tidal tails.

*Table 1.13: Local Group galaxies accessible to GAIA, based on Table 2 of Mateo (1998). E(B-V) indicates the foreground reddening, and (m-M)$_0$ is the true distance modulus. $V_{\rm lim}$ is the brightest star in the galaxy. $\mu_{v_t-v_r}$ is the estimated proper motion, assuming the transverse velocity equals the observed radial velocity. * denotes observed values.*

Galaxy	l (°)	b (°)	$E(B-V)$ (mag)	$(m-M)_0$ (mag)	Distance (kpc)	$V_{\rm lim}$ (mag)	N(stars) (V<20)	V_r (helio)	$\mu_{v_t-v_r}$ (μas/yr)
WLM	75.9	−73.6	0.02 ± 0.01	24.83 ± 0.08	925 ± 40	16.5	∼500	−116	26
NGC 55	332.7	−75.7	0.03 ± 0.02	25.85 ± 0.20	1480 ± 150	15.0	10's	129	18
IC 10	119.0	−3.3	0.87 ± 0.12	24.58 ± 0.12	825 ± 50	15.0	10's	−344	83
NGC 147	119.8	−14.3	0.18 ± 0.03	24.30 ± 0.12	725 ± 45	18.5	10's	−193	56
And III	119.3	−26.2	0.05 ± 0.02	24.40 ± 0.10	760 ± 40	20	-		60
NGC 185	120.8	−14.5	0.19 ± 0.02	23.96 ± 0.08	620 ± 25	20	-	−202	69
NGC 205	120.7	−21.7	0.04 ± 0.02	24.56 ± 0.08	815 ± 35	20	-	−241	62
M32	121.2	−22.0	0.08 ± 0.03	24.53 ± 0.08	805 ± 35	16	$\sim 10^4$	−205	54
M31	121.2	−21.6	0.08	24.43	770	15	$\gg 10^4$	−297	81
And I	121.7	−24.9	0.04 ± 0.02	24.53 ± 0.10	805 ± 40	21.7	-		
SMC	302.8	−44.3	0.08	18.82	58	12	$> 10^6$	158	900*
Sculptor	287.5	−83.2	0.02 ± 0.02	19.54 ± 0.08	79 ± 4	16.0	100's	110	360*
LGS 3	126.8	−40.9	0.08 ± 0.03	24.54 ± 0.15	810 ± 60			−277	72
IC 1613	129.8	−60.6	0.03 ± 0.02	24.22 ± 0.10	700 ± 35	17.1	100's	−234	71
And II	128.9	−29.2	0.08 ± 0.02	23.6 ± 0.4	525 ± 110	20	-		
M33	133.6	−31.3	0.08	24.62	840	15	$> 10^4$	−181	46
Phoenix	272.2	−68.9	0.02 ± 0.01	23.24 ± 0.12	445 ± 30	17.9	$\sim 10^2$	56	27
Fornax	237.1	−65.7	0.03 ± 0.01	20.70 ± 0.12	138 ± 8	14	100's	53	81
EGB0427+63	144.7	−10.5	0.30 ± 0.15	25.6 ± 0.7	1300 ± 700			−99	16
LMC	280.5	−32.9	0.06	18.45	49	12	$> 10^7$	278	1150*
Carina	260.1	−22.2	0.04 ± 0.02	20.03 ± 0.09	101 ± 5	18	$\sim 10^3$	229	478
Leo A	196.9	+52.4	0.01 ± 0.01	24.2 ± 0.3	690 ± 100			20	6
Sextans B	233.2	+43.8	0.01 ± 0.02	25.64 ± 0.15	1345 ± 100	19.0	10's	301	47
NGC 3109	262.1	+23.1	0.04 ± 0.02	25.48 ± 0.25	1250 ± 165			403	68
Antlia	263.1	+22.3	0.05 ± 0.03	25.46 ± 0.10	1235 ± 65			361	62
Leo I	226.0	+49.1	0.01 ± 0.01	21.99 ± 0.20	250 ± 30	19	10's	168	142
Sextans A	246.2	+39.9	0.03 ± 0.02	25.75 ± 0.15	1440 ± 110	17.5	10's	324	48
Sextans	243.5	+42.3	0.03 ± 0.01	19.67 ± 0.08	86 ± 4			230	564
Leo II	220.2	+67.2	0.02 ± 0.01	21.63 ± 0.09	205 ± 12	18.6	100's	90	95
GR 8	310.7	+77.0	0.02 ± 0.02	25.9 ± 0.4	1510 ± 330	18.7	10's	214	28
Ursa Minor	105.0	+44.8	0.03 ± 0.02	19.11 ± 0.10	66 ± 3	16.9	100's	−209	1000*
Draco	86.4	+34.7	0.03 ± 0.01	19.58 ± 0.15	82 ± 6	17	100's	−281	1000*
Sagittarius	5.6	−14.1	0.15 ± 0.03	16.90 ± 0.15	24 ± 2	14	$> 10^4$	140	2100*
SagDIG	21.1	−16.3	0.22 ± 0.06	25.2 ± 0.3	1060 ± 160			−77	16
NGC 6822	25.3	−18.4	0.26 ± 0.04	23.45 ± 0.15	490 ± 40			−57	25
DDO 210	34.0	−31.3	0.06 ± 0.02	24.6 ± 0.5	800 ± 250	18.9	10's	−137	36
IC 5152	343.9	−50.2	0.01 ± 0.02	26.01 ± 0.25	1590 ± 200			124	16
Tucana	322.9	−47.4	0.00 ± 0.02	24.73 ± 0.08	880 ± 40	18.5	10's		
Pegasus	94.8	−43.5	0.02 ± 0.01	24.90 ± 0.10	955 ± 50	20	-	−183	40

1.8.3 Rotational Parallax Distances to the LMC, M33, and M31

Rotation curves of disk galaxies are customarily derived from radial velocity data, correcting for disk inclination by determining the component of (major axis) rotation velocity projected into the radial velocity. Rotation curves can also be determined from proper motions, by measuring the transverse (minor axis) motion. Since proper motions are distance-dependent, and radial velocities are distance-independent, determination of a rotation curve using both methods provides a distance measurement. This can be done for the disk galaxies within the Local Group, and a few other spirals (M81, NGC 253, NGC 55). It is a priority for SIM, which plans to measure a few stars in each galaxy. Determination of the mean proper motion of a sample of stars near the minor axis of M31 to 1 μas yr^{-1} provides a distance to an accuracy of 2 per cent. This would provide a direct calibration of the Tully-Fisher distance scale, independent of all intermediate steps.

GAIA has a crucial contribution to make, in two ways. First, by measuring a larger sample of stars, albeit with lower individual accuracy, it will be possible to extend the analysis beyond a simple rotation-curve fit to determine a map of the two-dimensional kinematic field in each galaxy. In cases such as M33, where the warp is expected to dominate the projected kinematics in the most interesting outer parts of the galaxy, such information will be essential for meaningful analyses. As Figure 1.37 shows, 4300 M33 stars brighter than $V = 20$ mag can be observed. Second, these

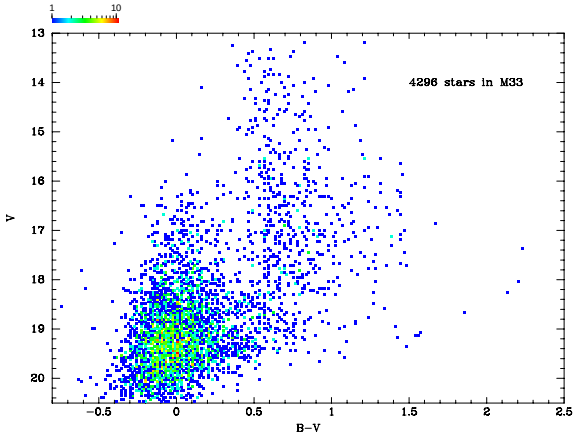


Figure 1.37: Colour-magnitude diagram for stars in M33. This shows 4300 stars brighter than $V = 20$ mag (courtesy of Phil Massey).

galaxies contain many globular clusters (M31 has a few hundred), whose unresolved cores will be measurable astrometrically (these are of apparent magnitude 16–17 mag for M31). In M31 the number of stars in the young disk ranges from 0.3 arcmin^{-2} at $I = 16$ mag to 30 arcmin^{-2} at $I = 20$ mag, allowing a substantial study. The GAIA proper motions will provide a three-dimensional determination of the halo mass in the outer parts of these galaxies.

1.8.4 Warps in M31 and M33

The most significant warp in terms of its observable effect outside the Milky Way is that of M31. Within $R \sim 25$ kpc the HI disk is flat, but then develops an 'integral-sign' warp that has carried the disk more than 3 kpc from the plane of the inner disk by $R = 30$ kpc. The HI disk of the third-largest Local Group disk galaxy, M33, is so extremely warped that at many points our line of sight intersects the disk twice. In M33 the warp starts at $R \sim 6$ kpc, and develops so rapidly that the gas at $R = 10$ kpc is rotating about an axis inclined at 40° to the axis of the inner disk.

The rotation curve amplitude of M31 (~ 265 km s^{-1}) corresponds to a transverse proper motion of ~ 100 μas yr^{-1}. In the warped outer parts, inclination effects reduce this to ~ 20 μas yr^{-1}. The brightest Population I stars (supergiants, Cepheids) have $I \sim 16 - 17$ in M31, so that the systematic warp effects are comparable to the measuring precision. Thus, not only must the warp be considered explicitly when deriving the M31 rotation curve parallax with GAIA (Section 1.8.3), but it is probable that the warp signature itself is detectable statistically. This second case will be critical in interpretation of the kinematic mapping and dynamical interpretation of the Galactic warp. For M33 the brightest stars are of comparable apparent magnitude, while the reduced rotational velocity is almost compensated by the smaller inclination angle. Thus, similar amplitude kinematic signatures are expected as in the M31 case.

For all kinematic and dynamical studies of galaxies of the Local Group, it will be necessary to make full use of the accuracy of the link of the GAIA reference frame to the extragalactic reference system described in Section 1.8.11.

1.8.5 Orbits in the Local Group: Gravitational Instability in the Early Universe

The orbits of galaxies are a result of mildly non-linear gravitational interactions, which link the present positions and velocities to the cosmological initial conditions. Non-gravitational (hydrodynamic) or strongly non-linear gravitational interactions (collisions, mergers) are sometimes significant. It is uniquely possible in the Local Group to determine reliable three-dimensional orbits for a significant sample of galaxies, in a region large and massive enough to provide a fair probe of the mass density in the Universe. Such orbital information provides direct constraints on the initial

spectrum of perturbations in the early Universe, on the global cosmological density parameter Ω, and on the relative distributions of mass and light on length scales up to 1 Mpc.

The present set of distances and radial velocities has been analysed by many authors (Kahn & Woltjer 1959; Gunn 1974; Gott & Thuan 1978; Lynden-Bell 1982; Kochanek 1995; Peebles 1971; Peebles 1996) among others. Peebles (1996) details the analysis methods, which are essentially Hamiltonian dynamical analyses for gravitating particles in the weak-field low-velocity limit of General Relativity, appropriate for growing perturbations in a region small compared to the Hubble length. Given suitable data, these (few-body) analyses can be supplemented by full numerical simulations of structure growth in model Universes, analogous to those of Moore et al. (1998), which consider resolved structure inside cold dark matter halos. Full analyses of this type consider the gravitational growth of structure from an initially almost smooth matter distribution. It is a self-consistency requirement that galaxian peculiar velocities are generated by the same gravitational perturbations which create the galaxies.

Only when real orbital information is available can one distinguish various possibilities for the spectrum of perturbations, later non-gravitational processes, local correlations between baryonic and dark matter ('bias'), and the true importance of mergers on various scales. That is, a map of the orbits of nearby galaxies provides a direct test of the dynamical history of the Local Group since recombination.

The required measurements are distances and transverse velocities for the relatively isolated members of the Local Group, those more distant than ∼100 kpc from another large galaxy, listed in Table 1.13. Improved distances will be derived from the GAIA-calibrated standard distance indicators, such as Cepheids and RR Lyraes, as described in Section 1.4.9. Radial velocities are known. The missing key information is the transverse motion, derivable uniquely from the GAIA proper motion. Table 1.13 lists the expected proper motion, assuming the transverse velocity equals the radial velocity. In each case the required accuracy is set by random motions, which here is about 100 km s^{-1}, or approximately one-half the signal. The signal is typically 10–20 μas yr^{-1}, attainable with measurements of a few tens of stars at $V = 19$ mag, or 100 stars at $V = 20$ mag. This is possible for more than 20 galaxies within 1.5 Mpc, and for a few large spirals out to 2.5 Mpc. The number of galaxies for which this experiment is possible increases extremely rapidly with improved limiting magnitude: a few additional examples are also listed in Table 1.13.

1.8.6 Galaxies

Growth of structure in the Universe is believed to proceed from small amplitude perturbations at very early times. Their growth from the radiation-dominated era to the present has been extensively studied, particularly in the context of the popular hierarchical clustering scenario. Many aspects of this picture are well-established. Others are the subject of active definition through redshift and imaging surveys of galaxies, and the microwave background experiments. There are several aspects of this research which require very wide area imaging surveys with high spatial resolution, to provide high-reliability catalogues of galaxies and quasars extending to low Galactic latitudes. Here GAIA will contribute uniquely, by detecting and providing multi-colour photometry with ∼0.35 arcsec spatial resolution for all sufficiently high-surface brightness galaxies crossing the astrometric focal plane (Vaccari & Høg 1999b; Vaccari & Høg 1999a; Høg & Arenou 2000; Lindegren 2000). This provides a valuable and unique data set at two levels: for statistical analysis of the photometric structure of the central regions of many tens of thousands of galaxies; and for study of the large-scale structure of the local Universe.

Distribution of Galaxies The primary scientific requirement is for a very wide angle survey, reaching into the 'zone of avoidance' at low Galactic latitudes, with a well-defined selection function. Such data are not available from ground-based imaging surveys, as star-galaxy separation to the requisite reliability cannot be achieved without high spatial resolution imaging.

Simulations indicate that the number of galaxies which can be detected and for which useful broad-band photometry profiles can be derived is about $10^6 - 10^7$. The large number of compact galaxies around 20 mag will not be detected in very large numbers given the short integration times and relatively high read noise (Lindegren 2000). Detected galaxies will nevertheless directly provide a measurement, through deconvolution of the measured angular power spectrum, of the spectrum of fluctuations well beyond the expected peak. Such data are both a natural complement to the ongoing redshift surveys, and also provide an input catalogue for future extension of those surveys. The very great volume surveyed locally makes the survey both an important local normalisation, and also potentially allows study of the largest scale lengths, without evolutionary complications.

A primary science case for such studies arises from the difficulty in understanding the peculiar motion of the Local Group. It is well-established that the Local Group has a peculiar motion of about 600 km s^{-1} towards (ℓ, b)=(268°, 27°). If our understanding of the gravitational instability picture for the growth of structure, and measurements of Ω_0 and biasing are valid, this must be explicable as acceleration by identifiable nearby galaxy clusters, or massive single galaxies. The largest of these sources, especially the Great Attractor and Perseus-Pisces, remain poorly mapped, being at low Galactic latitudes.

A second scientific goal concerns the amplitude, shape, and length of structures in the Local Universe. Large filaments, 'walls', and the Supergalactic Plane dominate the nearby galaxy distribution. All are lost with present data within 20° of the Galactic Plane. It is not even yet clear if the Supergalactic Plane is a plane at all. If we are to understand local large-scale structure, a reliable near all-sky galaxy survey is essential. At low Galactic latitudes random errors in star-galaxy classification, due to seeing, convolved with the numerical predominance of stars, prevent construction of such a catalogue. The combination of GAIA's spatial resolution and multi-colour photometry will allow substantially improved analyses.

Galaxy Photometry with GAIA The GAIA Galaxy Survey will also provide multi-colour information for individual galaxies, allowing detailed multi-colour photometric studies of their central regions (cf. Figure 6.5). This will include those galaxies for which redshifts are being obtained directly or in follow-up programmes (e.g. with the Sloan, 2dF, 2MASS and DeNIS surveys), directly linking morphology and spectra. Detailed analysis of the inner luminosity profiles of a large sample of galaxies will define the true incidence of core structures, and complex morphologies ($m = 1, 2, \ldots$). Inner colour gradients map recent star formation and dust lanes. Central luminosity cusps may indicate massive black holes.

Those galaxies with bright cusps within the central GAIA point spread function can be analysed as astrometric targets, in the same way as stars. In this case, which will include many active galactic nuclei, astrometric 'jitter' may be detectable, if a significant contribution to the flux is spatially variable. While specific modeling will be required on a case by case basis, correlation of motion of the photocentre with optical variability in Seyfert nuclei and quasars can test if nuclear starburst supernovae are a significant luminosity source. In cases where no 'jitter' is seen, these sources define the reference frame. The relative location of the optical nuclei and the active nuclei can also be measured with high angular precision.

Photometric studies of bright galaxies allow detailed analyses of internal structures, such as spiral arms and star formation regions. For fainter, and less well-resolved galaxies, simpler analyses, of such parameters as bulge-to-disk ratio, and central photometric cusp or nucleus structure, are appropriate (Okamura et al. 1999). For the faintest useful images, model-based two-dimensional image analyses, typically involving maximum likelihood comparison of the image data with a set of simple models, have been developed and applied, especially to surveys such as the HST Medium Deep Survey (e.g., Ratnatunga et al. 1999). These methodologies are capable of reliable analysis of galaxy images with signal-to-noise ratios comparable to those GAIA will produce for galaxies to magnitude V$\sim$17 mag. A crucial advantage of the GAIA galaxy survey over other studies is that it will be automatically all-sky and a complete magnitude-limited survey, with a very well-defined selection function.

The scientific value of this huge and homogeneous database will impact all fields of galaxy research, naturally complementing the several redshift surveys, and the deeper pencil-beam studies with very large telescopes. Among the most important unique GAIA science products will be determination of the colour and photometric structure in the central regions of a complete, magnitude-limited sample of relatively bright galaxies. Recent studies of early-type galaxies (Faber et al. 1997) and disk galaxies begin to address the wealth of structures in the central regions of galaxies. Early-type galaxies are crudely distributed into cores which are flat or are steeply rising, perhaps indicative of the effects of massive central black holes. Late-type galaxies have an extreme diversity of central structures, probably providing the key to bulge and central disk formation. Figure 1.38 shows the range of surface brightness of complex core structures over 15 magnitudes. GAIA will provide three key elements: high and uniform spatial resolution, a large sample, and multi-colour data.

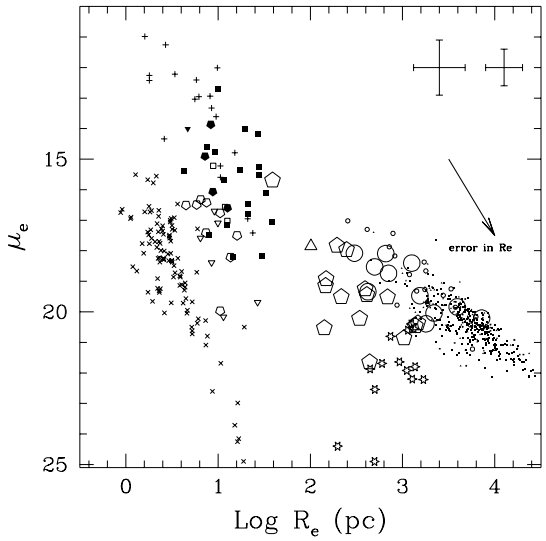


Figure 1.38: Surface brightness μ_e (mag arcsec^{-2}) in the visual band (F606W) at effective radius r_e for a sample covering all types of galaxies (from Carollo 1999).

Simultaneous multi-colour light curves, albeit sparsely sampled, will naturally be obtained for every galaxy observed. Thus the statistical incidence of active galactic nuclei and related variability will be determined, as a function of the photometric structure of the host bulge or inner disk. Other types of variable source will also be detected, ranging from novae in the Local Group, through Cepheids and luminous blue variables in more distant galaxies, to supernovae, and possible gamma-ray bursts. Rapid analysis of the photometric data during the mission will allow identification of variables for dedicated followup by other telescopes.

1.8.7 Supernovae

GAIA will detect all objects brighter than $I = 20$ mag, so that in principle supernovae can be detected to a modulus of $m - M \sim 39$ mag, i.e., to a distance of 500 Mpc or $z \sim 0.10$. Simulations show that in 4 years, GAIA will detect about 100 000 supernovae of all types (Høg et al. 1999d). This will allow determination of the supernova rate as a function of galaxy type, as well as discovery of supernovae in low surface-brightness galaxies, which generally are excluded from present surveys.

Rapid analysis of the photometric data during the mission is foreseen (see Section 6.7.3). This will allow identification of variables, including supernovae, for dedicated follow-up by other telescopes. GAIA detections of each point source will be compared with all previous GAIA observations of the same position on the sky to decide whether the star is new. A star will appear in the other Astro telescope within a couple of hours due to the overlap of consecutive scans. Analysis of the astrometric data obtained within the field of view will reveal whether an object has a fast motion and therefore minor planets will be discriminated. A comparison with epoch photometry from the first 12 months of the mission will suffice to distinguish supernovae from most variable stars, especially at Galactic latitudes $|b| > 15°$.

The velocity field of matter in the Universe is dominated by the Hubble expansion, and the three-dimensional structure is characterized by sheets with high density of visible matter separated by large voids. A superimposed smaller-scale streaming of the luminous matter is not easy to measure, because distance determination independent of the Hubble law is difficult. Type Ia supernovae are very accurate distance indicators, ±5 per cent, but they have been too sparsely observed to date.

The typical sequence of GAIA observation epochs is a group of about five observations in four broad-band colours at one hour separation. Such a group is obtained about every month, which is too sparse to produce a useful supernova light curve from the GAIA data alone, but rapid ground-based follow-up monitoring of the events can provide absolute magnitudes for about 50 000 cases (Høg et al. 1999d). These can be combined with the redshift of the parent galaxy to provide as many streaming velocities.

The absolute magnitude of a Type Ia supernova at maximum can be determined if it is detected not later than 10 days after maximum and its light curve is followed for a few weeks. The Type Ia stays within 1 magnitude from maximum from 7 days before maximum (the most useful cases) to 15 days after. About 50 000 of the detected supernovae will be of Type Ia and be observable for such a time from the ground.

1.8.8 Photocentric Motions from Relativistic Jets of Active Galactic Nuclei

GAIA astrometry will provide a completely new way of studying relativistic jets of active galactic nuclei. Ground-based radio interferometry shows these are compact (down to scales of 100 μas) and contain much structure on small scales (e.g., Zensus 1997). VLBI observations trace plasma knots which are moving in these jets and find apparent velocities in excess of c, indicating relativistic speeds and small inclinations from the line-of-sight. The knots emit synchrotron emission. In a few jets of nearby radio galaxies direct optical imaging is possible with HST. In these jets a close similarity between radio knots and optical emission is found. This suggests that the relativistic particles which are accelerated in the jets reach sufficiently high energy to emit optical synchrotron radiation. The motion of plasma knots can hence be studied at optical frequencies as well.

In well-sampled cases the trajectories were found to be curved, indicating that the plasma knots are not moving ballistically. Motion on curved trajectories implies that the angle between the line-of-sight and the local velocity vector changes with time. Since the knots are moving with relativistic velocities, changes in the viewing angle also change the effective Doppler beaming and hence the observed brightness of the emitting region. In at least one object (3C 345) the motion of synchrotron-emitting plasma clouds can be described by plasma moving on a helical path. Within projected velocities of a few tenths of a mas per year, the trajectories of these knots can be traced with VLBI observations taken every few months.

The most crucial parameter is the contrast in brightness with respect to the true nucleus. Estimates on this ratio can be obtained from variability studies: all of the radio-jets with superluminal motions detected by VLBI experiments are ejected from active galactic nuclei which are strongly variable at optical and radio frequencies. A close temporal correlation between flares in these two frequency regimes has been found. Babadzhanyants & Belokon (1986) realized that the epochs of major flares coincided with the ejection of radio knots. This has now been confirmed in several cases. Whereas the evolution of the knot at radio frequencies can be studied directly with VLBI, optical photometry so far only measures the total flux received from the core and the ejected plasma cloud. This total luminosity closely follows the light-curve expected from a plasma cloud moving on a helical path and illustrates that a large fraction of the variability may be due to changing Doppler amplification. This idea was suggested by Camenzind & Krockenberger (1992) and Wagner et al. (1993), and was modeled in detail for the case of 3C 345 (Schramm et al. 1993) and PKS 0420–014 (Wagner et al. 1995). The amplitudes of the optical variations exceed one magnitude, indicating that more than half the total optical flux is actually emitted by the knot which is moving away from the core on a helical path with projected velocities of up to a few tenths of a mas per year. At optical wavelengths the moving knot contributes a much higher fraction to the total flux than at radio frequencies. The optical photo-centre of the combined core-plus-knot is hence expected to change over a few months to years with amplitudes of up to 300 μas yr^{-1}, closely correlated with changes in total flux.

GAIA will resolve the amplitude of jet photocentric motions only in nearby nuclei, where the effect will exceed a mas within a few months. It will be statistically significant when compared to the astrometric accuracies of nearby nuclei of comparable brightness. The expected correlation between changes in luminosity and photocentric motions will be an unambiguous characteristic in those cases where the astrometric effects are of the order of the astrometric accuracies. Such studies will allow unprecedented investigations of the trajectories of plasma knots in general (with a temporal sampling exceeding those achieved in VLBI experiments). In particular it will allow detailed studies of the synchrotron cooling of the highly relativistic particles responsible for the optical emission.

1.8.9 Quasars

The astrometric programme to $V = 20$ mag will provide a census of $\sim 500\,000$ quasars at intermediate to high Galactic latitudes (Table 1.14). The mean surface density of $\sim 25\,\mathrm{deg}^{-2}$ at intermediate to high Galactic latitudes will allow a direct link between the GAIA astrometric reference system and an inertial frame, which was not possible (directly) for Hipparcos (Section 1.8.11). It will also provide a substantial, uniformly selected quasar sample. The quasars will be confined to a relatively small subset of the astrometric catalogue due to the absence of proper motions.

The GAIA broadband filter system, with coverage over the full optical waveband $300 - 900\,\mathrm{nm}$, is well-suited to the identification of quasars via established multi-colour selection techniques (see

Table 1.14: Magnitude-redshift relation for quasars in an area of 4000 square degrees. The table gives the expected number of quasars per magnitude and per interval of 0.5 in z, versus the central magnitude and redshift values (courtesy of D.P. Schneider).

Redshift	12	13	14	15	16	17	18	19	20	Total
0.25	4	8	10	12	113	1160	3355	238	0	4900
0.75	0	0	0	4	56	747	7013	16523	9481	33824
1.25	0	0	0	3	32	339	3646	16954	28430	49404
1.75	0	0	0	4	42	336	3019	13648	23667	40716
2.25	0	0	0	2	32	206	1709	8468	15948	26365
2.75	0	0	0	0	19	160	965	4651	8767	14562
3.25	0	0	0	0	9	63	364	1714	3579	5729
3.75	0	0	0	0	8	36	137	514	1207	1902
4.25	0	0	0	0	0	16	45	136	373	570
4.75	0	0	0	0	0	4	15	50	119	188
5.25	0	0	0	0	0	1	4	5	38	48
5.75	0	0	0	0	0	0	0	1	8	9
6.25	0	0	0	0	0	0	0	0	2	2
Total	4	8	10	25	311	3068	20272	62902	91619	178219

The numbers are based on a simulation which uses the measured quasar luminosity function, the observed distribution of continuum slopes and the observed emission line properties. The simulated quasars were 'observed' with the Sloan filters, and the results are shown for the i band. The turn-down of the counts at low redshifts and faint magnitudes is caused by the quasar cutoff at $M_B = -23$. This, of course, is an arbitrary cutoff; if one includes the AGN-phenomenon at lower luminosities, the counts continue upwards at fainter magnitudes. The table shows that GAIA will observe about 500 000 quasars over about 20 000 square degrees of the sky.

Figures 1.39). The vast majority of the quasars will have redshifts $0.2 < z < 3$ and absolute magnitudes $-28 < M_V < -23$. Redshifts to a precision of a few percent can be obtained for the brightest objects by cross-correlation of the GAIA spectral energy distributions from medium-band photometry with a small library of existing reference quasar spectra. The GAIA catalogue will include a factor of 50 more quasars and AGNs than the quasar catalogue planned following the completion of the Sloan survey (for the discrimination of quasars and other special objects by means of Sloan photometry, see Krisciunas et al. 1998; Lenz et al. 1998; Newberg et al. 1999).

The photometric depth of the bluer pass-bands are not by themselves adequate to allow identification of rarer sub-populations, such as quasars at redshift $z > 5$ (e.g., Figure 1.39). However, the combination of the GAIA astrometric and photometric data will eliminate the principal stellar contaminants, M-dwarfs. The result will be a sample containing the majority of the expected number of a few hundred quasars at $z \sim 5$, with $I < 20$ mag, over the whole sky. Follow-up studies would be particularly useful for the properties of the inter-galactic medium at high-redshift and the evolution of damped Lyman–α and other high-column density absorbers.

The all-sky nature of the catalogue, excluding the Galactic plane, ensures that a relatively high surface density background-grid of quasars are available for a wide range of projects. These may include probing the interstellar medium of nearby galaxies, the gas content of the halo regions of more distant galaxies, the gas content of central regions of rich galaxy clusters, and the large-scale clustering properties of high-redshift intervening magnesium and carbon absorbers, particularly in regions where complementary investigations of large-scale structure, via galaxy and cluster redshift surveys for example, have been undertaken.

Lensing Existing ground-based studies of gravitational (macro) lensing among the quasar population are restricted to resolutions of ~ 1 arcsec. The Hubble Space Telescope has allowed a few hundreds of the brightest quasars to be examined at resolutions of ~ 0.2 arcsec, but examples of multiply-imaged objects are extremely rare and there is little prospect of acquiring a statistical sample of adequate size. GAIA will provide sensitivity to multiply-imaged systems with separations as small as ~ 0.2 arcsec. For the brighter quasars, $V < 18$ mag, with a surface density of ~ 1 deg^{-2}, where examples of lensing are most common, GAIA's sample of $\sim 50\,000$ quasars represents an increase of two orders of magnitude over existing surveys. Pushing the sensitivity to image separations of a few tenths of an arcsec will access systems where most of the lensing due to individual galaxies is expected. In particular, the GAIA survey will provide new constraints on

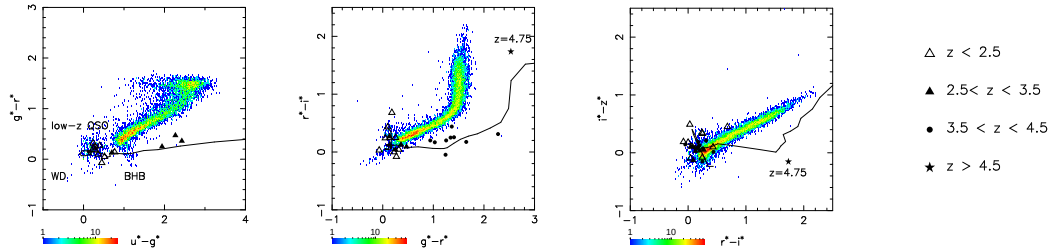


Figure 1.39: An example of the use of broad-band filter photometry to supplement proper-motion searches for quasars. The combination of zero proper motion and colour data will substantially improve the detection efficiency of quasars, and hence their utility for clustering studies. This figure uses Sloan Digital Sky Survey data, and illustrates the redshift ranges of quasars discovered in the test phases of the Sloan survey (from Fan et al. 1999).

lensing by the bulk of the galaxy population, including spiral galaxies, rather than the high-mass tail of ellipticals to which existing surveys are predominantly sensitive.

GAIA also offers intriguing possibilities in the field of gravitational microlensing of the quasar population. The first microlensing event was discovered in an extragalactic system, the quadruple gravitational lens Q2237+0305, and a number of examples of the influence of microlensing by objects of stellar mass in galaxies responsible for macrolensing quasars are now established. The photometric precision and variable sampling intervals provided by GAIA are well-suited to the identification of microlensing events in quasars, although the relatively short mission lifetime means that statistical studies rather than detailed investigations of particular systems are favoured. Statistical analysis of the photometric variability properties of subsets of the quasar catalogue projected within the halos of samples of galaxies of various Hubble types and behind rich galaxy clusters would provide constraints on the space density and mass function of compact objects, including stars, over a wide range of masses.

Gravitational microlensing produces multiple images with separations below the resolution limit of GAIA. However, the changing proportion of light contributed by the individual images during a microlensing event produces a small shift in the centroid of the summed micro images, i.e., in the apparent position of the quasar. For micro-lenses of stellar mass at cosmological distances the predicted image shifts are of the order of microarcsec and in the most favourable instances may reach several tens of microarcsec (see Scholz et al. 1997, and also Section 1.8.11). Such shifts should be detectable by GAIA in brighter quasars macrolensed by galaxies. A statistical analysis of image centroid shifts in existing and new (GAIA-discovered) multiply-imaged quasars could constrain the characteristic lensing mass and the size of the quasar continuum emitting region, as well as certain key cosmological parameters.

Assuming that the data transmission allows identification of quasars showing a complex structure within a field of 3 arcsec, the observations will lead to the detection of a complete sample of several thousand gravitational lenses. This homogeneous sample would provide decisive astrophysical information, including constraints on the cosmological parameters $\Omega_{\rm M}$ and Ω_Λ. To estimate the numbers of quasars affected, foreground galaxies have been modelled as 'singular isothermal sphere' lenses. A population of galaxies comprising 30 per cent of E/S0 ($\sigma^* = 225$ km s^{-1}, corresponding to a value of the efficiency parameter for macro-lensing $F = 0.025$) and of 70 per cent of spiral ($\sigma^* = 144$ km s^{-1}, $F = 0.007$; cf. Fukugita & Turner 1991) has been assumed. Given a galaxy with luminosity L, corresponding to a value σ for its velocity dispersion (Tully-Fischer or Faber-Jackson), located at a redshift z_l, and under conditions of perfect alignment, a background quasar will be imaged as an Einstein ring having an angular diameter $2\theta_E = 8\pi D_{\rm ds}\sigma^2/D_{\rm os}c^2$. If the alignment is not perfect, the ring will break into two lensed images having an angular separation also equal to $2\theta_E$, whose magnitude difference Δm increases as the alignment between the observer, the lens and the source gets worse (but with $\theta_S < \theta_E$). Adopting a uniform cosmological distribution of lenses, the expected angular image separations for a quasar at redshift z_q can be estimated. Realistic assessments have been made using appropriate angular resolution and dynamical range.

Approximately 3500 multiply imaged quasars are expected over the whole sky down to $V < 21$ mag, assuming $\Delta\theta < 3$ arcsec, and a 'less favourable' Universe with $\Omega_{\rm M} = 1$, $\Omega_\Lambda = 0$. This represents 96 per cent of the total number of expected lenses. (If the field of view is further reduced to 2 o 1 arcsec, only 83 per cent or 46 per cent of these will be detected, respectively). As expected, the contribution due to E/S0 galaxies remains the dominant one. In a Universe with $\Omega_{\rm M} = 0$, $\Omega_\Lambda = 1$, the total number of multiply imaged quasars increase by a factor of 13. All these results are independent of the value of H_0. From the observed fraction of gravitational lenses identified among quasars as a function their apparent V magnitude, GAIA thus offers high prospects to constrain the values of cosmological parameters very efficiently.

1.8.10 Galactocentric Acceleration

The Sun's absolute velocity with respect to a cosmological reference frame causes the dipole anisotropy of the cosmic microwave background. The Sun's absolute acceleration can be measured astrometrically: it will result in proper motions of quasars.

The Solar system's orbital velocity around the Galactic centre causes an aberration effect of the order of 2.5 arcmin. All measured star and quasar positions are shifted towards the point on the sky having Galactic coordinates $l = 90°$, $b = 0°$. For an arbitrary point on the sky the size of the effect is $2.5 \sin \eta$ arcmin, where η is the angular distance to the point $l = 90°$, $b = 0°$. The acceleration of the Solar system towards the Galactic centre causes this aberration effect to change slowly. This leads to a slow change of the apparent positions of distant celestial objects, i.e., to an apparent proper motion.

Assuming a Solar velocity of 220 km s^{-1} and a distance of 8.5 kpc to the Galactic centre, the orbital period of the Sun is 250 Myr, and the Galactocentric acceleration has the value 0.2 nm s^{-2}, or 6 mm s^{-1} yr^{-1}. A change in velocity by 6 mm s^{-1} causes a change in aberration of the order of 4 μas. The apparent proper motion of a celestial object caused by this effect always points towards the direction of the Galactic centre. Its size is $4 \sin \zeta$ μas yr^{-1}, where ζ is now the angular distance between the object and the Galactic centre. The above holds for all objects beyond about 200 Mpc, and in particular for quasars, for which it can be safely assumed that the intrinsic proper motions (i.e., those caused by real transverse motions) are negligible. A proper motion of 4 μas yr^{-1} corresponds to a transverse velocity of 30 000 km s^{-1} at $z = 0.3$ for $H_0 = 75$ km s^{-1} Mpc^{-1}. Thus, all quasars will exhibit a distance-independent streaming motion towards the Galactic centre. Within the Galaxy, on the other hand, the effect is hidden in the local kinematics: at 10 kpc it corresponds to only 200 m s^{-1}.

The determination of the acceleration of the Solar system barycentre has been included in the simulations undertaken for the radio/optical reference frame link (see Tables 1.15 and 1.16).

In principle, the displacement of the Sun can be used to measure secular parallaxes with GAIA for extragalactic objects out to distances of ~ 20 Mpc. Cusped nuclei beyond 200 Mpc will behave like the quasars.

1.8.11 The Radio/Optical Reference Frame

The major improvement in the precision of proper motions by GAIA implies that they should be referred to a non-rotating reference system. The GAIA reference frame materializing the reference system must be such that the biases introduced by its inaccuracy should be significantly smaller than the random errors of the phenomena that are referred to it.

At present, the International Celestial Reference System (ICRS) is primarily materialized by the International Celestial Reference Frame (ICRF) consisting of positions of 212 extragalactic radio-sources (Ma et al. 1998) with an rms uncertainty in position between 100 and 500 μas. The extension of the ICRF to visible light is the Hipparcos Catalogue with rms uncertainties estimated to be 0.25 mas yr^{-1} in each component of the spin vector of the frame (ω) and 0.6 mas in the components of the orientation vector (ε) at the catalogue epoch, J1991.25 (Kovalevsky et al. 1997). The GAIA catalogue will permit a definition of the ICRS more accurate by one or two orders of magnitude than the present realizations.

The essential quality of a reference system is that it remains unchanged as long as there are no theoretical reasons to consider that a new definition will remove some features that prevent it from being quasi-inertial or non-rotating. This is the case at present, and the assumption that the ensemble of distant extragalactic objects has no global rotation is sufficiently solid to remain the rule for many decades. As a consequence, the successive realizations of the ICRS are such that there is no rotation or shift in origin within the accuracy level at which they are constructed.

Linking the GAIA catalogue to the ICRS can be thought of as proceeding in the following three steps: First, the GAIA observations are reduced to an internally consistent catalogue of positions and proper motions. The orientation and spin of this provisional catalogue is however arbitrary, since the measured arc lengths between objects are invariant to frame rotation. Secondly, the (apparent) proper motions of quasars in this catalogue are analysed to determine the quasi-inertial spin vector ω of the catalogue with respect to the extragalactic frame. The catalogue is then 'stopped' by applying a correction corresponding to $-\omega$ to all the proper motions. Thirdly, the resulting positions of the optical counterparts of radio sources in the ICRF are compared with the radio positions, to give the orientation vector ε of the optical catalogue with respect to ICRF. The final GAIA catalogue results from applying a correction corresponding to $-\varepsilon$ to all the positions. (In practice these steps are incorporated in the iterative astrometric core processing.) The accuracy by which the final GAIA catalogue represents the ICRS is given by the uncertainties of ω and ε.

The spin vector can be determined very accurately by means of the many thousand faint quasars picked up by the astrometric and photometric survey. Several observational properties of the quasars can be combined to extract samples that are very clean (i.e. without stars), but not necessarily complete (for this purpose): colour indices, photometric variability, a negligible parallax and proper motion. Simulations using realistic quasar counts show that an accuracy of better than 0.4 μas yr^{-1} will be reached in all three components of ω (Tables 1.15 and 1.16).

If an object has an apparent tangential velocity equal to the Hubble recession velocity, the apparent proper motion would be 14 μas yr^{-1}, independent of the distance, while if it is equal to the mean velocity dispersion of galaxies, it would be significantly smaller than 0.1 μas yr^{-1} beyond a distance of 1 Gpc. The detection of anomalous cases is therefore straightforward.

As for the determination of the frame orientation (ε), the only possible procedure is to compare the positions of the radio sources in ICRF (and its extensions) with the positions of their optical counterparts observed by GAIA. The number of such objects is currently less than 300 and the error budget is dominated by the uncertainties of the radio positions. Assuming current accuracies for the radio positions, simulations show that the GAIA frame orientation will be obtained with an uncertainty of ~ 60 μas in each component of ε. The actual result by the time of GAIA may be significantly better, as the number and quality of radio positions for suitable objects are likely to increase with time.

Condition equations The spin vector should be determined simultaneously with the apparent streaming motion produced by the acceleration $\mathbf{a}$ of the Solar system Barycentre with respect to the cosmological reference frame (Section 1.8.10). The condition equations for the Galactic components of ω and $\mathbf{a}$ are:

$$\mu_{l*} = -\mathbf{q}'\omega + \mathbf{p}'\mathbf{a}/c \quad (1)$$

$$\mu_b = \mathbf{p}'\omega + \mathbf{q}'\mathbf{a}/c \quad (2)$$

where c is the speed of light and $\mathbf{p} = (-\sin l, \cos l, 0)'$, $\mathbf{q} = (-\sin b \cos l, -\sin b \sin l, \cos b)'$ are the unit vectors along $+l$, $+b$ tangent to the celestial sphere at the position of the quasar.

Simulations Numerical simulations were made of the least-squares solution of ω and $\mathbf{a}$, with the following assumptions. The cumulative number density of quasars as function of B magnitude was taken from Hartwick & Schade (1990). Only redshifts $z < 2.2$ were included, which gives some underestimation of the actual available numbers. Quasars were randomly distributed with this density over the sky, except in the Galactic belt $|b| < 20°$, where zero density was assumed. Depending on magnitude, only a fraction P of all the quasars were used; this simulates the use of various photometric and astrometric criteria to reject possible stars. The Galactic coordinates were transformed to the ecliptic system, and the standard errors in $\mu_{\lambda*}$, μ_β were computed as a function of β and the GAIA magnitude $G = B - 0.2$ according to Section 7. A separate least-squares solution was made for each interval of one magnitude in B from 14 to 20, and one solution for the whole magnitude range. Only the covariance matrices of the solutions are of interest; they were transformed back to the Galactic system, yielding the accuracy estimates in Tables 1.15 and 1.16.

To account for source instabilities (see below), a quantity σ_0 was added in quadrature to the formal proper motion uncertainties from Section 7. For Table 1.15, a fairly optimistic value of $\sigma_0 = 10$ μas yr^{-1} was assumed, while for Table 1.16 the assumption was a pessimistic $\sigma_0 = 100$ μas yr^{-1}. It is seen that sub-μas accuracy is reached in both cases due to the large number of faint sources. The accuracy is slightly lower about the third axis (normal to the Galactic plane) than in the other two coordinates, due to the zone of avoidance.

It was found that the solution for the acceleration $\mathbf{a}$ is practically orthogonal to that of ω and of equal accuracy, when expressed in comparable units ($\mathbf{a}/c$ has the dimension of proper motion, with 1 μas yr^{-1} corresponding to 4.606×10^{-11} m s^{-2}). The galactocentric acceleration of the Sun is expected to be 2×10^{-10} m s$^{-2} \sim 4$ μas yr^{-1}, and should thus be measurable at 5 to 10 per cent relative accuracy.

Source instabilities Apparent source instabilities may arise from a number of effects, among which the following have been considered: (1) macrolensing by intervening galaxies may cause apparent proper motions of several μas yr^{-1}, but only if the impact parameter is close to the critical value (of the order of 1 arcsec) where significant magnification occurs (Kochanek et al. 1996). The fraction of affected quasars is of the order of 1 per cent (Kochanek 1996), and they usually have additional structure (multiple images, arcs) on scales that are resolved by GAIA. For larger impact parameters the proper motion of the single deflected image is smaller than the proper motion of the lensing galaxy, i.e. $\lesssim 0.2$ μas yr^{-1} for a lens at $z \sim 0.1$; (2) microlensing by stars in the Milky Way Galaxy is extremely improbable for any given quasar, but all quasars will be subject to weak microlensing causing random, variable displacements of the order of 1 μas (Sazhin et al. 1998). The typical effect on the mean proper motion over the GAIA lifetime will be $\lesssim 1$ μas yr^{-1}; (3) most of the optical emission from a quasar comes from a region of $\lesssim 1$ pc, corresponding to $\lesssim 200$ μas at 1 Gpc. Assuming that the photocentre moves randomly within this region, a mean proper motion of $\lesssim 50$ μas yr^{-1} may result over a 4-year period. Motion of the photocentre is also induced by a variable nucleus in combination with the much fainter, but much larger galaxy. This effect could reach some 100 μas yr^{-1}, but extreme cases might be recognised by the correlation between position and brightness; (4) quasar spectra have strong emission lines potentially at any wavelength depending on redshift. This could make the chromaticity correction difficult, as it is will normally be based on broad-band photometry and calibrated mainly

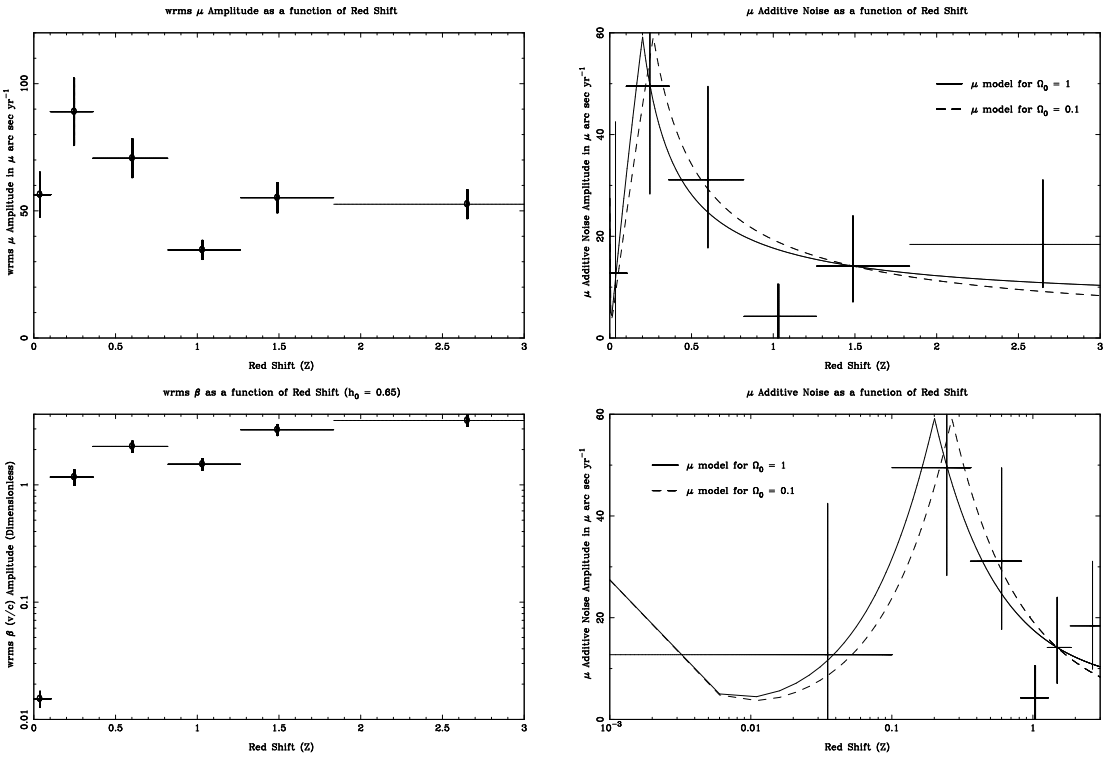


Figure 1.40: Proper motion residuals, σ_μ, computed according to different hypotheses about the source motion, and for different cosmological models (from Eubanks, private communication). See text for further details.

by means of stellar spectra. This, in combination with the variability of the quasars, could generate spurious proper motion of instrumental origin. The effect has not been evaluated in detail, but the results from a study of the chromaticity calibration for stars (Vannier 1998a) suggests that it could be ~ 10 μas yr^{-1}. In conclusion, the most important instabilities are probably due to variable source structure and residual chromaticity. The likely range of the combined effects for typical quasars may be 10–100 μas yr^{-1}, as assumed in Tables 1.15 and 1.16. Figure 1.40 shows σ_μ computed according to different hypotheses about the source motion, and for different cosmological models (from Eubanks, private communication). While the specific reasoning behind these specific models may be faulty, they do provide a reasonable fit for the observed axcess proper motion scatter, and it would be straightforward to adopt this model to estimate the ability of GAIA to determine the extragalactic reference frame and parameters (such as shear, rotation, and the gravitatioal radiation flux) that depend on it.

1.9 Fundamental Physics: General Relativity

The reduction of the Hipparcos data necessitated the inclusion of stellar aberration up to terms in $(v/c)^2$, and the general relativistic treatment of light bending due to the gravitational field of the Sun (and Earth). The GAIA data reduction requires a more accurate and comprehensive inclusion of relativistic effects, at the same time providing the opportunity to test a number of parameters of General Relativity in new observational domains, and with much improved precision. The rigorous formulation of the GAIA data analysis problem is considered separately in Section 9.3.

1.9.1 Light Bending in the Solar System

The dominating relativistic effect in the GAIA measurements is gravitational light bending. The possibility of accurately measuring the parameter γ of the Parametrized Post-Newtonian (PPN) formulation of gravitational theories is of key importance in fundamental physics.

Table 1.15: Residual spin of the GAIA reference frame estimated from a simulation of quasar observations. The columns contain, for each range of B magnitudes: P = assumed probability that a quasar is unambiguously recognised as such from photometric indices; $N_{\rm QSO}$ = expected number of recognised quasars with $z < 2.2$ and $|b| > 20°$; $\sigma_{\mu,tot}$ = mean standard errors in the proper motion per object and coordinate, including an assumed contribution of $\sigma_0 = 10$ $\mu as\ yr^{-1}$ from source instability; $\sigma(\omega_i)$ = resulting precision of the spin components about the Galactic axes ($i = 1$ towards the Galactic centre, $i = 3$ towards the Galactic pole); $\sigma(a_i/c)$ = resulting precision of the acceleration of the Solar system Barycentre along the Galactic axes.

B (mag)	P	$N_{\rm QSO}$	$\sigma_{\mu,\rm tot}$ (μas yr^{-1})	$\sigma(\omega_1)$	$\sigma(\omega_2)$	$\sigma(\omega_3)$ (μas yr^{-1})	$\sigma(a_1/c)$	$\sigma(a_2/c)$	$\sigma(a_3/c)$ (μas yr^{-1})
14 – 15	1.0	40	12	2.2	2.2	2.7	2.2	2.2	2.7
15 – 16	1.0	230	14	1.05	1.05	1.28	1.05	1.05	1.26
16 – 17	0.9	1 230	18	0.59	0.59	0.73	0.60	0.60	0.71
17 – 18	0.8	11 500	27	0.28	0.28	0.35	0.28	0.29	0.34
18 – 19	0.6	60 000	44	0.20	0.20	0.24	0.20	0.20	0.24
19 – 20	0.3	97 000	78	0.27	0.27	0.33	0.27	0.27	0.32
all		170 000		0.13	0.13	0.16	0.13	0.13	0.16

Table 1.16: Same as Table 1.15, but assuming a contribution of $\sigma_0 = 100$ $\mu as\ yr^{-1}$ from source instability.

B (mag)	P	$N_{\rm QSO}$	$\sigma_{\mu,\rm tot}$ (μas yr^{-1})	$\sigma(\omega_1)$	$\sigma(\omega_2)$	$\sigma(\omega_3)$ (μas yr^{-1})	$\sigma(a_1/c)$	$\sigma(a_2/c)$	$\sigma(a_3/c)$ (μas yr^{-1})
14 – 15	1.0	40	100	19	19	23	19	19	23
15 – 16	1.0	230	100	7.67	7.65	9.21	7.67	7.65	9.21
16 – 17	0.9	1 230	101	3.31	3.29	3.92	3.32	3.29	3.92
17 – 18	0.8	11 500	103	1.11	1.12	1.35	1.12	1.12	1.35
18 – 19	0.6	60 000	109	0.52	0.52	0.62	0.52	0.52	0.62
19 – 20	0.3	97 000	127	0.47	0.47	0.57	0.47	0.47	0.56
all		170 000		0.33	0.33	0.40	0.33	0.33	0.39

The Pound-Rebka experiment verified the relativistic prediction of a gravitational redshift for photons, an effect probing the time-time component of the metric tensor. Light deflection depends on both the time-space and space-space components. It has been observed, with various degrees of precision, on distance scales of $10^9 - 10^{21}$ m, and on mass scales from $1 - 10^{13} M_\odot$, the upper ranges determined from the gravitational lensing of quasars (Dar 1992). GAIA will extend the domain of observations by two orders of magnitude in length, and six orders of magnitude in mass.

Consider a star at angular separation χ from a body with mass M at a distance r from GAIA. To first order in M/c^2, and neglecting departures from spherical geometry, the deflection angle is:

$$\delta\chi = \frac{(1+\gamma)GM}{c^2 r \tan\chi/2} \tag{3}$$

where γ is the PPN-parameter, equal to unity in General Relativity, and c is the velocity of light. The deflection caused by the quadrupole term of the gravitational field is given by:

$$\delta\chi = \frac{2(1+\gamma)GM}{c^2 R}\, J_2\, (1 - \sigma_z^2 - 2d_z^2)\Big(\frac{R}{d}\Big)^2 \tag{4}$$

where R is the radius of the deflector, J_2 its dynamical flattening, d the impact parameter of the light ray, σ_z the component along the spin axis of the body, and d_z the same for the impact parameter vector.

Table 1.17: Light deflection by masses in the Solar system. The monopole effect dominates, and is summarized in the left columns for grazing incidence and for typical values of the angular separation. Columns $\chi_{\min}$ and $\chi_{\max}$ give results for the minimum and maximum angles accessible to GAIA. J_2 is the quadrupole moment. The magnitude of the quadrupole effect is given for grazing incidence, and for an angle of 1°. For GAIA this applies only to Jupiter and Saturn, as it will be located at L2, with minimum Sun/Earth avoidance angle of 35°.

Object	Monopole term					Quadrupole term		
	Grazing μas	$\chi_{\min}$ μas	$\chi = 45°$ μas	$\chi = 90°$ μas	$\chi_{\max}$ μas	J_2	Grazing μas	$\chi = 1°$ μas
Sun	1750000	13000	10000	4100	2100	$\leq 10^{-7}$	0.3	–
Earth	500	3	2.5	1.1	0	0.001	1	–
Jupiter	16000	16000	2.0	0.7	0	0.015	500	7×10^{-5}
Saturn	6000	6000	0.3	0.1	0	0.016	200	3×10^{-6}

Table 1.17 gives the magnitude of the deflection for the Sun and the major planets, at different values of the angular separation χ, for the monopole term (Equation 3) and the quadrupole term (Equation 4). While χ is never smaller than 35° for the Sun (a constraint from GAIA's orbit), grazing incidence is possible for the planets. With the astrometric accuracy of a few μas, the magnitude of the expected effects is considerable for the Sun, and also for observations near planets.

The parametrization of the gravitational deflection with the dimensionless parameter γ belongs to the 'parametrized post-Newtonian' (PPN) formalism, with which a wide class of metric theories of gravitation and their experimental consequences are described in a multi-dimensional parameter space. In the simplest case this space has two dimensions, γ and β, with General Relativity at (1,1). Different experiments are governed by different combinations of these two parameters: for example, for the secular advance of the periastron we have $(2 - \beta + 3\gamma)/3$; for the violation of the equivalence principle the combination is $(4\beta - 3 - \gamma)$ (Nordtvedt 1995), which occurs in lunar theory if the contributions of the internal binding energy of the Earth to the inertial and gravitational masses are different. Although most alternative theories of gravity have been discarded, the Brans-Dicke theory (Brans & Dicke 1961) remains a conceptually attractive competitor containing, beside the metric tensor, a scalar field ϕ and an arbitrary coupling constant ω, related by $\gamma = (1 + \omega)/(2 + \omega)$. The present limit on γ (Nordtvedt 1995) gives the constraint $|\omega| > 500$. Other generalized scalar-tensor theories have been proposed in which $\omega = \omega(\phi)$, and which therefore changes in time with the evolution of the Universe. Because of recent developments in cosmology (e.g. inflationary models) and elementary-particle physics (e.g. string theory and Kaluza-Klein theories), these scalar-tensor theories are considered as interesting alternatives to General Relativity. A large class of such theories contain an attractor mechanism towards General Relativity in a cosmological sense (Damour & Nordtvedt 1993a; Damour & Nordtvedt 1993b); if this is how the Universe is evolving, then today we can expect discrepancies of the order of $|\gamma - 1| \sim 10^{-7} - 10^{-5}$ depending on the theory (Damour & Nordtvedt 1993b). This kind of argument provides a strong motivation for any experiments able to reach these accuracies. Additional terms caused by gravitational effects on light rays at the μas level include the 'frame-dragging' effects of the motions and rotations of the Sun and the planets (Soffel 1989). The post-PPN term for a grazing ray is 7 μas for the Sun. It can be neglected for the planets, and also for the Sun at the large angles where GAIA will observe.

The astrometric residuals can be tested for any discrepancies with the prescriptions of General Relativity. This provides a constraint on the PPN term γ. Figure 1.41 compares the determination of γ derived from the final Hipparcos data with previous determinations by other means. Although the strong correlation between the deflection and parallactic displacement limits the precision, γ was found to agree with the General Relativistic value to within ± 0.003. A corresponding analysis indicates that the GAIA measurements will provide a precision of about 5×10^{-7} for γ, based on multiple observations of $\sim 10^7$ stars with $V < 13$ mag at wide angles from the Sun, with individual measurement accuracies better than 10 μas. In comparison with Hipparcos, the individual observations are much more precise and the Solar avoidance angle is smaller, which at the same time increases the signal and decreases the correlation with the parallax. As described above, this accuracy is close to the values predicted by theories which predict that the Universe started with a strong scalar component, which relaxes to the general relativistic value with time (e.g., Damour & Nordtvedt 1993a; Damour & Nordtvedt 1993b).

Space experiments purely dedicated to the measurement of γ with a precision of about 1 part in 10^6 have been

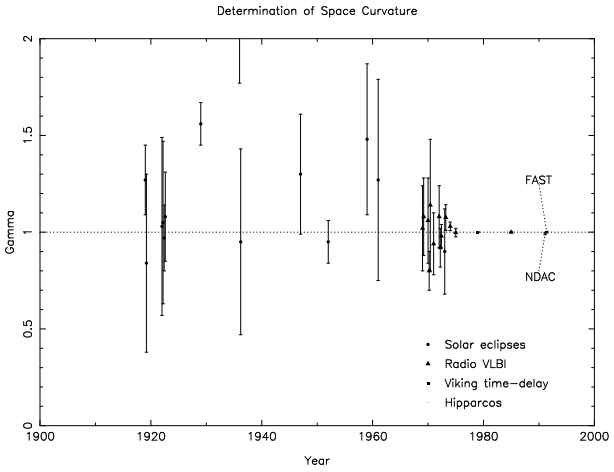


Figure 1.41: Determinations of the post-Newtonian parameter γ, representing the deviation of the gravitational light bending from Newtonian theory. According to General Relativity, $\gamma = 1$. Non-Hipparcos data are taken from Soffel (1989), and the Solar eclipse observations include the most recent measurements of Jones (1976). The value derived in the Hipparcos sphere solution process is shown, together with other determinations based on Solar eclipse observations, VLBI observations and the Viking spacecraft (Shapiro) time-delay. The Hipparcos results are derived from observations at large Solar angles, while all other metric determinations have been based on observations within a few $R_\odot$ of the Solar limb.

proposed (Veillet et al. 1993). Doppler measurements of the Solar gravitational deflection provide alternative routes to γ (Bertotti & Giampieri 1992; Giampieri 1996; Bertotti & Giampieri 1998), with experiments to be carried out with the Cassini mission in 2002–03 expected to provide a value of γ with a precision of $\sim 10^{-5}$ (Iess et al. 1999). Gravity Probe-B, a relativity gyroscope experiment, is expected to improve the accuracy in γ to about 6×10^{-5}, while plans for measurements with ESA's Mercury orbiter aim at levels of $\sim 10^{-5}$.

Observations of a star nearly aligned with a giant planet will be relatively rare, and while the effect must be accounted for in the data reduction, it will not put very strong constraints on space curvature.

1.9.2 Perihelion Precession of Minor Planets

GAIA will observe and discover several hundred thousand minor planets during its five year mission (Section 1.7.1). Most of these will belong to the asteroidal main belt, with small orbital eccentricity and semi-major axes close to 3 AU, a circumstance not favorable to see significant relativistic effects in their motion. This is true for the main belt asteroids, but not for the Apollo and Aten groups, which are all Earth-orbit crossers, and include objects with semi-major axes of the order of 1 AU and eccentricities as large as 0.9. The Amor group have perihelia between 1–1.3 AU, and approach the Earth but do not cross its orbit.

The relativistic effect and the Solar quadrupole cause the orbital perihelion of a Solar system body to precess at the rate (Will 1981; Shapiro 1989):

$$\Delta\varpi = \frac{6\pi\lambda G M_\odot}{a(1-e^2)c^2} + \frac{3\pi J_2 R_\odot^2}{a^2(1-e^2)^2} \tag{5}$$

where $\lambda = (2\gamma - \beta + 2)/3$ is the PPN precession coefficient, and the rate is given in radians per revolution. For the main belt, the precession is about seven times smaller than for Mercury in rate per revolution, although more than a hundred times in absolute rate.

Three cases of earth-crossing asteroids are considered in Table 1.18 giving perihelia precession larger than Mercury's, due to a favorable combination of distance and eccentricity. The diameters are of the order of 1 km for Icarus and Talos and 4 km for Phaeton. Observed at a geocentric

Table 1.18: Perihelion precession due to General Relativity and the Solar quadrupole moment.

Body	a AU	e	GR mas/rev	GR mas/yr	$J_2(=10^{-6})$ mas/rev
Mercury	0.39	0.21	102	423	0.30
Asteroid (main belt)	2.7	0.1	15	3.4	0.006
1566 Icarus	1.08	0.83	114	102	0.34
5786 Talos	1.08	0.83	114	102	0.34
3200 Phaeton	1.27	0.89	148	103	0.57

distance of 1 AU, these objects have a magnitude between $V = 15 - 17$ mag and an angular diameter of 4 mas and 1 mas respectively. Thus the astrometric measurements will be of good quality, virtually unaffected by the finite size of the source. A determination of λ with an accuracy of 10^{-4} is a reasonable goal, with a value closer to 10^{-5} probably attainable from the statistics on several tens of planets.

A dedicated simulation has to be performed to assess the real capabilities of GAIA in this field and to determine the number of objects that could be used during the mission for this purpose. With γ known (also from GAIA) to a much better accuracy, this translates into a value of $\sigma(\beta) \sim 3 \times 10^{-4} - 3 \times 10^{-5}$, one or two orders of magnitude better than the best determination from lunar laser ranging, either from the direct fit to the data or from the non-detection of a violation of the equivalence principle (Williams et al. 1996).

The independent determination of the Solar quadrupole moment J_2 requires good sampling in $a(1-e^2)$, and one can expect a result better than 10^{-7}. Today, an upper bound for J_2 of 3×10^{-6} is found from the effect (null experiment) of the Solar J_2 on lunar librations (Bois & Girard 1998) and planetary motions. Ground-based and SOHO measurements of the Solar oblateness $(9 \pm 0.6 \times 10^{-6})$, combined with a model of the Solar interior and of the differential rotation, give $J_2 < 2 \times 10^{-7}$ (Sofia et al. 1994; Lydon & Sofia 1996). The Picard mission in 2003–2004 will do much better for the oblateness. Despite this improvement, the quadrupole moment will remain model dependent, and a direct determination from dynamics with GAIA will be worthwhile and feasible, although the accuracy is difficult to assess without extensive simulations.

1.9.3 Secular Change of the Gravitational Constant

The possibility of a time variation of the constant of gravitation, G, was first considered by Dirac (1938) on the basis of his large number hypothesis, and later developed by Brans & Dicke (1961) in their theory of gravitation. Variation could be related to the expansion of the Universe, in which case $\dot{G}/G = \sigma H_0$, where H_0 is the Hubble constant, and σ is a dimensionless parameter whose value depends on both the gravitational constant and the cosmological model considered (Will 1987), with the standard model having G = constant and $\sigma = 0$. Revival of interest in the Brans-Dicke-like theories, with a variable G, was partially motivated by the appearance of superstring theories where G is considered to be a dynamical quantity (Marciano 1984). A scale-dependent gravitational constant could mimic the presence of dark matter (Goldman et al. 1992) and could enter discrepancies between the determinations of H_0 at different scales (Bertolami et al. 1993). The way in which the massive census of white dwarfs will impact the determination of time-dependent changes in G was described in Section 1.4.7. There are two aspects of the use of white dwarfs to this problem. First, when they are cool enough, their energy is entirely of gravitational and thermal origin, and any change in G modifies the energy balance which in turn modifies the luminosity. Second, since they are long-lived objects, with life-times of order 10 Gyr, even extremely small values of $\dot{G}$ can become prominent.

Astrophysical constraints on $\dot{G}$ have been obtained using observations of lunar occultations and eclipses, and planetary and lunar radar-ranging measurements (Will 1987), the evolution of the Sun (Pochoda & Schwarzschild 1964; Demarque et al. 1994; Guenther et al. 1998), primordial nucleosynthesis (Olive et al. 1990; Walker et al. 1991), gravitational lensing (Krauss & White 1992), and the white dwarf luminosity function (García-Berro et al. 1995). Best constraints are currently from helioseismology (Guenther et al. 1998) giving $\dot{G}/G \leq -1.6 \times 10^{-12}$ yr^{-1};

distance measurements to the Viking landers (Hellings 1987; Reasenberg 1983) with upper bounds in the range $\dot{G}/G \leq -(3 \pm 0.6) \times 10^{-11}$ yr^{-1}, and limited by the accuracy of the mass and distribition of minor planets, which will be improved significantly by GAIA; from the binary pulsar PSR 1913+16 (Damour et al. 1988) with a limit of $\dot{G}/G \leq -(1.10 \pm 1.07) \times 10^{-11}$ yr^{-1}, and limited by the proper motion adopted for the pulsar; and from white dwarf studies (García-Berro et al. 1995) giving $\dot{G}/G \leq -(1 \pm 1) \times 10^{-11}$ yr^{-1} if earlier stratified models are adopted.

Estimates of the upper bound on $\dot{G}/G$ that GAIA can achieve have two aspects. First, the error on the upper limit on the rate of variation of G given by García-Berro et al. (1995) essentially depends on the number of white dwarfs in the lowest luminosity bin. Since GAIA will detect numerous white dwarfs in this bin the errors can be reduced by a factor of roughly 5. Second, in order to reduce the upper limit itself, an age of the Solar neighbourhood independent of the white dwarf luminosity function is required, since the age of the disk is used in obtaining this upper limit. Assuming this is available, the upper limit could be decreased to $10^{-12} - 10^{-13}$ yr^{-1}, much better than the upper limit derived from the Hulse-Taylor binary pulsar (Damour et al. 1988).

The possibility that other physical constants vary has received attention in the context of unified theories like string theory, in which additional compact dimensions of space may exist. Although these theories do not require traditional constants to vary, they allow a rigorous description of any variations to be provided, using a self-consistency approach because of the extra dimensions of space in these theories (see, e.g., Drinkwater et al. 1998 and references therein). Recent detections of high-redshift absorption by both atomic hydrogen and molecular gas in the radio spectra of quasars have provided one such powerful tool for measuring possible temporal and spatial variations in the proton g-factor, g_p, and the fine-structure constant, α, with Drinkwater et al. (1998) giving $|\dot{g}_p/g_p| < 1 \times 10^{-15}$ yr^{-1} and $|\dot{\alpha}/\alpha| < 5 \times 10^{-16}$ yr^{-1} respectively between the present epoch and $z = 0.68$. Any change in the value of α will also impact white dwarf cooling rates, an effect which has not addressed in present considerations. Further studies are required to assess their implications on the values of $\dot{G}/G$ and, more generally, whether the GAIA census can hope to place independent limits on the variation of these quantities.

1.9.4 Cosmological Shear and Rotation, Mach's Principle

Astrometric VLBI data now determine proper motions for more than a hundred sources to better than 50 μas yr^{-1}, and permit metric terms to be estimated (Gwinn et al. 1997). By averaging over many sources, and taking account of individual statistically significant motions, presumably due to structural changes, Eubanks (private communication) reports the bounding of large-scale deformations providing the following interesting cosmological constraints: (i) most nearby sources are subluminal, with the smallest transverse motion, for M81, being 600 ± 900 km s^{-1}; (ii) our peculiar acceleration is $< 5 \times 10^{-10}$ m s^{-2}, constraining the mass of our Galaxy to $< 2.4 \times 10^{11} M_{\odot}$; (iii) the local velocity shear is $< 3 \times 10^{-10}$ yr^{-1}; (iv) the local vorticity is $< 9 \times 10^{-11}$ yr^{-1}; (v) the local gravitational radiation flux is $< 0.11 h^{-2}$ times the closure density for periods > 20 years. No detailed evaluation of the corresponding cosmological parameters from the GAIA observations has been made, but the very large number of proper motions at this accuracy level or better, suggests that important contributions will be made to cosmology in these areas by GAIA. Studies related to the detection of galactocentric acceleration (see Section 1.8.10) are also related to these issues of large-scale structure.

1.9.5 Gravitational Waves

The relevance of gravitational waves for high-precision astrometry has been examined in a number of recent publications (e.g., Braginsky et al. 1990; Braginsky & Grishchuk 1985; Fakir 1994b; Fakir 1994a; Durrer 1994; Kaiser & Jaffe 1997; Pyne et al. 1996; Gwinn et al. 1997; Damour & Esposito-Farèse 1998; Kopeikin et al. 1999). In particular, Gwinn et al. (1997) have already carried out a gravity wave search using VLBI measurements of quasar proper motions, which typically have accuracies of ~ 30 μas yr^{-1} (comparable to GAIA's).

The basic theoretical inputs to such a search are as follows. Consider the position measurement of an uncollimated point source (such as a star or quasar) by a freely falling telescope. Gravitational waves passing over the telescope will cause a time-varying shift in the apparent position of the source; i.e., the waves cause apparent proper motions. The angular displacement is order of h, the magnitude of the gravitational wave field at the telescope. For distant sources (distance $D \gg \lambda$, where λ is the gravitational wavelength), the effect from the waves passing over the telescope is far

greater than from either the waves passing over the source or over the photon trajectory between the source and telescope (Pyne et al. 1996; Damour & Esposito-Farèse 1998). The fact that the apparent motions are determined by the local gravitational wave field at the telescope implies that the motions are coherent across the whole sky; the relative motion of two nearby sources is proportional to their angular separation. That is, the angular shift for any individual star is of order h, but the relative shift for two nearby stars, separated by angle ϵ, is of order $h\epsilon$.

Fakir (1994b); Fakir (1994a) and Durrer (1994) claimed that when the photon trajectory passes particularly close to a gravitational wave source (within of order the gravitational wavelength) the measured deflection angle could be as large as $h(b)$, where b is the impact parameter, and conceivably this could have been detected by GAIA; however more complete calculations by Damour & Esposito-Farèse (1998) and Kopeikin et al. (1999) have shown these claims to be erroneous.

For example, consider the effect of planar gravitational wave passing over the Earth in the z-direction. Let the wave be '+' polarized, with frequency ω and magnitude h_+. For this case, Pyne et al. (1996) have shown that a source located at (θ, ϕ) undergoes an apparent sinusoidal displacement $\vec{\delta}$ given by:

$$\vec{\delta}(t) = \tfrac{1}{2} h_+ \cos\omega t[\sin\theta\big(\cos 2\phi\hat{\theta} - \sin 2\phi\hat{\phi}\big)] \tag{6}$$

where $\hat{\theta}$ and $\hat{\phi}$ are the unit vectors in the θ- and ϕ-directions. Equation (6) in fact summarizes everything needed to know about apparent motions from gravitational waves, since the waves bathing the Earth can be decomposed into a sum of plane waves with different directions and polarizations, and since any direction and polarization can be obtained from the one in the above example just by a rotation.

In their analysis of VLBI data, Gwinn et al. (1997) advocate expanding the proper motion field $\dot{\vec{\delta}}$ into vector spherical harmonics, as follows:

$$\dot{\vec{\delta}} = \sum_{l,m} \big(a_{l,m}\vec{\nabla}Y_{lm} + b_{l,m}\hat{r}\times\vec{\nabla}Y_{lm}\big) \tag{7}$$

where $\hat{r}$ is the radial unit vector, and where the reality of $\dot{\vec{\delta}}$ demands that:

$$a_{l,+m} = (-1)^m a^*_{l,-m} \quad b_{l,+m} = (-1)^{m+1} b^*_{l,-m} \tag{8}$$

Assuming that waves from different directions are uncorrelated and have random polarization, then $\frac{5}{6}$ of the 'power' in this expansion is in the $l = 2$ harmonics. So, ignoring higher harmonics, there are 10 independent (real) coefficients to try to measure. In Equation 7 attention is implicitly restricted to waves with periods larger than the observation time, so the expansion coefficients are essentially constant during the observation. That is, attention has been restricted to gravity wave frequencies $f < 1/T$, where T is of order the observation time. This is the frequency regime of interest for both VLBI and GAIA, because they examine only a tiny fraction of the sky at any instant, and because one must rely on the coherence of proper motions across the sky to build up signal-to-noise. It follows that T should be something like the time needed to map a good fraction of the sky to high accuracy, which for GAIA translates to at least half the mission lifetime. So GAIA will be most sensitive to gravity waves in roughly the frequency range $10^{-11} < f < 10^{-8}$ Hz. The lower end of this frequency range corresponds to the distance to a 'typical' source, taken to be ~ 1 kpc.

What are present upper limits on the gravity wave background in GAIA's frequency range, and how do they compare to levels that GAIA would be able to detect? It is useful to state the limits in terms of $\Omega_{\rm gw}(f)$, the ratio of the energy density in gravitational waves to the energy density needed to close the Universe:

$$\Omega_{\rm gw}(f) \equiv \frac{1}{\rho_c}\frac{d\rho_{\rm gw}(f)}{d\ln f} \tag{9}$$

Here $\rho_{\rm gw}(f)$ is the energy density in gravitational waves and ρ_c is the closure density. The characteristic gravitational wave amplitude at frequency f, $h_c(f)$, is related to $\Omega_{\rm gw}(f)$ by:

$$h_c(f) = 3\times 10^{-10}\sqrt{\Omega_{\rm gw}(f)}h_{100}\,\frac{10^{-8}\text{ Hz}}{f} \tag{10}$$

where $H_0 \equiv h_{100}100$ km s^{-1} Mpc^{-1}. Defining $\Omega_{\rm gw}$ to be the total integrated gravitational wave energy in all bands, $\Omega_{\rm gw} \equiv \int \Omega_{\rm gw}(f) d\ln f$, and $\Omega_{\rm gw}^{GAIA}$ to be that part of $\Omega_{\rm gw}$ that is inside the GAIA band, then, as shown in Pyne et al. (1996), the rms proper motion that GAIA would detect is given by:

$$\langle \dot{\vec{\delta}}^2 \rangle^{1/2} = H_0 \, (\Omega_{\rm gw}^{GAIA})^{1/2} = 21 \, \mu{\rm arcsec/yr} \, h_{100} \, (\Omega_{\rm gw}^{GAIA})^{1/2} \tag{11}$$

Big Bang nucleosynthesis gives a strong constraint on the total gravitational-wave energy in all bands: $\Omega_{\rm gw} < 10^{-5}$ (Allen 1997). This assumes the waves are generated in the early Universe, before Big Bang nucleosynthesis. The strongest limits on $\Omega_{\rm gw}(f)$ come from millisecond pulsar timing: $h_{100}^2 \Omega_{\rm gw}(f \sim 4 \times 10^{-9}$ Hz$) < 10^{-8}$, or (roughly) $h_c < 5 \times 10^{-14}$ at $f \sim 4 \times 10^{-9}$ Hz (Thorsett & Dewey 1996). (In order of magnitude, $h_c < 5 \times 10^{-14}$ comes from 10^{-6} s residuals divided by total observation times of $\sim 10^8$ s.) Limits from millisecond pulsar timing are much weaker for frequencies lower than 4×10^{-9} Hz $\sim 1/$(total observation time), since on longer time scales the gravity wave effect gets absorbed into the fit for the pulsar period and period derivative. At lower frequencies, one obtains the following limits from measurements of the orbital decay of PSR 1913+16: $h_{100}^2 \Omega_{\rm gw}(f) < 0.04$ for $10^{-11} < f < 4.4 \times 10^{-9}$ Hz and $h_{100}^2 \Omega_{\rm gw}(f) < 0.5$ for $10^{-12} < f < 10^{-11}$ Hz (Thorsett & Dewey 1996). However, the most stringent results presently available are based on B1855+09, and limit $\Omega_g h^2 < 2.710^{-4}$ in the frequency range $10^{-11} < f < 4.4 \times 10^{-9}$ Hz (Kopeikin 1997). VLBI astrometry currently gives the constraint $h_{100}^2 \Omega_{\rm gw}(f) < 10^{-1}$ in (roughly) the range $10^{-17} < f < 10^{-9}$ Hz (Gwinn et al. 1997). The VLBI constraint extends to extremely low frequencies because the radio sources are quasars, with distances of order the Hubble length. This constraint might improve by a factor 10–100 in the next 10 years.

The bounds from Big Bang nucleosynthesis models indicate that gravitational-wave proper motions, for any individual source, will be less than 0.1 μas yr^{-1}, and pulsar timing measurements strongly suggest that they will be less than ~ 0.002 μas yr^{-1}. Thus gravitational waves will not significantly affect any individual position measured by GAIA. However, the entire GAIA data set could be used to put the strongest limit on $\Omega_{\rm gw}$ in the band $10^{-12} < f < 4 \times 10^{-9}$ Hz. Roughly, the GAIA limit on $\langle \dot{\vec{\delta}}^2 \rangle^{1/2}$ should scale like $1/\sqrt{N_s}$, where N_s is the number of independently measured point sources across the sky. So the limit on $\Omega_{\rm gw}$ scales like $1/N_s$. Assuming $N_s \sim 10^8$ with a 'typical' accuracy of 100 μas yr^{-1}, it follows that GAIA could set an upper limit of roughly $\Omega_{\rm gw} < 10^{-6} - 10^{-7}$ in this frequency band. That is better than the Big Bang nucleosynthesis limit, and much better than the limit from VLBI or the binary pulsar. Of course, this upper limit assumes that other 'noise' in GAIA's proper-motion map does not have coherent large-scale correlations that look at all like $l = 2$ vector spherical harmonics–at the 0.01 per cent level. If 1 per cent of the errors have large-scale correlations looking like $l = 2$ vector harmonics, then GAIA's upper limit on $\Omega_{\rm gw}$ would be $\sim 10^{-3}$ (better than pulsars or VLBI, but not as strong as the Big Bang nucleosynthesis limit).

On the assumption that coherent errors are not a problem, GAIA could set, in the $10^{-12} < f < 10^{-10}$ Hz band, the best upper limit on $\Omega_{\rm gw}$. Of course, in principle the fact that GAIA could set the best upper limit also means that GAIA could make the first detection of a stochastic gravitational wave background. How likely is detection? The three favorite theoretical scenarios for generating a gravitational wave background in the early Universe are inflation, cosmic strings, and colliding bubbles during first-order phase transitions (such as the electroweak phase transition). Both inflation and cosmic strings predict a flat spectrum for $\Omega_{\rm gw}(f)$, so for these sources the millisecond pulsar limit already rules out detection by GAIA. Phase transitions give a strongly peaked spectrum, but almost certainly peaked at frequency much higher than the GAIA band. Thus, from current theoretical ideas about gravity wave generation in the early Universe, it seems much more likely that GAIA would set the best upper limit, not make a detection.

Nanoarcsecond Astrometry Finally, what are the implications of gravitational waves for astrometry missions beyond GAIA, e.g., for nanoarcsec astrometry? From the above, it is clear that if $\Omega_{\rm gw}$ is near the present upper limits, then the gravity wave background could cause ~ 1 nanoarcsec yr^{-1} proper motions of individual source positions. It does not seem likely that $\Omega_{\rm gw}$ is really near the upper limits, so this may not be a serious design consideration. What is perhaps more interesting is the fair chance nanoarcsec astrometry would have of making the first detection of a cosmological gravity wave background. The standard inflationary model predicts the existence of a stochastic background and gives its spectrum, but not its overall amplitude. If one normalizes the amplitude by the COBE observations (assuming that a reasonable fraction of the observed $\delta T/T$ is due to 10^{-18} Hz gravitational waves), then $\Omega_{\rm gw}(f)$ is predicted to be roughly 10^{-13} for frequencies $10^{-16} < f < 10^{10}$ Hz. This is four orders of magnitude too small to be observed by even the Advanced LIGO detectors (expected to come on-line around 2008), but would probably be measurable by a 'super' GAIA with nanoarcsec yr^{-1} accuracy; 10^6 sources with proper motions measured to this level would permit detection of a stochastic background down to the level $\Omega_{\rm gw} \sim 10^{-14}$ (always assuming that the gravity wave effect is larger than other noise sources that mimic its large-scale coherence).

1.10 Scientific Topics Beyond GAIA

In this section we collect some specific scientific issues which could in theory be tackled by high-accuracy astrometric measurements, but which remain inaccessible to GAIA's capabilities.

Cores of Globular Clusters The stellar densities in the inner regions of many of the galactic globular clusters are larger than GAIA can handle. HST will in principle be able to get the proper motions in many of these, but in a very small field (e.g., King et al. 1998). This is then complemented by GAIA for the outer regions.

Galactic Centre Ground-based proper motion studies by Genzel (1998), and Ghez et al. (1998) have put the strongest constraints yet on the mass of the black hole in the Galactic centre. GAIA cannot improve on these, as the reddening at I is too large.

The Curvature of Stellar Orbits Stars do not move along straight lines but have curved orbits in the Galactic potential. GAIA will measure the angular displacements of stars in the sky due to the relative motion between stars and the Sun. Does the curvature of the relative motion lead to a measurable effect in the astrometry? For nearby stars this difference will always be negligible and thus the curvature of their orbits is of no consequence to GAIA.

The motion of a star can be described by its position vector as a function of time, $\mathbf{r}(t)$, which can be found from the velocity $\mathbf{v}(t)$ and the acceleration $\mathbf{a}(t)$. Because of the very short time-span during which the relative motion between Sun and star will be measured the acceleration will remain constant and the relative motion can simply be described as: $\Delta\mathbf{r}(t) = \Delta\mathbf{r}_0 + \Delta\mathbf{v}_0\Delta t + \Delta\mathbf{a}\Delta t^2/2$. The acceleration term in the relative motion will only give rise to a significant additional angular displacement on the sky (in ~ 5 years time) for nearby stars. However for these stars $|\Delta\mathbf{a}|$ is very small. For example, if we take the Allen & Santillán (1991) model for the Galactic potential (convenient because all quantities can be evaluated analytically), the maximum value of $|\Delta\mathbf{a}|$ in a $200\times200\times200$ pc^3 volume centred on the Sun is $\sim 10^{-12}$ pc yr^{-2}. This leads to an extra displacement due to acceleration in 5 years of $\sim 2.5\times10^{-11}$ pc, which corresponds to ~ 0.05 μas at 100 pc. This is well below the measurement accuracy of GAIA. Another way to look at this problem, without resorting to the approximation of constant acceleration in time, is to realize that the relevant parameter is the curvature of the relative orbit of the star and the Sun. The curvature is given by: $\kappa = |\Delta\mathbf{a}\times\Delta\mathbf{v}|/|\mathbf{v}|^3$, where $1/\kappa$ is the radius of curvature of the relative orbit. Because the magnitude of $\Delta\mathbf{a}$ is very small for nearby stars, κ will also be small and the relative orbit will be a straight line.

The Effect of Local Density Enhancements The local Galactic potential is not completely smooth. There are spiral arms, molecular clouds, open clusters, etc. Will any local density enhancements have a measurable effect on the precision at which GAIA will perform its measurements? It turns out that this lumpiness is not measurable with GAIA.

The effects of local density enhancements such as molecular clouds can be estimated in a simple manner, as it is the relative acceleration of the Sun and other stars that matters. Consider the Sun, a star and the density enhancement represented as a Plummer sphere (molecular cloud, cluster) or a point mass (a possible black hole). Assume then that all objects are stationary. This assumption is justified, because even at relative velocities of hundreds of km s^{-1} the relative displacements of the objects will be at most $\sim 10^{-3}$ pc during 5 years. So for any reasonable configuration the relative positions of the objects will not change appreciably. Due to the symmetry around the Sun-density-enhancement line it is straightforward to map out acceleration difference between the Sun and the star due to the density perturbation. In the case of a nearby molecular cloud with a typical size of 45 pc one can use a Plummer sphere with a core radius of about 10 pc to model the cloud. The nearest known giant molecular clouds are the Taurus clouds at 140 pc. Assuming they have a mass of 2×10^5 M$_\odot$ (they are in fact 10 times less massive), the maximum extra angular displacement of a star perpendicular to the line of sight is ~ 0.1 μas. The same number is found for a giant molecular cloud of 10^6 M$_\odot$ at 500 pc.

In the case of a black hole in the Solar vicinity a star has to pass within 0.01–0.05 pc of a black hole of 100 M$_\odot$ (placed between 10–100 pc from the Sun) in order for the extra acceleration to amount to microarcsec-level effects. These numbers obviously change as a function of the black hole's mass. The chance that such an encounter between a star and the black hole may take place depends on the density of black holes in the Solar neighbourhood. So the question then is: how many black holes are there in the Solar neighbourhood and what mass-range do we need to consider? This issue was reviewed by Carr (1994), who showed that black holes are ruled out as important baryonic dark matter components over their whole mass-range. His Figure 3 shows that dynamical constraints (such as disk heating and the existence of wide binaries) exclude the existence of black holes outside the stellar mass range (1–100 M$_\odot$) in the Galactic disk. The existence of wide binaries excludes any objects of mass larger than 2 $M_\odot$ as candidates for disk dark matter. The same figure also shows that the density of stellar-mass black holes in the

disk has to be lower than 0.001 times the critical density of the Universe. This translates to a number density of less than $\sim 2 \times 10^{-12}$ pc^{-3} for 100 M$_\odot$ black holes. A much higher number on the density of stellar-mass black holes is obtained from considerations of their birth-rate, which depends on the stellar initial mass function and star formation history in the Galaxy. Shapiro & Teukolsky (1983) calculate that there are 8×10^{-4} black holes per pc^3 with masses larger than 10 M$_\odot$. Combining this number with the 'impact parameters' estimated above means that only in a fraction $\sim 5 \times 10^{-8}$ of the accessible volume can encounters between stars and black holes occur such that a signature in the GAIA astrometry may be seen. Hence a few of such events may occur within a couple of hundred parsec from the Sun. But most likely these interactions are of no consequence to GAIA given the stricter bound on the number density of stellar-mass black holes given by Carr (1994). Very massive black holes ($\sim 10^6$ M$_\odot$) will have an effect out to much larger distances and they may exist in the halo. But assuming all these black holes to be 10^6 M$_\odot$ and using the upper limit of 0.1 times the critical density for the halo black holes (Carr 1994) leads to a number density less than $\sim 2 \times 10^{-14}$ pc^{-3}, again making it very unlikely that GAIA will detect the astrometric signature of a star passing close to a black hole.

Geometric Cosmology Future astrometric missions, at levels of accuracy very much better than 1 μas, could determine the transverse motions of external galaxies and quasars routinely, and determine their kinematic properties independently of a dynamical model of the Universe.

Part II
Technical Design

2 Overall Design Considerations and Concept Development

As demonstrated in Section 1, an experiment which could provide extremely accurate three-dimensional position and velocity measurements as well as colours for every object in the sky down to faint magnitude levels would have extremely broad scientific appeal. With 10 μas accuracies, distances of objects throughout the Galaxy would be measured, with a 10 per cent accuracy even as far away as the Galactic Centre, and space velocities would be acquired with an accuracy of a few km s^{-1} even at 20 kpc. The colours allow derivation of fundamental astrophysical parameters of the stars, so that a complete description of the stellar populations in the Galaxy and its nearest neighbours would be achieved.

This Section describes the requirements that follow from the science goals, considering in turn the astrometry, the radial velocities, the astrophysical diagnostics and the need for on-board detection of the objects. It concludes with a summary of the conceptual development that has led to the present design of the GAIA payload and spacecraft.

2.1 Astrometry

Astrometry provides the only fully reliable way of measuring distances in the local Universe. A space astrometry mission aiming at the widest range of astrophysical problems must be able to perform *global* measurements. This means that positions, and changes in positions caused by proper motion and parallax, are determined in a reference system consistently defined over the whole sky. In practice this requires direct and accurate measurement of large angles (of the order of 1 rad). Hipparcos demonstrated that this can be achieved with mas accuracy by means of a continuously scanning satellite which observes two directions simultaneously. When combined with newly-developed technology, this same measurement principle can be used to gain, simultaneously, a factor of more than 100 improvement in accuracy, a factor 1000 improvement in limiting magnitude, and a factor of 10 000 in the numbers of stars observed.

2.1.1 From Milliarcseconds to Microarcseconds

The measurements conducted by a continuously scanning satellite can be shown to be almost optimally efficient, with each photon acquired during a scan contributing to the precision of the resulting astrometric parameters (Section 2.1.2). Every object crossing the field of view above a certain signal-to-noise threshold can be observed and will be ultimately contained within the final catalogue, irrespective of its distance, its space motion (i.e., including minor planets), or its photometric variability (i.e., including highly variable or burst sources). The conceptual appeal of being able to devote more observing time to a particular object of high scientific interest by means of a payload which can 'stop-and-stare' at a given region of sky is completely outweighed by the extraordinarily high accuracy that is achievable on such a large number of objects.

Pointed observations cannot provide the over-riding benefit of global astrometry using a scanning satellite, which is that a global instrument calibration can be performed in parallel, and the interconnection of observations over the celestial sphere provides the rigidity and reference system, immediately connected to an extragalactic reference system, needed for the kinematical interpretation of the individual observations themselves. The number of distinct observations per star is very much larger than the number of variables characterising the stellar motion. This allows an accurate and unambiguous calibration of the instrumental geometry, and simultaneous determination of the standard errors of the astrometric parameters which provide a realistic indication of the true errors. In this sense the overall instrument is self-calibrating, and the resulting astrometric parameters are determined along with estimates of their standard errors and correlations. For the rigorous scientific exploitation of the astrometry such confidence in the error estimates is crucial.

A further important feature of global astrometric data is the capability of determining the astrometric parameters of double and multiple systems, including brown dwarfs, and a vast number of extra-solar planetary systems. These aspects pose a considerable and continual challenge to the instrument design, the data acquisition, the data analysis, and the final catalogue production. The scanning satellite concept allows a semi-continuous sampling of the double or multiple star geometry, which is again directly placed within the overall reference frame of the global catalogue. The wealth of information contained in the Hipparcos results on 10^5 stars provides an insight into the importance of double and multiple systems within the context of a catalogue of 10^9 objects with μas accuracy (Section 1.5).

The wide separation of two individual viewing directions is a fundamental pre-requisite of the payload, since this leads to the determination of absolute trigonometric parallaxes, and thereby circumvents the problem which has plagued ground-based parallax determinations, namely the transformation of relative parallaxes to absolute distances.

Absolute parallaxes The trigonometric parallax is the most direct measure of stellar distance. The traditional (ground-based) method to determine the parallax is to measure the annual oscillations of the star's apparent motion, as caused by the Earth's revolution around the Sun, against the background of distant stars. These measurements are confined to the field of view of a long-focus telescope, usually within 10–20 arcmin. The measurable quantity is then the *relative* parallax, i.e., the difference between the parallax of the target star and the mean parallax of the background stars. To obtain the true or *absolute* parallax of the target star one must add a correction corresponding to the mean parallax of the background stars; this is often estimated simply from their mean magnitude and Galactic coordinates. The size of the correction is typically 1–2 mas, with a (model dependent) uncertainty of a substantial fraction of a mas (van Altena et al. 1988). This procedure is clearly inadequate when it comes to μas astrometry. The inestimability of absolute parallaxes in a small field depends on the circumstance that the parallax factor f, being proportional to the sine of the angle from the Sun to the star, is nearly constant in the small field. The relative displacement of two stars with parallax π_1 and π_2 is consequently given by $f \times (\pi_1 - \pi_2)$. From a series of measurements made at different times of the year (with different f) it is possible to determine $\pi_1 - \pi_2$, but not the individual parallax values.

The situation is radically different if wide-angle measurements can be made with the same accuracy as in a small field. The relative displacement must now be written $f_1\pi_1 - f_2\pi_2$, where in general $f_1 \neq f_2$ for stars separated by a large angle (~ 1 rad). Measuring this displacement at different times of the year clearly allows to determine π_1 and π_2 individually, in other words, to obtain the absolute parallaxes. This principle of absolute parallax determination by wide-angle measurements was implemented for the first time in the Hipparcos mission, and has been demonstrated to be a complete success: several independent tests on preliminary Hipparcos parallaxes have shown that their zero-point errors are much less than 1 mas, and possibly smaller than 0.1 mas. The same principle will apply to the GAIA mission, which will therefore produce parallaxes that are absolute at the sub-μas level.

Number of viewing directions The accuracy of a reference frame depends not only on the total variance of the astrometric errors, but also on their spectral distribution in terms of different spatial frequencies or angular scale lengths—a total standard error of 10 μas may be scientifically unacceptable if most of the power is in the very low-frequency components of the error spectrum. The problem has been addressed by Lindegren (1997b). Low-frequency error components in the astrometric reference frame originate from low-frequency errors in the angular coordinates along the scanned circle (this is true whether the data reductions are modeled on a great-circle type approach, or whether a more global-type solution is envisaged). Systematic errors in the measurements result in systematic errors on the sphere, although damped by the large number of scans across any given point. The low-frequency part of the (random or systematic) abscissa error can be written as a Fourier series as a function of the abscissa υ, with the general term $A\sin(n\upsilon + \phi)$ representing the n-th harmonic. For a given number of viewing directions, of width w, the mean value of the abscissa error, μ, and the dispersion of the errors about the mean, σ, can be calculated. As the viewing windows move around the circle, the mean variance $\langle\sigma^2\rangle$ indicates how much this particular error harmonic will influence the observations.

Detailed studies show that the case of a single viewing direction gives (as expected) a very poor detection of the low frequencies: such an instrument is good for relative measurements in small areas of the sky but not for global astrometry. The use of two viewing directions with a large basic angle γ gives a dramatic improvement, for example by almost two orders of magnitude in the amplitude of the first harmonic. At the same time, the basic angle must not be a simple fraction of 360°; thus $\gamma \simeq 60°$ makes it very difficult to detect the error component with $n = 6$. $n = 1$ is particularly important for the parallax determination (Lindegren 1997a). Further suppression of low-frequency harmonics can be achieved by the use of three or four viewing directions, but at the expense of a considerably more complex design. However, more recent studies by Makarov (1998) have shown that very good suppression is achieved, even for two viewing directions, through the global combination of all the scans throughout the mission. Thus, an adequately low level of the large-scale error components is expected already for the baseline concept, retaining just the two viewing directions.

2.1.2 Theoretically Achievable Accuracy

The ultimate accuracy with which the direction to a point source of light can be determined is set by the dual nature of electromagnetic radiation, namely as waves (causing diffraction) and particles (causing a finite signal-to-noise ratio in the detection process). Consider the observation of a distant monochromatic point source by means of an optical telescope or interferometer equipped with an idealized detector. The instrument generates a diffraction image in the focal plane and the detector records the precise location of each detected photon in the diffraction pattern. If λ is the wavelength and D the overall size of the instrument aperture (diameter or base length), then the characteristic angular size of features in the diffraction pattern that can be used to localise the image is of order λ/D radians. If a total of N detected photons are available for localizing the image, then the theoretically achievable angular accuracy will be of order $(\lambda/D) \times N^{-1/2}$ radians.

A realistic size figure for non-deployable space instruments is of order a few metres, say $D \sim 2$ m. Operating in visible light ($\lambda \sim 0.5$ μm) then gives diffraction features of order $\lambda/D \sim 0.05$ arcsec. To achieve a final astrometric accuracy of (say) 10 μas it is therefore necessary that the diffraction features are localised to within 1/5000 of their characteristic size. Two obvious requirements follow: firstly, that at least some 25 million detected photons are needed to beat down the statistical noise by this factor; secondly, that extreme care is needed to achieve such a huge improvement in practice. Elementary calculations show that the first requirement (number of photons) can be satisfied for objects around 15 mag with reasonable assumptions on collecting area and bandwidth (see below). The second requirement is clearly a technical challenge, but the conclusion from the present report is that this condition, too, can be met (see also Section 6.8).

Angular Accuracy Detailed consideration of the (aberration-free) diffraction images resulting from arbitrary apertures, and of statistically optimal algorithms for estimating the location of these images in the presence of photon noise, provides a rigorous theoretical limit for the angular accuracy σ, given by:

$$\sigma \geq \frac{\lambda}{2\pi B_{\rm eff}\sqrt{N}} \tag{12}$$

(Lindegren 1978), where $B_{\rm eff}$ is twice the rms extension of the aperture in the measurement direction. The 'effective baseline' $B_{\rm eff}$ is here adopted as a measure of the aperture size, because in the simple case of an interferometer with two small and identical apertures, $B_{\rm eff}$ equals the distance between the apertures. A filled rectangular pupil of dimension D in the measurement direction has $B_{\rm eff} = D/\sqrt{3}$, while a filled circular pupil of diameter d would have $B_{\rm eff} = d/2$.

The numerical factor 2π in Equation (12) means that the requirement in terms of number of photons is somewhat relaxed compared with the previous order-of-magnitude estimate. On the other hand, Equation (12) gives the limiting accuracy of a single angular measurement in the instrument frame. The positions, proper motions and trigonometric parallaxes of all the objects must however be determined by a least-squares combination of all such measurements collected over the whole mission. The accuracy of the parallax of an individual star depends on the total weight of the measurements of the star, and a certain factor determined by the geometry and temporal distribution of the measurements. For the parallax accuracy σ_π, by a slight generalisation of Equation (12):

$$\sigma_\pi = g_\pi \frac{\lambda_{\rm eff}}{2\pi B_{\rm eff}\sqrt{N}} \tag{13}$$

where g_π is the mean geometrical factor for the parallaxes. With the scanning law envisaged for GAIA, $g_\pi \simeq 2$ (Section 7). With $B_{\rm eff} \simeq 2$ m and $\lambda_{\rm eff} \simeq 0.7$ μm it follows that $N \simeq 4 \times 10^6$ is needed for $\sigma_\pi = 10$ μas. This can be reached for a $V = 15$ mag star with a total observing time of ~ 500 s, assuming a collecting area of 1 m^2 and an equivalent bandwidth $\int T_\lambda Q_\lambda \, {\rm d}\lambda \sim 100$ nm.

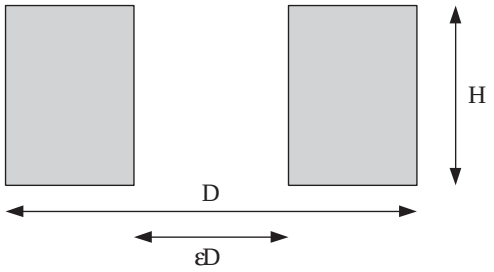


Figure 2.1: Parametrization of the pupil for the astrometric instrument in terms of the total length D in the scan direction and the obscuration ratio ϵ. The perpendicular dimension H is set by the (fixed) total area $A = (1 - \epsilon)DH$.

Available observing time per object For a continuously scanning instrument, surveying the whole sky, the total time $\langle\tau\rangle$ spent inside the field of view by an average object is given by:

$$\langle\tau\rangle = \frac{\Omega}{4\pi}L \tag{14}$$

where Ω is the solid angle of the field of view and L the (effective) mission length. Provided the detector allows simultaneous measurements of all the objects of interest within the field of view at any given time, this is also the total observing time on an average object. Equation (14) holds independent of the adopted scanning pattern and the spin period of the instrument. While a slower spin increases the time it takes an object to cross the field of view, it also reduces the (expected) number of such crossings by the same factor, leaving $\langle\tau\rangle$ unchanged. Similarly, the scanning law may affect the large-scale distribution of observing time between different parts of the sky, but not the total time on the average object.

For $\langle\tau\rangle \sim 500$ s and $L \sim 4$ years, Equation (14) gives $\Omega \sim 0.2$ deg^2 as a *minimum* field size. The need to get a reasonably uniform sky coverage and as many observations as possible of each object implies additional constraints on the field size. With the adopted revolving scanning law and a 3-hour spin period, the field will trace out a complicated path on the sky, with a maximum displacement of $0\overset{\circ}{.}5$ between successive traces across any small area of the sky. To ensure that no object escapes observation by falling between two successive scans, the height of the field perpendicular to the scan direction must be at least $0\overset{\circ}{.}5$. Thus we need to consider a field of at least $0\overset{\circ}{.}5 \times 0\overset{\circ}{.}5$ size.

These considerations are based on a targeted accuracy of 10 μas at 15 mag. In principle, a significantly higher accuracy might be achieved by the use of an even larger field, the longest baseline compatible with the Ariane 5 launcher, and the widest possible wavelength band. Accuracies of a few μas at $V = 15$ mag seem achievable in principle, but require large and efficient detectors with extremely high spatial resolution: for the immediate future this remains a hypothetical option.

As opposed to the grid option considered in the proposal by Lindegren et al. (1993b), the use of direct fringe detection with a continuously scanning interferometer implies the use of a detector system with a very large number of resolution elements or pixels. Direct fringe detection requires that the fringes are sampled at least at the Nyquist frequency. On the other hand, substantial oversampling is generally not advantageous, and the optimum sampling may in practice be around 1.5 times the Nyquist frequency. For given effective baseline $B_{\rm eff}$ and wavelength $\lambda_{\rm eff}$ this implies an angular pixel width of $\Delta\xi \sim \lambda_{\rm eff}/3B_{\rm eff}$. The astrometric accuracy σ_π is proportional to $\lambda_{\rm eff}/B_{\rm eff}\sqrt{\tau}$, if τ is the observation time per object. Combined with Equation (14), valid for a continuously scanning satellite, this gives $\sigma_\pi \propto \Delta\xi/\sqrt{L\Omega}$. But $\sqrt{\Omega}/\Delta\xi$ is roughly the number of pixels across the field of view in the scanning direction. Other things being equal, such as mission length and quantum efficiency, σ_π is thus inversely proportional to the number of pixels across the field of view.

It is especially worth noting that a longer baseline does not necessarily improve the accuracy, namely if the improved resolution leads to a proportionally smaller field of view due to a constraint on the linear size of the detectors (and therefore on the number of pixels). The ultimate practical limitation to the astrometric accuracy is thus set by the characteristics of realistic detectors, rather than by optical configuration and photon statistics alone. In the next section the resolution properties of the detector are therefore brought into consideration as well.

Table 2.1: Reference parameters used for a provisional optimisation of focal length and pupil shape (Figures 2.2 and 2.3). Pixel cross talk (MTF degradation) is not included. Sky background is negligible for a 15 mag star.

Parameter	Value	Meaning (see also Section 7)
f_λ	100 ph m^{-2} s^{-1} nm^{-1}	flux distribution ($V = 15$)
$\lambda_{\rm eff}$	550 nm	effective wavelength
$\Delta\lambda_{\rm eff}$	400 nm	equivalent bandwidth ($= \int T_\lambda Q_\lambda \, {\rm d}\lambda$)
N_i	2	number of instruments
A	1 m^2	total pupil area per instrument
D	variable	maximum along-scan dimension of each pupil
ϵ	variable	along-scan obscuration ratio of each pupil
F	variable	focal length
α	$\leq \alpha_{\rm max} = 0.25$ m^2	detector area per instrument
Ω	$\leq \Omega_{\rm max} = 0.25$ deg^2	detector solid angle per instrument
L	4 years	effective observation time (mission duration = 5 years)
l	9 μm	linear pixel width ($= F\Delta\xi$)
g_π	2	geometrical factor for parallax

2.1.3 Matching Optics and Detector

The use of CCD detectors for direct recording of diffraction images has important consequences for the overall instrument design. On a continuously scanning satellite, the CCDs must be read out in TDI mode, where the charge is shifted along the CCD at the same speed as the images move along it, until it is read out at the end. A practical constraint is set by the performance of real CCDs as a function of linear pixel size. While small pixels are obviously desirable, quantum efficiency and MTF (due to pixel cross-talk) tend to degrade seriously with decreasing pixel width. A reasonable approach is to choose the smallest pixel size for which the degradation is not too serious, and adjust other parameters (such as focal length) accordingly. The pixel width chosen for the present baseline design, $l = 9$ μm (along scan), is probably close to optimal in this respect (Section 3.7.3). For the purpose of illustrating the dependence of astrometric accuracy on basic instrument parameters, the linear pixel width as well as some other parameters are set to the reference values in Table 2.1.

In TDI mode, the CCD readout is clocked synchronously with the image motion along scan, with the image being integrated until it reaches the serial register. This operational mode has been used from the ground for both imaging and spectroscopic observations. For example, Schneider et al. (1994) carried out a survey to search for faint, high-redshift QSOs at the Mount Palomar 200-inch reflector, by performing low-resolution slit-less spectroscopy of all objects drifting across the field of view at the sidereal rate. The Sloan Survey also uses the same principle for the photometric sky scans (Kent et al. 1994).

The effect of the finite pixel width l on the astrometric performance can be described in terms of the smearing of the diffraction image or, equivalently, the attenuation of its high-frequency content. In the TDI observation mode, there is actually a double smearing: first by the 'binning' of photons into pixels of width l; and second by the motion of the optical image relative to the pixels during the TDI period. The combined effect (if the charges are transferred in steps of a whole pixel) is to reduce the amplitude at spatial frequency f by the factor $\mathrm{sinc}^2(\pi l f)$. To a first approximation, the astrometric error (σ_π) is inversely proportional to the amplitude of the fringes or PSF. Thus, Equation (13) should be amended by the corresponding factor. The question is only which spatial frequency to be used in the sinc^2 factor. A natural choice is $f = B_{\rm eff}/\lambda_{\rm eff}$, since $B_{\rm eff}$ is the aperture dimension relevant for the astrometric accuracy (recall that $B_{\rm eff}$ could be defined for arbitrary pupil shapes). With F denoting focal length, Equation (13) gives:

$$\sigma_\pi = \frac{g_\pi \lambda_{\rm eff}}{2\pi B_{\rm eff}\sqrt{N}} \left[\mathrm{sinc}\left(\frac{\pi l B_{\rm eff}}{F\lambda_{\rm eff}} \right) \right]^{-2} \qquad (15)$$

Here N is, according to Section 2.1.2, proportional to the solid angle Ω of the detectors. However, it is not realistic to keep the solid angle fixed independent of focal length: for sufficiently large F, the linear size of the focal plane will limit the available solid angle. In Table 2.1 this is reflected in having maximum values defined both for the angular and linear field sizes. A breakpoint occurs at the focal length $\sqrt{\alpha_{\max}/\Omega_{\max}} = 57.3$ m. With notations as in Table 2.1 the total number of photons detected over the mission is:

$$N = N_i A f_\lambda \Delta\lambda_{\text{eff}} L \times \frac{\min(\Omega_{\max}, \alpha_{\max}/F^2)}{4\pi} \tag{16}$$

Consider a rectangular pupil with maximum dimension D in the scanning direction, linear obscuration ratio ϵ, fixed area A, and consequently the dimension $H = A/(1-\epsilon)D$ normal to the scan (Figure 2.1). The effective baseline is:

$$B_{\text{eff}} = D\sqrt{\frac{1+\epsilon+\epsilon^2}{3}} \tag{17}$$

It is now possible to compute the astrometric accuracy as a function of the remaining 'free' parameters F, D and ϵ (Figures 2.2 and 2.3). Variations with pixel size are shown in Figure 2.4, but here it should be remembered that pixel cross talk and QE degradation (not taken into account in the figure) will drastically increase the errors towards the smaller pixel sizes.

Some important conclusions can be drawn from the figures:

- with the chosen pixel size, focal lengths of 40–50 m or more are required to provide astrometric results comfortably below the 10 μas level;
- unless extremely long-focus systems are considered, the optimum obscuration ratio is $\epsilon = 0$, i.e., a filled pupil (no interferometer);
- provided a focal length around $F = 50$ m is feasible in terms of accommodation and aberrations, the optimum pupil size is $D \simeq 2$ m;
- again assuming that $F = 50$ m is feasible, the potential gain from using smaller pixels (even disregarding the increased pixel cross talk) is relatively marginal.

Although based on simplified assumptions, the present analysis clearly points in the direction of *a moderately large filled aperture, as opposed to an interferometer with maximum possible baseline.* This conclusion is strengthened when considering the effects of the wide wavelength band, which will further reduce the fringe contrast for a split pupil, and is confirmed by the more complete accuracy analyses carried out during the present study.

2.1.4 Astrometric Requirements

The parameter space that is accessed by GAIA involves both accuracy, limiting magnitude, and numbers of objects. The requirement of 10 μas at 15 mag, as specified by the Survey Committee, follows directly from the scientific requirements of accessing a very significant fraction of our Galaxy, while at the same time being consistent with order-of-magnitude accuracy estimates derived from present day detector, optics, and launch constraints. Slightly higher accuracies would be possible in principle, but at the expense of further satellite complexity. A significant degradation of the target accuracies would significantly and dramatically weaken the scientific case, and it is debatable whether a future mission in space astrometry could be justified on those terms.

Accepting such target accuracies, restricting GAIA to a limiting magnitude of 15 mag, or a subset of all objects down to its detection limit, would provide a reduction in the down-link telemetry rate, but little or no change in the other main aspects of the payload design. These are driven simply by the photon noise budget required to reach a 10 μas accuracy at 15 mag. The faint magnitude limit and the number of target objects follow from the accuracy requirement.

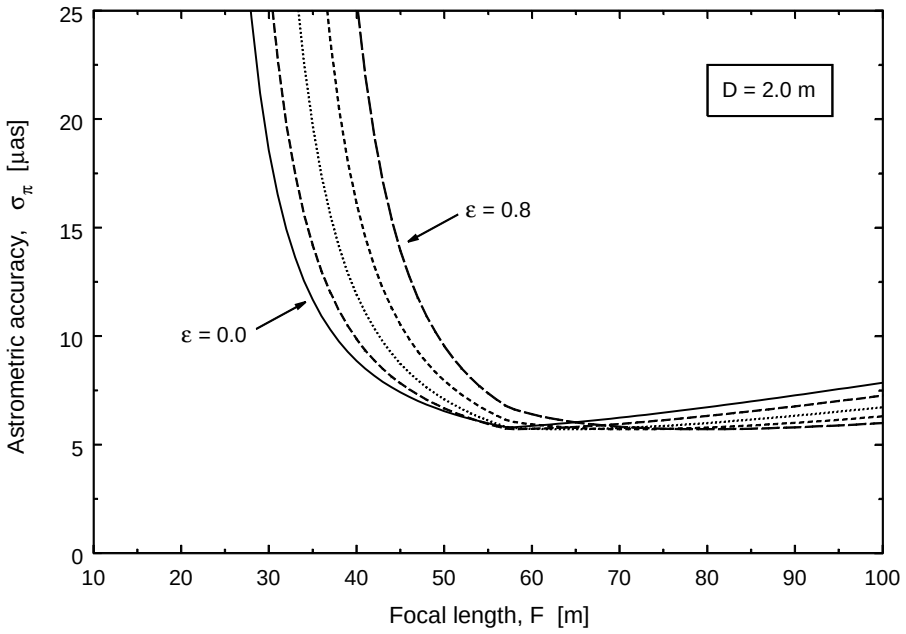


Figure 2.2: Final astrometric accuracy versus focal length for pupil obscuration ratios $\epsilon = 0$, 0.2, 0.4, 0.6 and 0.8. The total length of the pupil is $D = 2$ m and the total area $A = 1$ m in all cases. Other parameters are as given in Table 2.1.

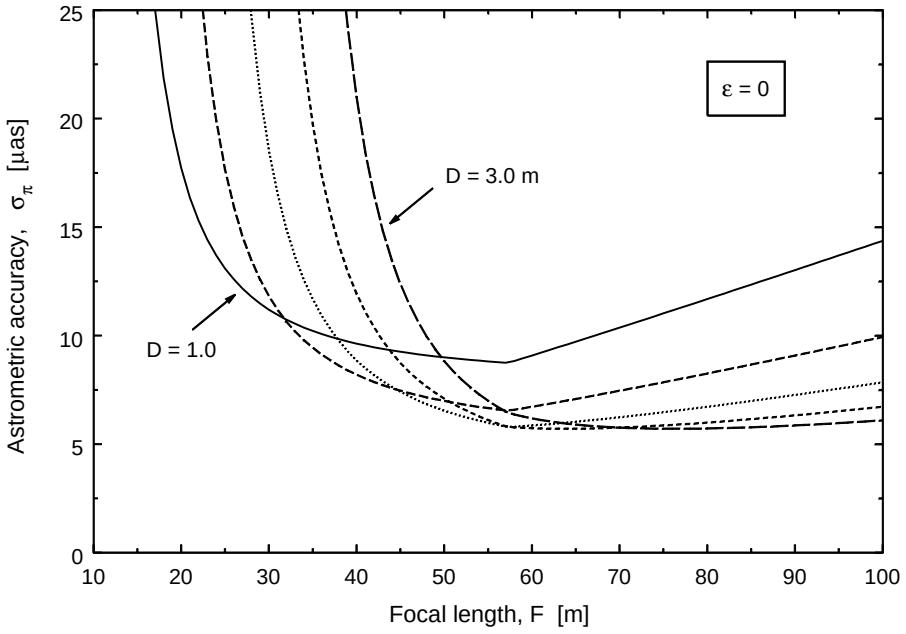


Figure 2.3: Final astrometric accuracy versus focal length for pupil lengths $D = 1.0$, 1.5, 2.0, 2.5 and 3.0 m. The obscuration ratio is $\epsilon = 0$ and the total area $A = 1$ m in all cases. Other parameters are as given in Table 2.1.

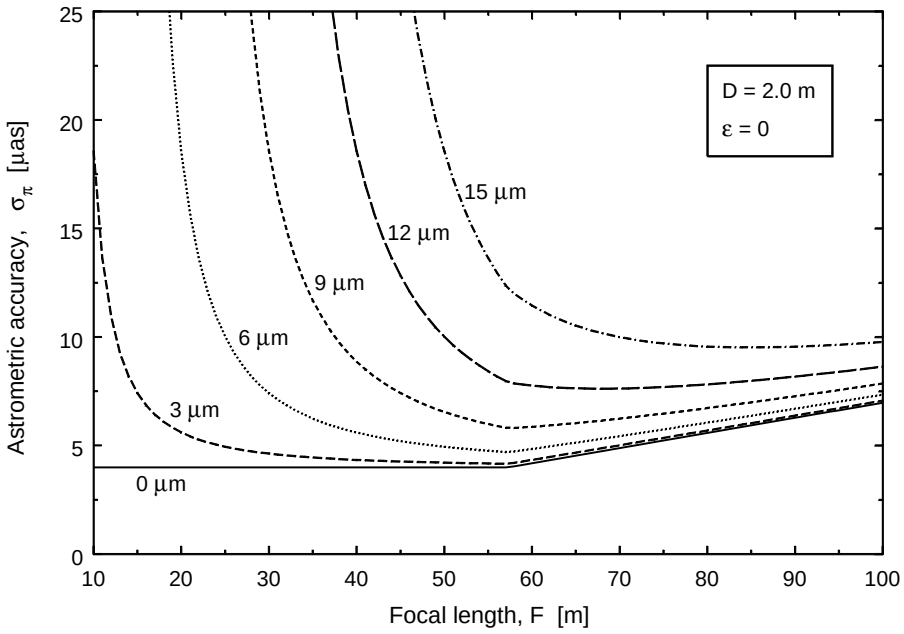


Figure 2.4: Final astrometric accuracy versus focal length for pixel sizes l = 0 to 15 μm. The calculations do not include the increasing QE and electronic MTF losses for pixels smaller than ~ 9 μm, and the curve for $l = 0$ is purely theoretical. The pupil size is $D = 2$ m and the obscuration ratio $\epsilon = 0$ in all cases. Other parameters are as given in Table 2.1.

Nevertheless, it should be stressed that the scientific objectives of faint star astrometry are paramount in the GAIA scientific case: for example, as presented in Section 1 a significant number of fundamental Galactic kinematic tracers *only* become accessible at $V > 17 - 18$ mag. The large number statistics are also a key element of the scientific case, and correspond to one of the main scientific differences between GAIA and, say, SIM (Appendix A): again, to take one specific example: the Galaxy models used to evaluate the GAIA mission suggest that GAIA will observe some 200 000 disk white dwarfs, and some 1000 halo dwarfs. Observation of the latter would be a straightforward observing programme for SIM, except for the fact that such halo dwarfs will be extremely difficult to identify *a priori*. By observing very large numbers of objects in our Galaxy, and characterising the through detailed photometric measurements, GAIA will identify very significant numbers of rare but important objects in luminosity and in dynamical phase space, which will go undetected other than through such observational methods. The rich scientific harvest from GAIA will arise from its combination of unprecedented accuracy, faint limiting magnitude, and very large number statistics.

2.2 Radial Velocity Measurements

There are three motivations for considering the parallel acquisition of radial velocities with GAIA: (i) the astrometric measurements supply only two components of the space motion of the target stars (those in the plane of the sky). The third component, the radial velocity, is directed along the line of sight, and is needed for proper kinematical or dynamical studies (Section 1); (ii) measurement of the radial velocity at a number of epochs is a powerful method for detecting and characterising binary systems (Section 1.5; see Figure 2.10); (iii) at the GAIA accuracy levels, the problem of 'perspective acceleration' is at the same time both a complication and an important observable quantity (Section 2.2.1). On-board acquisition of radial velocities with GAIA is not only feasible, but is relatively simple (see Favata & Perryman 1995, and Section 3).

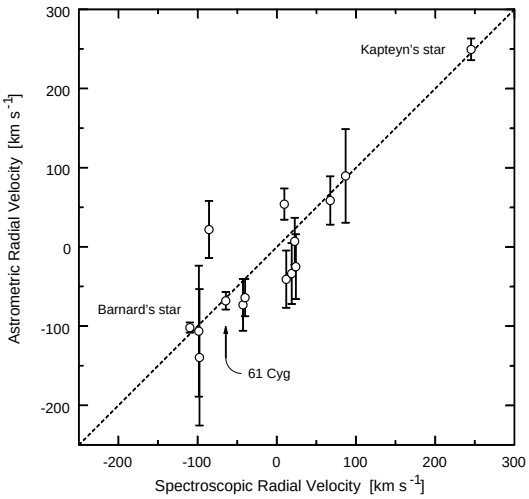


Figure 2.5: A comparison between spectroscopic radial velocities and astrometric radial velocities derived from a combination of Hipparcos and Astrographic Catalogue results (Dravins et al. 1999).

2.2.1 Perspective Acceleration and Related Effects

If the distance between an object and observer changes with time due to a radial component of motion, a constant transverse velocity is observed as a varying transverse angular motion. This perspective acceleration due to a radial velocity V_r is given by:

$$a = -2.0 \times 10^{-9} \pi \mu V_r \tag{18}$$

where a is the acceleration in mas yr^{-2}, π is the parallax in mas, μ is the proper motion in mas yr^{-1}, and V_r is in km s^{-1}. For a mission duration of t years, the integrated positional error is:

$$e = \frac{1}{2} a (t/2)^2 \text{ mas} \tag{19}$$

reckoned from the central measurement epoch. The penalty for not incorporating the effect will be misleading proper motions for an unknown subset of the stars observed, and the classification of apparent acceleration (e.g., due to binarity) to an otherwise linear motion of the photocentre. The astrometric measurements cannot provide an estimate of the perspective acceleration, and hence deduce the radial velocity of the star: the very high fraction of double and multiple systems, especially amongst the nearest stars, means that the disentangling of true orbital acceleration and perspective acceleration will demand an independent knowledge of the radial velocity (Section 1.5). This is particularly important for mass determinations and related orbital studies of binary systems (including planetary companions).

Assuming a mission duration of 5 years, the radial velocity at which perspective acceleration becomes a significant effect is a factor of 100–1000 times smaller than for Hipparcos, for any given product of μ and π, decreasing further as the square of the mission duration. Estimates of π and μ from Hipparcos can be used to estimate the number of objects affected. At the level of 1 μas, some 500 Hipparcos objects would be affected if their $V_r > 10$ km s^{-1}, and some 7000 if $V_r > 100$ km s^{-1}. Estimating the completeness of the Hipparcos Catalogue within 100 pc at around 20 per cent, the number of GAIA stars affected would be a factor of 5 larger. Requiring a correction at 0.1 μas, some 35 000 stars would be affected if their $V_r > 10$ km s^{-1}, and some 200 000 if $V_r > 100$ km s^{-1}. The number of stars actually affected clearly depends on the kinematic properties of the various populations sampled by the mission: however, provision must be made to ensure that the higher velocity stars (in particular, both the spheroidal component objects, and the white dwarfs and the long-period variable stars within the disk-population) are not subject to an astrometric bias due to their unknown radial velocities.

For Hipparcos, the effect of perspective acceleration was generally negligible, except for a few nearby, high-velocity stars. Nevertheless, attempts were made to eliminate systematic errors arising from the effects of perspective acceleration (by incorporating known radial velocities of affected stars) at the level of 0.1 mas, in view of a nominal mission accuracy of 1 mas. Figure 2.5 shows the results of a comparison between the extrapolated Hipparcos position/proper motion for 15 stars with positions from the Astrographic Catalogue (epoch around 1910) and,

from the difference, a determination of the radial velocities of these stars. It includes a value of 68 ± 11 km s^{-1} for 61 Cyg, in good agreement with spectroscopic determinations, by combing the Hipparcos results with those obtained by Bessel in 1838. The precision is not impressive by spectroscopic standards, but the fact that the method works is another indication of the high quality of the Hipparcos data, and an indication of the relevance of these considerations for GAIA. All stars display a corresponding effect of secular parallax evolution, or change of parallax with time due to their radial velocity, although Schlesinger (1917) had already concluded that the effect is small for every known star. In the case of Barnard's Star, the expected parallax derivative is $+34$ μas yr^{-1}, a tiny effect but one which will be measurable with GAIA.

Astrometric Radial Velocities Dravins et al. (1999) have demonstrated how the classical moving cluster method can be 'inverted', using proper motions and parallaxes to determine an astrometric radial velocity, rather than the other way around. This can be used to probe the difference between a star's astrometric and spectroscopic radial velocity. This is interesting because such shifts can arise from a variety of physical processes, including line asymmetries, atmospheric pulsations, surface convection, and gravitational potential. Astrometric radial velocity determinations will themselves be biased by any deviations from rectilinear motion: in the case of Proxima Centauri, for example, by 170 m s^{-1} as a result of acceleration towards the Galactic centre caused by the smoothed Galactic potential in the vicinity of the Sun, which amounts to about 2 Å s^{-2}. Light-time effects for typical Population II stars are at the level of 100–200 m s^{-1}, effects which will become relevant at the microarcsec level.

2.2.2 Accuracy Requirements

While in principle it is desirable to obtain as accurate as possible radial velocities for all targets, in practice a faint and a bright regime can be distinguished. The faint targets will mostly be distant stars, which will be of interest as tracers of Galactic dynamics. The uncertainty in the tangential component of their space motion will be dominated by the error in the parallax. Hence a radial velocity accuracy of $\simeq 5$ km s^{-1} is sufficient for statistical purposes. Stars with $V \lesssim 15$ will be of individual interest, and the radial velocity will be useful also as an indicator of multiplicity and for the determination of perspective acceleration.

The method foreseen for the measurement of the radial velocities is digital cross-correlation between the observed spectra using an appropriate template (either the high signal-to-noise spectrum of a star of similar type or a synthetic spectrum). Experience with ground-based spectroscopy shows that correlation accuracies of $\simeq 0.1$ pixel can be obtained on low (i.e., $\simeq 5$) S/N spectra, at least for late-type stars with a spectrum rich in lines, to a precision corresponding to about 0.1 pixel (Batten 1985, Carney et al. 1987). Accuracies 5 to 10 times better (i.e., $\simeq$ 0.01–0.02 pixels) can be obtained with S/N $\gtrsim 30$. Working in the red region of the spectrum (where the sensitivity of current digital detectors is greater) this translates into a spectral dispersion of about 10 km s^{-1} per pixel, or about R = 30 000, i.e., the typical resolution of an echelle spectrograph.

In practice (see Section 8.3) the present design allows (for red Population I stars of any luminosity class) to determine radial velocities to $\sigma_v \simeq 5$ km s^{-1} at $V = 18$ mag.

2.2.3 Choice of Spectral Range

Most GAIA stars are intrinsically red, and made even redder by interstellar absorption. Thus, a red spectral region is to be preferred for the GAIA spectrograph. To maximize the radial velocity signal even for metal-poor stars, strong, saturated lines are desirable. Broad lines also allow the radial velocity to be accurately derived from moderate-resolution spectra (as one only needs to sample the lines to derive the radial velocity; oversampling leads to marginal improvements in accuracy). Redwards of the Hα line, three strong spectral features are present in late-type stars: the K I doublet near 768 nm, the Na I doublet at 819.4 nm and the Ca II triplet near 860 nm (Figure 2.6). The latter is ideal for radial velocity determination in many types of stars (Munari 1999a).

The K I doublet has moderate equivalent width ($W = 0.01$–0.04 nm) and complicated luminosity dependence. It also is a strong interstellar absorption feature (which complicates its interpretation in the case of significant reddening), and it lies at the core of a strong TiO band (which makes it difficult to use for late spectral types).

The Na I doublet is stronger than the K I doublet (see Figure 2.6), but again has a complex luminosity dependence, which makes it weak in giants. The two lines in it are also quite close in wavelength ($\simeq 0.16$ nm), so that high spectral resolution is needed to observe them, and they are also affected by TiO bands. Also, both the K I and the Na I region have no lines in earlier (B and A) stars, so that no radial velocity information can be derived from them.

The Ca II lines are strong along most of the H-R diagram, with $W \geq 0.3$ nm for all dwarfs from F8 to M8, and $W \simeq 0.6$ nm in supergiants. The lines appear around B8 and dominate throughout M stars, with a positive and strong luminosity effect. Being non-resonant, there are no problems with contamination by interstellar components. They are so strong that useful radial velocities can be derived by cross-correlation even on low S/N spectra (like the spectrum at S/N = 7 in Figure 2.7, from which a velocity can be measured to better than $\simeq 10$ km sec^{-1}).

A further strong advantage of the Ca II IR triplet is that, in early-type stars (B, A and early F), where the Ca II lines disappear, the Paschen series of H lines appear, which are also strong, and allow for the radial velocity of these stars to be also determined (although O stars are out of reach). The Paschen lines are also important for classification (Frémat et al. 1996). Finally, the spectral region is free from telluric absorption lines (unlike the other two discussed above), so that follow-up (or complementary) observations from ground can be done in a homogeneous way. This spectral region also contains numerous features of intrinsic astrophysical interest (Figures 2.8–2.11):

- the region is not affected by strong molecular bands, but at the same time several metallic lines are present (N I 872.89 nm, Si I 874.26 nm, Mg I 873.60 nm, Ti I 873.47 nm and He I 873.34 nm), which allow a detailed abundance analysis and quantitative spectral classification for the brighter stars based on ratios of equivalent widths (Figure 2.8);
- stars with peculiar spectra are also easily detected in this spectral range. Figure 2.9c shows the Ca II IR spectrum of the proto-typical mass-loosing star P Cyg. The characteristic line-profile is evident in the Paschen and He I lines, and is easily detected even in low S/N epoch spectra. Classical Be stars are easily found (Figure 2.9a), as well as classical T Tau stars (which have strong Ca II emission). Active late-type stars will display emission cores in the Ca II triplet;
- while in the Ca II triplet region there is no resonant interstellar line, a medium-intensity diffuse interstellar band (DIB) is present at 862 nm. Munari (1999a) has shown that the correlation of its equivalent width with absorption (Figure 2.11) is rather tight, so that it might be used as one of the indicators to build a detailed reddening map, especially for high values of A_V. These results have been confirmed over a large sample of stars over a wide range of distances and Galactic longitudes (Figure 2.11). However, part of the DIBs are sometimes of circumstellar origin and may behave differently (Krełowski & Greenberg 1999). A survey of the DIB would in any case provide insights into the variations in the interstellar medium.

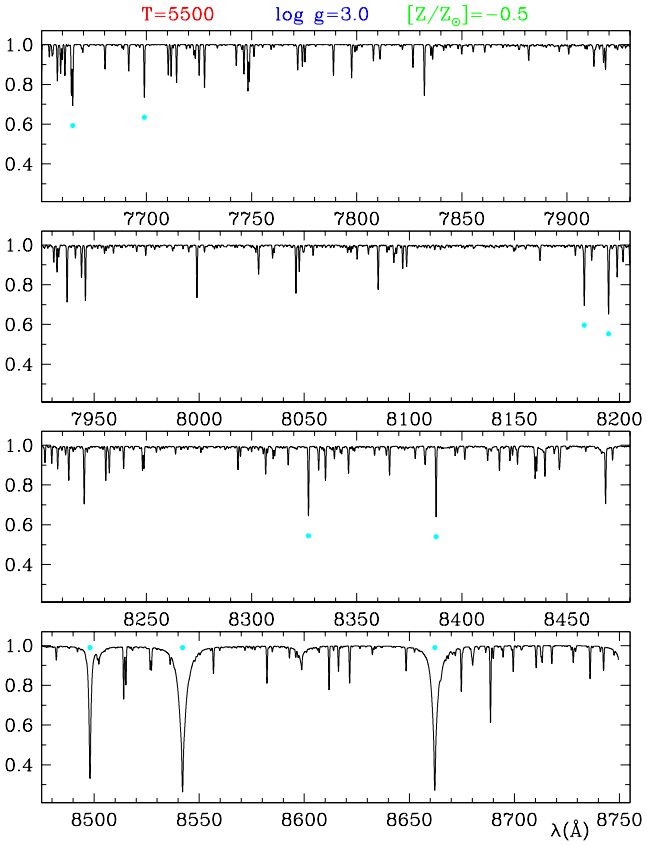


Figure 2.6: High-resolution spectrum of a disk G giant, showing the regions of the K I, Na I and Fe I doublets (indicated by the solid circles) and the Ca II triplet. The advantage of the Ca II triplet is evident (Munari & Castelli 1999).

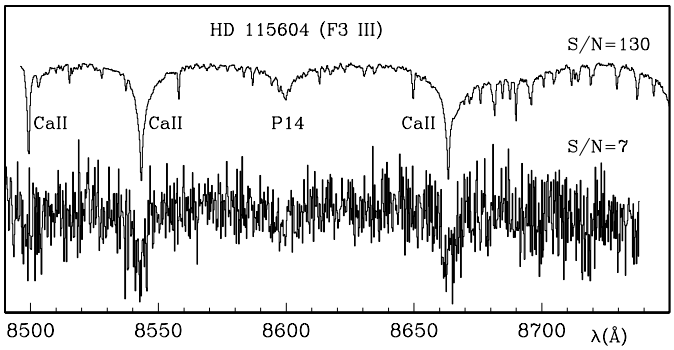


Figure 2.7: High and low S/N view of the Ca II triplet region for an F3 giant (from Munari 1999a). The $S/N = 7$ spectrum is representative of what can be detected in a single GAIA transit (yielding an 'epoch radial velocity'), while the $S/N = 130$ spectrum is representative of the summed spectrum accumulated throughout the mission (yielding a 'mission-integrated radial velocity') Notice how the lines are still clearly visible even at very low S/N.

2.2.4 Acquisition On-Board or On-Ground

The problem of parallel acquisition of non-astrometric data in the framework of a deep astrometry mission was considered by **??**, and subsequently by Bastian (1995). Acquisition of 'auxiliary' data (including photometric data considered in the following section) on the scale demanded by GAIA would be highly challenging. The possibility of acquiring the necessary spectral data through a dedicated ground-based multi-fibre spectroscopic facility was considered. Such programmes might be feasible down to magnitudes of 14–15 mag, but logistically highly complex and very expensive (cf. the Sloan Survey). There is a particularly significant additional scientific return in acquiring a large number of measurements, and doing so not only well spaced in time but also, preferably, simultaneously with the astrometric measurements. Given the large number of targets which will be observed, it is not currently feasible, or at least practical, to acquire such radial-velocities from ground-based facilities. Superior accuracy due to the absence of atmospheric scintillation, and an improved control of systematics, also argue for their acquisition from space.

2.3 Derivation of Astrophysical Parameters

Unsupported by appropriate diagnostic data, GAIA will yield immense numbers of positions and velocities of objects whose astrophysical nature will be unknown. The core science case requires measurement of luminosity, effective temperature, mass, age and composition for the stellar populations in the Galaxy and its nearest neighbours. These quantities can be derived from the spectral energy distribution of the stars, i.e., through multi-band photometry and spectroscopy. Acquisition of this astrophysical information is an essential part of the GAIA payload.

A broad-band magnitude, and its time dependence, can be obtained from the analysis and rigorous calibration of the primary mission data (i.e., by determination of the amplitude as well as the phase of the focal plane image). When combined with the parallax (and an estimate for the interstellar absorption), this gives an estimate of the absolute magnitude, with little or no requirements on the design of the astrometric part of the payload. In addition, multi-band photometry needs to be separately obtained for each target to obtain the necessary astrophysical measurements.

Photometric measurements are required also for the derivation of accurate astrometric parameters: the raw positional measurements need to be complemented by the spectral energy distribution of each astrometric target, in order to compensate for the residual chromaticity. This requires measuring the spectral energy distribution of each target to $V \simeq 20$ mag in four broad bands spanning the same wavelength range as the CCDs of the astrometric focal plane (Section 3.2.4).

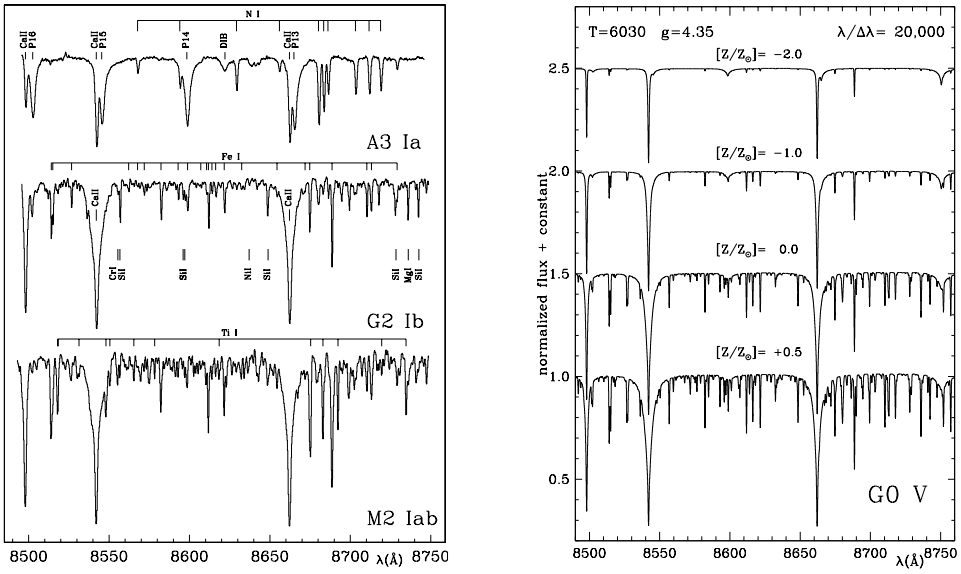


Figure 2.8: Spectral sequence in the Ca II triplet region, taken at a resolution of $\simeq 0.025$ *nm per pixel. The left panel shows the effect of temperature from A to M stars, the right panel the effect of metal abundance in G stars. From Munari (1999a).*

2.3.1 Photometric Requirements

For essentially every application of the GAIA astrometric data, high-quality photometric data will be crucial. Additionally, the photometric measurements will be indispensible in providing the basic tools for classifying stars across the entire HR diagram, as well as in identifying specific and peculiar objects. This makes it necessary to observe a large spectral domain, extending from the UV to the far-red. It must be able to determine (i) temperature and reddening at least for OBA stars (needed both as tracers of Galactic spiral arms and as reddening probes) and (ii) effective temperatures and abundances for late-type giants and dwarfs, assuming the reddening is known. To be able to reconstruct Galactic formation history the distribution function of stellar abundances must be determined to ~ 0.2 dex, while effective temperatures must be determined to ~ 200 K. Separate determination of the abundance of Fe and α-elements (at the same accuracy level) will be essential for mapping Galactic chemical evolution. The same accuracy allows separation of stars belonging to different populations (i.e. disk and halo stars, and also thin and thick disk populationas). These requirements translate into a magnitude accuracy of $\simeq 0.02$ mag for each colour index.

For $V \lesssim 15$ mag, the spectra obtained for the radial velocities will also provide accurate values for the chemical composition, surface gravity, ionization and excitation temperatures, axial rotation and will detect duplicity (from binary motion) and spectral peculiarities (Section 2.2.3). For these stars, photometry integrates the energy distribution over a wide wavelength range (thus providing near-bolometric fluxes), supplies a colour temperature (to be compared with the ionization and excitation temperatures of spectroscopic origin) and gives an estimate of the reddening by confronting observed and expected energy distributions. Very accurate absolute magnitudes (errors $\lesssim 0.1$ mag) and luminosities will be derived from the parallax (after correction for reddening, and a new derived bolometric correction). The spectroscopically derived parameters will also supply a 'training set' for the calibration of the photometric indicators for the fainter stars.

For stars with $V \gtrsim 15$ mag, the astrophysical diagnostics must be derived exclusively from the photometry, as the spectra in the Ca II triplet region will have insufficient S/N. For these fainter (but very numerous) stars, the approach taken will depend on the spectral type: the spectrum of early-type stars (OBA) is such that it is possible (given an absorption law) to simultaneously derive their astrophysical parameters as well as to derive the amount of interstellar absorption. This can

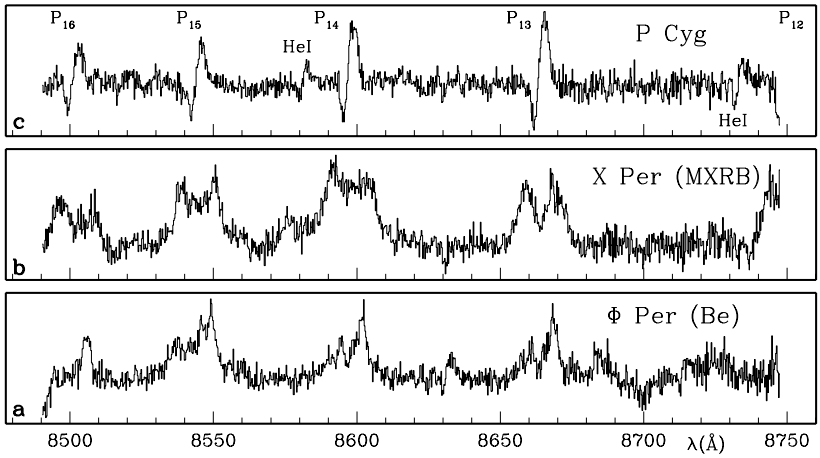


Figure 2.9: Ca II triplet/Paschen spectra ($S/N \simeq 30$) of (top) the prototype mass-loss system P Cyg, (mid) the massive X-ray binary X Per and (bottom) a hot star with an extended circumstellar envelope (ϕ Per). (Munari 1999a).

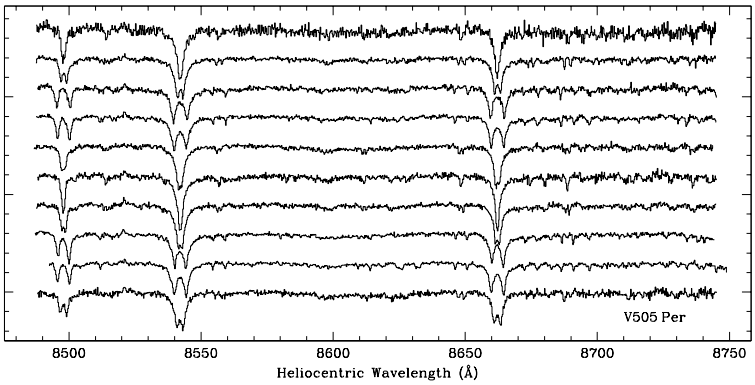


Figure 2.10: The periodically splitting Ca II triplet lines in a spectroscopic binary (V505 Per, F5V+F5V), showing the power of spectroscopy for the detection of close binary systems.

be achieved through the use of reddening-free indices, i.e. the so-called 'Q-Q diagrams', which map stars of a given type in the same colour region independently of the amount of reddening. One such diagram is shown in Figure 8.12. Thus for early-type stars the photometry will allow to simultaneously derive the main astrophysical parameters as well as the interstellar absorption.

For sufficiently hot (OB) stars the colour excess ($(\text{F47–F57})_{\text{obs}}$ – $(\text{F47–F57})_{\text{intr}} = 0.303 \times A_V$) can also be derived directly by the star's position in a colour-colour (F33–F41 versus F47–F57) diagram (see Section 2.3.2), allowing the reddening to be determined with very high accuracy: the mission-integrated accuracy will be $\sigma(A_V) \simeq 0.05$ mag at $V = 20$ for $A_V \simeq 1$. If blue horizontal branch (BHB) stars are used, at high and intermediate galactic latitude, this will allow reddening to be accurately traced throughout the Galactic plane, with $\sigma(A_V) \simeq 0.1$ mag at 25 kpc.

For later type stars this approach will not be feasible: as shown in Figure 8.12, for red stars the Q-Q diagrams become degenerate already for F and G stars, and the spectral characteristics of the stars are such that (especially for K and M stars, both giants and dwarfs) small photometric errors translate in large errors in the determination of both the astrophysical parameters and the interstellar reddening. For these stars (F and later) then a different approach will be used: a three-dimensional reddening map will be built, combining the information obtained from the brighter stars, the OBA stars, as well as existing H I and CO maps, and COBE/IRAS/DeNIS data.

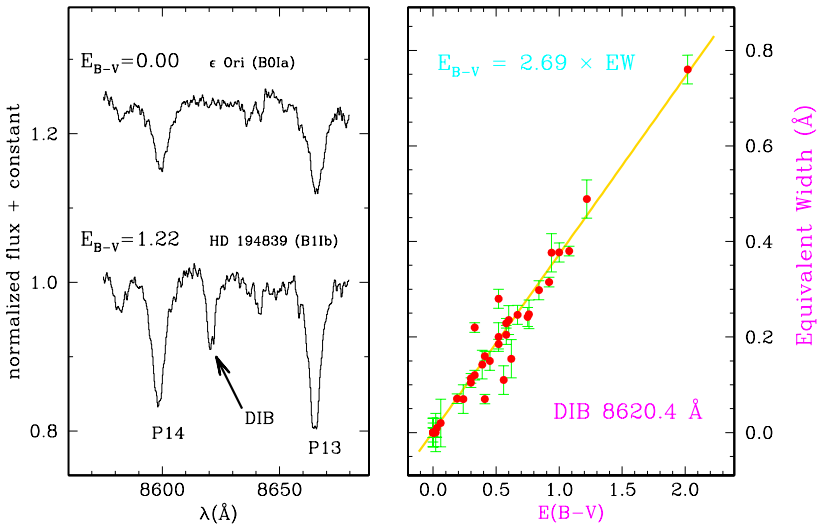


Figure 2.11: The diffuse interstellar band (DIB) at 862 nm. Left panel: spectra of un-reddened and reddened stars of similar spectral type. Right panel: dependence of DIB equivalent width on reddening. From Munari (1999a).

2.3.2 The Baseline Photometric Filter System

Many ground-based photometric systems exist (see Sterken 1992 for a review), but none satisfies all the requirements described in the above. Portions of the spectrum limited on-ground by terrestrial atmospheric O_3 and H_2O bands are opened up to GAIA. In addition to its central role in stellar classification and investigation, photometric measurements will be required for corresponding applications in quasar and galaxy photometry, Solar system object classification, etc. Considerable effort has therefore been devoted to the design of an optimum filter system for GAIA, taking into account the spectral energy distribution of the main Galactic stellar populations, as derived from model atmospheres and spectrophotometric observations (both from the literature, e.g. Gunn & Stryker 1983 and Pickles 1998 and obtained specifically for this purpose, e.g. Munari 1999b), as well as the experience with existing photometric systems (e.g., Straižys & Høg 1995; Munari 1999b; Straizys 1999; Grenon et al. 1999a; Grenon et al. 1999b). The result of this effort is a baseline system which provides the required accuracy for chromatic correction of the astrometric observations, and a good accuracy in the astrophysical information. The baseline system is considered to be a preliminary choice, and refinements with the same number of bands are being studied. The system features four broad wavelength bands (the Broad Band Photometer, BBP; Figure 2.12, left) and eleven bands of medium to narrow width (the Medium Band Photometer, MBP; Figure 2.12, right). The bands span the complete spectral range from $\simeq$ 280 to $\simeq$ 1000 nm. Their characteristics are listed in Table 2.2. The exposure time allocated to each filter has been chosen so as to achieve a balanced response and S/N across the whole wavelength range for the spectral types of interest.

The photometric filters are placed in two different locations, depending on their main purpose: the four broad bands are mainly dedicated to supplying the information necessary for the chromatic correction (and are somewhat redundant in terms of astrophysical information content); they are thus located on top of the trailing CCD chips in the astrometric fields of view, thus ensuring strict simultaneity, as well as the use of identical CCD chips and same pixel resolution as the astrometric observations. The bands whose main purpose is to yield astrophysical information are located within the dedicated spectrometric viewing direction, as shown in Figure 2.13, where the different image scale allows for longer integration times and thus higher S/N. The notation adopted is such that Fxx refers to the filter whose central wavelength is located at xx0 nm. A trailing 'B' indicates a broad band, located on the Astro focal plane.

F33 band The UV domain contains the most useful information on gravity and interstellar reddening for OBA-type stars, and on metallicity for F, G, K stars. In early-type stars, the Balmer continuum is devoid of absorption features in the 280–376 nm wavelength range, allowing for metallicity-independent determination of $\log g$ and absorption, while in G-K stars very many atomic lines are present, making UV flux a key metallicity indicator: for late G dwarfs, the line blocking in the 280–376 nm UV band is $\simeq 2.7\times$ larger than in the violet 376–430 nm domain. Ground-based observations cannot access the same band, as the atmospheric bandpass is limited by the saturated O_3 bands at $\simeq 300$ nm. Thus, a uniform, unbiased set of measurements in this band can only be obtained with a telescope in space. Due to its importance both for the determination of the astrophysical parameters and for the chromatic correction this band is duplicated across the Astro and Spectro telescopes.

F41 and F47 bands The F41 filter measures the Balmer lines for late B and A-type stars from the Balmer jump up to H_γ. For F-type stars and cooler, it measures the break short of 430 nm, due to a strong concentration of atomic and molecular lines (belonging to iron peak elements) in the interval 375–435 nm. CN bands with heads at 388 and 422 nm are also present, as well as the Ca II H&K lines at 393 and 397 nm, and the Q and R branches of CH radical, from 419–431 nm. The P branch, less intense, extends from 432–441 nm. The F47 band measures a pseudo-continuum, in a region where absorption by atomic and molecular lines is minimum, which is used with the F75 flux to derive $T_{\rm eff}$. The perturbation by H_β at 481 nm is small. There is a partial balance of the absorption by CH in the F41 and F47 bands, minimizing the effect of C abundance anomalies. The sensitivity to [M/H] of the colour index F41−F47 begins at F0, reaches a maximum around K0 then declines to zero around K7 for dwarfs, while for giants its sensitivity increases up to K5. This index is a sensitive metallicity estimator for evolved late-type stars. Since line blocking is dominated by the effects of the iron peak elements, the deduced [M/H] is closely related to [Fe/H]. For cooler stars, F41−F47 is no longer a metallicity index because of the contamination of the F47 band by TiO bands.

F51 and F57 bands The MgH+Mg b complex is a sensitive gravity indicator for late G to early M stars, whose intensity reaches a maximum around K7V. The MgH band intensity is nearly independent of [Mg/H] for G and K dwarfs and sub-giants, and its absorption in the F51 band is measured relatively to the fluxes in F47 and F57 bands. The F57 band measures a broad pseudo-continuum with very low absorption by metallic lines, except by Na D which becomes very strong in super metal-rich stars. For stars later than K4, the F57 band is contaminated by TiO. As the F47 band is also contaminated by TiO, the F47−F57 index remains a useful temperature parameter in addition to the F57−F75 index. In very metal-poor low-luminosity dwarfs, the MgH band remains strong whereas TiO bands vanish completely. Red colours, low TiO and strong MgH are thus typical signatures of halo counterparts of M dwarfs.

F67 and F65N bands The broad interval, 620–720 nm, features strong TiO bands in late-type stars. The depression in this domain, relative to the pseudo-continua in the F57 and F75 bands, is a good indicator of TiO abundance for M dwarfs. The main purpose of the F65N band is the detection (when present) of emission features in the spectrum, i.e. from the Hα and [N I] 655/658 nm spectral features, thus allowing to detect peculiar objects (such as Be stars, T Tau stars, etc.). The band size is chosen to allow detection of both Hα and [N I] 655/658 nm even in high-radial velocity stars.

F75, F78, F82 and F89 bands These measure the temperature and the TiO and the CN abundances of red stars with $T_{\rm eff} < 4200$ K. The F75 and F89 bands measure portions of the spectrum where the absorption by TiO bands is minimum, determining a pseudo-continuum. In B, A, F stars the F89 band measures the Paschen jump and lines, whereas the continuum is measured by F75, F78 and F82 bands. The F82 band measures either the continuum bluewards of the Paschen jump (in early-type stars), or a strong CN band for R, N, C stars. For M stars, F82 measures a spectral domain with weak absorption by TiO. These bands allow to distinguish between M and C stars, even if heavily reddened: at a given temperature, the fluxes are similar in the F78 and F89 for O-rich stars (the M sequence) and for C-rich stars (the C sequence), but very different in the F75 and F82 bands, namely because of strong CN bands developing redwards of 787 nm. These spectral regions are characterized by strong absorption bands of water, so that—similarly to the F33 band—uniform and unbiased measurements in these bands can only be obtained with a space-based telescope such as GAIA.

F83B, F63B, F45B and F33B bands These broad-band filters (placed on the Astro focal plane) cover the ranges of the F75, F78, F82 and F89 filters (F83B), the F57 and F67 filters (F63B), the F41, F47 and F51 filters (F45B), and the F33 filter (F33B). While the information they provide is thus redundant from the astrophysical point of view, their broader band is supplying a higher S/N and higher angular resolution for chromatic correction purposes. The F33B filter has the same width as F33 and can thus be used to improve the accuracy of the UV flux measurement.

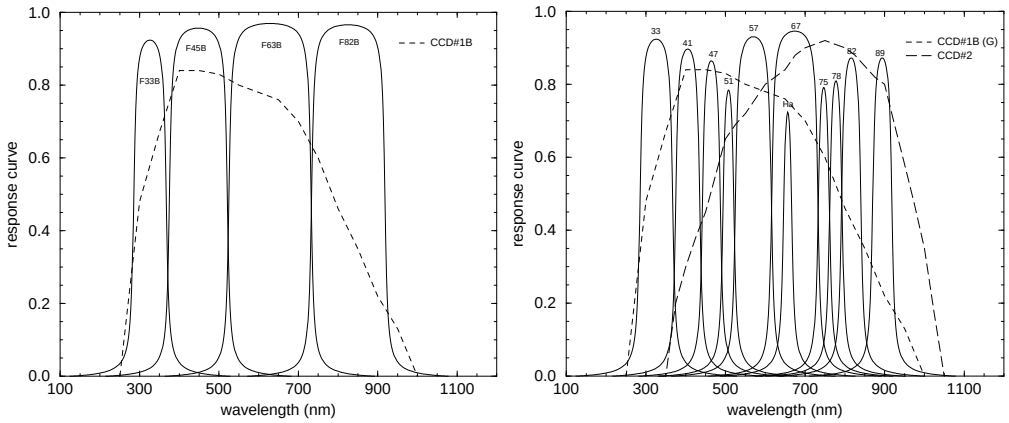


Figure 2.12: Filter transmission curves and CCD response curves for the broad-band and medium-band photometric systems. The left panel shows the broad-band filters located in the Astro telescopes, together with the response for CCD#1B used in that instrument. The right panel shows the medium-band filters located in the Spectro telescope and the response curves for CCD#1B and CCD#2.

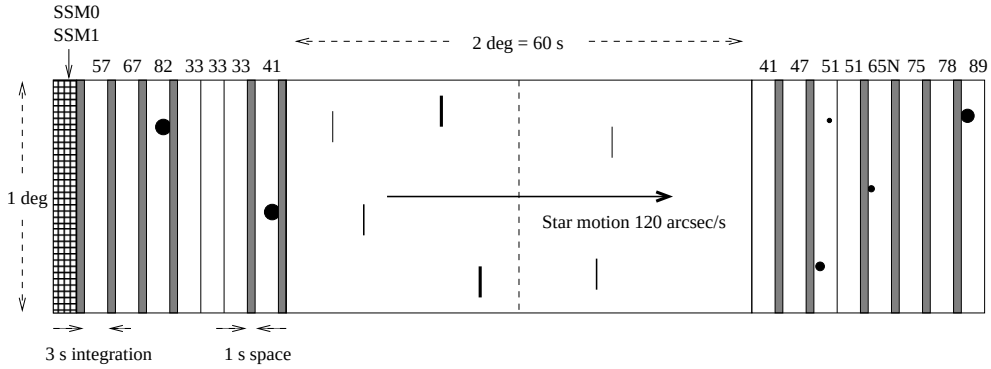


Figure 2.13: Layout for the medium-band photometer focal plane, showing the location and characteristics of the filters. The dedicated sky mapper and medium-band photometer lie in the outer parts of the field, with the radial velocity spectrometer (RVS) occupying the optically superior central region of the field of view. The baseline width of the RVS is 1 degree, but an extension to 2 degree (as shown) is targetted. The MBP filters are arranged in 'units' of 3 s integration time; different filters require a loss of ∼1 s between them. SSM0/1 are the spectrometric sky mappers.

Using the bands described above, different colour indices are defined, from which the astrophysical parameters required for each stellar population are derived. The characteristics of early-type stars (OB, up to A2) will be derived using 'reddening-free parameters', e.g. by their position in a plane defined by the indices $Q_{33-47-57-75}$ = F33 – F47 – 1.50×(F57 – F75) and $Q_{33-41-41-47}$ = F33 – F41 – 1.65×(F41 – F47). This (together with the DIB and H I maps) will allow local reddening to be determined, building the map to be used for the later-type stars. For them, effective temperatures can be determined (in a metallicity-independent way) through the index F47 – F57 for FG stars, F57 – F75 for K stars, and through the index F75 – F89 for M stars. The abundance of α elements is measured through the MgH index ($I_{\rm MgH}$ = F47 – 2×F51 + F57 + 0.11×[F57 – F75]) in the F and G stars and through the TiO index ($I_{\rm TiO}$ = F75 – F78 – F78 – F89 + 0.26×C57 – F75 – 0.08) for later (K and early M) stars. The global metallicity, [M/H], is obtained through the index F41–F47, which measures the break at 430 nm. This index needs to be corrected, for later spectral types, by the TiO abundance, so that the abundance indicator is defined as $I_{\rm Z}$ = F41 – F47 – 0.6×$I_{\rm TiO}$. $I_{\rm CN}$ = (F75 – F82) – (F82 – F89) is used to measure the N abundance of red stars with $T_{\rm eff} < 4200$ K. The approach to the derivation of the astrophysical parameters is summarized in Table 2.3, and has been discussed by Cayrel et al. (1999).

Table 2.2: Overview of the filter bands for the GAIA photometric system. The table lists, for each band, the central wavelength λ_c, the full width of the band $\Delta\lambda$, the peak transmission of the filter T_{max}, and the integration time τ for a single transit in both the Astro and Spectro focal planes. The first 11 filters occupy the Spectro focal plane. The last four, with suffix B indicating broad-band, occupy the Astro focal planes.

Band	F33	F41	F47	F51	F57	F65N	F67	F75	F78	F82	F89	F33B	F45B	F63B	F83B
λ_c (nm)	326	405	465	508	570	656	674	747	778	816	894	326	448	629	826
$\Delta\lambda$ (nm)	82	60	45	27	90	4	116	28	31	48	48	82	146	208	186
T_{max}	92	90	86	78	93	80	94	79	81	87	87	92	96	97	97
τ (s)	3×3	2×3	3	2×3	3	3	3	3	3	3	3	0.86	0.86	0.86	0.86

Table 2.3: Summary of the derivation of the astrophysical parameters. See main text for the definition of the indices.

	O–B–A2	late A – dFG	dK – dM	gK	gM
A_V	Q-Q diagram	3-D map	3-D map	3-D map	3-D map
$T_{\rm eff}$		F47−F57	F57−F75	F57−F75	F75−F89
$[Fe/H]$		I_Z	I_Z	I_Z	I_Z
$[\alpha/H]$		$I_{\rm Mg}$	$I_{\rm Mg}$	$I_{\rm Mg}$	$I_{\rm TiO}$
$\log g$	Q-Q, π, A_V	π, A_V	π, A_V	π, A_V	π, A_V
peculiar	$I_{\rm H\alpha}$	$I_{\rm H\alpha}$	$I_{\rm H\alpha}$	$I_{\rm H\alpha}$	$I_{\rm H\alpha}$

Width of the photometric bands The evaluation of the optimal size of a photometric bandpass, once its position is fixed, requires careful balancing of two effects: on one hand the use of medium bands closely placed next to each other (one measuring the given spectral feature, the other a nearby continuum) maximizes the sensitivity to the desired spectral feature (e.g. the MgH band); on the other hand narrow bands will not reach as faint a limiting magnitude as broader ones. Broadening the bands does not however automatically yield a gain: while fainter magnitudes will be reached, this will happen by collecting (mostly) additional continuum photons (which carry no information about the desired spectral feature). Also, by distancing the effective wavelengths of the two bands, the indices become more sensitive to reddening. Thus, the final size of the photometric bandpasses must carefully balance these two effects.

2.3.3 Options Not Retained in the Proposed Configuration

Low-resolution spectrometry Low-resolution spectrophotometry was considered as an alternative or complement to medium- and broad-band photometry, e.g. in the context of an astrometric interferometer using dispersed fringes (Høg et al. 1997). Introduction of a low-resolution spectrograph for spectrophotometry (Favata 1998; Høg 1998f) was considered as well, using an optical configuration similar to that of the radial-velocity spectrometer (Section 3.5). Overlap of spectra would be a significant problem especially at low galactic latitudes. The available focal area for filter photometry would be reduced and the instrument complexity and telemetry rate considerably increased. A low-resolution spectrometer working in a more restricted wavelength range would facilitate detection of objects with spectral peculiarities such as QSO's and Be stars. If the region 380–500 nm is covered at a resolution of about 5 nm/pixel, i.e. 24 samples per spectrum, it would for instance be possible to detect emission lines in stars brighter than 20 mag. The required raw data rate, about 1600 kbits/s, would however be a very significant load.

2.4 On-Board Detection

While there is some scientific merit in transmitting to ground the entire data stream from the focal plane CCDs, certain issues conspire to make this approach unrealistic. Full focal plane read-out in TDI mode would result in date rates of the order of several Gbit s^{-1}, compared with a realistic upper limit on the sustained telemetry rate from the L2 orbit of some 1–2 Mbit s^{-1}. Even if there were no telemetry rate limitations, the high read-out rates would result in a high read-out noise, and a correspondingly degraded signal-to-noise ratio per pixel, or in an excessive number of video chains. A detailed analysis has shown that full focal plane read-out is of little value: for most of

the sky, especially out of the Galactic plane, the fractional area covered by stars is very small, even at 20 mag. The down-link of small patches of the detector field of view, centred on each detected object, and extending a little beyond the Airy disks, is a fully satisfactory solution. Required data rates then drop to manageable levels (below 1 Mbit s^{-1}), and the read-out scheme can be designed to bring read-out noise per pixel down to adequate levels for a reasonable number of video chains, even for the faintest stars.

Given that the focal plane data on all objects down to about 20 mag can be read out and telemetered to ground, the central problem is how to identify the patches of sky containing the objects. Two options have been evaluated: (i) use of an input catalogue; and (ii) the on-board detection of targets. The issue is not a purely operational one: a clear definition and understanding of the selection function used to decide which targets to observe (for example the completeness as a function of magnitude) is a crucial scientific issue, strongly driving the final scientific output of the mission. The solution adopted is to detect each and every prospective target on-board, by means of the 'astrometric sky mapper'. This has the advantage that the detection will be carried out in the same wave-band, and at the same angular resolution, as the final observations. It can be made sufficiently robust to ensure that the same stars are reliably observed on a large number of different great circle scans. The approach also allows some quick-look facilities to be built into the ground-processing segment, to identify and follow-up on time-variable and burst sources.

Several problems make the construction of an input catalogue unfeasible: (i) such a catalogue, complete to the required magnitude limits, is unavailable: existing (ground-based) all-sky surveys are confusion-limited in crowded regions (i.e., essentially most of the Galactic plane) at magnitudes much brighter than the projected limiting magnitude for GAIA. This is true even of the recent digital surveys. Furthermore, especially in view of the following point, such a catalogue could almost certainly not be available even in the foreseeable feature; (ii) the GAIA passband is both broad and rather red. Available extensive deep ground-based material is essentially photographic, and thus mostly in a bluer band. Current IR all-sky surveys have both relatively bright limiting magnitudes and are also confusion-limited in crowded areas. More importantly, using an input catalogue built in a given band to observe in another band does not satisfy the requirement of a clearly known selection function, and this consideration alone would call for an input catalogue constructed strictly from homogeneous material in terms of survey wavelength; (iii) the same argument applies to the angular resolution of such a survey: unless acquired at the same high, sub-arcsec, angular resolution as the GAIA astrometric observations, the measurements would be biased to certain scale-lengths of resolved (double star) structure; (iv) such an input catalogue, even if possible in principle, would necessarily omit several important classes of target objects: variable stars; burst sources (supernovae and gamma-ray bursters); and the very large number ($10^5 - 10^6$) of rapidly moving Solar system objects down to 20 mag. Thus, a useful and robust input catalogue seems to be quite unrealistic. A further alternative, to build an input catalogue the first time a sky patch is observed, and then use that source list to ensure that the same targets are re-observed, was investigated. This has few merits, especially in view of variable stars, burst sources, and Solar system objects.

2.5 Concept Development

In the preceding parts of this section, the requirements for the various measurements have been presented. For completeness, we outline here the major steps that were followed in converging on a system design which fulfills all of these criteria.

2.5.1 Starting Point

The present study started from the overall concept proposed by Lindegren & Perryman (1996): since the measurement principle requires a high image localisation accuracy in one direction only (the scan direction), the natural idea was to make use of interferometry by considering two relatively small apertures (50–60 cm diameter) separated by a significant distance (the baseline length, $B_{\rm eff} \sim 2.5$ m). The optical system could be viewed as a 'three-mirror anastigmatic' telescope, each mirror being segmented in two parts. The high spatial frequency content of the Young's fringe pattern produced in the focal plane (fringe period $\lambda_{\rm eff} F/B_{\rm eff}$, where F is the focal length) provided the required image localisation accuracy. The proposed detection system was initially based on a pupil detection method; this was abandoned at the beginning of the Concept and Technology Study in favour of a direct fringe detection approach, motivated by a search for higher astrometric accuracy, as well as the problem of faint star confusion.

The configuration proposed by Lindegren & Perryman (1996) suffered from several drawbacks, notably: (i) the focal length (11 m) and the proposed baseline (2.5 m) led to a fringe period of about 2.5 μm in the focal plane, and therefore would require a pixel size in the sub-micron range. Such detector technology, providing simultaneously a high quantum efficiency, does not exist, and is unlikely to be available in the foreseeable future; (ii) the segmented mirror concept requires stringent alignment mechanisms (accuracy range of a few nm) necessary to achieve good visibility; this is feasible, but was shown to be unnecessary; (iii) the optical concept implied a stacked configuration for the interferometers. The overall opto-mechanical configuration suffered from a lack of symmetry which was not favourable to the optical stability and would not facilitate the basic angle monitoring in orbit; (iv) the large number of fringes, about eight per Airy pattern, implied a high data rate, in conflict with the preference to use the L2 Lagrange orbit. At the same time, the large number of fringes would preclude the use of a wide wavelength band (desirable for improved photon statistics) because fringe blurring strongly reduces the spatial frequency content of the image, except for the central fringe.

Fringe dispersion In the original proposal a restricted optical bandpass, $\Delta\lambda$, was required to maintain good modulation of the interference fringes across the Airy disk. Requiring that the fringe contrast does not fall more rapidly than the mean intensity results in the condition $\Delta\lambda/\lambda_{\rm eff} \lesssim D/B$: increasing the bandwidth increases the number of photons that can be used for the phase determination, but also reduces the modulation of the signal. At the same time, the requirements for multicolour photometry still applied. One possible solution was to disperse the fringes in the spectral domain, leading to 'fan-like' fringes. Simulations showed that the accuracy gain was modest for bright stars, and absent for the faintest stars, whilst calling for very low read-out noise at high read rates ($< 3\,{\rm e}^-$) so as not to avoid loss of precision. Smearing of the spectrum along the dispersion, and overlapping/confusion of the dispersed fringes (for faint stars, high-density regions, and binary systems including planetary candidates) would further degrade the performance.

In the original concept, radial velocities were obtained by a separate payload element. Photometric measurements were performed throughout the whole focal plane, with the central astrometric CCDs covered with different wide-band filters. A medium-band filter set (resembling the Strömgren system) were placed at the edge of the field. The proposed solution had several disadvantages, including the loss of photons caused by the use of filters on the astrometric CCDs (necessary for an interferometer). At this stage, the photometric system had not been optimized.

2.5.2 First Iteration

The aim of the design improvement study performed during Phase 1 was to maintain the fundamental ideas behind the initial concept (one-dimensional measurements, use of white light, etc.) while removing the above-mentioned drawbacks, and searching for the simplest conceptual design which could meet the required performance for the lowest possible cost by: (i) relying on available detector technology; (ii) simplifying the payload alignment control, and minimising (or suppressing) high-accuracy (nm-level) mechanisms; (iii) reducing the data rate; and (iv) maintaining a simple accommodation and interface with the radial velocity instrument. The work was guided by a parametric performance analysis based on the method outlined in Section 7.

The first step involved searching for an optical design which could preserve some symmetry for the interferometer accommodation and simultaneously provide a large focal length (~ 50 m) and a flat field for increasing the pixel size to a reasonable value, say above 6–7 μm (although such a pixel size would not be compatible with available technology, it was considered as a realistic starting point potentially achievable in the near future, such small devices being available now, albeit without the required performances in particular in terms of quantum efficiency and MTF). The parametric performance analysis showed that the useful field of view should be larger than about 0.25 deg^2. Full benefit of the whole allowable volume of the payload module could be taken by arranging the optics about the spin axis, rather than having a planar configuration with stacked interferometers.

The large focal length and flat focal plane condition led to a 'three-mirror anastigmatic' design, with an intermediate focus between the secondary and tertiary mirrors. The large focal length led to a design in which only the primary mirror remained segmented in two parts. A useful field of view (with WFE $< \lambda/30$ at $\lambda = 600$ nm) of up to 0.5 deg^2 was achieved, with a corresponding linear focal plane dimension of about 60×60 cm^2, providing sufficient margin for the detector

filling factor, star detection, and broad-band filters. The pixel size requirement led to rather severe constraints on the interferometer baseline in the range 1–1.5 m. Performance analysis showed a good compromise with a baseline of 1.3 m, and subaperture dimensions of 0.7×0.7 m^2, with a total collecting area of about 1 m^2. Having a large focal length implies that the focal plane is rather large, and must be made of several tens of CCD arrays, each of approximate size 6×3 cm^2. Almost coincidentally, it emerged that such sub-division was also demanded by TDI-phasing requirements in the presence of the optical distortion. Large focal planes of comparable size have been, or are currently being, manufactured for ground-based astrometry (e.g., Sloan, ESO OmegaCam, etc).

Wavelength range considerations Many of the (fainter) GAIA targets have intrinsically red colours (red giants, low mass dwarfs). Also, interstellar absorption will further redden many of the targets, especially close to the Galactic plane. This makes it desirable to have a detection passband with a strong response in the red or even near infrared. This had to be weighed against the decrease in the intrinsic maximum accuracy with increasing wavelength (thus lowering the maximum achievable precision for bluer objects), as well as with the available detector technology and characteristics. Given that CCD detectors (because of their maturity, and the need for TDI operation) appear to be the only current option for the GAIA focal plane, the final wavelength optimisation had to take into account, in addition to these 'astrophysical' considerations, the decrease of CCD QE at longer wavelengths, especially for smaller pixel sizes, in order to preserve MTF. The 'best compromise' central wavelength is $\simeq 700$ nm.

Although the revised optical design was compatible with a three-interferometer configuration, it was recognised that the advantages provided by a third interferometer (mainly an improvement in the control of large-scale reference frame distortions at the μas-level) could not outweigh the numerous advantages of a two-interferometer configuration. Including a third interferometer results in an additional optical system, focal plane and electronics system (affecting cost and complexity, also during verification and testing), an increased data rate by a factor of 3/2, a more complex mechanical and thermal configuration, and additional complexity in accommodating the radial velocity instrument. A two-interferometer configuration was therefore adopted as baseline, with a basic angle of $\gamma = 106°$. Compared with the stacked interferometer design the payload module height, including the radial velocity instrument, was reduced from about 3 m to less than 2 m. The sunshield diameter was thereby reduced from 13 m to 8.5 m, conveniently corresponding to the required diameter of the deployable solar arrays. In the process, the sunshield and solar array functions were merged into a single deployable structure of acceptable size and complexity. The payload opto-mechanical symmetry and performances were improved, and the critical alignment mechanisms were limited to a piston and tip-tilt control of the primary mirror parts. Simultaneously, the number of fringes in the Airy disk was reduced from 8 to 3; although the signal was spread over about 16 pixels, the performance analysis showed that most of the astrometric accuracy was contained in the central fringe, and that the star localisation process was only marginally degraded by sending only 6 pixels per star to the ground.

Beam combiner trade-off In the original GAIA proposal (Lindegren & Perryman 1996), and in the present design, each field of view is observed by a separate interferometer. An alternative option is to combine two fields of view in a single focal plane, using a 'beam combiner' (as adopted for Hipparcos). The advantages and disadvantages of the two approaches were evaluated during the course of the study. The advantages of a beam combiner element are: (i) it results in a single focal plane assembly (although with the two fields superimposed); (ii) it yields a possibly more compact design (each instrument simultaneously services two fields of view), thus reducing the total spacecraft height, with the potential positive implications on the design of the sun-shield; (iii) the beam-combiner would constitute a physical realization of one of the basic quantities which affect the GAIA observations, i.e., the basic angle. This could perhaps make the metrology easier, although this point remains debatable; (iv) by concentrating the basic angle in a restricted volume, thermo-mechanical variations should have a smaller influence. The disadvantages of the beam combiner option are: (i) by superimposing two separate fields of view, the beam-combiner will double the chance of source confusion. This would become a particular problem in the most crowded areas of the sky; (ii) given that the foreseen scanning law involves a (minor) transverse motion of the stars across the field of view, the presence of a beam splitter introduces an additional element of complication, as the stars belonging to the two fields of view would move with different transverse velocities in the field of view. While the magnitude of this differential transverse motion is not very large, its impact on the proposed focal plane design and operation would complicate on-board operation, especially in conjunction with the adopted approach of on-board source selection; (iii) such a large beam combiner would probably be rather massive. The corresponding mass penalty must be weighed against any saving derived by the reduction in the number of instruments.

Once the optical design evolved away from an interferometer, the placement of broad-band filters on the astrometric detectors was neither necessary nor desirable. Instead, it was proposed to use the edges of the astrometric field of view for broad-band photometry (with as main purpose to provide the information necessary for the chromatic correction), and to move the medium-band

filters (with as main purpose to provide astrophysical diagnostics) to the edges of the spectroscopic field of view. At this stage the Stromvil photometric system was considered for this purpose. This approach did not make use in an integrated way of all available photometric information (by keeping a rather artificial distinction between narrow bands and broad bands), and did not result in a well balanced focal plane. Also, the photometric system was not optimized for the redder stars, which are the main GAIA kinematic tracers. Therefore at the same time an ab initio study of the spectro-photometric characteristics of the target GAIA population was carried out, which resulted in a proposal for a set of photometric bands optimized for the main GAIA populations. The desire to optimize the S/N across the whole wavelength range resulted in a total integration of the two functions of photometry (chromatic correction and astrophysical information), leading to broader photometric bands, making optimal use of the red spectral region. By obtaining all the information from a unified focal plane, differential aging effects are minimized.

2.5.3 Second Iteration

The analysis performed during the first iteration provoked two rather fundamental questions about the optical design: (i) since the primary mirror was now the only segmented mirror, and since the gap between the two mirrors was (only) 0.6 m, was it of interest to fill the gap between the two mirrors, and thus to deal with a monolithic primary mirror? (ii) was it still useful to preserve a primary reflector made of two separated mirrors producing a pattern with only 3 fringes, since the useful astrometric information was mainly contained within the central fringe? The performance analysis produced results that were perhaps not entirely intuitive: by filling the gap between the two apertures, the astrometric accuracy improved significantly, from about 11 μas to about 7 μas at 15 mag for blue and yellow stars, due to the larger number of collected photons, and due to the concentration of the photons in the central (now unique) fringe. Conversely, it was possible to reach the original astrometric accuracy using a monolithic primary mirror of smaller along-scan dimension than the edge-to-edge diameter of the segmented mirror.

This final step, of making the primary reflector monolithic, with dimensions 1.7 m (along scan) by 0.7 m across scan, provided a further breakthrough in terms of minimising satellite complexity and cost: (i) the primary reflector could now be made stiff enough that no active control would be required in orbit, and no high-accuracy mechanisms would be needed; (ii) the reduction of the overall reflector diameter along scan slightly enlarges the Airy pattern along scan, and allows further increase of the pixel size to 9 μm along scan without performance degradation—this is of considerable importance since large, high-efficiency backside illuminated CCD arrays can be built with such pixel size with available technologies (in the USA and in Europe); (iii) the decrease of the overall aperture size minimizes the overall payload diameter (easing launcher accommodation) and required sunshield diameter; (iv) the increase of the collecting aperture, from 0.98 m^2 to 1.19 m^2, is favourable to the on-board star detection process. Additionally, discarding the interferometric option has one very significant advantage for the science case: the limiting magnitude of the observations will reach a full 2 mag fainter (to 20 mag rather than 18 mag), due to the improvement in signal-to-noise ratio with the signal spread out over fewer detector pixels.

This presented the Science Advisory Group with an interesting choice: should it pursue the design of an interferometric payload in order to satisfy the original Horizon 2000 'green dream' proposal to fly an optical interferometer in space, even though it could be shown to be scientifically and technologically both undesirable and unnecessary for the adopted goals? Or should it take the scientific goals established by the Horizon 2000+ Survey Committee (10 μas global astrometry) and present a greatly simplified mission which could achieve these goals, responding to demands for missions meeting the criteria of 'better, cheaper, faster'? The final decision has been to pursue the simplest and lowest risk technical design consistent with the scientific objectives, and to adopt minimum payload design parameters (collecting area, etc) consistent with these objectives. The enormous capabilities of GAIA in terms of limiting magnitude ($\sim$ 20 mag) and numbers of objects ($>$ 1 billion) follow from the simplified design proposed in this document. If an interferometric design should be demanded for other reasons, the trade-offs are now fully evident.

3 Payload

3.1 Measurement Principles

3.1.1 Overview

Taking account of the basic design considerations presented in Section 2, the general principles of the proposed mission can be summarized as follows: (i) it is a continuously scanning instrument, capable of measuring simultaneously the angular separations of thousands of star images as they pass across a field of view of about 1° diameter. Simultaneous multi-colour photometry of all astrometric targets is a necessary and integral part of the concept; (ii) high angular resolution in the scanning direction is provided by a monolithic mirror of dimension ~ 1.7 m, or by employing a small interferometer with apertures separated in the scan direction (in the latter case, the large field of view implies the use of Fizeau type interferometers, with the maximum baseline limited by the requirement for a non-deployable payload compatible with the Ariane 5 launcher); (iii) the wide-angle measuring capability is provided by two or more viewing directions at large angles to each other and scanning the same great circle on the sky. The precise 'basic angle' between two viewing directions is determined from the 360° closure condition on each great-circle scan, while short-term (< 3 hours) variations are passively controlled, and monitored by internal metrology; (iv) the whole sky is systematically scanned in such a manner that observations extending over several years permit a complete separation of the astrometric parameters describing the motions and distances of the stars. A longer temporal baseline permits the determination of additional parameters, for example those relevant to orbital binaries, or to the detection of extra-solar planetary companions.

Within this general outline many options exist which have been extensively explored, optimized and weighed against each other. These include, for instance, the number and optical design of each viewing direction, and the choice of wavelength bands, detection systems, detector sampling strategies, basic angle, metrology system, satellite layout, and orbit. The proposed payload design (Figure 3.1) consists of:

(a) two astrometric viewing directions. Each of these 'Astro' instruments comprises an all-reflective three-mirror telescope with an aperture of 1.7×0.7 m^2, the two fields separated by a basic angle of 106°. Each astrometric field comprises an astrometric sky mapper (ASM), the astrometric field proper (AF), and a broad-band photometer (BBP). Each sky mapper system provides an on-board capability for star detection and selection, and for the star position and satellite scan-speed measurement. The main focal plane assembly employs CCD technology, with about 250 CCDs and accompanying video chains per focal plane, a pixel size 9 μm along scan, TDI operation, and an integration time of ~ 0.9 s per CCD;

(b) an integrated radial velocity spectrometer and photometric instrument ('Spectro'), comprising an all-reflective three-mirror telescope of aperture 0.75×0.70 m^2. The field of view is separated into a dedicated sky mapper (SM), the radial velocity spectrometer (RVS), and a medium-band photometer (MBP). Both instrument focal planes are also based on CCD technology operating in TDI mode: with at least 3 large CCDs butted together for the radial velocity spectrometer; and two large CCDs, with a total of 11 medium-band filters, for the medium-band photometer.

(c) the opto-mechanical-thermal assembly comprising: (i) a single structural torus supporting all mirrors and focal planes, employing SiC for both mirrors and structure. There is a symmetrical configuration for the two astrometric viewing directions, with the Spectro system accommodated within the same structure, between the two astrometric viewing directions; (ii) a deployable sun-shield to avoid direct Sun illumination and rotating shadows on the payload module, combined with the solar array assembly; (iii) control of the heat injection from the service module into the payload module, and control of the focal plane assembly power dissipation in order to provide an ultra-stable internal thermal environment; (iv) an alignment mechanism on the secondary mirror for each astrometric instrument, with micron-level positional accuracy and 200 μm range, to correct for telescope aberration and mirror misalignment at the beginning of life; (v) a permanent monitoring of the basic angle, but without active control on board.

3.1.2 Accuracy Requirements

The primary requirement for GAIA is to reach a 10 μas rms positional accuracy for stars of magnitude $V = 15$ mag (comparable figures follow for the parallax and proper motion components), although there are a number of complementary considerations: (i) the field height must provide adequate overlap between two successive great circles for the data processing: with the selected scanning law, this requirement leads to a field height of at least $0\overset{\circ}{.}6$; (ii) the goal is to obtain a catalogue largely complete to as faint a magnitude limit as possibly, currently estimated at $V = 20$ mag, corresponding to about 1.3 billion objects. Simultaneously extending the observations to the brightest stars implies an extremely broad dynamic range, and is probably of less profound scientific relevance. According to the present design, stars as bright as $V = 6 - 7$ mag can be observed, and this could be extended to brighter stars if demanded scientifically; (iii) no particular requirement has been adopted for the specification of accuracy at other than 15 mag: for fainter magnitudes, the accuracy falls to about $20 - 40$ μas at $V = 17 - 18$ mag, and to $100 - 200$ μas at $V = 20$ mag, entirely due to photon statistics. For $V < 15$ mag, higher accuracy is achieved, but will be limited by systematic effects at about $3-4$ μas for $V < 10-11$ mag; (iv) colour information is to be acquired for all observed objects. The main motivation is to provide astrophysical information about the objects for which distances and space motions are derived, although calibration of the instrument's chromatic dependence is a key secondary consideration. Four wide-band filters are located in the main astrometric fields, and 11 medium-band filters are located in the photometric field; (v) while the astrometric instrument provides five astrometric parameters (angular positions, proper motion and parallax), an integral objective of the mission is to provide the sixth astrometric parameter, namely the radial velocity, by measuring the Doppler shift of selected spectral lines; (vi) raw data representing the star profile along scan must be sent to ground, with on-board data reduction not considered because of the irreversible loss of information.

3.1.3 Performance Assessment

The astrometric accuracy can be separated into two independent terms, the random part induced by photo-electron statistics on the localisation process accuracy, and a bias error which is independent of the number of collected photons. The random part decreases as $N^{-0.5}$, where N is the number of detected electrons per star, while the bias part is independent of N and represents the ultimate capability of the system for very bright stars. In order to provide high accuracy for brighter stars, residual biases are targeted as significantly lower than 10 μas.

The constant part of the geometrical transformations and all variations on time scales exceeding a few times the spin period of 3 hrs can be fully calibrated as part of the scientific data reductions. Consequently the requirements for alignment and metrology are determined by the following two main goals: (1) to maintain sufficient image quality at all times throughout the field of view; and (2) to control or measure the variations of the geometrical transformation (including the basic angle) on time scales shorter than about 5 hrs to a precision compatible with the overall mission goals, i.e., about 10 microarcsec. Therefore, while bias errors often dominate in measurement systems, few are relevant to the GAIA error budget. Examples of fixed biases which might affect the localisation error for a single star, and which will be determined by and corrected for during the global sky data processing, are inhomogeneities in the pixel quantum efficiency, geometric errors of the CCD, detector alignment residual errors, fixed (or slowly variable) fluctuations of the optical point spread function over the field of view, and optical distortions affecting the correspondence between a point of the focal plane and the sky angular coordinates. The satellite attitude is fully self-calibrated at low frequencies due to the simultaneity of the star measurements. Systematic terms which must be considered explicitly are: (i) basic angle fluctuations, considered in detail in Section 4.2.3; (ii) chromatic shift of the diffraction peak due to optical aberrations, considered in detail in Section 3.2.4: (iii) variations of the charge transfer efficiency of the CCD, considered in detail in Section 3.7.6.

Random errors are due to photon statistics and affect the localisation process. The effects which have been included in the model for the performance analysis include: (i) pupil geometry and diffraction effect; (ii) CCD quantum efficiency (amplitude and spectral distribution); (iii) pixel geometry and MTF; (iv) smearing effects due to optical distortion, TDI rate errors and telescope line-of-sight jitter; (v) effect of optical aberrations on the diffraction peak loss (Strehl ratio); (vi) discrete sampling of the signal and number of samples used; (vii) reduced detection probability for the faintest stars. Details of the accuracy analysis are given in Section 7.

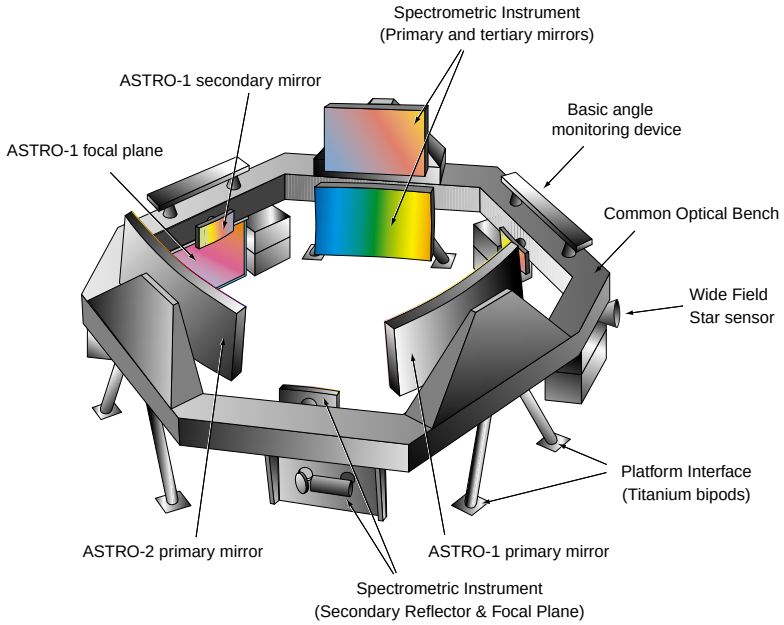


Figure 3.1: The payload includes two identical astrometric instruments (astro-1 and astro-2) separated by the 106° basic angle, as well as a spectrometric instrument (a radial velocity measurement instrument and a medium-band photometer) which share the focal plane of a third viewing direction.

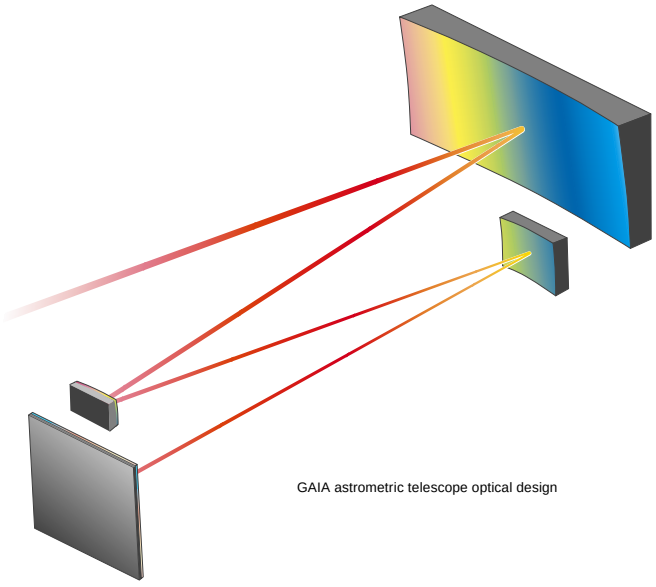


Figure 3.2: The light path of the astrometric instruments.

3.1.4 Scanning Law and Pointing Performance

In contrast to point-and-stare missions as exemplified by the Hubble Space Telescope and SIM projects, the observation mode adopted for GAIA is based on continuous sky scanning. The rationale for this particular choice, which deeply affects nearly all technical, operational and scientific aspects of the mission, may be summarized as follows: (a) it is optimally suited for a global, survey-type mission with very many targets; (b) astrophysical research primarily oriented towards the physics of stars and of the Galaxy as a whole is best served by a survey-type mission encompassing large and well-defined samples of a wide variety of objects; (c) from a technical viewpoint, a continuously scanning satellite appears to be particularly efficient, and probably optimal, in terms of observing time utilization and instrument stability and calibration, in particular in terms of the overall rigidity of the resulting reference frame; (d) it follows a proven concept, namely that of the Hipparcos mission.

The continuous scanning approach has a variety of important design consequences: a constant geometry with respect to the Sun is possible; that critical instrument parameters such as the basic angle, scale value and geometrical field distortion are obtained from the closure conditions on each complete satellite rotation; that these calibrations are part of the normal observations and therefore do not require any observational overhead; that no overhead is required for re-pointing the telescope; and that many objects are observed strictly simultaneously.

The satellite scans the sky according to a pre-defined pattern in which the axis of rotation (perpendicular to the three viewing directions) is kept at a nominally fixed angle ξ from the Sun, describing a precessional motion about the solar direction at constant speed with respect to the stars. Resulting satellite pointing performances are determined from operational and scientific processing requirements on ground, and are summarised in Table 3.1. In the table, 0.05 Hz is the maximum frequency that can be identified after measurement post-processing.

A revolving scanning law similar to that used for Hipparcos has been adopted for GAIA: no better scanning law has been found despite several alternative suggestions. On the other hand, the revolving scanning has some well-known shortcomings concerning the non-uniformity of sky coverage, in particular the relatively low density of observations in the ecliptic region, and the anisotropy of accuracy and absence of along-ecliptic scans in the same region. The constrained scan geometry for objects close to the ecliptic reflects negatively on the ability to resolve and reconstruct complex objects such as multiple stars or overlapping images in crowded regions. These deficiencies are however inherent to the constraint on the solar aspect angle (ξ, between the Sun and the nominal spin axis), which derives more directly from requirements of shading and straylight suppression, and is not specific to the revolving scanning law as such. These drawbacks may be minimised, although not eliminated, by adopting a larger solar aspect angle than the $\xi = 43^\circ$ used for Hipparcos—the value $\xi = 55^\circ$ is the baseline for GAIA.

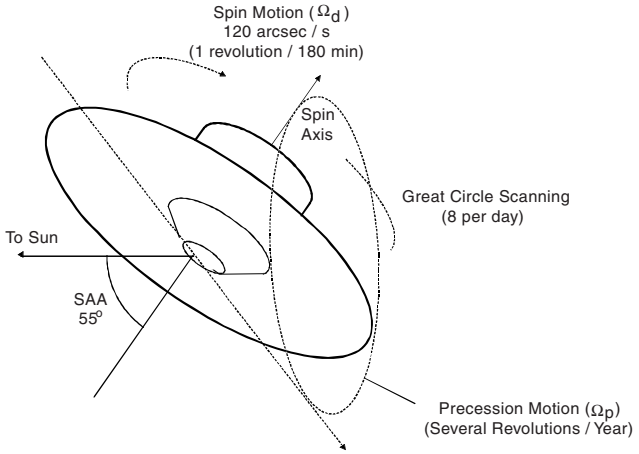


Figure 3.3: Characteristics of the observation strategy. The satellite rotates slowly about its spin axis, which itself performs a slower precessional motion at a fixed angle of 55° *to the Sun.*

Table 3.1: Scanning law and pointing requirements summary.

Parameter	Value
Satellite scan axis tilt angle	55° with respect to Sun direction
Scan rate	120 arcsec s^{-1}
Absolute scan rate error	1.2 arcsec s^{-1} (3σ)
Precession rate	0.17 arcsec s^{-1}
Absolute precession rate error	0.1 arcsec s^{-1} (3σ)
Absolute pointing error	5 arcmin (3σ)
Attitude absolute measurement error	0.001 arcsec (1σ)
High-frequency disturbances:	
power spectral density at 0.05 Hz	≤ 1000 μas^2 Hz^{-1}
for $f > 0.05$ Hz	decreasing at least as f^{-2}

Additionally, the accuracy of parallax determinations depends directly on this angle, with $\sigma_\pi \propto (\sin\xi)^{-1}$. Larger values of ξ have an impact on the design of the solar shade required to protect the payload against direct Sun illumination and rotating shadows, with some repercussions on the overall size and mass, the solar cell efficiency and resulting size of the solar arrays, and the design of the service module including its accommodation within the launcher fairing envelope.

Independent of the adopted angle ξ, the scanning law must be defined such that sky coverage is maximised with minimal precessional motion of the spin axis. This means that the number of revolutions per year around the solar direction should be $K \simeq 262°/\xi$, or $K \simeq 4.8$ for $\xi = 55°$, and that the speed of revolution must be modulated to maintain a nearly constant speed of the nominal spin axis with respect to the stars. The resulting precessional speed of the nominal spin axis, equal to the maximum nominal transverse attitude speed, is approximately 0.17 degree hr^{-1} (or 0.17 arcsec s^{-1}) for $\xi = 55°$, and depends only weakly on ξ.

A scan rate of 120 arcsec s^{-1} (cf. 168.75 arcsec s^{-1} for Hipparcos) has been selected. A longer spin period would offer certain advantages by improving the limiting magnitude performance of the astrometric instrument because more photons would be collected per pixel or per integration period for any given star. On the other hand, it would not change the final astrometric performance since the latter is based on the total quantity of photons cumulated per star over the whole observation period (the total quantity of photons will remain approximately the same for a given star, with fewer observations of longer period). It would lower the focal plane readout rate (and improve the readout noise), because each star would take more time to cross the same field of view (i.e., the instrument focal plane). Consequently, the on-board processing rate would be lower, as well as the science telemetry down link transmission rate, assuming that the limiting magnitude is maintained at its nominal value. On the other hand, a longer spin period would require an increased field height normal to the scan and put additional constraints on the attitude control and measurement subsystem (which would have to guarantee the same attitude and rate stability requirements over a longer period), the thermal control (which would have to guarantee the same temperature stability over a longer spin period), and more subtle effects such as the CCD charge transfer efficiency. For the attitude and thermal control, preliminary analyses show that the performance should still be met with a spin period which would be twice longer. In conclusion, a feasible design is proposed with a scan rate of 120 arcsec s^{-1}, although consideration of a longer spin period is still possible.

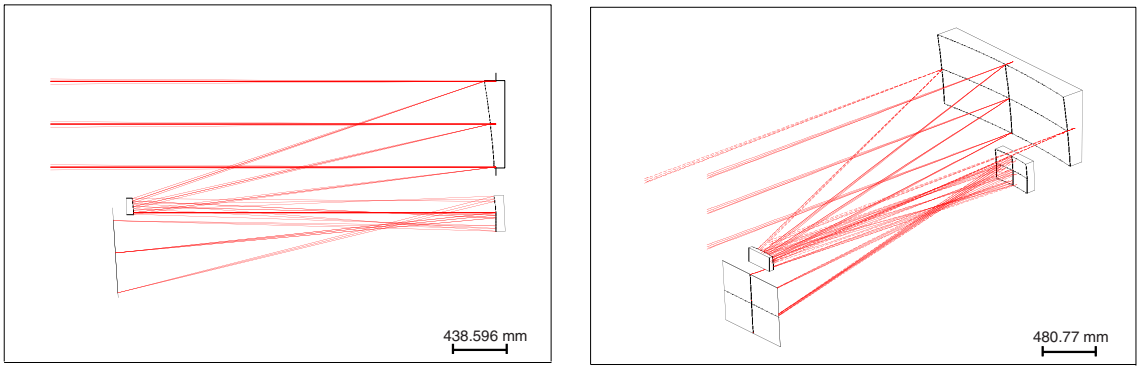


Figure 3.4: Astrometric instrument optical design.

3.1.5 Mission Lifetime

A mission length of 5 years is adopted for the satellite design lifetime, which starts at launcher separation and includes the transfer phase and all provisions related to system, satellite or ground segment dead time or outage. Although the many advantages of the space environment permit very significant measurements to be made in a relatively short time, some of the more complex motions cannot be properly explored on a time scale of 2–3 years, the very minimum acceptable mission duration from a design viewpoint. This concerns, in particular, many binaries with periods up to several years, and the detection of possible planetary and brown dwarf companions; the instrument is ideal for planetary detection, through its screening of a very large number of stars for their photocentric motions, but measurements extending over a significant fraction of the orbital period are mandatory.

Additional motivations for a longer temporal baseline are for the dynamical studies of asteroids, and for photometric variability studies. The design lifetime is used for the assessment of the astrometric and photometric accuracies, and for the evaluation of all instrument and system performances, including the reliability analysis. The design lifetime was subject to a major system trade-off, since a longer design lifetime could actually minimise the satellite complexity. An extended design lifetime of 6 years has also been considered, during which the satellite's design features will permit additional measurements to be made, although the nominal measurement performance requirements (e.g. pointing performance, astrometric measurement accuracy, etc.) no longer formally apply. The extended lifetime has been used for the sizing of all consumables (propellant) as well as for the sizing of items which suffer from aging and radiation effects (solar cells, thermal surfaces, etc.).

3.2 Optical Design

3.2.1 Overview

A number of important properties of the astrometric instrument are reflected in the adopted optical design (Figure 3.4). The astrometric telescopes have a large focal length, necessary for oversampling the individual images. A pixel size of 9 μm in the along-scan direction was selected, with the 50 m focal length allowing a 5-pixel sampling of the diffraction image along scan at 600 nm. The resulting optical system is very compact, fitting into a volume 1.8 m high, and within a mechanical structure adapted to the Ariane 5 launcher, for which the inter-mirror distance must be less than 3 m. Deployable payload elements have been avoided. The overall height has been minimised to reduce the mass and dimensions of the payload module, and an 'embedded' configuration is adopted, in which the two astrometric instruments occupy the same level within the payload. The field of view has to cover more than 0.4 deg^2 in order to reach the target accuracies, with a minimum height of 0.°6 (the scanning law implies a transverse motion of 0.5 degree per 3-hour scan). Including some tolerance on the deviation from the nominal scanning sets a minimum requirement of about 0.°6 diameter for the unvignetted diffraction-limited field of view, in order to avoid gaps in the sky coverage.

System optimisation yields a suitable full pupil of 1.7×0.7 m^2 area with a rectangular shape. Optical performances which have been optimised are the image quality, characterised by the wavefront error (WFE), and the along-scan distortion, avoiding at the same time a curved focal plane in order to facilitate CCD positioning and mechanical complexity. The optical configuration is derived from a three-mirror anastigmatic design with an intermediate image. The three mirrors have aspheric surfaces with limited high-order terms, and each of them is a part of a rotationally symmetric surface. The aperture shapes are rectangular and decentered, while each mirror is slightly tilted and decentered. These optical properties, as well as the effects of chromaticity, straylight, and alignment, are considered in further detail in the following sections.

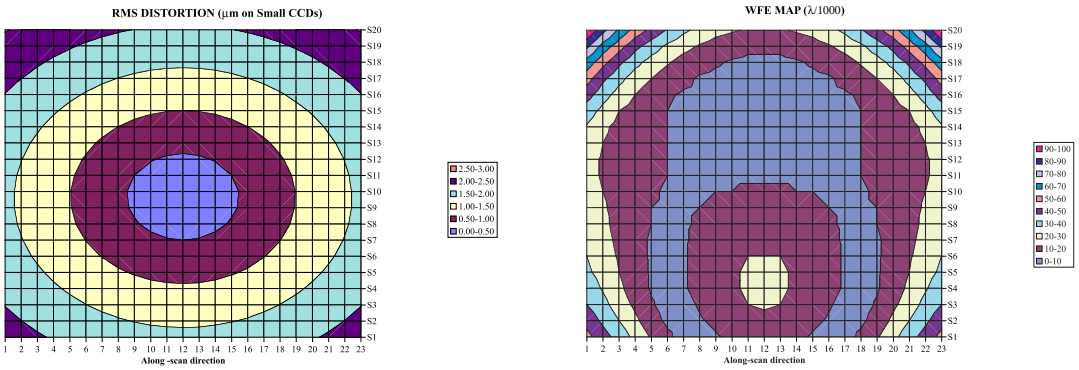


Figure 3.5: (a) Distortion performance for 29×29 *mm*^{2} *CCDs. Each grid corresponds to an individual CCD chip. (b) WFE variation over the field of view. The image quality is evaluated considering the full pupil. The sampling is* $0.^\circ033$ *in each direction, corresponding to a small CCD. Each point has been evaluated at the centre of a region of* 0.033×0.033 *sq. deg. The full image field is considered:* $-0.^\circ38$ *to* $+0.^\circ38$ *along scan, and* $-0.^\circ3$ *to* $+0.^\circ366$ *across-scan. Except for two small CCDs in the lower corners (where the rms WFE is* $\lambda/26$*), the image quality is better than* $\lambda/30$ *rms throughout the astrometric field. The extended field, including the ASM and the BBPs, has rms WFE* $< \lambda/10$*. The* $-0.^\circ3$ *across-scan limitation of the field is driven by mechanical considerations (distance between secondary mirror and focal plane assembly). The theoretically usable coherent field is 0.49 sq. deg., but is actually limited by size and complexity considerations.*

3.2.2 Optical Distortion

While almost any amount of optical distortion can be tolerated from the viewpoint of the *a posteriori* geometrical calibration made as part of the scientific data reductions, certain limits are set by operational considerations. In particular, the variation of scale across the field must not cause any significant blurring of the images during the TDI operation. A limit on the acceptable distortion depends on the size of the CCDs and the way in which they are clocked. In order to perform an integration over one CCD in TDI mode, the image spot motion has to be less than about 1/3 pixel rms (i.e. 3 μm). To achieve this, the along-scan dimension of each CCD has been limited to about 26 mm, roughly corresponding to the upper size of present devices (for scientific devices compatible with an acceptable technological yield), with a resulting integration time of ~ 0.9 s. The distortion calculations are performed using the chief ray position, and for each CCD the line with the worst rms distortion value is determined (see Figure 3.5a). Most of the astrometric field of view can be filled with 58×29 mm^2 CCDs (for which the resulting distortion over 1 scan line is < 2.8 μm), while the rest of the focal plane uses smaller CCDs of 29×29 mm^2 (for which the resulting distortion over 1 scan line is < 2.25 μm). The across-scan distortion causes a significant blur which affects the across-scan spatial resolution, and which results in a 0.3 mm parabolic motion across the astrometric field of view. With an individual rotational adjustment of each CCD, the effect is well compensated, resulting in a residual distortion of < 12 μm for the large CCDs, compared to the 27 μm across-scan pixel size. The distortion maps are evaluated by considering a linear correction term for the along-scan distortion for each CCD, implying individual TDI sequencing. The possibility of grouping adjacent CCDs has been examined, but the local focal length variation, even in the flat central zone, is non-negligible (1.7 μm motion over one CCD), and individual sequencing has been retained. Residual distortion is calibrated as part of the data processing, and should not impact on the astrometric accuracy.

Zero field curvature is obviously highly desirable, although a small curvature could probably be accommodated through the mounting of the CCDs. The 'best focus' position should be defined by maximising the MTF in the scanning direction, which could result in slightly different curvature in the scanning and transverse directions.

Table 3.2: Overall WFE budget for the telescope, at $\lambda = 0.6$ μm.

Item	WFE rms
Optical design	$< \lambda/30$
M1 polishing	$\lambda/30$
M2 polishing	$\lambda/50$
M3 polishing	negligible
Alignment	$\lambda/50$
Total	$\lambda/18$ rms

Table 3.3: Line-of-sight sensitivity to mirror displacements. Sensitivity to defocus and decenter are given in arcsec/μm, sensitivity to tilt in arcsec/arcsec.

Element	Defocus	Decenter	Tilt
M1	0.84	–0.06	2
M2	–0.88	0.046	–0.26
M3	0.05	0.018	0.12

3.2.3 Wavefront Errors

A detailed evaluation of the optical WFE is required for deriving the image point-spread function (PSF) and for evaluating the associated chromatic effects as a function of field location and the star's spectral energy distribution (dependent on spectral type, luminosity class, and reddening). The monochromatic PSF, $P_\lambda(\xi, \eta)$ at a specific point in the field is related to the corresponding wavefront error map $w(x, y)$ in the pupil plane through the diffraction formula, Equation 34 (Section 7). For evaluating the optical performance sensitivity to mirror alignment errors (decenters, defocus and tilts), a statistical analysis was performed using Code V. Alignment errors can be made negligible (WFE $< \lambda/70$ rms) provided that the mirrors are positioned with an accuracy of about ± 1 μm, a relatively stringent requirement considering the telescope size. A compensation mechanism is therefore included for optimizing the overall optical quality after launch (see Figure 3.5b).

The optical analysis showed that the simplest and most efficient compensator is on the secondary mirror. A statistical analysis was performed with Code V for defining the number of degrees of freedom to be controlled and the actuation accuracy, as well as the number of WFE measurement points in the field that are needed for recovering the optical quality over the whole field. The results are as follows: (i) five degrees of freedom are required for the secondary mirror: 3 translations and 2 tilts; (ii) this allows determination of the optical quality over the whole field of view (alignment WFE $< \lambda/70$ rms) even if all the mirrors are randomly misaligned by an amplitude ± 50 μm in all directions (for tilts, the 50 μm design reference is defined as a displacement at the mirror edge). The amplitude ± 50 μm is considered as satisfactory for coping with potential launch effects and on-ground residual alignment errors; (iii) the actuator accuracy requirement is ± 1 μm (not stringent with piezo-type actuators); (iv) the wavefront error measurement must be performed on at least three points of the field of view, preferably at the four corners of the field (for redundancy). The resulting residual alignment WFE after compensation is better than $\lambda/50$ rms over the whole field of view, including measurement errors. The compensation mechanism will also facilitate on-ground alignment. The overall WFE of the telescope is obtained by adding to the alignment and optical design WFE the polishing residual WFE for the three mirrors. A realistic WFE budget is provided in Table 3.2. The overall WFE corresponds to a Strehl ratio of 0.84 at 500 nm. The astrometric performance is not strongly dependent on the actual telescope WFE. The effect of aberrations is at first order an energy loss (equal to the Strehl ratio at first order) in the central diffraction peak, which is the only part of the PSF used for the star localization. The only reason if any for lowering further the overall WFE is related to the chromaticity effect and to its residual calibration error. Table 3.3 provides the telescope line-of-sight sensitivity (along scan) to mirror displacements. These results were coupled with NASTRAN FEM distortion analysis to derive the basic angle stability performance in orbit.

3.2.4 Chromaticity

Although the optical design only employs mirrors, diffraction effects with residual (achromatic) aberrations induce a small chromatic shift of the diffraction peak. This effect is usually neglected in optical systems, but was relevant for Hipparcos and becomes even more critical for GAIA. The chromaticity image displacement depends on position in the field, and on the star's spectral energy distribution (colour), but not on its magnitude. The overall system design must either reduce these chromatic displacements to levels below those relevant for the final mission accuracies, or demonstrate that they can be calibrated as part of the data analysis: one purpose of the broad-band photometry is to provide colour information on each observed object in the astrometric field to enable the chromaticity bias calibration on ground.

For a rough quantitative assessment of the effect, a chromaticity measure was defined, which approximately corresponds to the relative displacement (in microarcsec) of the diffraction peak between two stars of extreme spectral types (B3V and M8V). This measure can be calculated by means of a simple formula for any WFE map. The WFE maps calculated with Code V for different points in the field are thus transformed into a 'chromaticity map' showing the variation of the effect across the field of view for a given set of alignment and polishing errors.

A representative chromaticity map for the nominal system is given in Figure 3.6(a). It is truncated at -220 μas (the far extended field corner which has no astrometric function has a chromaticity error of 4 mas). In the astrometric field of view where the rms WFE is lower than $\lambda/30$, the chromaticity error is lower than 27 μas. The actual chromaticity error will in fact include all sources of WFE, i.e. optical misalignments and polishing residual errors. Figure 3.6(b) shows an example of chromaticity map obtained by including all sources of WFE, and assuming that polishing errors are $\lambda/30$ rms for the primary mirror and $\lambda/50$ rms for the secondary. For constructing this map, the polishing errors have been arbitrarily distributed over 3rd and 5th order Zernike polynomials. This is a worst case, since the actual polishing error will rather be distributed over a much larger number of polynomials (the actual spectrum depends on the polisher and on the polishing technique), and since high wavefront spatial frequencies contribute marginally to chromaticity. Nevertheless, this example shows that one should rather expect a chromaticity amplitude in the range of $100-150$ μas over the Astro field. Combined with the broad-band photometric data, this can then be accurately calibrated by the data processing on ground. At the same time, the polishing errors that have been considered are rather conservative: recent developments made on ion beam polishing have shown that polishing errors can be practically made negligible ($\lambda/100$ rms obtained on a SiC reflector of about 200 mm diameter). It is therefore likely that the chromatic shift can be reduced below a few tens of μas over the whole field.

Detailed analysis of the chromaticity effect requires accurate numerical evaluation of the PSF by means of the diffraction formula (Equation 34 in Section 7) as well as careful consideration of how the centroid of the image should be defined. The relevant observable for GAIA is the line-spread function, L_λ, in the scanning direction:

$$L_\lambda(\xi) = \int_{-\infty}^{+\infty} P_\lambda(\xi,\eta)\, d\eta \tag{20}$$

from which the polychromatic line-spread function is calculated as the flux-weighted mean for a given spectral distribution, f_λ, and instrumental characteristics, $T_\lambda\, Q_\lambda$:

$$L(\xi) = \frac{\int_0^\infty L_\lambda(\xi) Q_\lambda T_\lambda f_\lambda\, d\lambda}{\int_0^\infty Q_\lambda T_\lambda f_\lambda\, d\lambda} \tag{21}$$

While the precise centroiding algorithm to be used in the data processing is subject to optimisation, for the present considerations the centroid can be defined as the maximum of the cross-correlation between the line-spread function and a Gaussian function of standard width s. The centroid ξ_0 is then the solution to the equation:

$$\int_{-\infty}^{\infty} L(\xi)(\xi-\xi_0) \exp\left[-\frac{(\xi-\xi_0)^2}{2s^2}\right] d\xi = 0 \tag{22}$$

In practice, a realistic WFE map is defined by Zernike polynomial coefficients taking into account the theoretical optical design wavefront error, degraded by realistic polishing errors, and by assumed telescope alignment errors arising from on-ground assembly as well as launch effects.

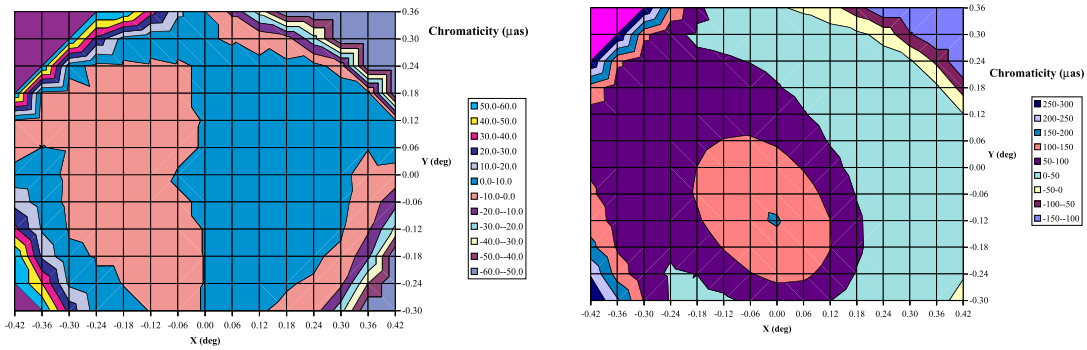


Figure 3.6: Chromaticity error maps: (a, left) only the intrinsic 'optical design' WFE is indicated, without effects of mirror misalignment or polishing errors. (b, right) Example of chromaticity error map including the optics misalignment errors and residual polishing errors. With the assumed WFE budget (polishing error $\lambda/30$ rms for M1, $\lambda/50$ rms for M2), the actual chromaticity amplitude is expected to be in the range of $100 - 150$ *μas over the astrometric field.*

To assess the chromatic effects, the monochromatic line-spread functions were computed by direct integration (with smearing by pixel width and TDI included at the same time), then folded with the set of 175 spectral distributions from Gunn & Stryker (1983) with extinctions of $A_v = 0, 2, 4$ mag (Lindegren 1998a). A detailed evaluation of the chromatic displacements (Vannier 1998a) has demonstrated that currently representative WFE maps, combined with colour information available from the broad-band photometric system, can be used to correct chromatic shifts by applying simple linear models to large samples of stars; the coefficients of these models depend on the precise form of the WFE, and would have to be determined empirically during the data processing. Such an approach, which was also adopted for the Hipparcos data processing, was shown to bring the rms scatter for an individual field crossing down to levels of about 6 μarcsec at 15 mag. Given that the present assignment of $\lambda/50$ rms polishing errors on the secondary are considered as conservative, a combination of correct emphasis on the mirror manufacture and alignment, the acquisition of adequate photometric information for each star, and a rigorous approach to the problem of image centroiding and chromaticity calibration, should result in adequate control and correction to compensate for all of these effects.

3.2.5 Straylight

The satellite concept and the selected orbit (L2) are such that the payload module is naturally protected against straylight: the sunshield protects the payload module from direct illumination from Sun, Earth and Moon. The major straylight level is expected to be due to residual sunlight, which can reach the focal planes by diffraction at the sunshield edge, followed by a diffusion on the payload module closing tent, and to direct illumination of the focal plane by starlight through the entrance apertures of the payload module. Both effects have been evaluated. For Sun-induced straylight, the worst-case configuration during the satellite rotation is when the Sun direction is in the plane of symmetry of an Astro telescope. Sunlight is diffracted at the sunshield edge and illuminates the upper part of the payload module closing tent (anti-service module side). Although the inner side of the closing tent is black painted, there is a residual diffusion of the light by the tent which can directly reach the focal plane. An evaluation of this straylight level has been made by using Fresnel diffraction theory and by assuming that the tent diffusion is lambertian with a total reflection factor of 0.2. The resulting focal plane illumination is about 20 e^- per pixel per TDI integration time of 0.86 s.

Although low, this induces a noise of about 4.5 e^-, non-negligible in comparison to the detector noise. It is therefore desirable to lower this level by about 1–2 orders of magnitude, which is achieved by taking advantage of the opening doors located on the payload module apertures (Section 4.3). A single vane located close to the baffle edge will directly block the light diffracted by the sunshield. The Sun straylight can then readily be reduced to virtually zero (say less than 0.1 e^- per pixel per integration time). In addition, one can baffle the focal plane for suppressing any direct path from the upper part of the tent (close to the telescope apertures) to the focal plane. This baffle is necessary in any case for suppressing star background straylight. Regarding the sky induced straylight, there is

a direct path from one Astro focal plane to the sky via the apertures of the other Astro telescope and of the RVS telescope. The dominant aperture is that of the other astrometric telescope. The solid angle seen by one focal plane and subtended by the other aperture is about 280 $\deg^2$. Considering that the total sky background corresponds to 460 stars of magnitude $V = 0$ (Allen 1973), and that a zero-magnitude star produces 40×10^9 e^- m^{-2} s^{-1}, one gets a sky induced straylight of 26 e^- per pixel per integration time. Therefore, the Astro telescope should preferably be baffled in such a manner that the focal plane does not directly see the apertures of the other Astro telescope and also of the Spectro telescope.

The best way for implementing this baffling is to take advantage of the Astro telescope design. Since the design features a real exit pupil of small dimensions (7.7×3.2 cm^2) located close to the tertiary reflector and at 2.3 m distance from the focal plane, a simple and straightforward way to baffle the telescope is to make use of a conical shape baffle, mounted on the focal plane, and subtended on one side by the focal plane edge and on the other side by the exit pupil. The baffle length is defined by the overall payload module geometry in such a manner that direct paths from the payload module apertures to the focal plane are blocked by the baffle. In practice, a baffle length of 1. m is sufficient for blocking both apertures (Astro-2 and RVS). It could be made by a CFRP shell (4 flat pieces glued together). By that means, both sky and Sun straylight are suppressed. In conclusion, the satellite concept and the selected orbit are such that the payload module is naturally well protected from straylight. Additionally, the concept allows an efficient baffling of the focal plane which suppress direct illumination from the sky.

3.2.6 Optical Alignment

A 5 degrees-of-freedom mechanism (two tilts and three translations) is implemented on the secondary reflector for ensuring the optical quality in orbit and recovering misalignments of the telescope optics which may be induced by launch effects. Since this mechanism is anyhow needed for optimizing the telescope WFE performance in orbit, one should take advantage of its capability for simplifying on ground optical quality verification.

The telescope will be aligned on ground in auto-collimation configuration and verified at the operating temperature. Although this alignment can be made very accurately in a given configuration, several uncertainties may affect the actual performance in-orbit, in particular gravity distortion of the torus structure, and prediction error of thermal gradients. These effects can be mastered or accurately modeled/verified on ground, but it is much simpler and more cost efficient to allocate part of the mechanism compensation capability for the correction of these residual errors in orbit at the beginning of the satellite life. Since the mechanism is able to recover a misalignment of the optics as high as ± 50 μm, a reasonable approach would be to allocate $15 - 20$ μm of the allowable misalignment to residual on-ground alignment errors such as gravity effects, the rest being allocated to possible launch effects.

Telescope re-alignment in-orbit is performed at the beginning of life by using the Shack-Hartmann wavefront measurement at four points of the field of view (e.g. at the field edges). The best secondary mirror position is derived from the WFE measurement and achieved using the alignment mechanism. The WFE measurement in orbit will also provide a good starting point for the chromaticity displacement determination, the ultimate accuracy being achieved through data processing on ground.

The Shack-Hartmann wavefront sensor works with point sources, and the measurements are possible with white light. The large telescope f-number ($f/30$ along scan and $f/70$ across scan) simplifies the wavefront sensor relay optics design, which may not require a dedicated calibration in-orbit. A preliminary design was used to define the wavefront sensor parameters, how it should be operated in orbit, and also for verifying that it does not induce severe accommodation constraints. The wavefront sensor parameters are given in Table 3.4. The wavefront sensor basically consists of a field lens which re-images the pupil onto a lenslet array, and of a collimating lens which transforms the converging beam into a parallel beam. The wavefront is then sampled by the lenslet array. Each lenslet produces a spot on the CCD and the spot position (centroid) provides the measurement of the wavefront average slopes over the corresponding sub-aperture. The number of sub-apertures is determined by the highest spatial frequency of the wavefront distortion one needs to measure. Since we are dealing with optics misalignments, the wavefront distortions are of low spatial frequency and a relatively low number of sub-aperture is required (3×5, corresponding to 15 modes). The relay optics focal length is large enough (150 mm) for simplifying the wavefront sensor element accommodation: the field of view of 1 $arcmin^2$ corresponds to 15×15 mm^2 in the focal plane, i.e. a fraction (1/4) of a small focal plane CCD. The field lens and a fold mirror only would be located in the telescope field of view, the rest of the elements (collimating lens, lenslet array and CCD) being accommodated outside the detection area. The integration time is selected in such a manner that the wavefront measurement can operate in staring mode although the satellite is rotating, on stars brighter than $V = 6$ mag. The practical measurement in orbit at the beginning of life would then consist in scanning a great circle where bright stars are identified. The wavefront distortion may also be measured from time to time during the whole lifetime, whenever a star brighter than $V \sim 6$ mag crosses the wavefront sensor field.

Table 3.4: Wavefront sensor parameters. The relay optics focal length is large enough for enabling a simple accommodation of the wavefront sensor elements. The integration time is selected so as to allow a wavefront measurement while the satellite is rotating (no TDI mode, use of off-the-shelf CCD) on sufficiently bright stars ($V \sim 6$ or brighter).

Parameter	Value
Number of sub-apertures	3 (across scan) $\times$ 5 (along scan)
Integration time	$\sim$ 4 ms
Detection	512 $\times$ 512 CCD
Pixel size	20×20 μm^2
Lenslet focal length	34 mm
Field of view	1×1 arcmin^2
Relay optics focal length	150 mm

3.3 Astrometric Focal Plane

3.3.1 Overview

Given the concept of GAIA as a scanning instrument, the focal plane is designed around a set of CCD chips operating in TDI (time-delayed integration) mode, scanning at the same velocity as the spacecraft scanning velocity and thus integrating the stellar images until they are transferred to the serial register for read out. To maximise the observing time (which is directly related to the final accuracy), as large a fraction of the focal plane as possible is covered with CCD pixels. The detailed design of the focal plane results from several competing requirements. It must have sufficient spatial resolution to allow extraction of the phase information contained in the image. At the same time, the need for as large a field of view as possible (to maximise the on-source time as well as the scan overlap) must be balanced against a focal plane assembly of manageable size and acceptable power and thermal characteristics. Pixels must be of a sufficient size to be feasible within the available CCD technology, and must provide adequate QE, MTF, and noise characteristics at the required read-out rates. The design of the focal plane is also linked with the requirements on the attitude stability in order to avoid image smearing during TDI integration. The foreseen attitude stability combined with the scanning velocity defines an 'elementary integration time', i.e., the maximum time along which one can integrate an image without significant degradation.

Three functions are assigned to the focal plane system (see Figure 3.7 and Figure 3.8): (i) the Astrometric Sky Mapper (ASM), which detects objects entering the field of view, and communicates details of the star transit to the subsequent astrometric and broad-band photometric fields; (ii) the Astrometric Field (AF), devoted to the astrometric measurements; (iii) the Broad Band Photometer (BBP), which provides broad-band photometric measurements for each object. The same elementary CCD is used for the entire focal plane, with minor differences in the operating modes depending on the assigned functions.

3.3.2 Astrometric Sky Mapper Field

Each astrometric field focal plane includes a dedicated 'astrometric sky mapper', preceding and covering the full height of the astrometric field (Figure 3.9 and Figure 3.10). Because of this on-board autonomous star detection discrimination function, the payload operates without the support of any pre-loaded star catalogue. The functions of the sky mapper are: (i) on-board detection and selection of stars to be measured in the astrometric field, including a discrimination for cosmic ray events; and (ii) scan rate determination about two axes, along and across scan, which are provided to the satellite attitude control subsystem. Information provided is sent to ground, in parallel, for attitude reconstruction. Each ASM comprises four vertical strips which operate sequentially as the stars enter the field of view.

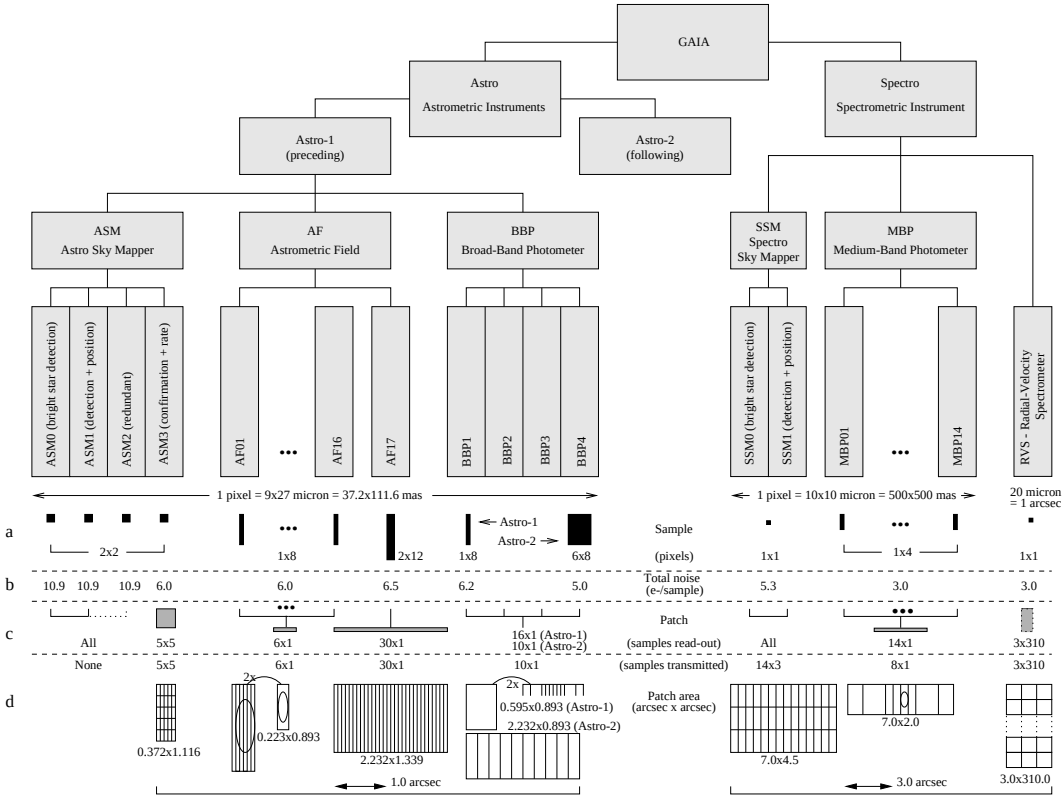


Figure 3.7: Overview of the overall payload instrumentation. There are two identical Astro fields (preceding and following), separated by the basic angle. Each includes an astrometric sky mapper region for object detection (ASM0–3), a main astrometric field, and a region dedicated to multi-colour photometry. The single Spectro field includes a dedicated sky mapper for object detection, and regions dedicated to photometry and radial velocity measurements. AF17 is identical to AF01–16 except for the enlarged sky region read out around each object. The lower part of the figure summarises (a) the sample sizes (in pixels) for each region; (b) the total rms read-out noise per sample (see also Table 3.16); (c) the patch sizes (in samples) actually read out on-board; and (d) the transmitted patch sizes in samples and in arcsec on the sky. The samples read out for photometry are numerically binned before transmission to ground.

A single vertical strip with a short integration time (to avoid saturation) of 0.6 ms, ASM0, is devoted to the detection of bright ($V < 12$ mag) stars. The next vertical strip, ASM1, employs readout of *all* pixels at the required high read-out rates (with correspondingly high read-out noise). ASM2 operates in cold redundancy, with the functions described below. Finally, ASM3 operates in windowing mode, centred on the objects detected by ASM1. It simultaneously provides cosmic-ray rejection logic in addition to its primary function which is the confirmation of objects detected in ASM1: it does so with an improved signal-to-noise ratio by means of the windowing mode operation (similar to that used for the main astrometric field, see Figure 3.12), the refined value of S/N being used to define the object selection threshold for the windowing mode operations as each object passes across the remaining Astro field of view. At the same time ASM3 in combination with ASM1 provides along-scan and across-scan satellite rate information, and, as appropriate, special treatment of high angular velocity Solar System objects. The width of the ASM1–ASM3 CCDs corresponds to a TDI period of 0.86 s.

ASM2 may be configured in-orbit to exercise the functions of ASM0, ASM1 or ASM3 in case of failures, with priority given to ASM1, which is the most crucial detection element for each row. Since it must be able to operate in all of these modes, it will have read-out circuitry (and hence read-out noise characteristics) appropriate to the fastest operating strip, ASM1. For the bright star

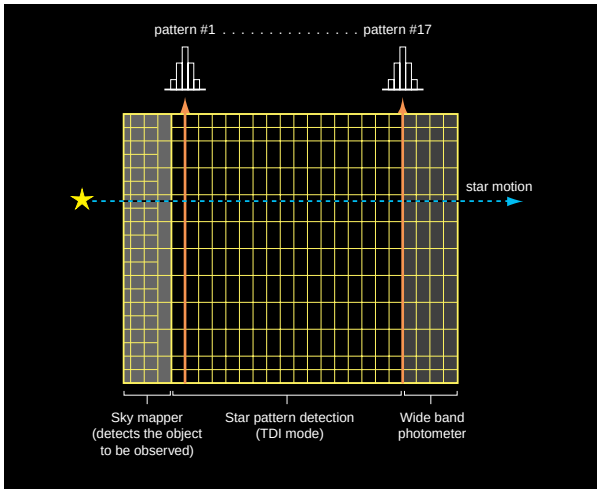


Figure 3.8: Overview of the astrometric focal plane (identical for Astro-1 and Astro-2). Star images cross the field, are detected by the astrometric sky mapper (at left), observed in windowing mode across the main CCD arrays, and finally across the BBP photometric field.

region, ASM0, a single strip is sufficient; the information is not used for scan rate determination and a low cosmic ray event rate is expected due to the smaller detector surface area.

The CCDs employed are identical to those used for the astrometric field, with identical pixel size, QE response, etc. This provides a good QE and a sufficiently high technological yield, the provision of the necessary signal to noise ratio even for the real-time detection of the faintest stars, and full compatibility with the telescope distortion constraints. Analogue binning of 2×2 pixels per sample is used in order to improve the signal-to-noise ratio for star detection, while still allowing adequate sampling of the optical image for centroid determination. Other operational details are provided in Table 3.16. The ASM array is built using 3-side buttable CCDs, with losses in the range of a few hundred μm on 3 sides, and in the range of 4 mm for the fourth side, allowing a global filling factor higher than 85 per cent.

The main driver for the computation of the star detection performance for the ASM is the achievable readout noise. Based on present state of the art readout noise, the total detection noise for ASM0 and ASM1, operated at 1 Mpixels s^{-1}, is equal to 10 e^- rms. For ASM3, with a useful pixel rate of 290 kpixels s^{-1}, the total detection noise will be about 6 e^- rms. Table 3.5 summarises the corresponding performances for stars with $V = 20$ mag, for an integration time of ~ 0.9 s. Estimates of cosmic ray event rates (Section 5.5), suggest that about 500 (primary) events will be generated in ASM1 per second, and with an energy comparable to the star signals, compared to a mean number of 300 stars observed per second. Cosmic ray discrimination is therefore mandatory.

The attitude subsystem requirement for scan rate determination is to provide at 1 Hz information for both axes with an accuracy better than 1 mas s^{-1}. Each astrometric field will typically process about 300 stars per second. The faintest stars will provide an individual accuracy in the $1-5$ mas along scan and $2-10$ mas across scan, improving significantly for brighter stars, with the outputs averaged for use by the attitude control subsystem.

The detection process includes a sky background estimate, a threshold function, a cosmic ray event discrimination function and a connectivity test which allows to discard noise and extended objects. Hardwired logic, based on FIFO memories, discrete logic circuits, ASICs or FPGAs, for each group of ASM0–3 CCDs, offers a satisfactory compromise in terms of real-time processing capacity and required resources (power, size, and mass). An alternative and more sophisticated approach based on digital signal processors (such as the analog ADSP 21020 device which offers a 10 MIPS processing capability with a power dissipation of a few watts) is also conceivable, depending on the complexity of the selected background estimate algorithm. It would offer some flexibility in the adjustment of the algorithm either on-ground (validation and verification phase) or in orbit.

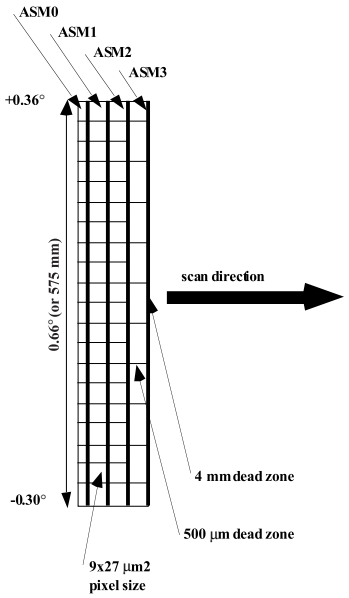


Figure 3.9: Astrometric sky mapper (ASM) topology. Stars enter the astrometric field from the left, and cross ASM0 (bright star detection), ASM1 (main star detection), ASM2 (redundant), and ASM3 (with windowing).

Table 3.5: Worst case achievable signal-to-noise ratio for stars with $V = 20$ mag plus corresponding centroiding noise. This worst case corresponds to a spot centred between four 2×2 binned samples.

Star type	Total signal	Signal per pixel (e^-)	SNR per pixel (worst case) ASM1/ASM3	Centroiding noise along scan (mas, rms)	Centroiding noise across scan (mas, rms)
B3V	320	80	5.8/7.5	5	10
G2V	390	98	6.8/8.5	3	8
M8V	2300	560	22/23	1	2

3.3.3 Astrometric Field

The main astrometric field follows the astrometric sky mapper, and comprises an array of CCDs operated in TDI mode (Figures 3.7 and 3.8). The size of the astrometric field is optimised at system level to allow the accuracy requirements to be achieved, with appropriate balancing of system complexity, for example in terms of optical quality, power and thermal budgets, etc. In the present design the astrometric field covers a field of about $0\overset{\circ}{.}5 \times 0\overset{\circ}{.}66$, with a filling factor larger than 85 per cent due to the use of 3-side buttable CCDs. The size of the individual CCD is a compromise between manufacturing yield, distortion, and integration time constraints. The pixel size is a compromise between manufacturing feasibility, detection performances (QE and MTF), and charge-handling capacity: a dimension of 9 μm in the along-scan direction provides full sampling of the diffraction image, and a size of 27 μm in the across-scan direction is compatible with the size of the dimensions of the point spread function and cross scan image motion. In addition, it provides space for implementation of special features for the CCD (e.g. pixel anti-blooming drain) and provides improved charge-handling capacity. The CCDs are each 29×58 mm^2, and are identical to those used for the astrometric sky mapper, although distortion considerations require that some CCDs near the borders have two video outputs per CCD chip instead of one (Figures 3.11, 3.12 and 3.13). The CCDs are slightly rotated ($< 0\overset{\circ}{.}8$, with a tolerance of a few arcmin) in the focal plane

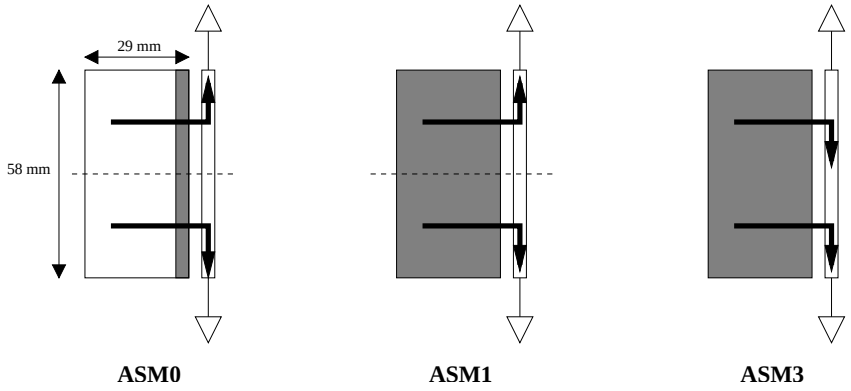


Figure 3.10: The same basic CCD will be operated with three different read-out configurations and associated timing diagrams for the ASM0–3 elements of the astrometric sky mapper. ASM2 is operated in cold redundancy, and can be reconfigured in orbit to serve the functions of ASM0, ASM1, or ASM3.

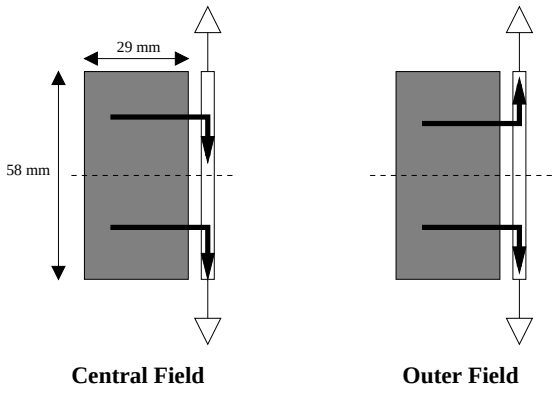


Figure 3.11: The two operating modes for the CCD used in the astrometric field. The selected mode is fixed, according to the location in the field of view, in order to accommodate the optical distortion.

and are individually sequenced in order to compensate for the telescope optical distortion. There is no analogue binning in the along-scan direction, while cross-scan binning of 8 pixels is implemented in the serial register for improvement of the signal-to-noise ratio. Each individual CCD features specific architecture allowing measurement of stars brighter than the normal saturation limit of about $V = 11 - 12$ mag: selectable gate phases (Figure 3.12) allow pre-selection of the number of TDI stages to be used within a given CCD array. The resulting astrometric error versus magnitude shows the effect of this discrete selection (Figure 3.14).

Columns AF01–17 (see Figure 3.7) are identical, with the exception of the possible sample and patch sizes of the last column, AF17. The function of AF17 is to read out a larger region around each target star, and to superimpose these during the on-ground processing in order to provide an astrometric and photometric mapping of the vicinity of each target star, extending to more than 1 arcsec, and ultimately characterising the surroundings of each object to about 23–24 mag.

Because of the very large number of CCDs and the relatively high image acquisition frequency (elementary integration time of approximately 0.9 s), the data rate at focal plane output is enormous, corresponding to a few gigapixels per second if all pixels were readout and processed. It is therefore mandatory to implement suitable filtering of the data flow on board in order to reduce the complexity of the on-board hardware and software, and indeed to achieve feasibility at all. A

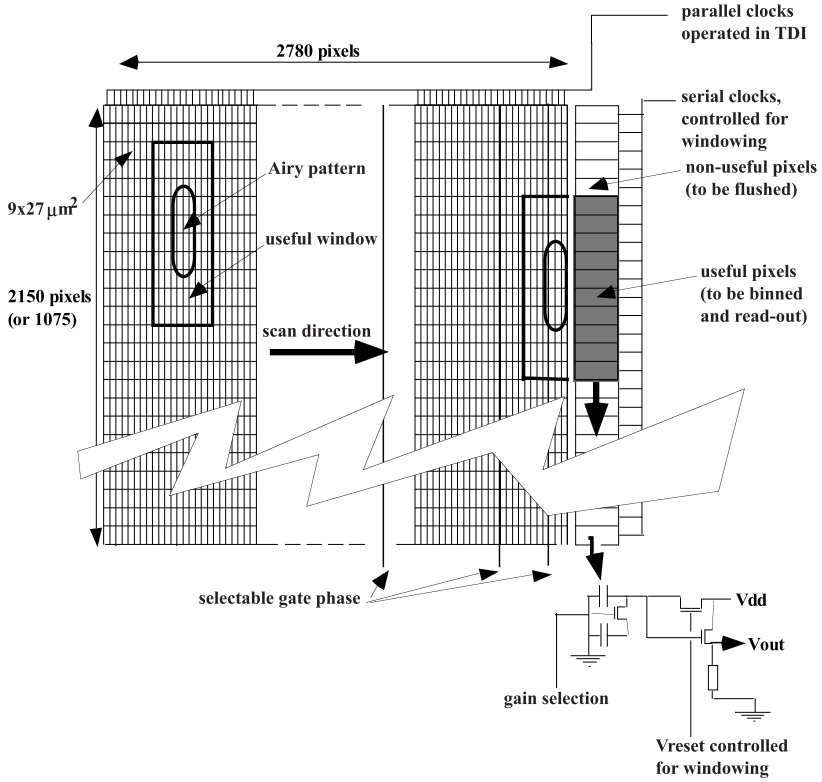


Figure 3.12: Operating mode for the astrometric field CCDs. The location of the star is known from the astrometric sky mapper, combined with the satellite attitude, a window is selected around the star in order to minimise the resulting read-out noise of the relevant pixels.

corresponding consideration requires minimising the quantity of data to be stored between two consecutive ground station visibility windows (with the objective of minimising the capacity and related mass, size and power of the solid state recorder); and to minimise the quantity of data to be transmitted to ground (with a direct impact on the communication subsystem design complexity). This on-board filtering process is acceptable from a scientific view point because most of the pixel data do not include any useful information; it leads to the windowing concept adopted for the astrometric field. The windowing mode, and the implementation of the bright star measurement mode, are both driven by the detailed star data derived from the astrometric sky mapper transits. For each detected star and each CCD, a window is defined which identifies the pixels to be binned and readout at the serial register and output stages; all other pixels are flushed at a higher frequency. The acquisition window includes appropriate margins to take into account the uncertainty on the scan rate and the focal plane geometry.

In the current design of the ASTRO FPA, a single type of CCD is used everywhere in the field of view, in order to minimise the development effort. Because of higher distortion effects at the border of the field (top and bottom lines of CCDs), it is suggested to split the large image zone of each CCD within these areas into two consecutive zones of smaller size and to read them via two separate video outputs. But this is valid only if we are able to drift the charges of the upper and lower parts of the CCD at two different rates, which is unfortunately not feasible with a single CCD. Two solutions might then be proposed: (1) to use the same and single type of large CCD everywhere in the field (including at the top and bottom) and to operate all CCDs in the same normal mode (full image zone). The performance at the border of the field will be slightly degraded but impact on final performance should be very low; (2) to develop a dedicated CCD for the edge of the field with an image zone which is half that of the central CCDs. This is of course much more expensive, and the first option is preferred.

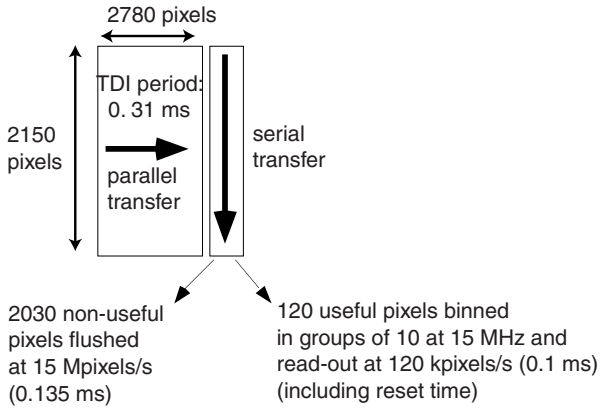


Figure 3.13: Readout operation and main timing parameters for a CCD located at the centre of the astrometric field. The parallel transfer requires 0.034 ms. The serial transfer process includes an overhead of 0.03 ms to account for 'clock feed-through'.

Table 3.6: System and payload parameters for the astrometric field.

Parameter	Value
Satellite design lifetime	5 years
Observation time	4 years
Number of telescopes	2
Basic angle	106°
Entrance pupil	1.7 m along scan; 0.7 m across scan
Focal length	50 m
Field of view	0.32 square degrees
Transverse field of view	0.°66
Sampling size	1 pixel along scan; 8 pixels across scan
Pixel size	9 μm along scan; 27 μm across scan
QE versus wavelength	see Table 3.15
Pixel MTF	see Table 3.15
TDI integration time per chip	0.86 s
Overall aberration WFE	0.036 μm rms (Strehl ratio 0.8 at 500 nm)
TDI errors	0.3 pixel rms
Optical transmission	0.9

3.3.4 Broad-Band Photometric Field

The broad-band photometric field provides multi-colour, multi-epoch photometric measurements for each object observed in the astrometric field. In the adopted configuration the photometric CCDs cover the full height of the astrometric field, so that photometric measurements are acquired at each measurement epoch. Four spectral bands are implemented within each instrument, with one vertical CCD strip per spectral band, each with an integration time of 0.86 s. The CCDs are identical to those used for the astrometric sky mapper and the astrometric field, and provide WFE $< \lambda/15$, and distortion < 5 μm over the region of interest. The readout is similar to the windowing mode selected for the astrometric field, and for ASM3, but with 1×8 or 6×8 pixels per sample binning to improve the signal-to-noise ratio and hence the photometric accuracy. The dead zones between CCDs will be larger than for the astrometric field to take account of the mechanical interfaces required for filter mounting. The selectable gate phases (see Section 3.3.3) will be used to obatain broad-band photometry of bright stars.

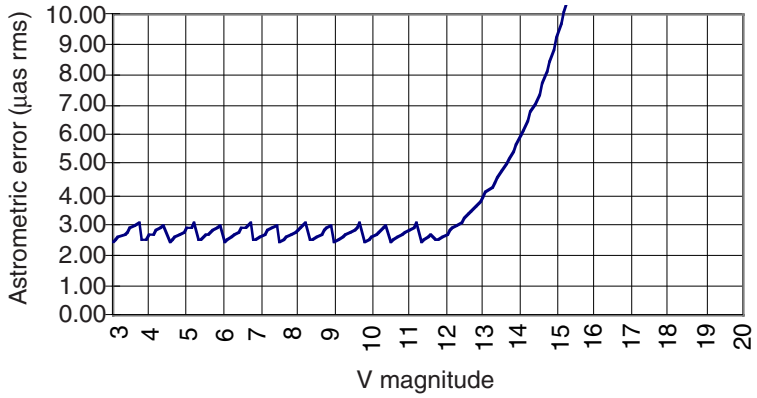


Figure 3.14: Nominal accuracy performance versus magnitude, for a G2V star. For $V < 15$ mag the astrometric performance first improves because the number of detected photons increases. This remains true until the detector saturation level is reached ($V \sim 11.6$ for G2V star). Once saturation is reached, the performance is practically independent of magnitude (2.5–3 μas rms floor) since the total number of detected photons per star no longer increases, due to the way that integration time is assigned for bright stars, based on ASM detections, to avoid saturation. For this curve, fixed biases have been added, namely 1.4 μas rms, shared between the basic angle short-term variation knowledge (1 μas rms) and the residual error on chromaticity (1 μas rms assumed).

3.4 Sensitivity Analyses

Figure 3.15 shows how the performance evolves with the pupil diameter along scan and the pixel size, all other assumptions being unchanged. The influence of the focal length can be derived from the same figures, since the astrometric performance is a function of the ratio (l/F) (all other parameters being unchanged), where l is the linear pixel size along scan and F the focal length. For example, increasing the pixel by 50 per cent has the same impact as increasing the focal length by the same proportion. The performance evolution with the pupil diameter across scan H (all other parameters being unchanged) is not represented, because it varies simply as $H^{1/2}$. Figure 3.16 represent polychromatic star images. The point spread function covers approximately 5 pixels for blue and red stars. The useful information (to be transmitted to ground) is always contained within 6 pixels. Figure 3.17 shows the effect of image motion during TDI integration.

3.5 Spectrometric Instrument

3.5.1 Telescope Design

A dedicated telescope, mid-way between the two astrometric viewing directions and employing a rectangular entrance pupil of 0.75×0.70 m^2, feeds both the radial velocity spectrometer and the medium-band photometer: the overall field of view is split into a central $1^\circ \times 1^\circ$ devoted to the radial velocity measurements, and two outer $1^\circ \times 1^\circ$ regions devoted to medium-band photometry. The focal length of the telescope is 4.17 m, giving a pupil magnification of 0.1, and a size of 8 cm for the grating. The telescope is a 3-mirror standard anastigmatic (without intermediate image) which can be accommodated within the available length of nearly 3 m. The mirror surfaces are coaxial conics. The all-reflective design allows a wide spectral bandwidth for photometry. The image quality at telescope focus allows the use of 10×10 μm^2 pixels within the photometric field, corresponding to a spatial resolution of 0.5 arcsec (compared with 9×27 μm^2 pixels for the astrometric field, and 20×20 μm^2 pixels for the radial velocity spectrometer). The geometric spot diameter containing 90 per cent of the energy is less than 7 μm; the diffraction spot diameter, for 85 per cent energy at 600 nm, is smaller than 13 μm (Table 3.8).

Table 3.7: Numerical parameters for the astrometric field.

Parameter	Value
Focal plane scale	1 arcsec = 242 μm Pixel along scan = 9 μm = 37 mas Pixel across scan = 27 μm = 111 mas
Diffraction radius at 550 nm	Along scan 16 μm (67 mas) Across scan 39 μm (162 mas)
Star speed along scan	120 arcsec s^{-1} Focal plane 29.1 mm s^{-1}
Star speed across scan (maximum value)	171 mas s^{-1} Focal plane 41.4 μm s^{-1}
Time/field correspondence	1 s = 120 arcsec = $0\overset{\circ}{.}033$ 1 pixel along scan = 0.309 ms
Field of view	Field height (across scan), $\Phi_y = 0\overset{\circ}{.}66$ Field width (along scan), $\Phi_x = 0\overset{\circ}{.}56$ Solid angle $\Omega = 0.32$ deg^2
Integration time (faint stars)	TDI integration time, $\tau_1 = 0.86$ s (2780 pixels) Integration time per passage $\tau_n = 14.6$ s Total over lifetime, $\tau = L\Omega/(4\pi) = 978$ s
Average total number of focal plane passages per star per instrument	$N = \tau/\tau_n = 67$ observations
Star number and flow	Average star flow = $N_s\Phi_y\omega = 550$ stars s^{-1} Average number of stars in field, $N_s\Omega = 8000$ Rate of processed stars, $N_s\Omega/\tau_1 = 9300$ stars s^{-1}

Table 3.8: PSF encircled energy for the spectrometric telescope.

	400 nm (%)	600 nm (%)	900 nm (%)
10 μm pixel	87–90	83–86	77–80
Diffraction	94.1	91.2	85.5

3.5.2 Radial Velocity Spectrometer

The radial velocity spectrometer acquires spectra of all sufficiently bright sources (to a magnitude limit imposed by the telescope pupil size, detector readout noise, focal plane size or integration time, and required spectral resolution) and is based on a slit-less spectrograph comprising a collimator, grating and imager, working at unit magnification. The two lens assemblies (collimating and focusing) are identical, compensating odd aberrations including coma and distortion.

The overall optical layout is shown in Figure 3.18. A transmission grating, with a grating step of 1.6 μm, has been chosen because it allows accommodation with a small pupil distance, and a good dispersion uniformity. The grating dispersion is in the across-scan direction, since the spectral scale variation over the field of view in the dispersion direction is not compatible with TDI operation over a full CCD line. The dispersive element is a grism, i.e. a prism with transmission grating, with an additional prism correcting the slit curvature. The prism plays a marginal role for the dispersion and is introduced to provide linear star motion in the focal plane at a given wavelength, therefore allowing TDI operation over the entire field of view. The lenses are refractive, using radiation resistant optical glasses. Each lens assembly has 6 elements with all-spherical surfaces. The design gives good image quality and convenient control of the distortion residuals. Four folding mirrors are inserted near both images and both pupil lenses to provide a more compact design. The maximum clear aperture is 14 cm. Table 3.10 summarizes the image quality of the entire radial velocity spectrometer instrument, including the telescope, collimator, disperser, and focusing lens.

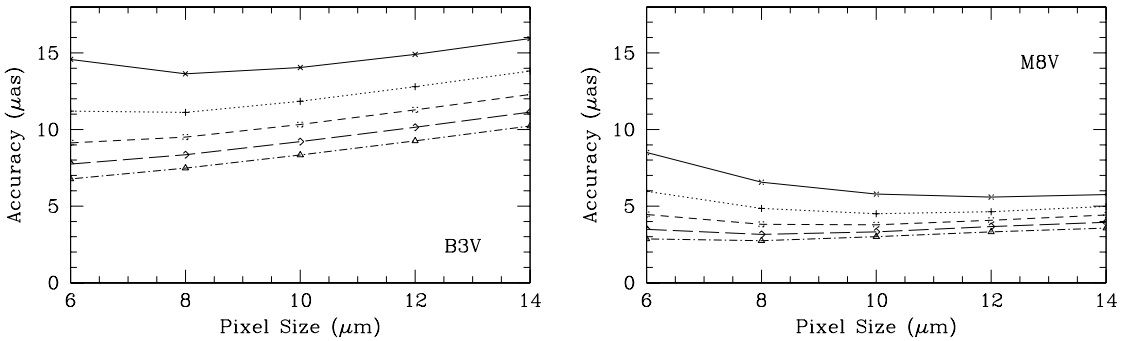


Figure 3.15: Astrometric performance evolution for B3V (left) and M8V (right) stars, versus pixel size and pupil diameter along scan. From top to bottom, the curves are for D = 1.3, 1.5, 1.7, 1.9 and 2.1 m. All other parameters are fixed. In particular, the number of pixels along scan per star image is 5, and the pupil diameter across scan is H = 0.7 m (the astrometric accuracy varies as $H^{1/2}$).

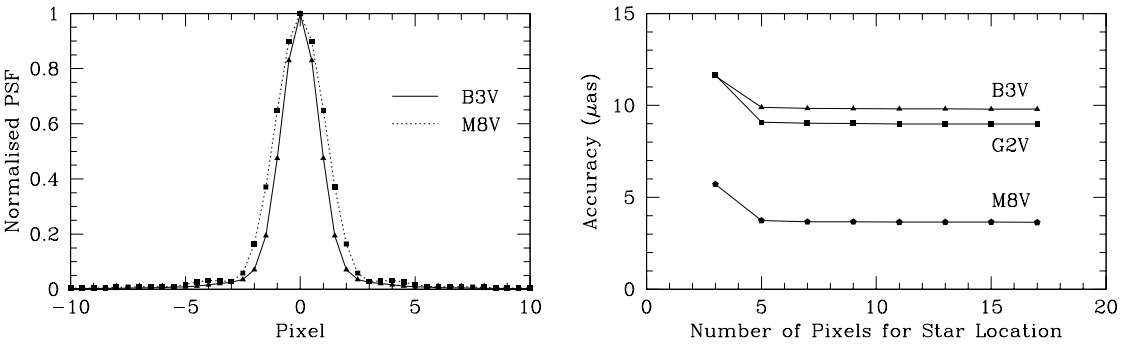


Figure 3.16: Star images and spot sampling: the left figure represents the global point spread function for blue and red stars (including pixel smearing and TDI image motion). The useful information is contained within 5–6 pixels. This is confirmed by the right figure, which shows the astrometric performance evolution as a function of the number of pixels selected for the star localisation process.

The dispersion direction is perpendicular to scan direction. The spectrometer has good dispersion uniformity: for $\Delta\lambda = 23$ nm the size of the spectra vary from 6.10–6.18 mm, with the spectral resolution also varying slightly between 0.0754–0.0744 nm over the whole field of view.

Across-scan image motion during TDI integration have been specified at less than $< 1/3$ pixel. The spectrometer has a very low maximum peak-to-valley error of 7 μm over 1 CCD line, worst case. Since this error is parabolic, the rms error is 2 μm, which is negligible (the spectral resolution is not affected). This motion has been reduced using a prism which reduces the grating's slit curvature.

Spatial resolution variations arise from three contributions: (i) the dispersive elements give an anamorphic component of 9 per cent. The across-scan image dimension is smaller than that along-scan yielding respectively 1.1 arcsec/pixel and 1 arcsec/pixel spatial resolutions in each direction. Since the spectral direction is across-scan, the useful area of the spectrometer CCD is nearly square, and corresponds to a rectangular area of dimension 73.3×73 mm^2; (ii) the telescope optics cause a trapezium-like distortion, which causes a linear line size variation depending on the scan line. The total line dimension variation is 53 μm peak-to-valley over the field of view, or 15 μm rms distortion considering a mean reference focal length. The CCD is split in three parts across scan for redundancy purposes, each with its own sequencing frequency, which further limits the distortion

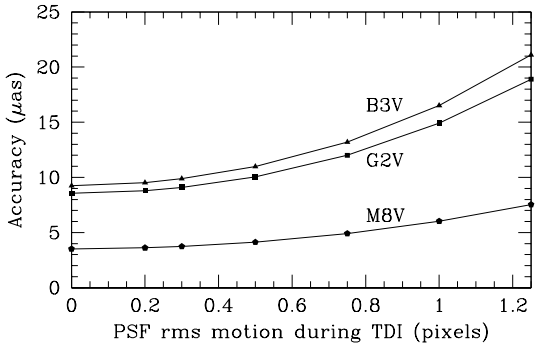


Figure 3.17: Image motion during the TDI operation will blur the image and degrade the astrometric accuracy. The figure represents the astrometric performance evolution versus the image motion expressed in along scan pixels rms. The selected value for the maximum allowable image motion is 0.3 pixels rms, or 3 μm rms for a pixel size along scan of 9 μm. This motion is shared by the residual optical distortion along scan ($<$ 2.9 μm rms), the attitude control ($<$ 0.5 μm rms, corresponding to a line-of-sight jitter $<$ 2 mas rms) and the TDI rate errors ($<$ 0.5 μm rms, corresponding to a speed error $<$ 2 mas s^{-1}).

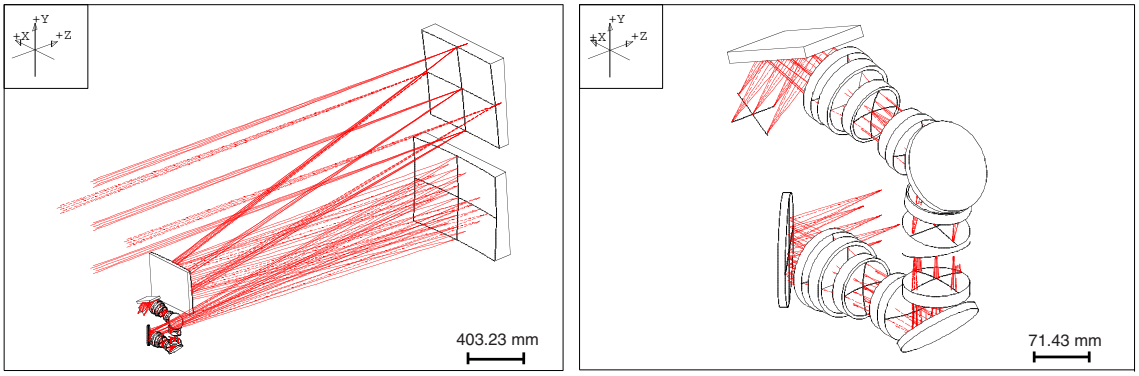


Figure 3.18: Optical configuration of the spectrometric instrument, consisting of a three-mirror anastigmatic telescope: the central part of the field is directed to the radial velocity spectrometer, while the outer parts are used for medium-band photometry.

to 6 μm. An alternative solution would be to use a cylindrical and prismatic field lens near the CCD, in which case the distortion could be reduced by a factor of 2; (iii) dispersive elements cause a rotation of the spectrum during the motion of a star across the CCD. This effect is well corrected by an adjustment of distortion of the spectrometer lens. The maximum rotation is –4.°1 to +4.°1 on the +0.°5 field line. Thus the resulting rms distortion is 2.4 μm for the extreme spectral wavelengths.

3.5.3 Radial Velocity Spectrometer Focal Plane

Figure 3.21 shows a provisional layout for the radial velocity spectrometer focal plane. The array covers a field height of 1°. Each 20×20 μm^2 pixel corresponds to an angular sampling of 1 arcsec and a spectral sampling of approximately 0.075 nm/pixel. The focal plane consists of three elementary CCDs mechanically butted together. Each CCD is operated in TDI mode with its own sequencing, which allows partial compensation for the optical distortion of the instrument.

The main operational constraint arises from the very small charge levels collected for the faintest stars, leading to the requirements for a maximum integration time (presently 30 s), together with

Table 3.9: The radial velocity spectrometer

Parameter	Value
Field of view	$1° \times 1°$
Wavelength range	850–875 nm $\sim$ 310 pixels
Spectral resolution	1 pixel = 0.075 nm
Spatial resolution	1 pixel = 1 arcsec
Aperture	0.75×0.70 m^2
Limiting magnitude	$\sim$ 17 mag
Pixel size	20×20 μm^2
Integration mode	TDI mode
Integration time	$\sim$ 30 s

Table 3.10: Image quality of the radial velocity spectrometer. The PSF diffraction encircled energy is calculated over the 20×20 μm^2 pixel size. The geometric spot diameter is calculated at 90 per cent energy. The resolution is not limited by the aberrations, and is compatible with the 20 μm pixel size.

Telescope + spectrometer	847 nm	859 nm	870 nm
PSF encircled energy	88–91%	88–90%	87–90%
Geometric spot diameter	< 14 μm	< 13 μm	< 14 μm

a minimal readout noise. With 8 readout stages implemented within the serial register of each CCD, a pixel readout rate as low as 18 kpixels s^{-1} can be achieved, with a corresponding total readout noise as low as 3 e$^-$ rms with the use of a dedicated 'skipper-type' multiple non-destructive readout architecture with 4 non-destructive readout samples per pixel.

The dark current should ideally be lower than that of the astrometric field, i.e. lower than 0.04 e$^-$ s^{-1} per pixel. Given that inversion mode is not compatible with intra-pixel TDI operation, two alternative approaches to the reduction of the dark current will be considered: either operation of the CCD in dithering mode, or by decreasing the CCD operating temperature to around 170 K. Based on a typical charge handling capacity of 400 000 e$^-$ for a 20×20 μm^2 pixel saturation is expected for stars with $V < 5.5$ mag, and no further management of the intrinsic dynamic range has been considered necessary.

One important and specific consideration for the radial velocity spectrometer is that the scanning law generates a periodic star speed across scan of amplitude ± 0.17 arcsec s^{-1} and period 3 hours. Consequently, a star crossing the focal plane will describe an oblique line, with a maximum transverse motion of $0.17 \times 30 = 5.1$ arcsec, or ± 5.1 pixels (± 100 μm). The effect of this transverse motion is a spectral blur, inducing a spectral resolution degradation up to 0.55 nm. The effect could be acceptable if the 30 s observation time would be divided into 10 sub-integrations of 3 s, although the penalty on the readout noise and data rate are significant, unless the spectral data are binned on board by an appropriate processing taking into account the across scan velocity of the star. The solution retained is to eliminate the transverse velocity component by slightly rotating the detector in order to re-align the CCD lines with the star motion line. This requires an amplitude of only ± 50 μm at the detector edges, or ± 4.7 arcmin, with a 3 hour period. This combination of low amplitude and low period requires that compensation needs to be performed only intermittently, typically every 5 min. The actuation would be driven in open loop using the attitude control data. The actuator motion amplitude (± 100 μm) and the positioning accuracy ($1 - 3$ μm) are similar to that of secondary mirror actuators of the astrometric telescopes, and the same inch-worm actuator can be used. The rotation by ± 4.7 arcmin of the CCD introduces a shift along scan of one end of the spectrum relative to the other end by up to ± 0.46 pixel. This must be taken into account in the sampling and data reduction.

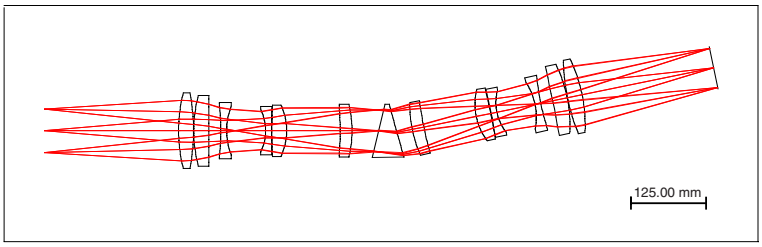


Figure 3.19: Radial velocity spectrometer optical design (unfolded).

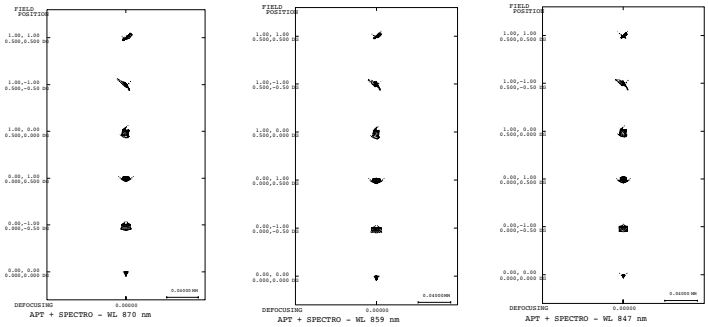


Figure 3.20: Spot diagrams in the radial velocity spectrometer focal plane.

3.5.4 Medium-Band Photometry Focal Plane

Figure 2.13 shows the preliminary medium-band photometric focal plane layout, which includes two separate fields, on either side of the radial velocity spectrometer field, comprising a dedicated sky mapper, and 11 medium spectral bands (see Table 2.2). The requirement for the CCDs are very similar to those of the astrometric field, including the use of TDI. A dedicated sky mapper is implemented in order to detect stars, to determine their location, and to drive the windowing readout process applied to the other fields of the CCD. The sky mapper works in white light and will detect stars down to the main field and medium-band photometric field limit of about 20 mag. The spectral bands will require spectral filters to be directly fixed onto the CCD array. Two video chains per CCD field, 32 in total, are required to read out the science data with the necessary readout noise performance. The saturation of the CCD fields for bright stars will depend on the final bandwidth and transmission characteristics of each spectral band. A preliminary assessment indicates that saturation occurs at about $G \sim 10$ mag, although it is expected that analysis of the image wings will extend the range to brighter stars.

3.6 Science Data Acquisition and On-Board Handling

Preliminary investigations have been carried out to identify the minimum set of data to be transmitted to ground to satisfy the scientific mission objectives; to identify some on-board data discrimination compression principles able to provide the targeted data compression ratio; to assess the feasibility and complexity of implementing such compression strategies and related algorithms on board; to assess the resulting compressed data rate at payload output, which are used for the sizing of the solid state memory and communication subsystem; and to derive preliminary mass, size, and power budgets for the on-board processing hardware. The proposed science data flow, derived from these considerations, is summarised in Figure 3.22. For estimating telemetry rates, a specific sampling of the CCD data has assumed. This is summarised in Figure 3.7 and described below. This sampling is not considered to be the optimal and final one, but represents a useful first approximation of what is required to achieve the main science goals of GAIA, and results from an extensive series of studies (see Section 9.4.2).

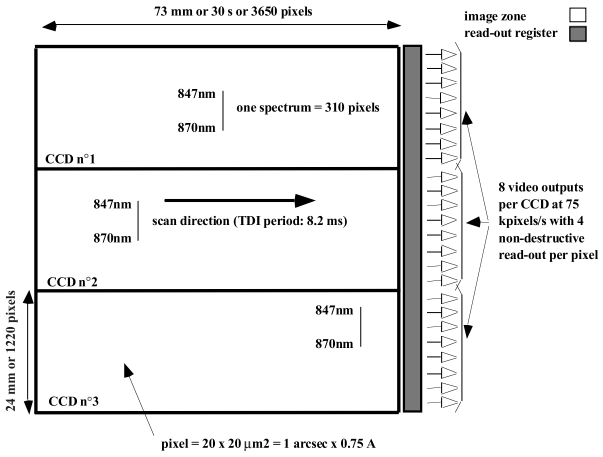


Figure 3.21: Preliminary layout for the radial velocity spectrometer focal plane.

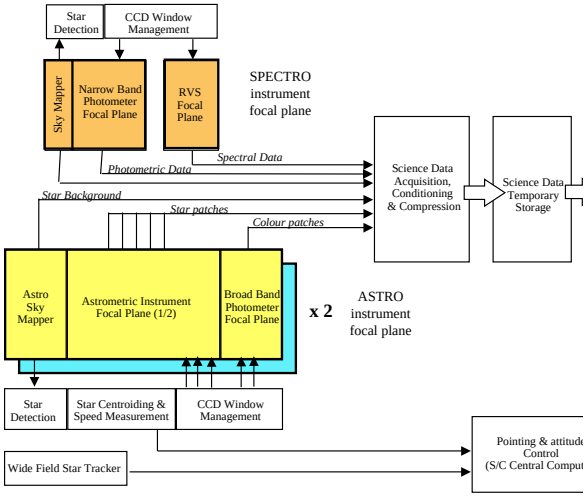


Figure 3.22: Payload science data flow diagram.

3.6.1 Astro Instrument

For each detected object (in principle all point-like sources with $G < 20$ mag), the following data are provided by the astrometric sky mapper: (i) the star position translated into an x, y position with respect to the focal plane reference frame, along with the timing reference; (ii) star scan rate (along and across scan) with respect to the focal plane reference frame, used on board for spacecraft attitude control and within each astrometric instrument for the CCD windowing mode operation; (iii) approximate star magnitude, one single quantity per star, used on board for control of the TDI integration period (either the standard integration period of 0.86 s or a reduced period for bright stars) and the gain of the CCD output stage; (iv) image shape and size information, used to extend the CCD window e.g. for double stars. Some of these data need to be transmitted to ground for *a posteriori* reconstruction of the star selection and sampling processes. Of the sky mapper data, only samples from the second sky mapper (nominally ASM3) are transmitted.

The part of the focal plane which is dedicated to the astrometric measurements is organised into 17 CCD strips, perpendicular to the scan motion. Each CCD is operated in TDI mode and provides a two-dimensional image for each star observed. This two-dimensional image is reduced to a one-

dimensional image, or star patch, through on-chip binning since the highest angular resolution is only relevant in the along-scan direction. Consequently, 17 star patches are generated by the astrometric field for any detected star crossing the instrument field of view, each patch consisting of 6 samples for a proper angular sampling of the pattern (or 30 samples for the extended mapping provided by AF17; see Figure 3.7). The minimum data to be transmitted to ground are these 17 star patches per detected star, completed with identification of the star with respect to the sky mapper data, and its time of measurement.

The broad band photometer includes four spectral bands per astrometric instrument. The requirement is to provide a star patch for each spectral band and each detected star which is 16 or 10 samples wide (along-scan direction; see Figure 3.7). It is completed with information compatible with the further identification of the star with respect to the astrometric sky mapper data and the retrieval of its position and time of measurement.

With reference to Figures 3.7 and 3.8, the focal plane of each astrometric instrument includes 250 CCDs organised in 10 rows of 25 arrays. Each row includes 4 arrays for the sky mapper, 17 arrays for the astrometric measurements and 4 arrays for the broad-band photometer. Because the scan motion of the spacecraft is parallel to the rows of CCDs, each star detected within the sky mapper field of a given row will cross the instrument field of view along the same row, and the star will be detected and measured by all the astrometric and broad-band photometric CCDs of the same row. Because of the parallel architecture of the astrometric focal plane (10 identical rows), it is possible to implement a parallel architecture for the electronics with one dedicated sequencing readout processing electronics per row of CCDs, i.e. 10 identical packages per astrometric instrument. This parallel architecture approach allows partitioning of the processing resources into packages of reasonable size, compatible with available technologies. It eases also the integration and verification activities. Finally, it results in some system robustness, since a failure on one row should not affect the operation on the other rows.

With a few tens of stars (on average) crossing a single CCD row of the field of view per second, a few hundred acquisition windows per second will be generated by the ASM processing electronics and transmitted to the corresponding Astro and BBP CCD sequencers and readout electronics. This moderate processing load may be handled by the same DSP based processing circuit as the one foreseen for the star position rate determination function. The readout of the Astro and BBP CCDs video signals requires one dedicated sequencer (see Figure 3.23) and one dedicated video chain per CCD. The local sequencer will generate clock signals consistent with the window acquisition commands sent by the ASM. The video chain will acquire, condition and digitise the corresponding pixel signals and transmit the 16-bit digital encoding to the payload data handling electronics for further compression and formatting.

There is no need to explicitly measure and store the position and time of measurement of each star crossing each Astro or BBP CCD. The knowledge of the initial position of the star when crossing the ASM field and the knowledge of its scan rate, combined with the geometrical characteristics of the focal plane measured on ground (which may be verified, if necessary, in orbit) will allow unambiguous identification of each star crossing each CCD and to *a posteriori* derive its position. The minimum set of data to be transmitted to ground for further star identification and positioning, for any star crossing the Astrometric instrument field of view, may be reduced to the CCD row number, the initial star position and time, and the scan rate value (which may be updated and maintained while the star is crossing the total Astro field of view).

3.6.2 Spectro Instrument

A dedicated sky mapper will be used to detect and localise all stars with $G < 20$ mag which cross the field of view of the spectrometric instrument, i.e. in principle the same objects as measured in the Astro instruments. The corresponding star position and timing information are used to drive the medium-band photometer. These star position identification data will be sent to the ground together with the MBP and RVS data.

The medium-band photometer includes 11 spectral bands (see Table 2.2) which provide photometric data on all detected objects. A patch of 14 samples is read from the CCD for each star and each spectral band. The samples are numerically binned to 8 samples per star and spectral band before

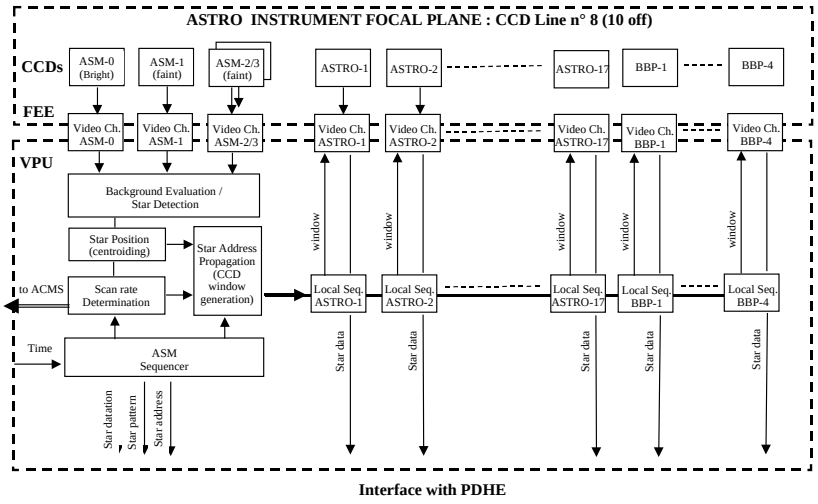


Figure 3.23: Astrometric data acquisition architecture.

transmission to the ground (see Figure 3.7). They are transmitted together with information needed to identify the star with respect to the sky mapper data.

The radial velocity spectrometer provides spectral data for each detected star with $V \lesssim 17$ mag. The total spectrum extends over about 310 CCD pixels. Across the spectrum, 3 pixels are required for full coverage (Figure 3.7). Thus $3 \times 310 = 930$ samples will be transmitted to ground together with information compatible with identification of the star with respect to the sky mapper data.

3.6.3 Telemetry Rate

The different elements of the acquisition chains of GAIA will deliver data on board in an essentially continuous mode. The instantaneous data rate will primarily fluctuate with the stellar density in each of the three fields of view. Thus the Galactic latitude of the lines of sight will be the main source of variation of the stellar density and thus of the instantaneous data rate produced by the observing instruments. In the present design, the capability to downlink data to the ground is limited by the direct visibility of the ground station, which is a minimum of about eight hours per day (Section 5.1). A large on-board storage will store a full day of observation ready to be sent to the ground at a rate three times larger than the acquisition rate (Section 3.6.5).

If N_s is the number of stars per square degree brighter than the limiting magnitude G_max in a particular direction on the sky, and Ω is the solid angle of the field of view under consideration, then the number of stars in this field of view is $N_\mathrm{s}\Omega$. With the field height Φ_y and the scan rate $\omega = 1/30$ deg s^{-1} the star flow per unit of time is $N_\mathrm{s}\omega\Phi_y$, which is equal to the number of stars to be processed per second in each column of CCDs. The relevant values are given in Table 3.11 for the relevant fields in the astrometric and spectrometric instruments.

Recent Galactic models give $N_\mathrm{s} = 25\,000$ stars deg^{-2} up to $G = 20$ mag with large variation with Galactic latitude and locally very dense regions in specific Galactic windows (see Section 6.4.4). Apart from scanning circles coinciding with the Galactic plane, in most cases the amount of data per day will be close to the the average due to the crossing of regions of different Galactic latitude over a scan period. The conclusion reached in this section are scaled to the above value of N_s. For each instrument the basic information consists of a patch of binned samples of CCD counts. The size of the patch and of the sample was in each case provisionally optimized for the relevant task (detection, astrometry or photometry), by taking into account the conflicting requirements of resolution versus on-chip binning (for lower readout noise and data rate). The assumed numbers and the resulting raw data rates are given in Table 3.11.

Table 3.11: Average stellar flow in the various fields of the astrometric and spectrometric instruments, and the resulting average telemetry rates. A limiting of $G = 20$ mag is assumed for the astrometric instrument (AF and BBP) and for the medium-band photometer (MBP), and $G = 17$ mag for the radial-velocity spectrometer (RVS). The sampling scheme in Figure 3.7 is assumed. Note that the 11 photometric bands in the MBP require 14 CCDs according to Table 2.2. The average rate at which the stars are read out can also be written $R = fN_{\mathrm{CCD}}$. It is assumed that each sample represents 16 bits of raw data. The resulting raw data rates are before compression and do not include overhead.

Parameter	Astro-1 and 2 (per instrument)				Spectro		
	ASM3	AF01–16	AF17	BBP	SSM1	MBP	RVS
Limiting magnitude, G_{max} [mag]	20				20		17
Average star density, N_{s} [deg^{-2}]	25 000				25 000		2900
TDI integration time per CCD, τ_1 [s]	0.86				3.0		30
Field width across scan, Φ_y [deg]	0.66				1.0		1.0
Star flow through Φ_y, $f = N_{\mathrm{s}}\Phi_y\omega$ [s^{-1}]	550				833		97
Number of CCDs along scan, N_{CCD}	1	16	1	4	1	14	1
Solid angle of CCDs, Ω [deg^2]	0.019	0.302	0.019	0.077	0.100	1.400	1.000
Number of stars on the CCDs, $N_{\mathrm{s}}\Omega$	473	7568	473	1892	2500	35 000	2900
Readout rate, $R = N_{\mathrm{s}}\Omega/\tau_1$ [s^{-1}]	550	8800	550	2200	833	11 662	97
Samples per star read out	25	6	30	16 or 10	42	14	930
Samples per star transmitted, n	25	6	30	10	42	8	930
Raw data rate, $16nR$ [kbit s^{-1}]	220	845	264	352	560	1494	1443
Raw data rate per instrument [kbit s^{-1}]	1681				3496		
Total raw data rate [kbit s^{-1}]	$2 \times 1681 + 3496 = 6858$						

Including overhead, the total raw science data rate is roughly a factor 7 higher than the mean (continuous) payload data rate foreseen in the telemetry budget ($\sim$ 1 Mbit s^{-1}, see Section 4.7). Data compression (see below) will reduce this discrepancy, but there remains roughly a factor two to be gained either by a smarter design of the CCD sampling (or, in the worst case, a more restrictive sampling leading to a moderate loss in accuracy), or by increasing the link capacity.

3.6.4 Data Compression

The science data flow produced by the instrument focal planes, when using the star detection discrimination capability of the sky mapper and the windowing mode for CCD readout (combined with some binning operations on chip), is already three orders of magnitude smaller than that corresponding to the readout of all CCD pixels. However, there is still room for some further reduction of the data flow. For example: (i) each sample data is coded on 16 bits at the level of the video chain in order to match the star magnitude dynamic range while providing the required coding resolution over this range of magnitude. But for a star of a given colour and magnitude, the dynamic range of the collected signal samples and the coding resolution required on these samples do not require 16 bits of coding but typically 10 bits or less; (ii) there is some redundancy in the scan rate values transmitted with each detected star; (iii) there is some redundancy in the information provided by the 17 star patches provided by the astrometric CCDs which are scanned by the same star (e.g. the shape of the star image is the same) and it should be possible to take benefit of this redundancy to save some bits. It was not the goal of this study to go deeper in the identification and definition of such further data compression processes. At this stage, a compression ratio of 3 on the astrometric and radial velocity data is considered feasible without major impacts on the processing hardware and software complexity. Further investigations will have to assess more accurately this additional complexity and to compare it with the saving in solid-state recorder and communication subsystem sizing.

The first goal of the data compression is to reduce the average science data rate so that it becomes compatible with the actual telemetry link. Actually, a better compression rate than just the one needed for the average conditions should be targetted, in order to be able to transmit data from scan circles close to the Galactic plane. Three basic approaches for the compression can be considered: (i) lossless compression, better referred to as one-to-one and reversible coding; (ii) weakly lossy compression, which amounts more or less to transmit a signal without the informationless noise; (iii) strongly lossy compression or preprocessing, based on additional assumptions on the properties of the objects observed (single or multiple star, constant or variable star, galaxy, etc). The third option is to be avoided for GAIA.

Lossless compression The core of the astrometric information in the Astro fields consists of the patches of six samples, each binned over eight vertical pixels of unit width. A coding over 16 bits per sample results in 96 bits per transiting star. However, the actual electron counts are of the order of a few hundred for the faintest stars and this can be stored in much less than 16 bits. These stars account for nearly one half of all the stars observed (Table 3.12). With 9 bits, one can also include all the stars fainter than $G = 18$ mag, that is to say more than two-thirds of the total. Therefore a straight coding allows nearly to reach a factor two in compression for more than half of the observation program. A similar conclusion applies to the broad-band photometry patches, in fact with a better compression as many samples contain only counts from the sky background, and the patch of 16 samples can be coded in much less than 1024 bits, with only a small dependance with the star brightness. Some additional compression should be achievable through the redundancy in the (largely identical) star profiles along the consecutive CCDs in the astrometric field: these form a small two-dimensional image of 16×6 samples, with a low spatial frequency content in one coordinate.

Weakly lossy compression In this mode one tries to retain only those bits with useful information, by rounding the numbers in such a way that the loss in resolution is within the noise level. The typical quantization consists of choosing a step-size of the order of the noise and then coding the resulting data. For a Poissonian variable N this typically amounts to the well-known 'square-root coding', which uses the fact that the standard deviation of $\sqrt{N}$ is roughly constant ($\simeq 0.5$) independent of the mean value of N. The additional noise introduced by discretization of $\sqrt{N}$ is therefore a constant fraction of the Poisson noise. The idea is readily generalized to the more realistic situation including analogue-to-digital conversion and additive readout noise. If g is the gain factor (in electrons per 'Data Number' or 'AD unit') and r the rms readout noise (in electrons), then the appropriate transformation of the Data Numbers n is $k = \lfloor c\sqrt{(n-n_0)g + r^2} \rfloor$. Here n_0 is the dark value and c a constant; the fractional increase of the total noise caused by the coding is $\simeq 1/6c^2$. Thus, a fractional increase by 4 per cent (which may be tolerable) is achieved with $c = 2$. For typical $r \simeq 6$ e$^-$ this gives the k values in the right part of Table 3.12, where $S = (n - n_0)g$. Coding in 6 bits is thus sufficient for the vast majority of stars, rather than the 16 bits assumed in Table 3.11. This can be further improved by searching for an optimal coding, knowing that the configuration space of the parameters is not uniformly populated with higher density associated to faint stars. A Huffman-like coding may be applied to allocate less bits to the most frequent occurrences.

3.6.5 On-Board Data Storage

Because of the use of a single ground station for the control of the satellite and the reception of the scientific data, the transmission period may be as low as 8 hours per day (minimum visibility period, with Perth assumed as ground station). Because the payload is continuously operating, it must be possible to transmit to ground in 8 hours the scientific data collected during the previous 24 hours. On the other hand, it must be possible to store on-board, between two consecutive transmission periods, the data continuously produced by the payload. Consequently, a solid-state recorder with a capacity of the order of 100 to 200 Gbits must be implemented in the payload data handling design.

Solid state recorders for space applications, with a capacity of a few tens of Gbits, are the subject of on-going development in Europe (Metop, Rosetta, etc). Some of these items are designed with a growth potential to a few hundreds Gbits. In the US, several products are already available for space applications with a few hundred Gbits capacity. Consequently, implementing such a solid-state recorder on board will be neither a major technical issue nor a significant cost driver at the time of the satellite development.

Table 3.12: A possible (square-root type) coding for weakly lossy compression of the astrometric CCD data. S is the typical signal size (in electrons) for stars in the indicated magnitude interval; k is the corresponding coded value, admitting a 4 per cent increase in the total noise.

G [mag]	fraction of all stars	S [e^-]	bits	k	bits
< 15	0.04	10000	13	200	8
$15 - 16$	0.04	5000	12	141	8
$16 - 17$	0.08	2000	11	90	7
$17 - 18$	0.14	750	10	56	6
$18 - 19$	0.25	300	9	36	6
$19 - 20$	0.45	150	8	27	5

3.6.6 Star Density Fluctuations

The average data rate is based on an average star population density (25 000 stars per square degree) or average star flow crossing the field of view of the instruments (550 stars s^{-1} for each astrometric instrument). The actual rate will vary significantly from one area of the sky to another, especially in regions close to the Galactic plane. A typical maximum star density of 155 000 stars per square degree has been considered during the study with a peak value in some local areas of 3 million stars per square degree, leading to a star flow which is 10 to 100 times higher than average. This significant increase of the star population density may lead to several saturation problems at payload and spacecraft level.

The hardwired logic in charge of the star detection process should be able to handle such an increase in detection rate (except that some more frequent overlaps between stellar objects at such high rates and a probably bright and inhomogeneous sky background may degrade the detection probability). The limitation will come first from the DSP based electronics in charge of the connectivity test, the star position and speed assessment and the acquisition window propagation. The proposed parallel architecture, with one DSP (10 MIPS) per row of CCDs, should be able to handle 10 times more stars per sec than the average rate but will saturate for higher values. An improvement may be obtained either with the use of advanced processing devices (performance will further improve in the coming years) or by splitting the processing electronics into specialised processors. This improvement may however have a noticeable impact on mass and power budgets.

In order to minimise potential fluctuations of the astrometric CCD power dissipation induced by a fluctuation of the star population density, it is considered appropriate to fix the sequencing of the astrometric field CCD (e.g., to fix the number of acquisition windows to be readout at a given time per CCD). The sequencing may be sized such that it is compatible with the highest expected star detection rate: assuming a maximum star density of 3 million stars per square degree, it leads to a maximum of 12 acquisition windows to be readout simultaneously per astro CCD. Some of these acquisition windows will be then filled with dummy data when the instrument is observing sky areas with a lower star density. The memory capacity of 100 Gbits presented in the previous section corresponds to an average star population density of 14 300 stars per square degree. This capacity may be increased in order to cover periods of observation (several hours) with a higher average star density but the problem will be then to find time to transmit this excess of data to ground (see below).

The sizing of the down link telemetry (1 Mbps), as implemented in the baseline spacecraft design, corresponds to an average star population density of 14 300 stars per square degree, and to a minimum visibility window of 8 hours per day. An increase of the quantity of data collected and stored over one day will require either a longer transmission period (but the transmission period or visibility period depends on the satellite elevation angle with respect to the Earth equator, and is not simply adjustable on request), or a higher transmission rate the design of the antenna subsystem is already critical because of the need of a large field of view combined with a high gain at a high data rate. The proposed electronically scanned phased array antenna concept described in Section 4.7 could possibly be oversized to transmit a data rate which is twice the current baseline (6 Mbps instead of 3 Mbps) but the complexity will increase significantly (number of radiating elements, thermal control design, etc.) as well as the power budget.

The spacecraft design is able to handle for a short time (a few minutes) a flow of stellar objects which is several tens times higher than the average one, with minimum adaptation of the hardware. For higher rates and longer periods, saturation will occur either at detection level or at memory transmission level. The design may be improved to handle such rates but this will be done at the price of the spacecraft complexity and cost. Several alternative options may be considered to solve the saturation issue to stop the measurements when scanning local crowded areas or close to the Galactic plane, to adjust the star detection threshold as a function of the sky area being observed (to limit the detection rate and to maintain a constant average rate over any great circle), etc.

3.7 CCD Details

3.7.1 Introduction

The CCD detectors form the core of the GAIA payload, and their development and manufacture will represent one of the key challenges for the programme. Although the CCD is well-known and widely used in optical astronomy, a large number of design possibilities must be collectively optimised. While separate characteristics can be individually tuned (for example, QE, red or blue response, MTF, read-out noise, read-out rate) figures for any individual parameter quoted in isolation can be highly misleading. Thus operating temperature may be lowered but at the expense of QE in the red; pixel sizes may be decreased to a few microns but at the expense of QE and MTF; read-out noise may be lowered considerably, but at the expense of read-out rate and signal dynamic range. Similarly, it is important to consider the effect of the total detection noise (including the effects of analogue video noise, quantization noise, and correct definition of the electrical bandwidth) and not just the CCD read-out noise.

During the Concept and Technology Study, detailed literature and off-the-shelf product surveys were conducted, contact was made with a number of major manufacturers (Philips, Thomson, EEV, Site, Dalsa, Lockheed Martin Fairchild Systems, Kodak, Hamamatsu, TI, Lincoln Labs, EGG, and Sarnoff Corporation), and this was followed by specific consultancy contracts with the aim of identifying a baseline operational configuration with associated performances. Consideration was given to the requirements on electro-optical behaviour, array size, buttability, pixel size, bright star handling, serial register performance, output amplifier, power dissipation (in the image zone, serial register, and output amplifier), trade-off between QE and MTF, photo-response non-uniformity, dark current, conversion factor and linearity, charge handling capacity per pixel, charge transfer efficiency in the image zone and serial register, minimization of residual images, anti-blooming efficiency, and packaging. The present baseline is summarised in Table 3.13.

Table 3.13: Properties of the CCDs foreseen for the GAIA astrometric instruments.

Feature	Corresponding Information
Array size	29×58 mm^2 with 25×58 mm^2 active area
Pixels per CCD	2150 columns $\times$ 2780 TDI stages
Dead zones	Top: 0.25 mm; sides: 0.6 mm; bottom: < 5 mm
Pixel size in image zone	9×27 μm^2 (9 μm along scan)
Phases in image zone	4
Pixel size in serial register	27×27 μm^2
Phases in serial register	4
Device thickness	10–12 μm
Si resistivity	In the range 20–100 Ωcm
Buried channel	Standard n-type channel used for scientific CCDs
Oxide thickness	Standard
Anti-blooming implementation	Shielded anti-blooming at pixel level
Notch channel	Implanted for all CCDs
Output amplifiers	2 per device (selectable), 2-stage
Conversion factor	Selectable between 3 and 6 μV/e$^-$
Additional gates	5–10 gates with external commanding (bright stars)
Power dissipation	< 560 mW
Photo-response non-uniformity	$< 1\%$ rms at high spatial frequency $< 10\%$ peak-to-peak over entire array
Mean dark current	< 0.5 e$^-$ s^{-1} pixel^{-1} at 200 K
Non-linearity	< 1 per cent over 0–2V; < 20 per cent over 2–3.5V
CTI in image area	$< 10^{-5}$ at beginning-of-life; $\sim 10^{-4}$ after major solar flare
CTI in serial register	$< 10^{-5}$ at beginning-of-life; $\sim 5 \times 10^{-4}$ after major solar flare
Quantum efficiency	See Table 3.15
MTF at Nyquist frequency	See Table 3.15
Read-out noise	See Table 3.16

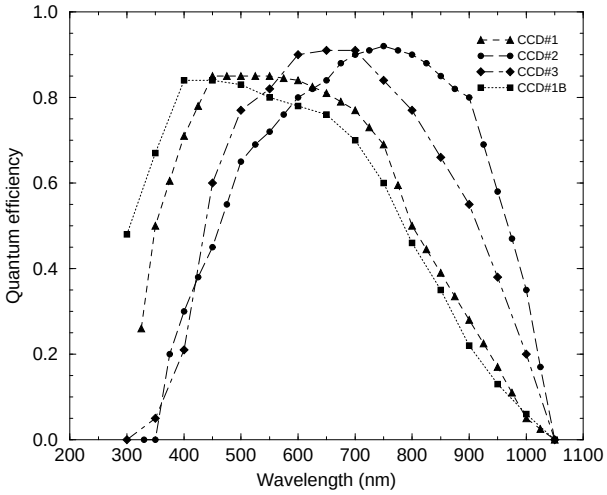


Figure 3.24: Quantum efficiency of various representative backside illuminated CCDs, including the baseline astrometric CCD response, CCD#1B.

3.7.2 Quantum Efficiency

Figure 3.24 represents three standard backside illuminated CCD QE responses: #1 enhanced in the blue range (EEV data); #2 enhanced in the red range (MIT/LL data); #3 standard epitaxial CCD (MIT/LL data). The performance impact is shown in Table 3.14, where it can be seen that these three CCD responses actually provide comparable astrometric accuracies as a consequence of working with a broad wavelength band (no filters).

In practice, for the astrometric field the MTF (see Section 3.7.3) must be optimised in parallel with the QE. The accuracy depends essentially on the integral of QE×MTF over the wavelength band, and consequently, the QE cannot be optimised in isolation. Details have been presented by Saint-Pé (1998), and the resulting QE and MTF values for the CCD selected for the astrometric field are summarised in Table 3.15. The resulting QE response curve is also shown in Figure 3.24, curve CCD#1B. This response has been used in the detailed astrometric accuracy analysis in Section 8.1.

The same response has been used for the assessment of the photometric accuracies (Section 8.2) for the broad-band photometric system (where the same focal plane CCDs are used), while the more red-sensitive CCD#2 has been adopted in the accuracy analysis for the red MBP filters and the radial velocity spectrograph, where the QE response versus wavelength can be optimised with less attention to the effect of MTF.

Table 3.14: Detected photons and astrometric accuracy for three types of CCD.

Quantity	CCD	B3V	G2V	M8V
Detected photons (e^- s^{-1})	#1 (blue)	37 510	36 060	156 380
	#2 (red)	30 500	42 300	317 910
	#3 (standard)	31 470	40 460	250 340
Astrometric accuracy (μas)	#1 (blue)	8.2	9.3	5.2
	#2 (red)	10.1	9.3	3.8
	#3 (standard)	9.7	9.3	4.2

Table 3.15: Baseline astrometric field CCD (CCD#1B), giving the along-scan MTF at Nyquist frequency for low ($< 10^5$ e^-) and high (saturation-level) signals, and the corresponding optimised spectral quantum efficiency. The QE includes a broad-band anti-reflection coating, and values are for operation at 200 K. These QE values are relevant for the astrometric field, where MTF performance is also relevant, and are as used in the detailed astrometric accuracy analysis. For the photometric and radial velocity CCDs, the QE response versus wavelength can be optimised with less attention to the effect of MTF.

λ (nm)	300	350	400	450	500	550	600	650	700	750	800	850	900	950
QE	0.48	0.67	0.84	0.84	0.83	0.80	0.78	0.76	0.70	0.60	0.46	0.35	0.22	0.13
$\mathrm{MTF_{low}}$	–	–	0.30	0.30	0.33	0.36	0.38	0.42	0.46	0.50	0.52	0.54	0.55	0.55
$\mathrm{MTF_{high}}$	–	–	0.13	0.13	0.14	0.17	0.19	0.23	0.29	0.34	0.38	0.40	0.41	0.42

3.7.3 Pixel Size and MTF

The pixel size, nominally adopted as 9×27 $\mu\mathrm{m}^2$ for the astrometric field, is an important design parameter. A smaller pixel size would decrease the telescope focal length, as well as the overall size of the Astro focal plane assembly, and consequently the overall size of the overall payload. However, such devices provide worse performance in a number of other areas. The trade-off between QE, MTF, and charge-handling capacity results in the design reference values presented in Table 3.15.

The manufacturing feasibility of CCDs with smaller pixels is not in doubt, and is demonstrated by CCD manufacturers even today. However, such devices have poor quantum efficiency, and/or poor MTF, and low charge-handling capacity. All of these parameters are key parameters for the ultimate astrometric performance. On the other hand, some manufacturers offer CCDs with very high performances but they do not offer small pixel size, one reason being that they use large design rules (typically $1.5 - 2\mu$m). The less expensive and most secure way to build up a payload of reasonable complexity with feasible CCDs of acceptable performance is to use the existing technology devoted to high-performance CCDs with the smallest achievable pixel size. The consultancy work carried out during the study concludes that CCDs with 9 μm pixels can be produced with good electro-optical performance and with acceptable risk. The feasibility of smaller pixels is much more questionable. Consequently, CCDs with 9 μm pixels have been considered and implemented within the baseline design. Further bread-boarding activities have been already identified which should demonstrate in a relatively short term the validity of this approach.

Assuming no gaps between pixels, the MTF due to the finite size of detector elements is:

$$\mathrm{MTF} = \left| \frac{\sin(\pi l f)}{\pi l f} \right| \tag{23}$$

where l is the pixel width and f is the spatial frequency. For two pixels per period (= Nyquist sampling, giving the minimum number of pixels to recover the full information in the image), MTF at the image spatial frequency is $2/\pi = 0.64$. Lateral diffusion of charges, fringing fields, and capacitances between electrodes causes one electrode to collect charges from another, degrading the MTF.

3.7.4 Readout Noise

The various steps in estimating the noise levels for each part of the Astro instrument consist of determining: (a) the target maximum star density used (the noise figures remain identical for regions of lower star density because of the fixed readout and flush frequencies adopted thereafter); (b) the number of useful pixels per CCD row (where a 'row' is perpendicular to the the scan direction); (c) the frequency at which the useful pixels are read-out from the CCD output stage; (d) the read-out noise and the analog noise, both depending on the read-out frequency; (e) contributions to the total noise that do not depend on the read-out frequency.

A 'worst-case' star density, corresponding to about 2.8×10^6 stars per square degree (about 19–20 mag in Baade's window) has been used in a detailed analysis by Vannier (1999), using numerical values from MMS and EEV (including consideration of the reset, binning and flush frequencies; and the TDI transfer and clock-feed through times). The total noise per sample includes the CCD

*Table 3.16: Summary characteristics of the various detection chains. ASM = astrometric sky mapper; AF = astrometric field; BBP = broad-band photometer (BBP1 for Astro-1, BBP2 for Astro-2); SSM = spectrometric sky mapper; RVS = radial velocity spectrograph; MBP = medium-band photometer. The last line gives the total readout noise assumed in current simulations and plots in this report. The values given for the MBP and RVS are conservative values, slightly higher than the computed ones because these are not yet finalised (indicated *).*

Parameter	ASM0/1	ASM3	AF01–16	AF17	BBP1	BBP2	SSM0/1	MBP	RVS
TDI lines/chip	2/2780	2780	2780	2780	2780	2780	*	730	3650
Columns/chip	2150	2150	2150	2150	2150	2150	*	7300	1220
Pixel size (μm)	9×27	9×27	9×27	9×27	9×27	9×27	*	10×10	20×20
TDI period (ms)	0.31	0.31	0.31	0.31	0.31	0.31	*	4.1	8.2
Sample bins (pixels)	2×2	2×2	1×8	2×12	1×8	6×8	*	1×1	1×1
Patch size (samples)	1×1	5×5	6×1	30×1	16×1	16×1	*	1×1	1×1
Patches per row	538	20	12	120	32	192	*	*	*
Useful pixels per row	1075	200	12	120	32	192	*	*	*
Parallel transfer (kHz)	30	30	30	30	30	30	*	20	175
Serial binning/flush (MHz)	–	15	15	15	15	15	*	15	–
Window mode	Full	√	√	√	√	√	*	√	Full
Total chips	10	10	136	34	40	40	*	7	3
Outputs/chip	2	1	1–2	1	1	8	*	2	8
Video chains	20	10	180	24	40	40	*	14	24
Conversion factor (μV/e$^-$)	6	3/6	3/6	3/6	3/6	3/6	*	3/6	6
Useful pixel read freq (kHz)	1200	290	120	370	331	*	*	*	75
Max. charge (10^3e$^-$/sample)	580	330	330	330	330	330	*	330	330
CCD readout noise (e$^-$)	9.4	5.2	4.1	5.7	5.5	4.3	*	*	1.8
Analogue video noise (e$^-$)	1.9	0.5	0.3	0.6	0.6	0.4	*	*	0.15
Quantization noise (e$^-$)	2.6	1.5	1.5	1.5	1.5	1.5	*	*	0.75
Total readout noise (e$^-$)	10.9	6.0	6.0	6.5	6.2	5.0	*	*	2.1
Total noise assumed (e$^-$)	10.9	6.0	6.0	6.5	6.2	5.0	5.3	3.0	3.0

read-out noise at the relevant read frequency, RON, analog-to-digital conversion noise, $N_{\rm q}$, and the video chain analog noise, $N_{\rm analog}$. The resulting total noise, including a margin of 1.1 for the non-linearity of the ADC signal, is determined as:

$$\text{Noise} = \sqrt{N_{\rm q}^2 + 1.1^2 \times (\text{RON}^2 + N_{\rm analog}^2)} \tag{24}$$

The present value of the relevant parameters for the various detection chains are summarised in Table 3.16.

All noise estimates are based on the 6 μV/e- conversion factor. On the other hand, the maximum charge (330 000 e$^-$) is based on the same figure for the conversion gain. For 3 μV/e$^-$ conversion factor (which is the baseline, except for the ASM0/1, the RVS and the MBP sky mapper), which can be selected when necessary using information from the ASM, the maximum charge would be equal to 667 000 e$^-$. The conversion factor of 6 μV/e$^-$ is mandatory for both RVS and MBP channels as the read-out noise has to be minimized. The addition of a switchable 3 μV/e$^-$ conversion factor will depend on the need for an increase in dynamic range (although it will be more difficult to get good information about the spectrum energy distribution given that the RVS pixels will be read out continuously). For the RVS, 8 video outputs are baselined at 75 kpixels s^{-1} with 4 non-destructive read-out (using a skipper mode CCD).

3.7.5 Dynamic Range

The goal is to cover as large a dynamic range as possible, given that very interesting but different science will be possible for both bright and faint objects. The projected faint limit of $G \simeq 20$ mag poses an immediate challenge for bright objects: even with the expected full-well capacity of 250 000 e$^-$, the implied dynamic range of $\simeq 10^4$ gives a bright limit of $G \simeq 10$ mag, with this estimate being somewhat on the optimistic side given the competing requirements of dynamic range and read-out noise. Various schemes have been considered to reach brighter magnitudes, such as subdividing the field of view in parts, either covered with chips with different pixel sizes,

or (partly) covered with filters. Each part of the field of view would thus cover a different part of a wider dynamic range. However, these schemes imply that the objects at the extremes of the magnitude range, which are thus only observed in parts of the field, would have lower exposure times. An alternative solution has been identified, in which bright stars are flagged in the ASM, and are thereafter observed for only a specific fraction of each CCD chip crossing, depending on their magnitude. The resulting effect on the accuracy leads to the saw-tooth accuracy profile for bright stars evident in Figure 3.14.

Bright objects will also cause blooming, which may affect a significant area nearby, and may (temporarily) change the sensitivity of the over-saturated pixels. Current anti-blooming CCD technologies should be sufficiently effective in preventing these effects.

3.7.6 Charge Transfer Efficiency

All CCDs used for both astrometric and photometric measurements are operated in the drift-scan or TDI (time delay and integration) mode. That is, charge packets are gradually built up while transferred from pixel to pixel at the same rate as the optical image moves across the detector. At the end of the chip the charges are transferred to the serial register and then shifted to the output node. Centroiding on the digitised output provides an estimate of the mean position of the optical image relative to the electronic transfer along the pixel columns. Demands on the precision (reproducibility) of the mean image position are strict. For instance, a standard error of 10 μas in the final trigonometric parallax translates to a required centroiding precision in the astrometric field, after geometric and photometric calibration of the detector, of 150 μas per CCD crossing, or 36 nm (1/250 of a pixel) in linear measure. Laboratory measurements undertaken to assess the feasibility of such sub-pixel centroiding accuracies are reported in Section 6.8. In addition, the averaging effect of TDI operation, plus backside illumination (no intra-pixel electrode effects) should further improve the centroiding accuracy.

A key parameter for achieving such high degree of reproducibility is the Charge Transfer Efficiency (CTE) of the CCD. In the present context it is more convenient to discuss the Charge Transfer Inefficiency (CTI) $\varepsilon = 1 - \mathrm{CTE}$, which is the fractional loss of charge from a charge packet as it is transferred from one pixel to the next. Very few charge carriers are actually lost (through recombination) during the transfer process; rather, some carriers are captured by 'traps' and re-emitted at a later time, thus ending up in the 'wrong' charge packet at the output; if short-time constant processes dominate, the main effect observed is that of image smearing. The CTI has an effect both on the photometric measurement (by reducing the total charge remaining within the image) and the astrometric measurement (by shifting charges systematically in one direction). Although CTI occurs both in the parallel (along-scan) and serial (normal to scan) transfer, the CTI during parallel transfer is particularly critical, since it affects the astrometric measurements in the direction where the highest precision is required, i.e. along the scan. In addition, CTI is worse along-scan due to the lower transfer rate.

While the effect is well-known in astronomy, detailed characterization is complex, and depends on variable such as material, supplier, depletion depth, read-out rate and signal level and, most importantly, operating temperature and trap density, the latter depending on irradiation levels. Recent literature describing the state of the art includes Holland et al. (1991); Dale et al. (1993); Hopkins et al. (1994); Gendreau et al. (1995); Hopkinson (1996), with a detailed report on specific modeling relevant to GAIA by Lindegren (1998b). The strong temperature-dependence of the effect can be appreciated by consideration of the time constants of the relevant mechanisms: the capture of signal electrons from the conduction band to the trap level and their subsequent emission back to the conduction band (Hopkins et al. 1994):

$$\tau_c = 1/\sigma_n v_{\mathrm{th}} n_s \tag{25}$$

$$\tau_e = \frac{\exp(E/kT)}{\sigma_n X_n v_{\mathrm{th}} N_c \chi} \propto T^{-2} \exp(E/kT) \tag{26}$$

where n_s = electron density within the buried channel; σ_n = capture cross section for mobile electrons; $v_{\mathrm{th}} = \sqrt{3kT/m_e}$ = mean thermal velocity for electrons; N_c = effective density of states in the conduction band; T = absolute temperature; k = Boltmann's constant; X_n = 'entropy factor' associated with the entropy change for electron emission from the trap; χ = field enhancement factor (accounts for enhanced emission due to the Poole-Frenkel effect or to phonon assisted tunneling); E = energy level of the trap below the conduction band.

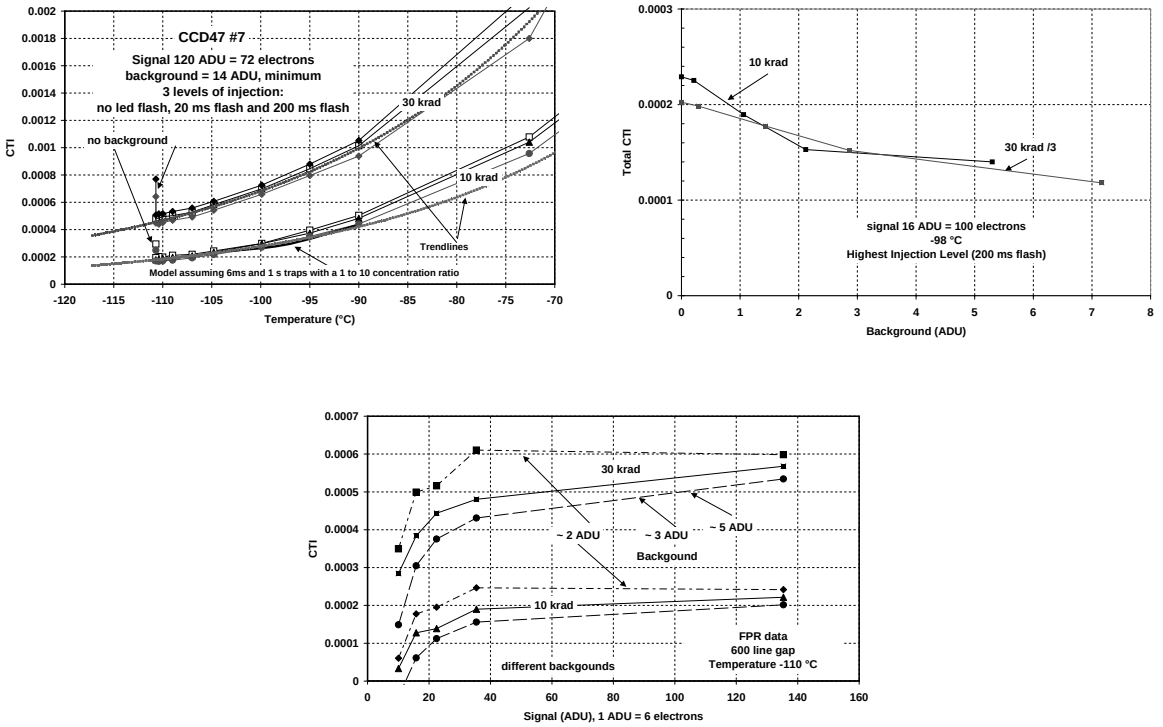


Figure 3.25: Results from the Sira investigations into the effects of radiation and charge injection on CTI. Top left: measurements made using the 'first pixel response' method show the dependence of CTI on irradiation level, and the improvement of CTI with decreasing operating temperature (the model parameters indicated apply at $-90°$). Top right: the CTI scales with irradiation level, and decreases with increasing background level, at any given operating temperature, due to the filling of trapping sites. Bottom: the CTI, at given irradiation levels and background level, appears to decrease with signal level, for reasons not yet understood.

The magnitude of the problem can be estimated as follows. Ignore for a moment the detailed physical causes of CTI and assume that a constant fraction ε of the charge is left behind while the fraction $1-\varepsilon$ flows into the next pixel. After N such transfers the fraction of charge remaining in the leading packet is $(1-\varepsilon)^N \simeq 1-N\varepsilon$, while the fraction of charge delayed by one pixel is $N\varepsilon(1-\varepsilon)^{N-1} \simeq N\varepsilon$, and so on. To first order in the (small) quantity $N\varepsilon$, the centroid of the output image has then been shifted by the fraction $N\varepsilon$ of a pixel. Actually this assumes a constant size of the total charge as it travels along the CCD. In reality, through the TDI observation mode, the signal builds up gradually along the chip, reaching its final size S (electrons) immediately before it arrives at the serial register. Detailed analysis of this process shows that, to first order in $N\varepsilon$, the number of electrons delayed by one pixel is $N\varepsilon S/2$. Consequently the expected centroid shift is $\simeq N\varepsilon/2$ pixels. The CCDs in the astrometric field of GAIA have $N = 2780$ pixels of size 9 μm along the scan. Assuming $\varepsilon = 10^{-5}$ (cf. Table 3.13) results in a centroid shift of 125 nm or about 500 μas.

To the extent that the centroid shift is constant (or rather reproducible under well-defined conditions), it can be calibrated as part of the normal data analysis and will not affect the final astrometric accuracy. The actual number of electrons left behind through the CTI is however a stochastic process. There is consequently also a random component to the centroid shift. The number of electrons delayed by one pixel (on average $N\varepsilon S/2$) can be assumed to vary by $\pm(N\varepsilon S/2)^{1/2}$. The corresponding uncertainty in the centroid shift is $\pm(N\varepsilon/2S)^{1/2}$ pixels. This is less than the required maximum uncertainty of 1/250 pixel if $S > 10^3$ e$^-$ (still assuming $\varepsilon \sim 10^{-5}$), i.e. for stars brighter than $V \sim 18$. At this magnitude the centroiding error due to the total photon noise in the image is several times larger. Since the latter error also varies approximately as $S^{-1/2}$, it follows that the CTI effect considered here contributes a roughly constant fractional increase in the total standard error. Given that the standard width of the point spread function in the astrometric field is about one pixel, it is seen that the fractional increase in the astrometric standard error is about $0.25N\varepsilon$. This is practically negligible (say, less than 5 per cent) if $\varepsilon < 7\times10^{-5}$. This value could however be exceeded after a major solar flare (Table 3.13).

The conclusion from this first appraisal of the CTI effects on the astrometric accuracy is that the effect is negligible for the expected nominal characteristics of the undamaged (beginning-of-life) CCD, but potentially serious for the degraded performance that may result after significant exposure to particle radiation in orbit. To proceed with a more careful evaluation of the effects and of possible remedies, it is necessary to consider in some detail the physical mechanisms of the CTE degradation in orbit.

In the analysis above the CTI was represented as a single number. However, it is well known that the charge loss in real CCDs is not simply proportional to the number of electrons in the charge packet, but depends on many other factors including the history of charge transfers across a given pixel. It is also well known that the CTI may be dramatically increased through particle radiation. This effect is believed to be caused by trapping states associated with lattice displacements produced by energetic particles such as protons. During the transfer of a charge packet across these traps, electrons are captured and thus removed from the charge packet. Depending on the temperature and the detailed characteristics of the trapping states, the captured electrons may be released significantly later and thus join a different charge packet. The net result is an increased CTI through 'charge deferral'.

The increasing use of CCDs in space missions has stimulated a lot of research into the causes of CTI and especially the role of radiation-induced traps. Laboratory measurements of the CTI in radiation-damaged CCDs have been made and it has been possible to represent the gross behaviour of the CTI as function of radiation dose, temperature, background illumination and signal size by means of a relatively simple physical model of the radiation-induced traps. For a recent example of such analysis, see Hardy et al. (1998). On a microscopic level the CTI is then described in terms of the time constants of capture and re-emission, as the charge packets move across the traps.

The dominant CTI effect in radiation-damaged buried-channel CCDs is attributed to the formation of E-centres (phosphorus vacancy complex) introducing trapping states at an energy level of about 0.4 eV below the conduction band. At the operating temperature considered for GAIA ($\sim$ 200 K) the typical emission time constant for the E-centre is of the order of 0.1 second. The capture time constant is inversely proportional to the instantaneous charge density at the trap, but is much shorter than the parallel transfer dwell time ($\sim$ 0.1 ms) even for a charge packet with only a few tens of electrons, if it passes right over the trap. Thus, whenever a charge packet in a stellar image travels across a trap, an electron will be captured (unless the trap was already filled). Furthermore, it is likely that the electron is re-emitted into a much later packet, rather than the next packet as was previously assumed. Qualitatively, this will affect the measurements as follows: (a) the charge loss is strongly non-linear, resembling a threshold phenomenon rather than a proportional loss — thus, faint stars are relatively more affected than bright stars; (b) the CTI depends strongly on the background level, since a weak, uniform illumination will tend to saturate the traps and hence minimise signal electron capture; (c) the relatively long emission time constant means that deferred charges are likely to be deposited into the general background field — this could reduce the CTI-related centroiding error, compared with the calculation above; (d) on the other hand, the short capture time constant means that trapping preferentially removes electrons from the leading charge packet in the electronic image of a star, again shifting the centroid towards the later packets.

The effects of the radiation-induced CTI on the GAIA measurements thus become complicated functions of factors such as the background (sky) level, object brightness and the density and distribution of objects. The temporal build-up of the effect must also be considered: the CTI is roughly proportional to the density of traps in the CCD bulk material, and therefore to the accumulated radiation dose.

Much of the CTI effects can be calibrated as part of the normal data analysis, i.e. the systematic centroid shift and signal size reduction as functions of magnitude, colour, background level and time. However, the stochastic effects related to the charge losses can never be eliminated by clever processing.

Experimental results A series of laboratory experiments was carried out during the GAIA study by Sira Electro-Optics Ltd (Hopkinson 1999), under a technology contract focusing on (i) an improved understanding of the effects of radiation damage on CTI; (ii) characterising the time constants of the trapping centres as a function of temperature; (iii) establishing the feasibility of filling the traps ahead of the signal passage using controlled light flashes or regular signal injection on time scales shorter than the emission time constant(s); and (iv) on providing inputs to an improved model of the effects of CTI and its impacts on the centroiding accuracies. Different CCD chips were irradiated at levels of 4, 10 and 30 krad (the former corresponding to the cumulative damage expected from one or two 'major' solar flares), and the captured charge measured by the technique of 'first pixel response' in which the deferred charge was estimated from the different signal levels measured in regions with and without background charges. Measurements were carried out down to $-90°$C, at various signal levels, and using LED pre-flashes (with various time delays) in order to fill the traps ahead of the signal passage (Figure 3.25).

Results are consistent with the presence of at least two types of trap, with time constants of 6 ms and 1 s at $-90°$C (the latter consistent with the known energy of the P-V centres), with the concentration of 1 s traps some $10\times$ that of the 'fast' traps. The temperature dependence of the processes is such that at $-110°$C the 1 s traps would be filled, with the emitted charge from these traps giving a small background of about 10 e^- at 4 krad and at $-90°$C. The suspected fast traps affect the image profiles, giving a small tail. The implications are that cooling to $-110°$C will be advantageous, while keeping the fast traps filled might require going a further 30–40° cooler. Future work should concentrate on (i) tests at a substantially colder temperature in order to provide clearer information on the relative populations, energies, and time constants of the traps, (ii) measurements in TDI mode, and (iii) irradiation at other proton energies to verify the reproducibility of the trap characteristics. Finally, during the project phase, analysis would be performed to assess the screening of the spacecraft, and to dimension additional thin aluminium screening panels to be fixed, for example, on the payload thermal structure.

3.7.7 Timing Considerations

The on-board timing requirements are currently specified as 1 μs over 10 s (relative) and 5 μs (absolute), originating from the requirements of the great-circle reductions, aberration corrections for solar system objects, and Earth-ephemeris definition. The observation method defined for the GAIA mission requires a high stability of the sequencing between CCDs and between focal planes. Therefore, the phase difference between the various sequencer commands will be less than 10 ns for the two focal planes. This is considered feasible with existing technology.

Typical design guidelines will employ: (i) a common core for the sequencer, with symmetrical architecture (typical 50 Ω coaxial cable delay is 5 ns/m); (ii) selection of adequate parts (ECL technology) for the specific functions (dispersion significantly less than 1 ns); (iii) use of pairs of CCDs from the same batch; (iv) re-synchronisation between command signals using a common master clock; (v) appropriate layout of the focal plane; (vi) if necessary, exchange of bi-directional synchronisation signals. The same recommendations apply for the exchanges with the video chains up to the digitization stage. In addition, it is understood that there is a requirement on the temporal stability of the sequencing over 100 s for the focal plane array sequencer, which will be better than $\pm$10 ns. This is also deemed feasible with existing technology; (vii) the temperature fluctuation over 100 s at sequencer level is expected to be low.

The required limit on field distortion also depends on whether all CCDs are clocked synchronously. If they are clocked individually then the distortion requirement may be significantly relaxed. Apart from these systematic effects the clock rate should ensure a shift of charges equal to the scan velocity with a 1σ error about 1×10^{-4}. A 3σ error will then give a smearing by 1 pixel width across the wide CCD containing 3330 pixels. Thus the real-time knowledge of scan rate should be about $0.5 \times 10^{-4} = 0.006$ arcsec s^{-1}. This implies a tolerance on acceleration about 0.006 arcsec s^{-2}. The requirement on the relative rate error means that there is no need to adjust the TDI sequencing during one integration period. An adjustment capability will be implemented, which will be used at least at the beginning of the observation phase, in order to adjust the sequencing frequency to the current spin rate.

3.7.8 Focal Plane Technology

During the study, an assessment of the feasibility of the Astro focal plane and of its achievable performances has been carried out by the industrial contractor with the support of several CCD manufacturers (EEV, Thomson, Philips). A review of the most innovative features of the Astro focal plane array is presented here, together with the conclusion of the feasibility assessment study.

The size of the elementary array foreseen is 29×58 mm^2. This size is comparable with existing 3-side buttable thinned backside CCD devices which are already used for on-ground astronomical projects. Larger CCDs, even if feasible with a satisfactory technological yield, are not of interest due to the optical distortion constraint.

The astrometric focal plane array represents a very large focal plane (about 0.7×0.6 m^2), with 250 elementary CCD arrays plus the corresponding proximity electronics. The power dissipation is expected to be about 100 W and the operating temperature is about 200 K. Construction of such a focal plane is challenging but the design is not so far from what is today developed for on-ground astronomical observatories. The Sloan Digital Sky Survey project and the Megacam project are two examples of such applications based on similar architectures. The Megacam focal plane, for example, uses 3-side buttable CCDs, with elementary CCD size of approximately 30×60 mm^2, covering a 260×280 mm^2 surface. Nevertheless, the detailed focal plane design will require further effort to select the most appropriate architecture and technologies for the opto-mechanical-thermal assembly and the most appropriate architecture and technologies for the front end electronics. For the problem of the alignment (co-planarity) of the CCD mosaic, see Tulloch (1996).

The following alignment requirements on the focal plane assembly are derived from the optical analyses: (i) the optical design provides a flat focal plane. Therefore, all the CCDs can be mounted on a mechanical flat support and there is no need for adjusting the CCDs in focus. The detector planarity is not critical (and is of order $<$ 0.05 mm peak-to-valley) due to the large f-number ($f/30$), with a 1 mm defocus corresponding to a WFE = $\lambda/14$ rms; (ii) the same remark holds for the focal plane tilts. The focal plane tilt adjustment should be such that the error does not exceed 0.05 mm at the focal plane edge. This corresponds to a tilt adjustment of about 0.6 arcmin, which will be achieved by measuring the WFE with an interferometer at several points of the field

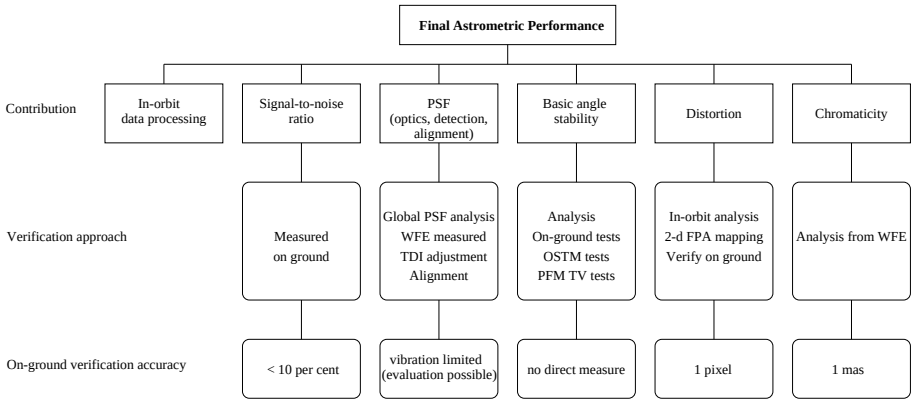


Figure 3.26: Verification philosophy and test method.

and shimming the focal plane; (iii) there is no specific optical requirement on the lateral position (translation) of each detector, except for having the best possible filling factor (> 86 per cent); (iv) each CCD detector must be aligned in orientation in such a manner that the star motion is made parallel to the detector columns (TDI direction). The orientation adjustment requirement should correspond to a motion of less than about one pixel across scan (~ 30 μm) at the detector edge, i.e. an angular adjustment accuracy of ~ 3 arcmin. An orientation offset, of maximum value $0\overset{\circ}{.}6$ (at the field edge) is also required for each detector for compensating across scan distortion. This offset is deterministic and is not critical because the distortion amplitude is rather insensitive to optical misalignment.

3.8 Performance Verification

The GAIA performance is driven by the astrometric accuracy. The major contributors to the final performance, and the proposed verification method for each parameter, is as follows (Figure 3.26):

- the point spread function of the whole instrument, including the contributions from the optical PSF, the detection effects (including TDI synchronisation), and the focal plane versus telescope alignment (defocus, tilt, etc.). The optical PSF will be computed from the wavefront error measurements performed on ground. As the system is purely reflective, a monochromatic interferometric measurement will be sufficient to give all the necessary information about the wavefront. A set of measurements in the field of view is necessary. The detection effects will be estimated by test and analysis. For the tests, the MTF of the detectors will be measured in representative conditions, i.e. in TDI mode with a moving stimulus. The evaluation of the TDI synchronisation error will be taken from the expected accuracy of the on orbit adjustment system. The alignment adjustment effect will be computed by analysis. The numbers used for adjustment values will be taken from the resolution of the tilt/focus adjustment device foreseen in orbit;
- detection signal-to-noise ratio: this can be measured on ground using radiometrically calibrated sources;
- basic angle stability: this is estimated by analysis. The thermo-elastic analyses will allow prediction of the expected in-orbit variations of this parameter, while the concept of the monitoring system will be verified by use of an on-ground proof-of-concept test bench;
- the stability of the structure, telescope and focal plane will be measured in some specific thermal cases during the payload module thermal test. During this test, the thermal loads will be amplified to allow a direct measurement of the stability as the final accuracy is not accessible. The obtained results will be fed back into the models. The accuracy of the proof of concept test bench will also be limited by the ground environment but it will help to verify that the proposed concept is working in principle;
- distortion: the distortion map is obtained after post processing. A starting value can be obtained for the telescope contribution using the results of the optical modeling because the first order distortion only depends on the optical design. The focal plane can also be mapped on ground at sub-system level. If necessary, the distortion in some chosen points of the field of view could be directly measured for the whole instrument on ground;
- chromaticity: this parameter is estimated by analysis using the wavefront maps measured on the telescope.

Estimates of the accuracy that could be obtained by direct on-ground measurements are given below, and show in most cases that the ultimate accuracy cannot be measured directly on ground, and that analyses have to be performed to assess this final accuracy. An important feature for the on-ground tests is that the detectors operate in TDI mode which implies a synchronously moving target. Tests with that type of target have already been performed on ground successfully at detector level. Others are foreseen at satellite level. The optical ground support equipment (OGSE) will therefore need, as baseline, a moving star object as a stimulus:

- point spread function of the whole instrument: at telescope level the wavefront can be measured in air using an interferometer, such as the Zeiss 'Direct 100' system which averages the turbulence. Tests have shown that the repeatability of the measurements can be in the 15 nm rms range or better when the air is properly mixed using fans. The test accuracy will also depend on the actual figure of the auto-collimation mirror but this is not critical. At focal plane level, it is assumed that TDI measurements of the MTF will have been performed by the detector manufacturer. It is therefore likely that the main contributors to the PSF parameter can be measured on ground with a sufficient accuracy. At instrument level the knowledge of the PSF is necessary with a spatial accuracy in the order of a few percent of the total image size. This corresponds to a tenth of a pixel. The constraints on the synchronisation of the TDI mode are thus not very stringent. The final limitation will be the influence of the external vibrations considering the very long focal length of the telescope and has not been evaluated today;
- signal-to-noise ratio: this parameter is not considered as critical. The standard accuracy for radiometric measurements can be taken in the 10 to 20 per cent range;
- basic angle stability: this cannot be measured on ground. The accuracy of the stability tests to be performed on the optical structural and thermal model will be limited by the vibration environment;
- distortion: the wavefront error measurement accuracy has been discussed above. For the focal plane, it is possible to perform a two-dimensional mapping of the CCDs. A realistic accuracy without too much complicating the test set up is in the range of 10 μm, which means roughly one pixel. If a better accuracy is mandatory, it seems possible to go down to 3 μm but at the expense of more costly test equipment;
- chromaticity: it is not foreseen to measure this parameter directly. A possible on ground test could be to use a fibre tip at focus of a collimator and to switch between two different colours. A post processing such as a centroiding on the data could give an accuracy of the order of 1/40 pixel (this value is evaluated from experience on static spots and with focal lengths in the meter range; the real accuracy in TDI mode for the GAIA focal length remains to be verified). However this results in an angular accuracy around 1 mas, far from the final 10 μarcsec. This test can thus only be foreseen for first order check but its interest is dubious.

A preliminary overview of the OGSE which will be needed for the test of the GAIA instrument includes:

- an interferometer needed for the wavefront error measurement of the telescope. The interferometer is a standard equipment. This wavefront parameter will be used during the telescope alignment and as an alignment success criterion. The measurement also needs an auto-collimation mirror covering the full pupil. The auto-collimation mirror is a flat mirror with dimensions equal to the pupil. Its absolute curvature is not a problem to first order as a refocusing mechanism is foreseen in orbit. Its wavefront (or the knowledge of it) will give a limit to the accuracy of the measurement. However mirrors with such a large size and good accuracy have already been manufactured and do not present feasibility problems;
- measurement at focal plane assembly level: the OGSE for measurements at focal plane level must simulate a scanning star with the required optical quality. However only the aperture ratio of the instrument is needed and the field of view of the OGSE can be small. The optical part is thus not a problem. For the scanning, there are two possibilities either to scan the (point source and optics) assembly or to scan the point source at the focus of the optical system. The trade off will have to be made in the design phase but this OGSE does not appear to be critical for the time being. Such OGSEs have indeed been already used by detector manufacturers. For the mapping of the focal plane, a mechanical arm with 3D capabilities and an accuracy in the μm range will be added. This part will be similar to the one used in 3D measuring machines;
- measurement at instrument level/scanning auto-collimation mirror: for measurement at instrument level, we propose to use a principle already used for Hipparcos and which avoids having a full pupil size collimator. This one is presently ruled out for cost reasons and because its size together with the payload module would not allow to fit within the existing vacuum test chambers. We propose to have pinholes used as sources in the focal plane (slits parallel to the scanning direction are also possible). Such a pinhole gives a plane wave at telescope output which can be sent back to the focal plane by a flat auto-collimation mirror. To be compatible with the TDI mode the mirror must be dynamically tilted. Therefore its structure and actuation mechanism will have to be properly studied but this does not appear as unrealistic. In this configuration, the optical beam will pass twice inside the telescope so that the contribution of the optical quality is doubled. However the limitations in accuracy will likely be given by the vibration environment. In any case this OGSE will allow to check all the functionality of the payload;

Two major performance tests can be performed on ground: (a) a payload optical structural and thermal model test: the hardware under test is made up of a real SiC structure (used for flight) with SiC blanks (i.e. un-polished mirrors) and a dummy focal plane. The aim of the test is first to correlate the thermal mathematical models but another goal is to measure the stability of the basic angle or at least to verify that the thermo-elastic behaviour of the instrument is in line with the predictions. It is not possible with the present design to implement the real basic angle monitoring system at that level. Indeed this implies to have polished sub-pupils on the mirror but it also requires an alignment and a representative device for image acquisition in the focal plane. The present idea is to put a number of metrology devices inside the payload to make distance or angular metrology measurements. Such devices could be laser gauges or autocollimators and are yet to be defined. As the predicted variations for the nominal environment are in the picometer range, we propose to amplify the thermal perturbations such as gradients so that the predictions give values which can be measured. The amplification factor should be around 1000 as this would give values comparable to the Hipparcos performances which were measured during the thermal vacuum tests; (b) a protoflight model payload module test: this test is performed in vacuum with the final flight hardware. To be compatible with existing vacuum chamber, the only OGSE which can be used for this test is the auto-collimation scanning flat mirror working with the focal plane pinholes and described above. For example, this is compatible with the vacuum chamber SIMDIA at INTESPACE. However this facility is not equipped with a vibration isolated bench and the vibration environment might be severe. In any case it is realistic to consider that the tests will be functional ones only. The proposed OGSE allows to perform those tests.

4 Spacecraft System

4.1 Overview

The spacecraft subsystems provide all necessary support to the payload instrumentation, and its primary features are summarised in the following sections. Generally, the subsystems follow well-established spacecraft engineering approaches, although specific innovative features are included, particularly within the mechanical and thermal configuration necessary to reach the high structural stability requirements of the payload, and within the telecommunication subsystem in order to provide the high telemetry rates required. Summaries of the resulting power, propellant, and mass budgets are provided.

(a) Mechanical design (Section 4.2): the main structure has a hexagonal conical shape, designed to avoid turning shadows on the sunshield. It is an aluminium structure with CFRP shear walls, and a central tube supporting 4 propellant tanks and a main engine. The deployable solar array is made of 6 panels (CFRP structure back insulated with multi-layer insulation, shape-memory alloy hinges) and completed with a sunshield made of multi-layer insulation sheets and provided with Kevlar cables for deployment post launch.

(b) Thermal control (Section 4.3): the thermal control is based on optical solar reflector (OSR) and multi-layer insulation (MLI) sheets on outer faces of the service module and on a black painted cavity completed with a heat pipe network for temperature homogeneity

(c) Propulsion and attitude control (Section 4.4): a chemical bi-propellant propulsion system for the transfer phase comprises one main (400 N) engine for apogee raising plus one set of redundant 10 N thrusters for attitude acquisition/spin control/mid-course corrections/L2 orbit injection/safe mode. In addition, one redundant set of FEEP (Field Emission Electric Propulsion) thrusters is provided for the control of the operational orbit and spin motion at L2, operated in continuous proportional mode. The total propellant mass amounts to 1010 kg (including margin). The attitude measurement and control involves spin-stabilisation during the transfer phase, with 3-axis control mode in operational phase. Three Sun acquisition sensors plus one gyroscope are provided for spin axis stabilisation during the transfer phase, with one large field of view star sensor plus use of the main instrument sky mappers for the operational phase.

(d) Payload data handling (Section 4.5): dedicated acquisition and processing electronics are provided for the science data (data discrimination/compression/temporary storage) including a 100 Gbits mass memory. The typical (continuous) payload data rate amounts to about 1 Mbps. There is a centralised platform processor for the overall control and monitoring of the satellite.

(e) Power and electrical subsystem (Section 4.6): this comprises 24 m^2 of solar array split into six deployable panels, with GaAs cells on a CFRP structure; one 14 Ah Lithium-Ion battery for launch and early-orbit operations; and an unregulated power bus.

(f) Communications (Section 4.7): telemetry and telecommand employs X-band up- and down-links with a few kbps capacity and an omni-directional coverage. An S/X band system is also possible if required. The science telemetry X-band down-link with ~ 3 Mbps capacity used during each ground station visibility period (> 8 hours per day), based on a set of electronically-scanned phased array antennae accommodated on the service module panels.

4.2 Mechanical Design

4.2.1 Payload Mechanical Design

The payload mechanical and thermal design aims at providing the required stability passively (the only mechanism in the astrometric telescopes is the secondary mirror mechanism, whose purpose is to compensate at the beginning of life potential optics misalignments due to launch and residual on-ground alignment errors). The selection of an L2 orbit combined with the sunshield thermal cover and the constant solar aspect angle allow minimising the external perturbations.

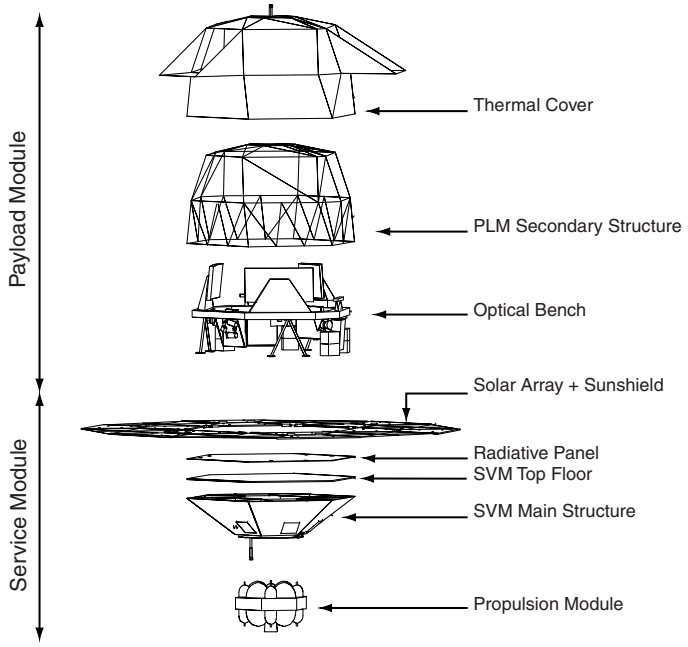


Figure 4.1: Satellite exploded view.

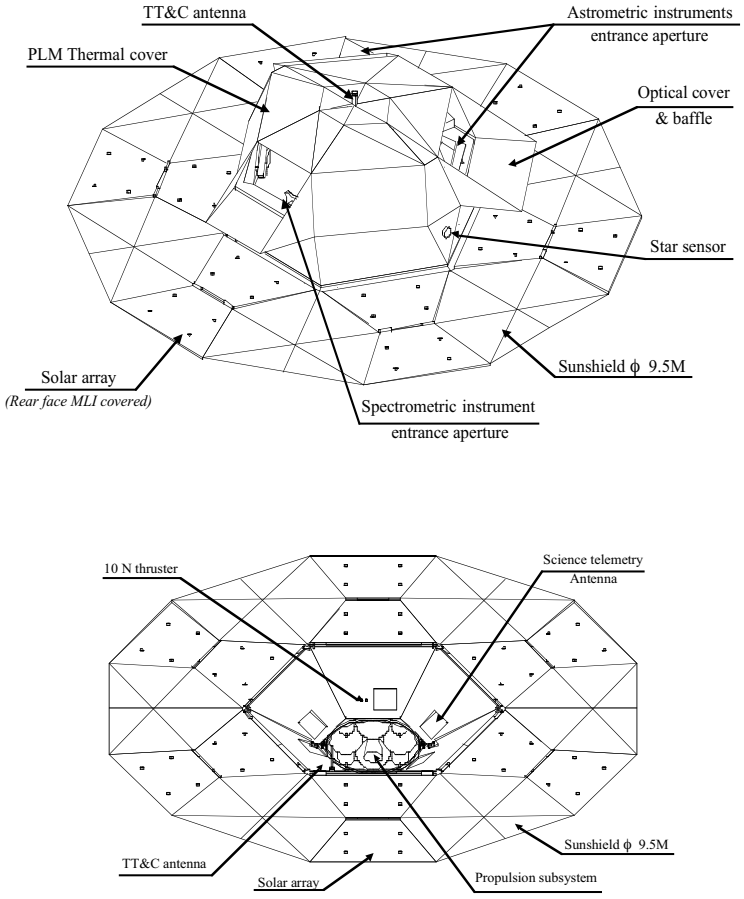


Figure 4.2: Satellite overview from the payload side (top), and from below (bottom). A deployable solar array/sunshield assembly protects the payload against direct Sun illumination and avoids rotating shadows on the payload module.

Table 4.1: Key properties of candidate materials for GAIA. For CFRP, an isotropic lay-up is considered with M55J/RS3M fibre/resin.

Parameter	Zerodur	Carbon-Carbon	Silicon Carbide	CFRP
Young's modulus, E (GPa)	90	60	420	117
Density, ρ (g/cm3)	2.53	1.56	3.15	1.65
Ultimate tensile strength (MPa)	20	190	350	240
Thermal expansion coefficient, α (ppm/K)	0.05	–0.1	2	0.46
Thermal conductivity, λ (W/m/K)	1.5	7	190	37
Light-weighting ratio, E/ρ	35	38	130	71
Thermal distortion ratio, α/λ	3.4	1.4	1	1
Polishing	√	×	√	×
Coating	√	√	√	√

To take full benefit of this environment, the payload material should have a low coefficient of thermal expansion and a high thermal conductivity for minimizing the thermo-elastic distortions due to variable heat loads; good structural and optical properties, allowing construction of mirrors and the primary structure with the same material, and therefore to have an alignment stability independent of uniform temperature variations; and a good light-weighting capability for mass reduction. Table 4.1 represents key properties of candidate materials for GAIA. Silicon carbide is appropriate for GAIA because it provides simultaneously good strength properties, good optical properties, an excellent light-weighting capability and a very good intrinsic stability (low coefficient of thermal expansion combined with a high thermal conductivity).

Silicon Carbide SiC100 material from SiCSpace (France) is selected for building the GAIA payload. This material has the remarkable property of combining the most interesting properties of homogeneous silica glass (homogeneity, polishability, coating) and of metals (thermal conductivity, high strength), while providing an excellent light-weighting capability. The properties of silicon carbide allows this material to be used both for the telescope mirrors and the payload torus structure providing a homogeneous, high conductive, athermal payload:

(1) the material manufacturing process is industrial and fully well-mastered. It is manufactured in serial production since many years for automotive industry and for heat pumps;

(2) the material exhibits an unprecedented light-weighting capability: a very high Young's modulus of 420 MPa, for a moderate density of 3.16 g cm^{-3};

(3) the material is homogeneous, and features a very low coefficient of thermal expansion (CTE), which drops from 2 ppm/K at room temperature to about 0.5 ppm/K at 100 K;

(4) the material can be polished like glass, and is used for visible applications. It can also be coated like glass, with the same coating process;

(5) the manufacturing process (pressureless sintering at high temperature) is such that the material exhibits very low in-built stresses, lower than 1 MPa;

(6) the material can be used over a very wide temperature range (0–1800 K). It is insensitive to very harsh environmental constraints (humidity, acid or alkali attack, radiation);

(7) the material thermal conductivity is very high, comparable to that of aluminium;

(8) the material features good strength properties (bending strength about 400 MPa) and can be used for the telescope structure, allowing an all Silicon Carbide fully passive athermal payload design. The material is currently considered as a serious candidate for the FIRST 3.5 m telescope and a spherical reflector of 1.35 m diameter was built in 1999 by MMS for demonstration (see Figure 4.4). For the purpose of FIRST, the demonstration model was deliberately made of nine segments brazed together. The material has also been selected for the OSIRIS telescope to be launched on ROSETTA, and for the SOFIA telescope secondary reflector. For both FIRST and OSIRIS, the telescope structure is also made of silicon carbide.

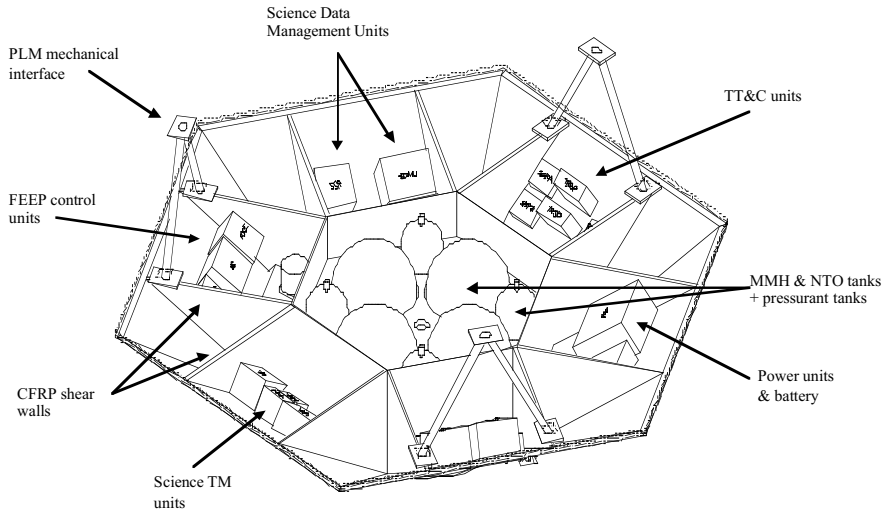


Figure 4.3: Service module overview. The service module has a conical shape in order to avoid rotating shadows onto the payload module. The structure is made of aluminium and CFRP. A dual propulsion system is implemented (unified bi-propellant system for the transfer phase with FEEP electrical micro-thrusters for the operational phase). Phased array antennas are accommodated on each of the six service module lateral panels to establish the link with the ground station despite the satellite spin motion.

For GAIA, the mirror sizes are compatible with present manufacturing capability and will be monolithic. As for OSIRIS, the mirrors will be overcoated by a thin layer of SiC CVD (Chemical Vapour Deposition silicon carbide, whose properties are very close to that of SiC100) of about 50 micron thickness prior to polishing. For minimizing chromaticity errors, the fine polishing will most probably be achieved with an ion-beam figuring technique, which gives excellent results on this material. As an illustrative example, an aspheric mirror of diameter 200 mm was polished by MMS in 1998 with a residual wavefront error as low as $\lambda/100$ rms in the visible. The torus structure will be made from individual parts bolted together.

4.2.2 Optical Bench Design

A 'cylindrical' type structure (torus) supports all the optical instruments and the focal plane (Figure 4.5). It is a simple architecture well-adapted to the symmetry of the two viewing directions, while the RVS/Photometer assembly is located on the symmetry axis. For the torus, an open-back design is chosen to improve manufacturing and integration. The open-back concept is the same as for mirrors. It has a rectangular section of 300×150 mm, with a thickness of 5 mm. The locations of the reinforcement are chosen with respect to the equipment and the interfaces. In order to simplify the manufacturing, the torus can be made of several parts which are then bolted together. A single material (SiC) is used for the reflectors and the torus. Then the optical alignment is insensitive to uniform temperature variations. All mirrors are in one part, isostatically mounted on the torus, and are made stiff enough for making negligible any mirror distortion effect on optical quality. A preliminary design of the primary mirror has been performed to assess the mass budget, considering a classical open back concept with a triangular reinforcement.

The focal plane electronics units are mounted on the torus. They are thermally decoupled using permaglass washers and MLI to avoid direct coupling with the torus—they are only coupled by their front side with the payload cavity. Their rear side is a radiator outside the tent. A fine power regulation is used to minimise perturbations. The video processing unit may have a wide variation of power. To avoid perturbations on the payload module they are mounted to the service module and radiate directly to space with dedicated radiators. The harness between

Figure 4.4: FIRST 1.35 m Silicon Carbide reflector, demonstration model made of nine brazed segments (courtesy MMS).

the focal plane units and the video processing unit is thermally stabilized with radiator and/or heaters. The stars trackers are also mounted on the torus. A specific insulation using MLI and permaglass washers is used to avoid thermal perturbation in the payload module cavity. They are completely outside the cavity.

The optical bench is linked to the service module by means of three titanium bipods. This naturally gives a conductive insulation. These bipods, which have been shown to have adequate resistance to stress, must provide a sufficient pseudo-isostaticity for the optical bench which will not be perturbed by the service module distortion. For the very low thermal fluctuations of the payload module ($\sim$25 μK), the current flexibility of the bipods is sufficient to provide the appropriate stability of the instruments. But the service module has a turning gradient ($\sim$60 K) between the Sun and anti-Sun directions. With a non perfect isostaticity of the payload module deformation of the bipods and of the optical bench will result. The proposed solution is to unlock two bipods after launch; the optical bench has only one interface point and is then naturally purely isostatic.

To insulate the payload module from space, a secondary structure is mounted on the service module to support the thermal foils. This tent is optimised to be under the protection of the Sun shield and has no connection with the payload module to improve stability. In the lower part of the tent (between service module and torus), 2×6 sheet MLI separated by the bars are used to improve the radiative isolation from the solar array subsystem. In the upper part, only a 6 layer MLI is used. The MLI is fixed with Velcro or inserts. Bars are in CFRP to improve mass. Baffles are deployed with the doors to avoid direct radiative coupling between the solar array subsystem and the payload module cavity.

4.2.3 Basic Angle Monitoring in Orbit

Although the satellite mechanical/thermal design (Section 4.3) meets the 10 μas stability requirements for the short-term variations of the basic angle *passively*, due to the stable thermal environment at the Lagrange L2 point and to the presence of a flat sunshield shadowing the whole payload module, it is necessary to design and implement a device for monitoring the relative line-of-sight variations of the two viewing directions. Indeed, the required line-of-sight stability (1 μas rms) is quite unusual and the short-term basic angle stability over the satellite revolution period (3 hours) is the only critical parameter so far identified which cannot be properly calibrated by on-ground data processing. A line-of-sight variation of 1 μas rms corresponds to an optical path

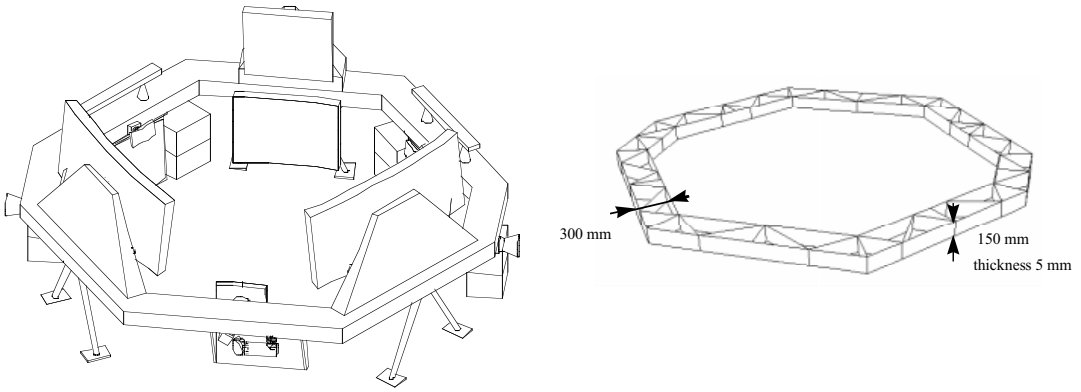


Figure 4.5: Open back optical bench concept. The torus structure which supports all payload mirrors (left) is comprised of struts of silicon-carbide (right).

difference over a distance 0.85 m (along scan on M1) of 4 pm rms or ∼16 pm peak-to-valley. The thermal/mechanical analyses show that a basic angle variation of 1 μas rms corresponds to thermal gradient variations of $\sim$ 25 μK in the payload module torus structure. The corresponding typical displacements of the reflectors are in the range of 20–40 pm. The proposed basic angle monitoring device is based on the following principles. All the mirrors are monolithic, and are made stiff enough so that the line-of-sight fluctuations due to thermo-elastic distortions of the mirrors can be assumed to be negligible (i.e. below 1 μas). The line-of-sight fluctuations of each instrument are therefore related to small rigid body motions of each reflector and of the focal planes, in a given (arbitrary) reference frame common to both viewing directions.

The requirement on the basic angle stability is that it should be within 1 μas rms over the satellite revolution period of 3 hours, or should be at least known with this accuracy. Variations over periods larger than 3 hours can be self-calibrated by the data processing on ground. Given that the payload structure and optics are made of the same SiC material, which provides a low thermal expansion coupled with a high thermal conductivity, and simultaneously good optical and structural properties, the overall payload is therefore athermal to first order, and line-of-sight fluctuations result only from thermal gradient fluctuations within the payload. Monitoring in-orbit should be able to verify that fluctuations are in the range of 1 μas rms, and should also be capable of providing the basic angle fluctuations should their actual value in orbit be for some reason higher than expected. As a starting point, the measurement device should withstand basic angle variations 100 times higher than expected, or 100 μas rms. In that case, the temperature variations are in the range of 2.5 mK, the optical parts displacements are in the range of 2–4 nm, while the optical path difference between the two sub-apertures is ∼1.6 nm. Finally, if we consider two bars located on the torus structure, at the opposite side of the primary reflectors, the relative translation of one bar with respect to the other is also in the range of 2–4 nm, while the relative rotation is found to be of the order of magnitude of the basic angle variations (i.e. 100 μas rms).

The optics manufacturing is not, in practice, over-constrained by the distortion conditions. Indeed, the reflector isostatic mounts must be designed, as for any telescope, for preserving the optical quality during the telescope integration therefore, the isostaticity must be such that a relative displacement of the mounts in the range of 0.1–0.2 mm (typical values) does not degrade the telescope wavefront error (WFE) by more than about $\lambda/25$ rms (the total WFE budget being $\lambda/14$ rms at 500 nm), or numerically WFE $\sim$ 20 nm rms. In that case, a relative mount displacement of 4 nm (i.e. with a margin factor of 100 included) will naturally correspond to a WFE ∼0.4 pm rms, or to a line-of-sight fluctuation $\sim$ 0.1 μas rms.

The proposed method (Figure 4.6) consists of illuminating simultaneously both primary mirrors by a single artificial point source S located on board. The point source is optically at infinite distance (artificial star) and generates two optical beams, one for each viewing direction. The set-up must be such that the relative angle along scan between the two beams is perfectly stable. Each beam propagates through the corresponding telescope and produces an image of S in the telescope focal

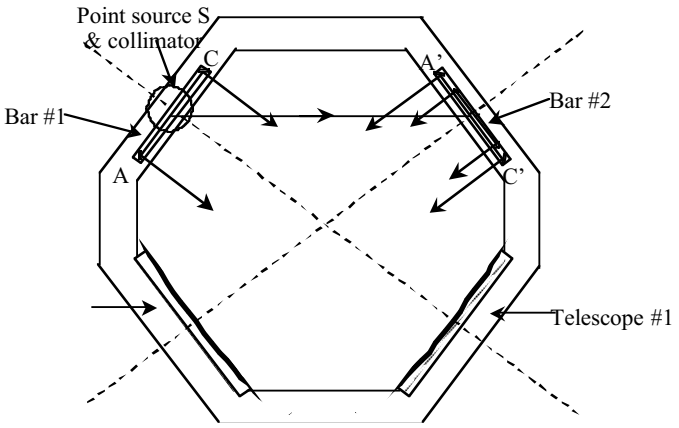


Figure 4.6: Interferometric measurement principle of the basic angle variations.

plane. The system is constructed such that the spatial frequency content of the point source image is similar to that of a star image. In that way, the same (or a similar) image localization process that is used for stars can be applied for the point source image, with the notable difference that the number of collected photons is driven by the laser power and can be made as high as needed. Thus the system performance is intrinsically equivalent to the best accuracy that can be obtained with GAIA on bright stars.

A practical way of generating an image pattern with a high spatial frequency content is to use interferometry, i.e. to illuminate the telescope by two small coherent beams separated by a large distance. For a given wavelength λ, the Airy diameter along scan is $2\lambda/D$, where D is the primary reflector diameter along scan. Let B be the distance between the two small beams (i.e. the baselength) and D' the beam diameter ($D' \ll B$). The two beams will produce two superposed Airy patterns of radius λ/D', with Young's fringes inside the pattern of period λ/B. Therefore, a single fringe is comparable to the telescope Airy disk provided that $B \sim D/2$, i.e. $B \sim 80$ cm. On the other hand, by taking $D = 20$ mm, one gets about 80 fringes in an Airy disk. Since the number of collected photons can be made as high as desired, the localization process accuracy obtained with a single measurement is equivalent to that obtained for GAIA with about 80 measurements on a bright star. This reasoning, combined with the fact that the measurement frequency can also be high, shows that the random measurement error due to photon or detection noise can be made negligible in comparison to that obtained with GAIA on bright stars.

The selected parameters for the basic single interferometeric measurement are: $\lambda \sim 0.8\ \mu$m, $B = 93$ cm, $D' = 20$ mm, pixel size $= 9 \times 27\ \mu$m (as for the main instrument). A line-of-sight variation of 10 μas rms corresponds to an optical path difference of 38 pm rms, and to a fringe motion of 2.4 nm rms. The fringe period in the focal plane is $f\lambda/B = 50\ \mu$m, i.e. one fringe is sampled by 5.6 pixels (approximately as for the main instrument with red stars). The Airy disk produced by the 20 mm subapertures is $2f\lambda/D' = 4$ mm diameter, and contains about 80 fringes. The achievable localisation accuracy is 1 μas rms with a laser diode power of 100 μW.

All the auxiliary optics (mainly small mirrors) are mounted on two SiC bars. The design is such that the measurement is insensitive to rigid body motions of the bars, to temperature variations of the bars, and to point source jitter. Since the measurement relies on differential optical path difference measurements, the laser stability requirement is also not critical ($\sim 1 - 10$ ppm). Therefore, the proposed design is rather robust, since it does not really rely on the actual thermal stability that is achieved for the payload module structure.

Test Results Due to the central importance of the basic angle monitoring concept, a specific experimental verification of the proposed set-up was designed and fabricated at TNO/TPD (Delft), as part of the study contract. Details of the experimental configuration and results are included here, and Figure 4.7 illustrates the quality of the fringes obtained in air.

The detector was a commercial CCD Kodak, operated in MPP mode, with a pixel size of 9 μm, a saturation corresponding to 104 000 e^-, and a dark image noise of $\sim 7\ e^-$ rms. The optical configuration used a baselength of 93 mm, and aperture diameter of 6.25 mm, and a focal length of 6.6 m in air (7.5 m in the 0.5 mbar vacuum, since the focal length is sensitive to the refractive index due to the telescope design and to the high magnification of the secondary reflector). The

fringe period corresponds to 4.6 pixels in air, and 5.3 pixels in vacuum, with one fringe period roughly corresponding, in pixels, to a star image for GAIA. The number of fringes was ~ 15 per Airy radius, or about 30 in total.

Two sets of fringes were produced on the CCD. The fringes were localized for each set and the OPD variations were deduced from the relative fringe displacement. A specific OPD could be generated for one of the bars, by rotating a flat piece of 10 mm BK7 glass working close to normal incidence. The glass is rotated by a piezo, with excellent accuracy on the knowledge of the OPD, because of quasi-normal incidence, of about 2 pm. The second bar could also be rotated to verify the insensitivity of the measurement to the bar rotation. The set-up was made with aluminium for reducing the cost. One measurement consisted of an image acquisition (one image = two sets of fringes), and a second image acquisition by introducing a given OPD (which can be zero) with the OPD generator. The second image was taken about 1.5 s after the first. The procedure could be repeated for quantifying noise and reproducibility, and/or for averaging.

The fringe localization algorithm is relatively simple and robust, based on a correlation technique. The biases are minimized (bias variation negligible) and the algorithm is insensitive to beam overlapping or fringe contrast. The measurement of OPD variations is also insensitive to pixel non-uniformities, because the same pixels are always used. The fringe period and rotation are determined by the algorithm. Since about 3×10^8 photo-electrons are detected, it should be possible to reach $\sim 10 - 12$ pm OPD noise per measurement, corresponding to 1/12000 of a pixel, by averaging few images. The standard deviation over 10 measurements is about 40 pm OPD, still above the expected accuracy due to photon noise and to the algorithm, but by about a factor 4 only. By averaging over 10 measurements, the accuracy is therefore about 13 pm. Since one pixel corresponds to an OPD of 120 nm, and considering that one OPD measurement requires two localizations of fringes, the measured accuracy corresponds to a physical localization of the fringes of 0.7 nm on the detector, i.e. to 1/13000 of a pixel and to 1/69000 of a fringe. Translated to GAIA, 0.7 nm is equivalent to 2.9 μas rms.

Tests were carried out to examine the dependence on the OPD generator inputs, using an OPD of $x/2$, x, $2x$, where $x = 100 \pm 15$ pm. Ten averaged measurements gave OPDs of 37 nm, 88 nm, and 176 nm respectively. The results are therefore perfectly correlated, within the measurement noise and the calibration accuracy of the OPD generator. Although the OPD generator is practically linear, the scale factor between the piezo command and the OPD is somewhat uncertain, mainly due to the measurement accuracy of the glass plate tilt offset. Nevertheless, the OPD generator could be accurately calibrated from the measurements.

Sensitivity to the bar rotation gave excellent results. By rotating the bar of 1 μrad, the measured OPD is less than 400 pm. Since the baselength is 93 mm, the attenuation effect is better than 1/200. The residual OPD is due in fact to the residual mismatch between the input and output baselengths on the bar. The measured value corresponds to a mismatch of 0.5 per cent, within the set-up accuracy. For GAIA, the attenuation factor will naturally be 10 times better, even if the same accuracy on the baselength realisation is retained, simply because the baselength is 10 times larger.

The results in vacuum can be summarized as follows : (i) a thermal stability of below 1 mK was achieved; (ii) the OPD measurements were highly reproducible, with a noise of 50 pm rms during the night for one measurement, and about 200 pm during the day. One fringe corresponds to an OPD of 632.8 nm, so that one pixel corresponds to an OPD variation of 120 nm. The measured noise corresponds to 1/2400 of a pixel, and to 1/12700 of a fringe. When translated to GAIA, the measured noise corresponds to a displacement of 3.8 nm on the detector, i.e. to about 16 μas rms; (iii) there was a reproducible correlation between the OPD introduced and the fringe motion; (iv) insensitivity to the bar rotation was verified: for a 'relatively large' rotation of the bar of 1 μrad, the observed wavefront rotation from OPD measurements was 0.5 per cent of the introduced value, demonstrating a rather efficient filtering, with the residual effect probably due to the input/output baselength adjustment. For GAIA, the corresponding filtering will be about 10 times better with the same baselength accuracy (about 0.5–1 mm) since the baselength is about 10 times larger; (v) an OPD drift of about 23 pm/s was observed, which could not be explained by thermal effects, and is probably attributable to the residual leakage of the chamber or to the temperature increase of the detector (which results in a thermal expansion of the detector, and to a change of the distance between the two fringe patterns) or to both effects. Neither effect will exist for GAIA.

The experiment has also demonstrated another central feature of the GAIA measurement concept, by showing that if biases are neglected (or if they are considered to be properly calibrated), a pattern on the CCD can be localized with an accuracy even better than the mission requirements.

4.2.4 Service Module Mechanical Design

The service module is designed to ensure a stable mechanical and thermal environment for the payload module. This leads to the implementation of a sunshield protecting the instruments from the major perturbation of the Sun. A conical shape has been adopted for the service module to avoid turning shadows on the Sun shield. This conical shape is due to the Sun aspect angle of 55°. The service module is structured around a central hexagon supporting the liquid apogee engine and the tanks. This hexagon is interfacing the launcher interface. The solar array is stowed against the lateral panel of the service module. Each wing is made of two panels. The sunshield, a simple

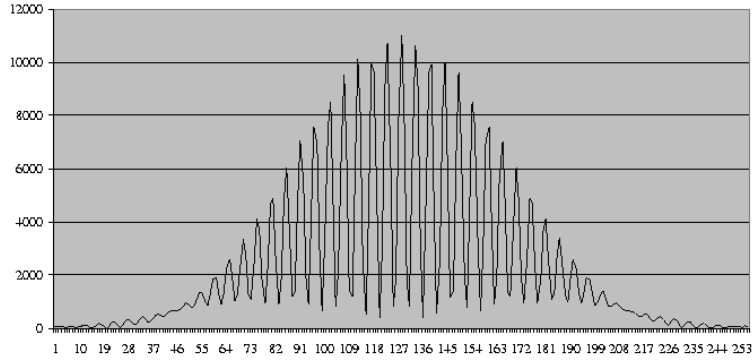


Figure 4.7: Fringes in air resulting from the basic angle monitoring tests carried out at TNO/TPD, Delft (September 1999).

MLI sheet, is deployed together with the solar array. In orbit, it is maintained between the wings. This design respects the service module/payload module modularity.

The major constraints on the service module are provided by the Ariane 5 fairing diameter, the payload module height (which imposes the sunshield diameter), the launcher interface, and the solar array surface (two panels per wing). A perfectly conical shape is not practical for the solar array and for the equipment interface. Flat surfaces are preferred and, after geometric analysis, a hexagonal shape has been adopted, giving a truncated pyramid shape. Out of plane frequencies of the lateral panels are axial modes. To avoid excessively stiff panels and to save mass, shear panels are added. They support the bipods of the payload module, the interface of the hold-down points of the solar array and add inertia to the service module lateral panel. CFRP is used for the shear walls to save mass. The CFRP/Al combination does not generate excessive stresses given the thermal environment. All electronic equipment is located on the lateral panel to improve thermal conduction with the radiator. They are localized with respect with their function power, data management, etc. No optimisation is done in order to have a well-distributed dissipated power. The propellant tanks are dimensioned with respect to the hexagonal tube and to the Ariane 5 prohibited interface area. The pressuring He is split into 2 or 4 tanks. The liquid apogee engine is between the tanks, and a dedicated panel is used for the propulsion system.

4.2.5 Sunshield

The sunshield is designed to provide thermal stability for the payload module performances. It is constituted by an MLI disk which cover the rear side of the solar panels and is deployed between them (Figure 4.8). Its deployment must be simple and it must not perturb the payload module with micro-vibration. Due to its large surface, the mass per square meter must be as low as possible. Its size is limited by the height of the service module. In the launch configuration, the sunshield is simply folded, and thereafter it is deployed with the solar panels. This deployment is undertaken in two steps: release of the first panel with the second panel still folded (the first fold of the MLI is deployed at the same time), followed by deployment of the outer panels (accompanied by the last folds of the MLI). Details of the deployment mechanism are given in the industrial report.

The following classical lay-out is considered for the solar array panels: cover glass: 250 μm; GaAs solar cells: 250 μm; CFRP at 0/90°: 2×60 μm; core structure: 20 mm; CFRP at 0/90°: 2×60 μm. Some masses are added for reinforcement, wiring and MLI (12 sheets for solar array and sunshield), resulting in a surface mass of 5.3 kg/m^2. Considering a typical hinge stiffness of 80 Nm/rad, the first frequency of a solar array wing is at 2.2 Hz ($>$ 0.1 Hz for AOCS). This frequency is out of the bandpass of the AOCS control although, if necessary, damping may be improved by the use of Lavotchkin passive dampers. There are four hold down points per wing linked to the shear panels. For the mechanical analyses only three were considered, the third one being near the antennae panel in the middle of the panel (worst case). Many techniques are feasible for the hold down points, including pyros, Kevlar with thermal cutter, and shape-memory alloy 'Frangibolt'. The criteria will be the load capability, the induced shock and the compatibility with the integrity of the sunshield. The sunshield is a 12-layer MLI which is attached between the solar array wings.

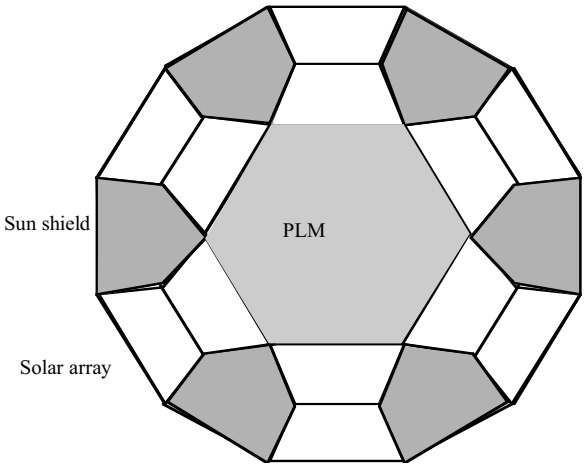


Figure 4.8: Sun shield design, viewed from below. When deployed, the gaps between the solar array panels (unshaded) are filled with sunshield material, providing complete protection of the payload from direct illumination by the Sun.

4.2.6 Mechanical Analysis

The mechanical model of GAIA consists of 31 300 degrees of freedom, and 7305 elements. The primary mirrors are modeled physically, while all other mirrors and electronic units of the payload module are modeled as point masses. Spacecraft equipment is considered to be equally distributed on all side panels. The apogee engine and tanks are modeled as distributed mass on a dedicated panel in the central tube. All the interface points are clamped. The MLI tent is modeled as a non-structural mass on the bars. The doors are modeled as point masses with an out-of-plane frequency at 20 Hz. The radiative panel is modeled as a non-structural mass on the closure panel of the service module. The video processing units are considered to be mounted on the torus. This is a conservative case for axial frequency and stress. Quasi-static load analysis, and dynamic analysis based on the Ariane 5 Users Manual demonstrate adequate margins in both cases.

4.2.7 Gravity and Inertia Effects in Orbit

Assessment has been made of the effects of the satellite rotation on the distortion of the payload structure, and of the gravitational accelerations and gradients at L2 on, for example, the basic angle fluctuations (Safa 2000). Realistic estimates of the satellite masses and mass balancing, toroid and bipod characteristics, and satellite rotation rate have been included.

A constant spin rate has no effect on the structural distortion, and only spin-rate fluctuations need be considered. With the FEEP control, the satellite attitude is maintained within about 2 $\mathrm{mas\,s^{-1}}$. At a spin rate of $\Omega = 5.8 \times 10^{-4}$ $\mathrm{rad\,s^{-1}}$, $\delta\Omega/\Omega \sim 1.7 \times 10^{-5}$, and the resulting flexure and torsion of the toroidal structure results in spin rate fluctuations of the basic angle of $\ll 1\,\mu$arcsec.

At L2, the gravity gradient due to the Earth and Sun amount to about 2.4×10^{-14} $\mathrm{g\,m^{-1}}$ and 0.8×10^{-14} $\mathrm{g\,m^{-1}}$ respectively. The forces acting on the satellite are of order 10^{-10} N, and with realistic values of the Young's Modulus and geometric parameters of the toroidal supporting structure, resulting distortions are of the order of 2×10^{-5} pm, which translates into negligable fluctuations of the telescope line-of-sight of the order of $10^{-5}\,\mu$arcsec.

4.3 Thermal Control

4.3.1 Payload Thermal Control

The payload module opto-mechanical stability is sensitive to residual thermal variations due to the rotation of the satellite, thermal variations of the service module, and the electronics power dissipation fluctuations within the payload module. Since the frequency and the amplitude of these thermal variations vary from one source to the other, the thermal control principle is based on a specific regulation and insulation for each source. The payload module reacts as a thermal cavity composed of the torus + mirrors + electronics. Its time constant (about 5 hours), is around the period of revolution and therefore it has a high intrinsic stability. The natural high emissivity of SiC increases the time constant and improves this stability. The internal cavity is insulated from space by MLI foils. The internal side is black painted for straylight purpose. The external side is VDA coated to minimise thermal exchanges with the sunshield. A part of the payload module electronic power dissipation is directly radiated to space to maintain the required level of temperature inside the payload module cavity and minimise electronics perturbations. This concerns the rear side of focal planes and the associated electronics.

The payload module service module interface is optimised to avoid heat load variations coming from the service module cavity and Sun illumination. To optimise the conductive interface, the extremities of the titanium bipods are maintained at a stabilised temperature by heaters. Due to the low conductivity of the titanium bipods, a control accuracy of $\Delta T = 0.5°\text{C}$ is sufficient for a time step of 20 s. To optimise the radiative interface, the payload module service module interface panel is insulated from the service module internal cavity with MLI. Five blankets are required to ensure an efficient insulation. An intermediate panel, conductively de-coupled from the service module closure panel with permaglass washers, support MLI blankets. An active control of the radiative panels with heaters would have lead to the use of a second order command law with a thermal regulation better than 0.01 K.

The payload module/sunshield interface is designed to avoid heat load variations induced by the Sun and the satellite rotation. Successive Sun lighting of service module radiators panels generates hot spots on the sunshield surface by reflection, due to the different inclination angles in the service module corners (59° instead of 55°). 12-layer MLI is used behind and between the solar arrays to protect the payload module from the Sun and the turning hot spots. The only residual couplings between the sunshield and the payload module are filtered by a double wall of MLI on payload module tent between torus and service module. In order to minimise the radiative couplings between the sunshield and apertures, baffles are defined with MLI and the shutters on payload module tent.

The MLI tent around the payload module payload ensures the thermal stability of the torus assembly. It is constituted with 6-layer MLI fixed on the truss. A double wall is used between torus and service module to filter sunshield perturbations. The external MLI coating is VDA, while the internal side is black painted. The torus assembly temperature is at 180–190 K. The focal plane temperature level is 250 K. Part of focal plane dissipation is evacuated inside the payload module cavity. The other part is evacuated directly towards space. Focal planes are decoupled from the torus with insulated washers. Dissipation of electronic units is directly evacuated towards space. They are decoupled from payload module cavity with MLI. They are fixed on service module structure, and therefore decoupled from torus.

4.3.2 Service Module Thermal Control

The service module thermal control design objectives are to maintain electronic units inside the temperature specifications, and to avoid Sun reflections, hot spots or turning shadows on the payload module cavity towards sunshield and service module interface.

The external sides of service module panels are covered with optical solar reflector material to minimise the temperature inside the service module cavity. The temperature level of units is acceptable with optical solar reflectors, but not with white paint (increase of 45°C with white paint). The internal sides are black painted to decrease the temperature gradient between the side facing the Sun and the others panels. The internal cavity is thermally decoupled from the payload module with MLI blanket (12 layers). The antenna is active when it is facing the Sun, and it tends to be hot. But the rotation of the satellite and the thermal capacity of units decrease and improve the thermal fluctuations during a spin. No Sun reflections and turning shadows are directed toward payload module cavity, except when the Sun rays light the edges of the panel which is opposite to the facing Sun panel (in corners, the angle is equal to 59° instead of 55°). The reflection on the sunshield surface is negligible when service module panels are facing the Sun, due to low radiative thermal exchange. Their impact on payload module cavity stability is reduced with the sunshield MLI concept. With optical solar reflector material the solar reflections on sunshield are less damaging than solar reflections with white paint. The objective of the sunshield is to provide adequate thermal stability for the payload. It is constituted of a 12-layer MLI disk which covers the rear side of the solar

panels and is deployed between them. The rear side of the GaAs solar panels are insulated from cold space. The temperature throught the transfer and operational modes is compatible with classical qualification values. Between solar panels, Kapton coating is required to avoid high temperatures. On the payload module side VDA coating gives the optimised performances for the torus assembly stability.

4.3.3 Thermal Analysis

The thermal model characterises all interfaces between the service module, the payload module and the sunshield, and uses about 200 nodes for the payload module, 100 nodes for the service module, and 100 nodes for the sunshield. The payload module is represented by 50 nodes for the torus, 30 nodes for focal planes and RVS assembly, 10 nodes for bipods, and 100 nodes for MLI. The service module is represented by 72 nodes for the radiative panels (one internal node and one external), 20 nodes for the internal structure (shear walls, cylinder, closure panel), and 7 nodes for the ground MLI. The electronics boxes are not individually modeled. The thermal dissipation and thermal capacity are directly applied on the service module panels. The internal radiative thermal exchanges with black paint are taken into account. The sunshield breakdown is detailed to be more representative from Sun reflections toward the payload module. It is represented by 24 nodes for the solar panels, 24 nodes for MLI under the solar cells panels, and 2×30 nodes for the MLI between solar panels (payload module and Sun side). Details of the assumptions used for the thermal modeling of both the payload and the service module (thermal dissipation of the electronic units, thermal conductivity and thermal capacity of the various materials, equivalent conductivity and and radiative coupling of the multi-layer insulation material, absorptivity and emissivity of the solar arrays, etc) are given in the industrial report.

4.3.4 Thermally-Induced Variations of the Basic Angle

The primary and most crucial aim of the payload thermal analysis is to assess the resulting basic angle fluctuations. The payload module stability has been evaluated by combining the following detailed models: (i) a service module/sunshield thermal model: this characterises the radiative interface between the service module and the payload module via the sunshield and assesses the fluctuations introduced by the spin motion on the payload module; (ii) a payload module/sunshield thermal model: this characterises the payload module and sunshield thermal behaviour and assesses the fluctuations introduced by the focal plane power variations. The output of the service module/sunshield model is also introduced in this model as inputs; (iii) a telescope optical model (Code V): this provides the relationship between the line-of-sight of each telescope and the mirror displacements (translations and tilts). Two field values are considered: the field centre (0°) and 0°.3 along scan which takes into account the variation of mirror curvature; (iv) a payload module/mechanical model: the output of the optical model (line-of-sight/displacement matrix) and of the thermal model (temperature map) are both included in the Nastran Finite Element Model, and the basic angle variations are then derived for each load case.

35 thermal nodes were used for the torus and the mirrors. For each thermal node, a conductive interpolation is performed in the mechanical model for a variation of 1 K, all the other nodes staying constant. This gives a contribution to the line-of-sight shifts. Finally, the basic angle variations is calculated by linear combination of the contributions for each time step. Only temperature variations around the average value need to be considered. Four load cases have been considered for the torus assembly stability: (i) satellite rotation to get the impact of the hot spots turning on the sunshield; (ii) thermal regulation of bipods associated with the satellite rotation. The time step used for calculation is 20 s, with a ΔT threshold of ± 0.25 K; (iii) efficiency of the service module/payload module thermal decoupling (MLI shield at payload module/service module interface), associated with the satellite rotation; (iv) fluctuations of power dissipation in the focal planes. Load cases are defined below for evaluating the sensitivity of the basic angle variations to such fluctuations. The fluctuations are co-phased for both focal planes to provide a worst case estimate.

The effect of the power variation in the focal plane electronics is expected to be negligible since no low-frequency variation over 3 hours is expected. However, for evaluating the sensitivity of the basic angle variation to such power fluctuation, three somewhat arbitrary load cases at different frequencies have been considered: (i) a power fluctuation due to the regulation of the power source. Typical stabilities are in the range of 0.1–0.01 per cent due to the use of adequate power regulators. There is no identified reason to have such power fluctuations at the satellite spin period (3 hour). It should be higher frequencies (noise) or lower (ageing). Nevertheless, a load case has been

Table 4.2: Basic angle variations resulting from various thermal fluctuations.

Period	Level (per cent)	Thermal variation (μK)	Basic angle variation (μas rms)
3 hours	0.02	20	3.2
20 seconds	1	4.6	0.2
1 second	5	1.1	negligible

defined by a power fluctuation of 0.02 per cent with a period of 3 hours (worst case); (ii) the satellite, and especially the payload module electronics, have a permanent and stabilized operation mode over a spin period of 3 hours. The highest frequency considered is related to the TDI period, that is 1 s. Again, for evaluating the sensitivity of the payload module stability to such fluctuations, a load case defined by 5 per cent power fluctuation at 1 Hz has been considered; (iii) finally, an intermediate load case is considered, where the power fluctuation is 1 per cent over 20 s, corresponding to the time needed for a star to cross the entire focal plane. Video processing units are coupled to the focal planes using cables with very low thermal conductivity (< 0.02 W/K). Calculations show that its impact is negligible with variations of about 20 per cent. Nevertheless, the extremity of this cable can be maintained at a constant temperature ($\pm 0.5^\circ$C) near the video processing unit.

The results provided in Table 4.2 confirm the robustness of the design with respect to focal plane power fluctuations, since significant power fluctuations are strongly filtered by the system inertia as soon as their time scale is below a few tens of seconds. The results show that under reasonable assumptions on service module thermal control, the basic angle stability is driven by the torus temperature variations and dominated by the service module torus coupling. This is not surprising since the period of this residual effect is naturally matched to the satellite revolution period. The sunshield torus coupling is efficiently minimized with a double wall of MLI and baffles on payload module tent, while focal plane power variations are efficiently filtered by the payload module inertia. The time scale of the thermal fluctuations plays a fundamental role for the payload module stability: large-scale thermal fluctuations with periods well above 3 hours do not harm the final performance, while 'fast' fluctuations at a period much smaller than 3 hours are naturally filtered by the payload module inertia.

The results obtained with the selected opto-mechanical concept show that a passive basic angle stability of 6.4 μas rms is attainable, while the assigned design goal was 10 μas rms. Further improvements can probably be obtained either by increasing the torus inertia (balance performance/mass) and/or by performing a detailed thermal optimization of the service module to reduce the temperature fluctuations inside the service module (the service module/torus coupling is the dominant contributor to the basic angle fluctuations).

4.4 Propulsion and Attitude Control

4.4.1 Launch and Early Orbit

Details of the launcher and launch strategy, injection into the transfer orbit, description of the transfer orbit, and a description of the operational orbit are given in Section 5.1. The proposed orbit is a Lissajous orbit of limited elongation (400 000 km $\times$ 100 000 km typically), which ensures a Sun-Satellite-Earth angle lower than 15° and a visibility period with Perth exceeding 8 hours per day. One or two manoeuvres are required for the injection from the transfer orbit into the L2 operational orbit. The selected orbit is free of eclipses for at least 5 years, and it is possible to extend this further using a specific eclipse avoidance manoeuvre performed during the nominal observation phase. After this manoeuvre, a time of at least 6 years without eclipse is available for observations. The preliminary ΔV budget assessed for these two sets of manoeuvre (injection into L2 orbit plus eclipse avoidance) is close to 200 m s^{-1}. The total ΔV budget required for the injection of the spacecraft from geostationary transfer orbit into the transfer orbit, for the transfer orbit itself, for the final injection into the L2 operational orbit and for the maintenance of this orbit (including an eclipse avoidance manoeuvre) is presented in Table 4.3. The ΔV budgets for the attitude control during the transfer phase and the operational phase, which are linked to the spacecraft design and attitude control strategy, are given in Section 4.4.8.

Table 4.3: ΔV budget outcome of the ESA/ESOC mission analyses excluding the ΔV budget required for the spacecraft attitude control during the transfer and operational phases.

Phase/Manoeuvre	Budget (m s^{-1})
Orbit Acquisition:	
– apogee raising manoeuvre	800
– gravity loss of perigee manoeuvre	52
– transfer mid-course corrections	30
– L2 orbit injection + eclipse avoidance	200
Sub-total	1082
Total (including 10 per cent system margin)	1190
Orbit Maintenance:	
– orbit maintenance (over 6 years)	18
Total (including 50 per cent system margin)	27

4.4.2 Attitude Control Overview and Requirements

This section summarises the requirements that apply to the spacecraft attitude control and measurement subsystem. The requirements apply to the transfer and/or to the operational phase at L2. Particular attention is given to the transition between these two phases, which has been identified as one of the design drivers of the attitude control subsystem.

The overall attitude control and measurement system architecture is presented in Figure 4.9. The following elements are presented below in further detail: (a) acquisition mode: star sensor (may be used for spin rate reduction) + Sun acquisition sensors (for rate reduction and Sun acquisition) + X-axis gyroscope + 10 N bi-propellant thrusters; (b) transfer mode: Sun acquisition sensors + X-axis gyroscope + 10 N thrusters + 400 N bi-propellant engine; (c) transition mode: star sensor + payload instrument sky-mappers + 10 N bi-propellant thrusters + FEEP thrusters; (d) operational mode: star sensor + payload instrument sky-mappers + FEEP thrusters; (e) slew mode: star sensor + 10 N bi-propellant thrusters; (f) safe mode: attitude anomaly detector + Sun acquisition sensors + gyroscope + 10 N bi-propellant thrusters.

The ACMS controls all orbit correction manoeuvres from launcher separation to injection into the final operational orbit around L2. Assuming a dual launch with a standard Ariane 5 into a standard geostationary transfer orbit, the total ΔV budget for these corrections is estimated as 1082 m s^{-1}, split as shown in Table 4.3.

It will be possible to carry out preliminary scientific observations towards the end of the transfer phase, when the Sun-spacecraft-Earth angle is low enough ($< 15°$) to make science communications with ground possible. Payload instruments are nominally switched off during the transfer phase. It means that they cannot be used as sensors for the ACMS subsystem. Consequently, the ACMS design will be based on dedicated ACMS sensors. The ACMS subsystem must satisfy three complementary sets of requirements: (a) those deriving from the spacecraft and payload design and which are necessary to meet the specified spacecraft measurement performance requirements; (b) those which come from the data reduction process on ground (*a posteriori* processing of the spacecraft data) and which allow to reach the ultimate accuracy on star position and velocity; (c) those which are linked to the maintenance of the operational orbit during the operational phase.

A stringent stability of the instrument line-of-sight with respect to the nominal scanning law is required, in order to avoid bluring during each sub-field of view integration time period of 0.9 s. A budget of one sixth of a pixel (peak-to-valley) is allocated to this line-of-sight motion. For a pixel size of 37 mas, it leads to a maximum motion of 6 mas in 1 s, or an rms motion of 2 mas over 1 s. In order to ensure a perfect synchronisation between the scan motion (i.e. star image motion at focal plane level) and the charge transfer operation within the CCDs (TDI operation) during each integration period, the spin rate must be stable during this integration period. A budget

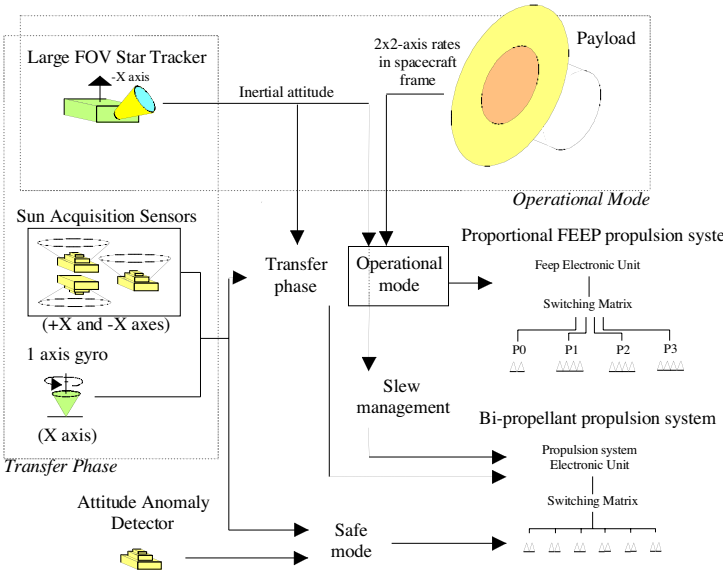


Figure 4.9: Architecture of the attitude control measurement subsystem (ACMS) and the reaction control subsystem (RCS).

equivalent to a shift of one sixth of a pixel (peak-to-valley) over one integration period is allocated to the rate. For a pixel size of 37 mas, it leads to a relative rate error of 6 mas s^{-1} peak-to-valley or 2 mas s^{-1} rms. Data reduction requirements are achieved with the proposed design.

In addition to the performance requirements specified above, the ACMS subsystem must be able to provide the necessary orbit maintenance manoeuvres during the operational phase. The orbit correction need is estimated to 3 m s^{-1} per year for gravity effect compensation and to 5 m s^{-1} per year for solar pressure effect compensation, resulting in a total of 48 m s^{-1} for a 6 years extended lifetime. During the operational phase, the two astrometric instruments are themselves used to measure and control the spin rate, via the detection of stars and the measurement of their displacement directly within the instrument field of view (sky mapper function). Consequently, the transition between the transfer phase (which uses dedicated ACMS sensors) and the operational phase (which uses the astrometric instruments as attitude sensors) will be designed such that it is possible, at the end of the transition, to acquire stars with the astrometric instruments. The main requirement which applies to the ACMS subsystem for the transition phase is so to be able to stabilise the spacecraft to its nominal scan rate with an accuracy of about 1 arcsec s^{-1} using only standard ACMS equipment, and this performance will be maintained for a time period which is long enough to allow the instrument acquisition.

This transition phase between the transfer phase and the operational phase, is a design driver for the ACMS measurement system, for both the transfer and the operational phases. Two options were evaluated during the study: with and without a dedicated star sensor. The solution with a star sensor has been adopted since it provides simplification at instrument level (no star catalogue, no star pattern recognition algorithm), at system level (lower number of equipment units), and at mission level (no specific calibration manoeuvre). All operations of the transition phase can be performed using a star sensor with a medium or large field of view. Sun sensors, as mentioned above, are kept on-board in any case at least for safety reasons, but the acquisition of the operational position and the payload acquisition during the transition phase can now be done using the star sensor. The spin stabilisation of the spacecraft during the transfer phase (which is the most robust design for the transfer phase) requires only standard Sun acquisition sensors completed with a simple 1-axis low-cost gyro on the spin axis.

Since the attitude pointing accuracy required during the transition phase is compatible with the performance of such a star sensor, there is now no need of using the payload instruments, except for rate measurement. This concept is consequently significantly simpler. Moreover, during the operational phase, an hybridisation can be performed through a Kalman filter between attitude information provided by the star sensor and rate information provided by the instrument. Then, a temporary lack of star in the instrument field of view can be compensated by propagating the rate information. The rate measurement accuracy is very high compared to the required absolute pointing error, and the unavailability of star sensor information can last about 20 hours still staying in the pointing accuracy specification. The star sensor has not been selected yet but star sensors with comparabale performance exist already, while sensors with improved performance and capabilities are the subject of ongoing qualification activities, and should be available very soon.

A detailed evaluation of the relative merits of using spin-stabilisation or 3-axis stabilisation during the transfer phase has been carried out, taking into account the overall system complexity, the required thruster configuration, and the necessary propellant consumption. The spin-stabilisation concept is adopted as baseline. The spacecraft inertia are fully compliant of a passive control (passive nutation damping) with an inertia ratio of 2. The required spin rate is 5 rpm around the X axis, to limit the nutation amplitude to 1° during the main engine (400 N) firing. This spin rate can be lower during the Sun-pointing phase, in order just to stabilise the spacecraft.

4.4.3 Transfer Phase Propulsion System

A wide variety of propulsion technologies, either off-the-shelf or under development, are available. Selection is based partly on the specific impulse (I_{sp}) of each system, which gives an indication of the efficiency of the thrust, being inversely proportional to the propellant mass required to realise a given acceleration. The apogee raising manoeuvre and the further orbit and attitude corrections require the implementation of a high thrust engine (400 N) together with a set of small thrusters (10 N). Several technical and performance issues have been considered when designing the propulsion system: (a) the system must produce 3-axis torques for operational attitude acquisition and a pure force, at least in one direction, for the apogee raising and orbit correction injection manoeuvres; (b) the system should minimise the time needed for the correction manoeuvres (Robbin's penalty impact, and crossing time of the van Allen radiation belts); (c) using the high thrust engine for the final L2 orbit injection, after a long stand-by period (i.e. transfer phase), adds complexity in the propulsion system design.

The solution which is proposed to bypass this latter issue is to use the 10 N thrusters for the L2 orbit injection manoeuvres. Such a solution is compatible with the L2 approach trajectory at the price of a small propellant mass penalty compared to the use of the 400 N engine. On the other hand, the initial perigee manoeuvres using only four 10 N thrusters would require at least 12 transfer orbits to limit the Robbin's penalty to 10 per cent of the required correction. This would significantly increase the time spent in the van Allen radiation belts.

The specific impulse of existing technologies varies between 60 s (cold gas) and 6000 s (electric/electrostatic propulsion). The drawback is that a high I_{sp} generally corresponds to a very low thrust level capability. The transfer phase propulsion system selection trade-off consists on the one hand to look for the highest specific impulse (in order to reduce the propellant mass, since the total correction is very large, 1190 m s^{-1}), and on the other hand to look for a high thrust level capability in order to minimise the losses during perigee burns (or 'Robbin's penalty') and the transfer phase duration. From the choice of cold gas, solid motor, electric propulsion with high or low specific impulse, or classical liquid propulsion systems, the unified bi-propellant MMH/NTO liquid propulsion solution has been selected. Its implementation, although well mastered, is more complex than the mono-propellant system but this additional complexity is compensated by the propellant mass saving provided by the improved specific impulse (mass saving greater than 200 kg with an $I_{\mathrm{sp}} = 290$ s against 230 s for hydrazine). It is also more attractive than the dual-mode solution both because of the propellant mass saving (greater than 100 kg) and because of the European background to the concept (dual-mode is a US technology).

4.4.4 Observation Phase Propulsion System

The attitude control system must meet stringent requirements on the instrument line-of-sight stability, as well as on the spin-axis pointing and rate measurements during the operational phase, whilst minimising the propellant mass. Impulsive solutions, as used on Hipparcos using 20 mN cold gas nitrogen thrusters, are based on a limit cycle control strategy, where only the pulse frequency and duration is controllable, not the thrust amplitude. They would result in high rate disturbance levels during actuation, loss of scientific measurements during the attitude correction period, and would also require re-tsrating of the star acquisition process with the sky mappers after each thruster pulse. At the same time, with $I_{\rm sp} = 70$ s, the corresponding propellant budget would be about 220 kg. Other impulsive thruster technologies, such as Teflon Pulsed Plasma Thrusters, may be considered but still the propellant mass budget is high (at least 60 kg).

In order to achieve the lowest possible level of disturbance, compatible with the stringent pointing and rate stability requirement, a proportional control thruster system is adopted. Thrusters offering a full range thrust modulation capability are quite suitable for GAIA, since they allow an accurate compensation of the disturbances over a wide frequency range. No high frequency dynamic disturbance is generated, and induced jitters can be made negligible. Since the disturbance levels encountered will be very small, thruster technologies with low thrust level are acceptable. The two most promising candidates are proportional He or N thrusters ($I_{\rm sp} = 70$ s) and 'field emission electric propulsion' thrusters (FEEP, $I_{\rm sp} = 6000$ s).

The FEEP solution has been selected as baseline due to its very high specific impulse which leads to a global propellant budget of only 2.6 kg for a 6 year mission. A thruster configuration which produces pure torques and pure forces in all directions has been identified, which also limits plume interactions between the thrusts and the sunshield. Three clusters of two (redundant) thrusters are accommodated under the sunshield, and one thruster is accommodated on the service module. In this preliminary design, one Caesium tank is associated with each cluster and the electronics is centralised. Orbit maintenance will be performed with the same set of FEEP thrusters, and at the same time as the attitude and rate control, since the system is also able to produce pure forces. Part of the correction can be computed on-board using a solar pressure force estimator, the other part will be computed on-ground and sent to the spacecraft.

In the FEEP thrusters Caesium flows through a 1 micron slit, forming a cylindrical surface at the exit of the slit. Under a strong electric field, the free surface of the liquid metal approaches a situation of local instability. A series of cusps (Taylor cones) are created at the surface of the liquid. The local electric field at the tip increases as the radius of curvature of the cusp decreases. When the electric field reaches 10^9 V/m, the atom at the tip spontaneously ionises and an ion jet is extracted. The quantity of Caesium to be expelled has to be heated to 28.6°C. It flows to the emitter by capillary and drag forces due to the electric field. A classical attitude control system using impulsive thrusters is based on a modulation unit, the role of which is to realise the commanded torque by modulating the thruster pulse duration. As a consequence, the commanded torque can only be realised when its value reaches a threshold compatible with the thruster 'minimum impulse bit'. In order to minimise the number of thruster actuation, the attitude pointing is maintained below a selected threshold and corrections are performed only when the threshold is reached. This is called a limit cycle. This is no longer the case when using proportional thrusters such as FEEP since no thrust modulation nor limit cycle are required. The control of the scan law is performed quasi-continuously by commanding a continuous torque profile with values comprised between 0 and a few mN (typically 1 mN). The scanning law will be defined in the inertial system ICRS, and controlled by the star sensor and instrument measurements, using a simple proportional derivative controller to stabilise the system with an efficient low-pass filter (such as a Cauer filter) to limit the measurement noise transmission. Using a command matrix, the torque commanded by the controller will be transformed into thrust levels commanded to each thruster.

4.4.5 Disturbance Torques

External perturbing sources are those induced by the Sun and the gravitational effects of nearby planets. The latter effect concerns only orbit drift and is estimated to amount to approximately 3 m s^{-1} per year. Solar pressures are created by solar radiation and by the solar wind, although the latter is 2–3 orders of magnitude smaller than the former and has not been considered further. Its effect is twofold: (i) it creates a force applied to the centre of pressure which makes the spacecraft orbit drift slowly. The force contains a constant part in the spacecraft frame along the spin axis

estimated to 180 μN, and a harmonic part at the 3-hour spin period in the sunshield plane estimated at approximately 120 μN. The total drift effect is estimated to amount to about 31 m s^{-1} over 6 years; (ii) since the centre of pressure is not identical to the centre of gravity of the spacecraft, it induces a torque which contains a constant part and a harmonic part at the 3-hour spin period. The solar pressure is subject to low-frequency variations which are modeled by an incoherent noise superimposed with the (seismological) 5 minute oscillations.

Although there are no mechanisms on-board activated during star observations (apart from the occasional small activation of the radial velocity aligment system), disturbances are injected into the instrument line-of-sight by a number of internal sources: by the propulsion system itself, the attitude and rate measurement system, and various other sources (plume impingement, centre of gravity variation, thermo-elastic effects, calibration residue). The following effects have been accounted for: (i) thruster disturbances: disturbances produced by the FEEP proportional system include a white noise below the actuation frequency (corresponding to the dispersion noise), a quantification noise which corresponds to the smallest possible thrust variation, and a sampling noise induced by the thrust amplitude refreshing frequency; (ii) measurement system disturbances: all measurement sensor errors are transmitted to the spacecraft attitude through the control bandwidth of the ACMS, the actuation system and the spacecraft dynamics behaviour. It concerns the star sensor 'noise equivalent angle', the star sensor random bias and the instrument (sky mappers) rate measurement noise. A first filtering of the measurement noise is performed by the attitude estimation filter (Kalman-like filter), a second one by the Cauer filter (high-frequency rejection), a third one by the controller (depending on its bandwidth) and a fourth one by the spacecraft dynamics; (iii) plume effect due to the interaction between particles emitted by the thrusters and exposed spacecraft surfaces such as the sunshield and the solar panels: these effects may be minimised using a specific configuration of the FEEP thrusters, ensuring a thrust direction separated by 45° from both the sunshield and the payload module, the FEEP ejection half-cone angle being 40°; (iv) thermo-elastic effects due to the 3-hour spin motion: this induces a variable misalignment at the spin period between the star sensor frame and the instrument frame. However it is expected to be much smaller than the pointing specification since both items are mounted on the same optical bench which is designed and controlled to ensure a 10 μas basic angle stability; (v) centre of gravity variation due to potential liquid motion on-board: this will be minimized by the use of dedicated trapping devices inside propellant tanks. A sponge, some blades and one or more surface tension screens will be installed in each tank. These are well-known methods already used on-board telecommunication satellites such as the Eurostar platform. Since capillary forces are dominant during the operational phase, the liquid will be trapped in a position that minimises its free surface, that is along and inside trapping devices. A specific acceleration manoeuvre using 10 N thrusters will be performed in order to set up liquids before the FEEP system is switched on and the final operational attitude acquisition is done; (vi) calibration error between the star sensor frame and the instrument line-of-sight: this has a direct impact on pointing accuracy. The initial misalignment bias will be identified via a measurement and comparison campaign between star sensor and instrument measurements. It is expected that most of the star sensor noise and random bias will be filtered and will have a negligible impact on the misalignment calibration. Effects of micrometeoroid impacts on the satellite attitude have been assessed based on the Hipparcos analysis and results (ESA SP-1200, Vol. 2, pp123-126) and are, similarly, not expected to degrade the mission performance.

4.4.6 Pointing and Stability in Operational Mode

Two approaches have been used to assess the ACMS performance at system level, the first involving a frequency analysis incorporating the effects of each identified perturbing source, the second based on time simulation. The preliminary performance assessment, using simplified noise and control models, demonstrate the capability of the proposed system to reach the specified performance (Figure 4.10). Since the control system design is driven by the solar pressure disturbance torques, the control bandwidth must be selected to control efficiently these disturbances in order to meet the pointing stability requirement. However the direct impact of an increase of the control bandwidth is also a decrease of the stability performance because the control itself introduces noise into the instrument line-of-sight. Another way of optimising the system would be to decrease the solar torques by decreasing the lever-arm between the satellite centre of gravity and centre of pressure. This could be done by a small inclination of the sunshield. It would then be possible to decrease the control bandwidth and the induced relative rate error.

The frequency approach involves computing the transfer function between the disturbance sources (measurement noise and external torques) and the instrument line of sight, and to apply these transfer functions to each source. The integration of the obtained power spectral density gives an evaluation of the system performance. The drawback of the method is that it applies only to noise modeling. Harmonics, constant torques and non linear contributors (such as sampling or quantification effects) are not modeled. Their impact must be evaluated and added to the previous estimation. The time simulations approach has been used to consolidate the frequency analysis by including non-linear effects such as sampling and quantification effects. A MATRIX-X simulator has been developed and includes

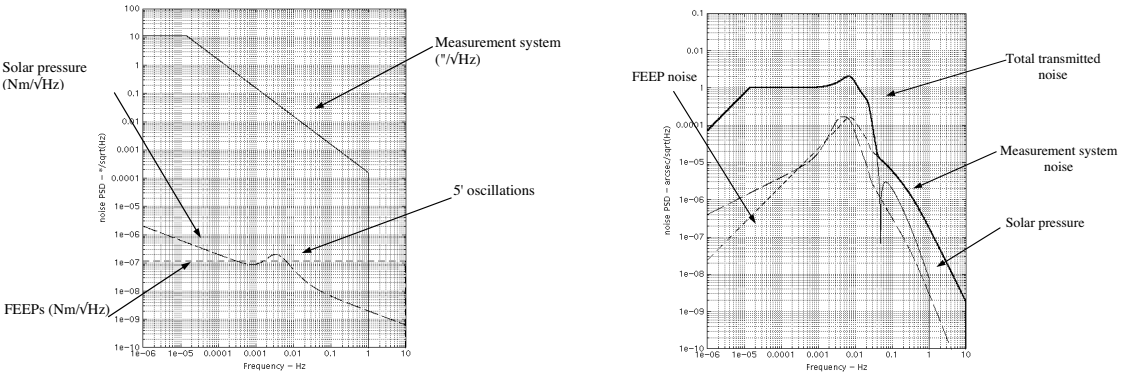


Figure 4.10: Results of the pointing stability analysis: (a, left) Noise power spectral density of main contributions. (b, right) Noise contribution to instrument line-of-sight stability.

all the previous disturbances as well as a dynamic model of the satellite, a Kalman filter model for the attitude measurement and a proportional derivative controller. Results are consistent with the previous budget. They confirm the results of the frequency approach and show that the quantisation effect has a small impact on the system performance

Table 4.4: Pointing error, stability, and restitution budget.

Contributors	Absolute pointing error (mas)	Relative pointing error over 1 s (mas)	Pointing restitution error (μas)
Measurement system noise	59	0.2	1.4
Star tracker random bias	25000	2.2	0
FEEP noise	0.4	0.015	4.9
Solar pressure noise	0.45	0.012	0.4
Harmonic torques	322	0.2	0
Constant torques	1931	0	0
Total (rms sum)	25.1 arcsec	2.2	5.1
Specification	5 arcmin	2	1000

4.4.7 Slewing Capability

Some large-angle slewing capability must be maintained during nominal scientific observations, possibly for scanning law adjustments, but more importantly as a transition to safe mode. The FEEP based reaction control subsystem offers a very low capacity for large manoeuvres because of the very low thrust level. Assuming a preliminary 2 mNm capacity (1 mN thrusters, 2 m lever-arm) a 55° change of orientation, required to reach a 'safe' Sun-pointing position from the nominal scanning law position, would take about 40 min; probably too long for a safe mode. Unexpected events such as micro-meteoroid impacts would also need to be rapidly compensated. The solution adopted is to use the 10 N thrusters already implemented within the spacecraft for the transfer phase. These thrusters are still available during the operational phase and can provide adequate thrusts, providing that a minimum quantity of propellant is kept on board for such purpose, and that sloshing effects are suppressed.

One way of minimising the slewing time would be to increase the maximum thrust level of the FEEPs. Thrusters with thrusts up to 8 mN should be available at the time of the GAIA mission. But the drawback of such an approach

is a significant increase of the electrical power requirement 60 W/mN. In safe mode, the spacecraft attitude must remain stable over a long period of time without any intervention from Earth. The most attractive and classical solution on such a distant orbit is a spin stabilised Sun-pointed safe mode. Once the Sun has been acquired, the spacecraft is spun about the Sun direction to ensure its stability. This manoeuvre is performed using the X-axis gyroscope already implemented for the transfer phase. Only equipment not used during the operational mode is used for the safe mode. The gyroscope lifetime is not a problem since it is only used during the transfer phase. For a higher reliability, solid-state and still low cost technologies such as a Ring Laser Gyroscope or a Fibre Optic Gyroscope can also be considered. Communications are ensured in safe mode via the omni-directional low-gain X-band antenna.

4.4.8 Propellant Budget

The total spacecraft propellant budget is summarised as follows: (i) orbit correction: for a spacecraft launch mass of close to 3 tons (see Section 4.8) and a total orbit injection correction ΔV requirement of 1190 m s^{-1}, 983 kg of propellant are required, that is about one third of the launch mass. This can be considerably reduced (to typically 200 kg) if the launcher is able to directly inject the spacecraft on the transfer trajectory to L2 (restartable Ariane 5); (ii) transfer phase manoeuvres: 12 spin or de-spin manoeuvres are typically required during the transfer phase, requiring a total propellant budget of 21 kg. Spin axis changes of orientation of 60° (9 manoeuvres and 3 kg budget) are also required for orbit corrections; (iii) attitude control during operational phase: the Caesium propellant budget for the FEEP thrusters in order to compensate for the Sun radiation pressure disturbing torques has been estimated, using time simulations, to 0.0005 mg s^{-1}, or 0.1 kg over 6 years. A supplementary 0.1 kg is added for scanning-law control; (iv) orbit maintenance during operational phase: the orbit maintenance includes gravity disturbance and solar pressure force compensation. The former effect is estimated at 3 m s^{-1} per year, the latter at about 5 m s^{-1} per year. The total requirement over 6 years is 48 m s^{-1} which corresponds to 1.6 kg of Caesium propellant. In conclusion, the total propellant budget for the bi-propellant system is 1007 kg (including 10 per cent margin) and 2.7 kg for the FEEP-based system (including 50 per cent margin).

4.5 Data Acquisition and Processing

The spacecraft electrical architecture may be split into two main parts: (i) a dedicated science data chain, attached to the payload, which includes all the units required for the acquisition, processing, formatting and transmission of the science data; (ii) a set of equipment units providing all the general services to the spacecraft and payload (power generation, regulation and distribution, telemetry and telecommand, attitude control, thermal control, etc.). Most of the novelties and critical areas are located within the science data chain, because of the complexity of the focal planes and of the quantity of data to be acquired and processed. Details of the architecture of the science data handling is given hereafter, with a description of the more general spacecraft, which are based on a more classical approach, given in Section 4.6. The electrical functions and the processing required on-board for the sequencing of the instrument focal planes and for the acquisition, discrimination and formatting of the observation data are detailed in Section 3.6. An overview of the corresponding science data chain functional architecture is presented in Figure 4.11. The science data handling chain consists of:

(a) Focal Plane Assemblies: there is one focal plane array per astrometric instrument (two in total, so-called Astro-1 and Astro-2), one for the Radial Velocity Spectrometer (RVS) and one for the Medium-Band Photometer (MBP). Each focal plane array includes CCD arrays and front end electronics. The latter includes the video preamplifiers as well as all the necessary bias and voltage filters and clock drivers which have to be implemented close to the detectors. The focal plane array of each astrometric instrument includes about 250 CCDs, corresponding to about 300 video chains. The detailed architecture of these focal plane arrays (electrical, mechanical and thermal) is consequently a major issue (mechanically buttable CCD devices, packaging for the front end electronics, minimum power dissipation) and will require further investigations in future studies.

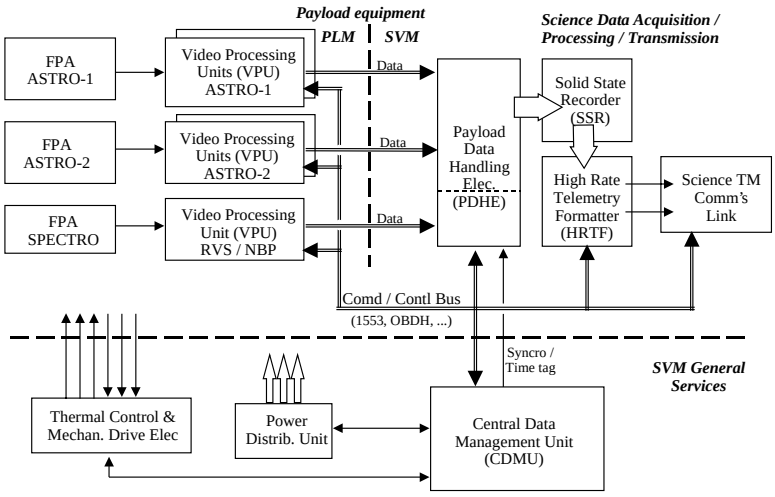


Figure 4.11: Payload module electrical architecture. Each astrometric focal plane assembly includes about 250 CCD arrays. An equivalent number of video chains are implemented within the video processing units. Dedicated data handling units, including a 100 Gbits solid-state recorder, acquire and process the data produced by the instruments. A dedicated X band telemetry link will transmit these data to ground.

The focal planes are not redundant, although the electrical architecture is organised in such a way that a local failure (on one CCD or one video chain) does not impact the entire focal plane array but only one or a few other CCDs. A similar graceful degradation concept may be also apply to the RVS focal plane (which is split in three CCD sub-arrays) and to the MBP focal plane (which is provisionally shared between two CCDs, each one including 3 spectral bands and a potentially redundant sky mapper).

A preliminary sizing of each astrometric focal plane gives a size of $740 \times 680 \times 260$ mm^3, a mass of 40 kg [CCD (11 kg) + front end electronics (9 kg) + structure and thermal (20 kg)], a power dissipation of 170 W, and approximately 2500 interface cables with the video processing unit. The Astro focal plane arrays are accommodated on the payload module optical bench (SiC torus), with a conductive decoupling between the focal plane array and the torus. The power dissipated by the focal plane array is partly radiated into the payload module optical cavity and partly toward cold space, making use of a dedicated radiative area at the back of the focal plane array. The electrical harness will first run from the focal plane array onto the optical bench. It will be then split into multiple flexible bundles connected to the video processing units.

(b) Video Processing Units: The video processing units include all the video chains up to the digitisation stage, the data discrimination function (star detection and discrimination by mean of a fixed programmable threshold), the localisation and datation of the detected events, the measurement of the scan rate (data extracted from the sky mapper field) and the transmission of these data to the spacecraft central computer for ACMS purpose, the multiplexing of the data and their transmission to the data handling and processing unit. The units also include the necessary secondary sequencers and secondary power supplies. A very stable power dissipation (over short and medium timescales) is required at the level of the focal plane array which requires high-performance voltage converters at video processing unit level, with a secondary voltage stability better than 0.01 per cent over a few hours, which should be achievable because of a stable operating temperature (that of the optical cavity) and a stable sequencing mode. The complexity of the Astro video processing units is driven by the number of video chains to be implemented (about 300 for each astrometric instrument) and by the associated sequencing (data detection and discrimination,

windowing mode control, etc.). Despite the foreseen use of hybrids and ASICs to improve the level of integration while minimising the power dissipation, four video processing unit boxes per astrometric focal plane are necessary for the hardware accommodation. Preliminary sizing of the 4 units per astrometric instrument gives a size of $400 \times 300 \times 250$ mm^3, a mass of 15 kg, and a power dissipation of 90 W. They are accommodated on brackets attached to the service module primary structure, close to the astrometric focal plane arrays. The dissipated power is radiated toward space via radiative areas embedded in the payload module secondary structure but thermally disconnected from it.

(c) Payload Data Handling Unit: The payload data handling unit collects the data provided by each video processing unit, and multiplexes and compress these data. They are completed with ancillary data coming from the payload (temperature, datation, threshold or gain level, etc.) or the spacecraft (attitude data, etc.) and then either stored into the solid-state recorder or directly transmitted to ground via the high-rate telemetry formatter and the science telemetry down-link RF subsystem. The payload data handling unit also includes the primary sequencer (which drives all the video processing units) and the primary power converters. A possible data compression scheme (compression algorithm and compressor architecture) to be applied to the data produced by the astrometric fields and by the RVS is not defined yet. A conservative compression ratio of 2 (lossless compression) has been considered which is already achievable with existing hardware. The processing rate remains moderate (about 10 000 stars per second on average) due to a preliminary discrimination of the data made at video processing unit level. Consequently, the payload data handling unit is not considered as a critical item.

(d) Solid-State Recorder: The solid-state recorder capacity is sized to be compatible with the temporary storage of the science data between two consecutive ground-station visibility periods: 100 Gbits should be adequate taking into account an average payload production rate of about 1 Mbps. Some oversizing of the capacity may be required for areas of the sky with a higher star density. Size and read/write rate requirements are nevertheless considered as moderate. Such a unit is not considered as a critical hardware item and should be available by the time of the GAIA development.

(e) High-Rate Telemetry Formatter: The high-rate telemetry formatter provides the necessary formatting and coding of the data as well as the interface with the science telemetry transmitter. Classical (Reed-Solomon and Viterbi) or TURBO coding may be applied to the data in order to reach the link budget and bit error rate requirements. TURBO coding, developed for ESA, has been selected as baseline because of a better efficiency.

4.6 Power and Electrical Subsystem

4.6.1 Overview

A centralised electrical architecture is proposed, in which a single computer (the central data management unit) provides all the necessary control to all service module units. It also provides the required command and control signals to the payload electronics (actuator commands, gain control, and parameter setting). Consequently, the software implemented within the central data management unit includes the ACMS functions, the thermal control functions, the standard command monitoring and reconfiguration functions, etc. This concept, which minimises the quantity of hardware and software, is already applied to the new generations of space platforms. It is a realistic design arising from the moderate complexity of the thermal control and attitude and pointing control functions, which means that the processing load and processing complexity will remain fully compatible with current processors. It is also foreseen to implement direct interfaces between the central data management unit and most of the various actuators and sensors instead of going through dedicated remote terminal units (except for the payload refocusing mechanisms, optical

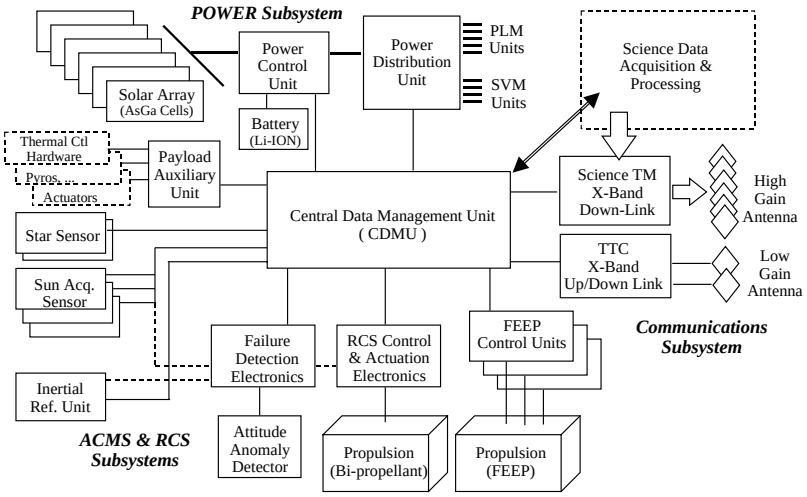


Figure 4.12: Service module electrical architecture. The service module design is based on a centralised computer architecture with one single unit responsible for the satellite operation, attitude and orbit control, thermal control, etc.

covers, etc., for which a dedicated auxiliary control unit is foreseen). This simplified approach corresponds to a general trend in the design of modern platforms. A functional overview of the service module electrical architecture is presented in Figure 4.12. It includes the following items:

(a) Power subsystem: comprising (i) six deployable Ga-As solar panels; (ii) a power control unit including the classical battery charge and discharge regulation functions, which should not be critical because of the possible avoidance of any eclipse during the operational lifetime; (iii) the power distribution unit which provides the primary power (typically 28 V DC unregulated power bus) to all service module and payload module units; (iv) one single battery, including extra cells for reliability, with a small capacity of typically 14 Ah (a state-of-the-art Li-ion battery is adopted as baseline because of its low mass, its compact size, and a lower sensitivity to the thermal environment, although a more conventional NiCd battery may also be considered).

(b) Thermal control hardware: comprising a set of thermistors, thermo-switches and heaters needed for the thermal control of the service module and payload module. Due to the passive design approach and to the stable thermal environment of the L2 orbit, very few heaters are required for the thermal control in observational mode, and the control laws correspond to achievable accuracy and stability requirements (e.g. 0.25 K accuracy and 20 s control frequency for the Titanium thrust between the payload module and the service module). The detailed power budget includes the heater power with margin for any complementary thermal control loops.

(c) RCS subsystem: comprising (i) control and actuation electronics, in charge of the actuation and control of the bi-propellant propulsion subsystem; (ii) a unified bi-propellant propulsion subsystem, including a perigee boost engine, is considered for the transfer phase and the transition to safe mode; (iii) the FEEP propulsion subsystem adopted for the observational phase consists of four clusters of FEEP micro-thrusters, driven by two FEEP thruster drive electronic units. These devices are the subject of on-going development under ESA contract.

(d) ACMS subsystem: comprising (i) a large field of view star sensor which includes within a single package the optical head and all the necessary electronics and processing capabilities (inclusive a star catalogue) to automatically retrieve the spacecraft attitude data. The accuracy required is compatible with existing equipment; (ii) classical Sun sensors providing a 4π steradian field of view used for the preliminary attitude acquisition as well as safe mode; (iii) inertial reference unit limited to a standard low-cost one-axis gyroscope used during the transfer phase and in safe

mode; (iii) Astro Sky Mapper (ASM): this equipment is included in the astrometric instrument design (specific field of the instrument focal plane) and is not formally part of the service module units. It is used for the measurement and control of the spacecraft attitude and spin rate; (iv) a classical failure detection electronics system which monitors the spacecraft behaviour and controls the transitions to safe mode; (v) classical attitude anomaly detectors used to detect any spacecraft abnormal attitude (e.g. Sun aspect angle out of acceptable range).

(e) Communication subsystem: comprising (i) the telemetry and telecommand transponder, including the diplexers and an additional amplifier in order to provide the required RF transmit power; (ii) two low-gain antennae accommodated on the spacecraft in order to provide omnidirectional coverage for the telemetry and telecommand links; (iii) a science telemetry transmitter, which includes the RF modulators, the RF distribution unit and a control unit for the control of the high-gain antenna; (iv) a phased-array (high-gain) antenna based on 6 planar arrays accommodated onto the service module lateral panels. Most of these units are accommodated in the service module, either on the inner faces of the service module lateral panel or within the service module central cone. The star sensors are mounted on the payload module optical bench to minimise the distortion and mis-alignment between the sensors and the instruments. A low-gain antenna is mounted on the top of the payload module secondary structure in order to provide the required field of view.

4.6.2 Power Requirements

The preliminary power budget has been computed in observational mode and with the spacecraft within ground station visibility, with all instruments and payload units switched on, and with the science telemetry down-link subsystem transmitting data to the ground station. The numbers presented at equipment level in Table 4.5 are current power budgets, with a 15 per cent contingency included.

Table 4.5: Satellite power budget, computed in observation mode with all instruments switched on and including telecommand and science communications with Earth. The total spacecraft power demand includes margins at equipment level, harness loss and margin at system level.

Module/Subsystem	Power (W)	Remark
General services:		
Communications	335	including telemetry and telecommand
Power generation/distribution	49	Sun lit, no eclipse
Reaction control system	166	FEEP control
Attitude control system	52	including failure detection electronics
Data handling	39	
Total service module	641	
Payload:		
Astrometric instruments	1242	including focal plane assemblies
Spectrometric instrument	52	including focal plane assemblies
Data processing	143	including data handling units
Thermal control	90	payload/service module interface
Total payload module	1527	
Satellite:		
Total current power budget	2168	
Harness loss	76	3.5 per cent of spacecraft current power
System contingency	224	10 per cent of satellite + harness loss
Total power requirement	2468	including equipment and system margins

4.6.3 Solar Array

The spacecraft power demand is 2468 W in observational mode, including the equipment and system margins presented in Table 4.5. The interface between the solar array assembly and the spacecraft power control unit introduces additional losses of the order of 6 per cent at the solar array interface level, taking into account a 28 V primary power bus. Consequently, the power required from the solar array assembly is increased by 6 per cent with respect to the spacecraft power demand, such that the specified solar array power is 2616 W at the end of 6 years. Detailed consideration of solar panel losses have been carried out, including effects intrinsic to the solar array design, the performance of the solar cells, losses induced by environment considerations during the specified extended lifetime UV radiation and micrometeorite damage), the Sun aspect angle and the thermal control constraints.

Ga-As technology should be the standard technology by the time of the GAIA development, resulting in reduced procurement cost compared to current costs. Ga-As cells provide a better conversion efficiency than more classical Si cells, and result in a reduced size and mass of the solar panels. Efficiencies of the order of 18–20 per cent are currently proposed by manufacturers (beginning of life, 25°C). A preliminary assessment of the cell efficiency at the end of the extended lifetime has been carried out, based on the assumption of 250 μm coverglass thickness (compatible with the moderate radiation environment at L2); normal Sun incidence; a total radiation dose of 10^{14} 1 MeV e^-/cm^2 over the 6 years extended lifetime; a cell operational temperature of 60°C (obtained in observation mode with 55° Sun aspect angle); and a 95 per cent fill ratio. This results in a Ga-As cell efficiency of 205 W/m^2 at end-of-life, yielding a requirement on the total solar array surface of 24.1 m^2.

4.6.4 Battery

Batteries are needed during the launch phase, until the deployment of the solar arrays, and during eclipses, which should only occur during the early transfer phase. With the hypothesis of an injection of the spacecraft by the launcher into a geostationary transfer orbit, it is assumed that the launch phase will not last more than 60 min, starting with the switching from ground power to spacecraft battery and ending with the solar array deployment. Within this time period, only the last minutes (after spacecraft jettisoning) will require some significant power (service module switch on and solar array deployment sequence). Eclipses should not last more than 30 min. For the purpose of a preliminary sizing of the battery and preliminary assessment of its criticality, a 100 per cent margin on this number is included, leading to a 60 min worst case requirement for the battery sizing.

Immediately after solar array deployment, the instruments and associated data processing electronics are off, the science telemetry down-link subsystem is off, the FEEP propulsion subsystem is off, while all other service module units are on. Assuming a 50 W provision for payload module thermal control, the total power budget in the launch and early-orbit phases is therefore of the order of 360 W, leading to a required energy of 360 Wh. Due to the low capacity and moderate operational constraints of the battery required (very low number of cycles), a standard NiCd battery may be considered. In order to save mass, and because the development will be mature at the time of the satellite development, a Lithium-ion battery is selected as baseline. Because of the very low number of eclipses during the spacecraft lifetime, a relatively high depth-of-discharge may be considered, typically 70 per cent or more. The required battery capacity is 14 Ah.

4.7 Communications

4.7.1 Ground Station Assumptions

A feasible technical solution has been identified for the communication subsystem which is able to transmit the few Mbps required for the science data based on a single ground station concept.

For the L2 orbit, the duration of the ground station coverage per day depends only on the geographical latitude of the station, and on the declination at which the spacecraft is seen from the centre of the Earth. Villafranca cannot be used as the single ground station because of a very low coverage, although it may be foreseen as a complement to Perth. Kourou is the best candidate in terms of coverage because it is close to the Earth equator, but the achievable performance is moderate. Perth presently offers the best compromise between coverage and performance and is

considered as the baseline. The performances which have been considered for the sizing of the telecommunication subsystem are summarised in Table 4.6. Use of a ground station network of 2–3 stations optimally situated on Earth, in order to maintain an uninterrupted link between the spacecraft and the ground, would not significantly change the concept and complexity of the communication subsystem, since the scanning motion of the spacecraft would still imply use of an antenna with a 360° field of view in azimuth. The only significant advantage would be a decrease of the transmission rate by up to a factor 3. Using more than a single ground station would increase the operational costs and would require a more complex ground segment architecture for the collection, archiving and dissemination of the satellite data.

Table 4.6: Perth ground station performance.

Parameter	Perth (32m)
S band G/T at 10°	37.5 dB/K
S Band EIRP	87–89 dBW
X Band G/T at 10°	50.1 dB/K
X band EIRP	97 dBW
Ka Band G/T at 30°	56 dB/K

4.7.2 Telemetry and Telecommand Link

A classical omni-directional coverage will be provided by the on-board antennae, to cope with any spacecraft attitude at any time from launch through to the end of the mission. The number of antennae to be accommodated in the spacecraft must take into account potential masking induced by solar arrays, sunshield, baffles, etc.

For the up-link telecommand, a data rate of 1 kbps (on station) and 2 kbps during launch and early-orbit phases is provisionally considered. Data are transmitted to the spacecraft when the spacecraft is in visibility of the ground station. It is not foreseen to up-load periodically any star catalogue. The only star catalogue which is foreseen on board is one implemented in the star sensor. It is a 'standard' catalogue, loaded on ground during the equipment integration and test activities.

A provisional data rate of 2 kbps is assumed for the spacecraft housekeeping telemetry. This data flow is permanently produced by the spacecraft. Consequently, it will be temporarily stored on-board in the central computer, and then transmitted to ground when the spacecraft is within ground station visibility. The required memory capacity is in the range of 180 Mbits corresponding to one full day. Taking into account a minimum visibility window of about 8 hours, the transmission data rate will be 3 times higher than the production data rate, i.e. 6 kbps.

A relatively high RF transmitted power is required (17 W) in order to provide the required recovery margins from the L2 orbit. This relatively high power level is not provided by current transponders, and an additional solid-state power amplifier will probably have to be implemented at the transponder output. This is a standard approach for high-rate telemetry subsystems. A preliminary performance budget is presented in Table 4.7. There is no critical area identified at this stage. Detailed link budgets have been established. A conventional PCM PSK PM modulation scheme is considered.

While S-band is the conventional frequency band used on spacecraft for the telecommand up-link and the telemetry down-link, it is expected to become very crowded in the coming years, at least for the telemetry down-link, because of the rapidly increasing demand of mobile users. Alternatives are the use of an X-band channel for the down-link combined with an S-band channel for the up-link, or the use of an X-band up- and down-link system. All alternative approaches (S/S, S/X and X/X) have been considered. An X/X design, although not optimal in terms of hardware availability and power demand, is adopted as baseline because it satisfies the potential need to go higher in frequency. However, a S/X design is also possible without impacting on the spacecraft design and cost.

The science telemetry link is operated only during the observational or pre-observational phases, i.e. once the spacecraft is on or close to its final operational orbit, and once the operational attitude

Table 4.7: Preliminary performance budget for the telemetry and telecommand link.

Performance Budget	Nominal case
Perth 32 m	EIRP = 97 dBW G/T = 50.1 dB/K
On-board antenna gain	–3.5 dBi at 90°
Polarisation	Circular (L+R) with combiner on ground
Solid-state amplifier RF power	17 W
Telecommand recovery margin (with ranging)	+2.6 dB (spec = +3 dB)
Telemetry recovery (with ranging)	+3.3 dB (spec = +3 dB)
Ranging accuracy	22 m (1σ)

and scanning law are reached. Consequently, the telemetry link antenna coverage must take into account the spacecraft attitude, the spacecraft scanning law and the spacecraft position with respect to the Earth throughout the observational phase. In addition to the standard parameters of the proposed scanning law (Sun aspect angle 55°, spin period 3 hours, precession period 73 days) the Sun-Satellite-Earth angle excursion over a six-month L2 orbit is ±15° (the smaller this angle, the simpler the antenna design). The resulting antenna coverage requirement, assuming a continuous transmission to Earth, may be defined in terms of azimuth and elevation coverage with respect to the spacecraft spin axis, with an azimuth coverage of 360° (about the spacecraft spin axis) and an elevation coverage of $55 \pm 15°$ (with respect to the spacecraft spin axis). A transmission time shorter than the spin period would require the same elevation coverage but a reduced azimuth coverage, depending on the ratio between this transmission time and the spacecraft spin period.

The data rate considered for the science telemetry down-link sizing depends on the instrument design, the on-board data handling design, and the transmission scenario. A preliminary assessment for the data rate produced by all payload instruments, including the on-board data handling, is about 1 Mbps. This is stored temporarily in a solid state recorder and transmitted to ground when the spacecraft is within ground station visibility. If the whole visibility window of approximately 8 hours is used for the transmission of the science data, then the required transmission data rate is three times higher than the payload production rate, or 3 Mbps. A communication scenario based on a shorter transmission time per visibility period would require a higher transmission rate. The communications scenario has been developed on the basis of no loss of scientific data.

Several antenna technology candidates have been identified and compared during the study: single beam fixed antenna, switched antenna network, electronically scanned phased array antenna, and single beam steerable antenna. Detailed studies of the antenna accommodation on-board (in view of the propulsion engine, for example), and stability implications of a continuously moving part on the satellite, have led to the adoption of the electronically scanned phased array (or 'conformal') antenna as baseline, with the single beam steerable antenna as a possible back-up solution. The phased array has no moving parts, which means that there is no source of dynamic perturbation for the attitude control and measurement subsystem, and that it is possible to acquire and transmit data simultaneously. The concept offers high reliability employing some graceful degradation (the required antenna directionality is still obtained with one failure in the RF path) or reconfiguration features. Although based on off-the-shelf technologies at elementary level, it will require significant engineering and development activities at assembly level, including the production of dedicated MMIC and hybrids, dedicated design and arrangement of the radiating elements, and the overall mechanical and thermal design.

The conformal antenna concept is the result of various studies and technology validation activities carried out by Alcatel for CNES and ESA for several years (see Figure 4.13). It is a semi-active antenna concept which exploits the geometry of the required antenna field of view to reduce the number of radiating elements to a minimum. The proposed radiating aperture is a single element included in a cone with a 35° half-angle, aligned with the spacecraft X axis. The antenna configuration is a truncated conical array with the radiating aperture being covered with sub-arrays placed along the cone surface and performing a 2-dimension scanning with only a 1-dimension control.

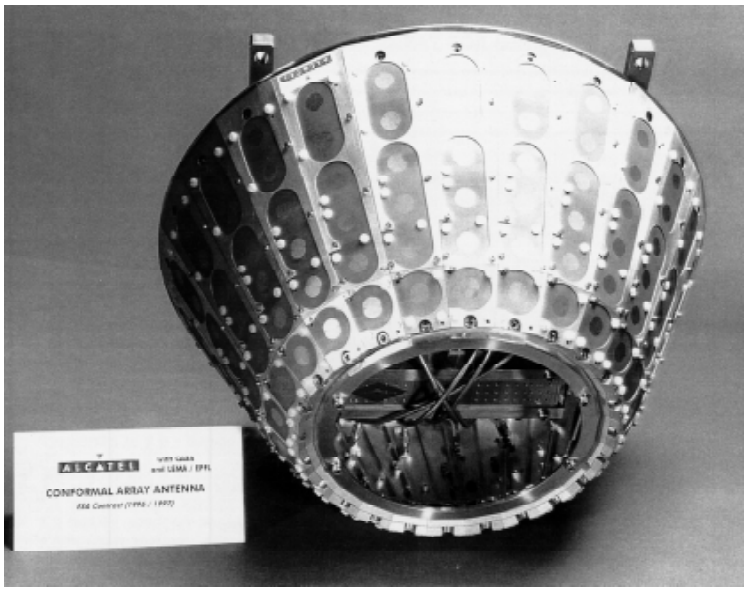


Figure 4.13: Example of a conformal array antenna built by Alcatel for ESA. A similar system is foreseen for the main GAIA antenna.

The most critical performance requirement is the antenna EIRP target of 32 dBW, required to transmit 3 Mbps from an L2 orbit. Distributed amplifiers (active MMIC modules) are placed at an intermediate stage of the beam feeding network, after the power dividers and phase shifters but before the Butler matrices. Current state-of-the-art shows that MMIC amplifiers in X-band can provide the required power per chip. The conical array geometry would include 24 sub-arrays with each array being made of 6 patches. In such a configuration, all solid-state power amplifiers are used always at full power, using distribution via Butler matrices to a suitable phase-shift control; there is no need to switch the power on and off from one set of sub-arrays to another. The resulting design is very compact, with an overall height of about 173 mm and a diameter of about 377 mm. Preliminary performance analyses have been carried out which show that the required directionality should be achievable throughout the specified field of view. Assuming 3.8 dBW solid-state power amplifiers, and taking into account relevant losses, leads to the specified EIRP of 32 dBW, even if one of the 24 RF paths fails. The physical implementation depends on the precise choice of the high thrust engines and their location on board.

Limitations on the Telemetry Rate If an increase by a factor of, say, 3 in telemetry rate is required, several solutions may be investigated and implemented to solve the communication issue: (a) increase the antenna gain: the conformal antenna design assessed by Alcatel offers a 17 dB gain over the required field. There is perhaps room for a 1–2 dB improvement; (b) increase the transmittor RF power level: in order to achieve the 32 dBW EIRP required in the baseline design, the conformal antenna radiates 36 W, which leads to some difficulties in terms of thermal design and power generation (the SSPA efficiency is low). Increasing this further (i.e. above 40 W) is questionable; (c) increase the ground station coverage: improving e.g. by 2–3 the station coverage is straightforward, providing that 2–3 ground stations instead of one with adequate location on Earth (e.g. one in the northern hemisphere and one in the south, or two at the same latitude and at opposite longitude, e.g. at the equator). These complementary ground stations would be used only for science data collection and neither for mission control nor for operations which means simple interfaces with the system. Good sensitivity (G/T) is required, however the data rate is still modest which means that it should be relatively easy in 2009 to transmit these data from complementary ground stations to any place in the world; (d) improve the data coding: the transmission signal modulation and coding strategy proposed as baseline (QPSK modulation + Turbocode) is based on existing techniques. Improved techniques (investigations are running for LEO systems with very high data rates) may allow to transmit a somewhat higher quantity of data with the same energy (but with some impact on the antenna bandwidth, receiver stage, etc.); (e) change to a faster antenna technology: the steerable antenna concept remains as a potential alternative able to transmit higher rates.

4.8 Mass Budgets

Masses given at unit or sub-assembly level are current mass budgets, including a contingency which is a function of the equipment design maturity. The combined payload and service module results in a current total mass of 1696 kg (Table 4.8). Including a system margin of 20 per cent results in a mass of 2035 kg, with the liquid apogee engine. The total satellite mass depends significantly on whether the liquid apogee engine is required (Table 4.8c).

Table 4.8: (a, left) Payload module current mass budget. (b, right) Service module current mass budget. (c, below) Satellite mass budget, and its dependence on the inclusion of the liquid apogee engine.

Sub-assembly/unit	Mass (kg)	
Astrometric instruments:		415
Mirrors and mounts	173	
Alignment devices	74	
Focal plane assemblies	96	
Video processing units	72	
Spectrometric instrument:		112
Mirrors and mounts	96	
MBP focal plane	2	
RVS spectrometer	6	
RVS focal plane	2	
Video processing unit	6	
Data handling electronics:		72
Structure and thermal hardware:		293
Primary structure	224	
Secondary structure/thermal	69	
Payload module current mass		893

Sub-assembly/unit	Mass (kg)
Telemetry and telecommand subsystem	15
Science telemetry	87
Power subsystem (without solar array)	39
Attitude and measurement subsystem	163
Data handling subsystem	47
Harness	36
Solar array	108
Sunshield	24
Thermal hardware (OSR, MLI, heaters)	30
Structure	253
Service module current mass	803

	With LAE	Without LAE
Spacecraft dry mass	2035	1975
Propellant (including margin)	1010	200
Launcher adaptor	92	92
Total launch mass	3137	2267

5 Orbit, Operations, and Ground Segment

5.1 Orbit Selection Requirements and Constraints

The two candidate orbits proposed at the start of the study were a Lissajous orbit around the Earth-Sun Lagrange point L2, and a geostationary orbit. These two orbits, as well as a low-Earth orbit, and orbits around the Earth-Moon libration points, were assessed and compared in order to select the most appropriate orbit for GAIA. The results of the trade-off exercise, summarised in Table 5.1, indicate that L2 is the preferred orbit, as confirmed by ESOC studies (Hechler 1997b; Hechler 1997a).

The L2 region provides a very stable thermal environment (in order to satisfy the stringent geometrical stability requirements of the optical payload); an absence or minimisation of eclipses (which would perturb the thermal environment and geometrical stability of the payload, and thus require a more complex power subsystem and more complex satellite operation); an absence of Earth or Moon occultations (which would introduce straylight, thermal fluctuations, or blooming at detector level); more stable perturbing torques (dominated by solar radiation pressure); and a lower radiation environment. The orbit is compatible with an Ariane 5 dual launch (the assumed ESA baseline launch strategy), and results in the lowest cost impact at system level, i.e. taking into account launch cost, induced satellite complexity, and operations cost. The orbit is consistent with a 6-year extended lifetime, with minimum mission outages due to the transfer phase, perturbations, eclipses, etc.

The major disadvantage of the L2 orbit is created by the problems of communication: visibility from a single ground station (e.g. in Europe) may reach only 6 hours per day; and the telemetry rate is strongly reduced for any given transmitter power and antenna profile. This potential problem is nevertheless surmounted with the proposed telecommunication subsystem design (Section 4.7).

Table 5.1: Operational orbit selection trade-off. '–', '', '**' and '***' indicate the preference hierarchy, with the latter representing the optimum choice with respect to candidate orbits. A single ground station is considered to be feasible in all cases. A Lissajous orbit at L2 is the most adequate operational orbit.*

Parameter	Low Earth	Earth-Moon (L4/L5)	Earth-Sun (L2)	Geostationary
Thermal environment	–	*	***	–
Radiation environment	**	***	***	–
Optical environment (occultations)	–	*	***	*
Eclipse avoidance	–	**	***	*
Dynamic environment	–	**	***	**
Injection Δv	***	***	***	**
Maintenance Δv	–	***	***	***
Communications	**	*	**	***
Launch mass	***	**	***	*
Transfer duration	***	**	*	***
Operations	**	**	**	***
Ground facilities	***	***	***	***
Lifetime	*	***	***	*
Overall complexity/cost	–	*	***	*
Recommendation	rejected	rejected	baseline	rejected

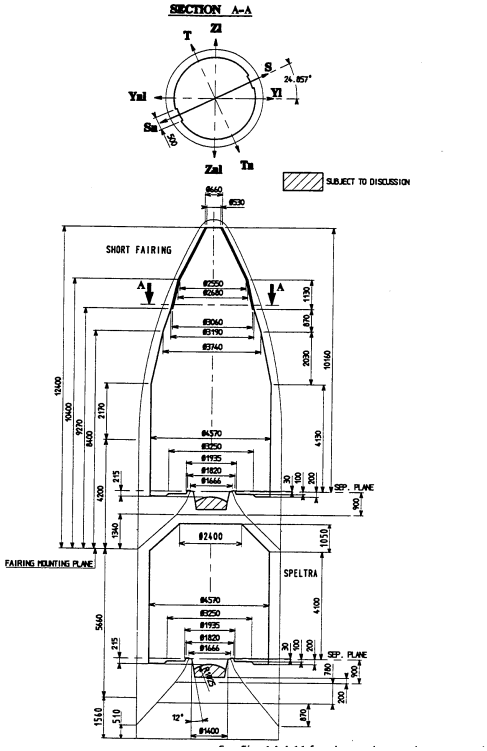


Figure 5.1: Spacecraft interface data with the Ariane 5 Speltra assembly.

5.2 Launch Requirements and Constraints

The launch strategy selected as baseline is based on a dual or multiple launch with Ariane 5 (Figure 5.1), completed with an injection of the satellite from the standard geostationary transfer orbit into the L2 transfer orbit via an autonomous propulsion system.

Two other options may be considered further, depending on technical developments and any directives arising from science programme launch policy: (a) 'restartable' Ariane 5: the baseline assumption of injection by the launcher into a standard geostationary transfer orbit leads to a number of design constraints at satellite level: (i) the need for an autonomous propulsion system able to provide the necessary Δv for the injection into the L2 transfer orbit (800–900 m s^{-1}); (ii) the need for a large quantity of propellant to generate this Δv, requiring the accommodation of large tanks; (iii) the need of a dedicated attitude control system to control the L2 injection phase. All this adds some complexity to the satellite design, increasing the satellite launch mass and decreasing the system reliability. An alternative strategy using a 'restartable' Ariane 5 (an Ariane 5 with a restartable upper stage) would allow the satellite to be injected directly into its transfer orbit to L2. Such a concept is already foreseen for Rosetta, and for FIRST/Planck. The scenario for GAIA would be different from those mentioned above. The first passenger of Ariane 5 would be delivered after a first burn of Ariane 5 into GTO, then the upper stage would be restarted, possibly after some coasting arc, to provide the additional energy from GTO to the L2 transfer. Resulting satellite design impacts are straightforward: the satellite reaction control subsystem no longer has to perform the preliminary perigee/apogee raising manoeuvres, although it must still provide the ΔV required for any change of inclination, mid-course corrections during the transfer phase, final L2 orbit insertion, for the operational orbit maintenance, and for the attitude control. Consequently the main engine can be omitted and the propellant tanks decreased in size. The saving in terms of satellite launch mass is about 800 kg. (b) alternative launchers: these may be considered for cost reasons, or if alternative launchers offer direct injection into the transfer orbit. Any candidate must still offer a fairing envelope large enough to accommodate the payload, and capability to launch 2 tons (direct injection) or 3 tons (injection into the geostationary transfer orbit). Proton would offer a large-enough fairing (4.35 m) together with an attractive launch cost, even in a single-launch configuration. While potential medium-sized launchers (Delta III, EELV Medium, Soyouz-Fregat) have fairing diameters (3.3–3.6 m) far below the Ariane 5 diameter of 4.5 m, a more compact design may be feasible, but with corresponding impact on the astrometric performance. Of the candidates noted, Soyouz-Fregat should be able to inject the satellite directly into its transfer orbit. Very small amplitude Lissajous orbits can possibly be reached with very little injection ΔV by employing lunar gravity assists.

5.3 Mission Analysis for the L2 Orbit

5.3.1 The Small Amplitude Lissajous Orbits around L2

Orbits around the co-linear libration points in the Earth-Sun system (L1, at about 1.5×10^6 km from the Earth towards the Sun, and L2, at about the same distance away from the Sun) have been used for space missions since the 1970s. ISEE-C was placed in a Halo orbit around L1, while SOHO is still operational at the same location. The orbits around L2 are particular interesting for astronomy missions, and are being considered for NGST, FIRST/Planck, and also GAIA.

A basic dynamic property of the colinear libration points is that they are inherently unstable, so that a spacecraft placed at those points, which theoretically are at gravitational equilibrium in the system rotating with the Earth around the Sun, will move away from that point because small perturbations of the orbit cannot be avoided. If a spacecraft could be precisely located at the libration point L2, then the Sun and Earth would be in a line at all times, and a spacecraft axis with a high-gain antenna could be permanently pointed toward the Earth. This scenario has, however, two drawbacks. First the spacecraft would be permanently in the Earth half shadow, and second, the ΔV required to position the spacecraft at L2 from the Earth transfer orbit would be about 500–600 m s^{-1} for Ariane perigee conditions, although cases which allow use of lunar gravity assists may make this easier.

There is a nevertheless a family of 'quasi periodic orbit' around L2 in the frame rotating with the Earth around the Sun. Denoting the ecliptic as the xy plane with the x-axis from the Sun to the Earth in the rotating frame, and z out of the ecliptic, then these orbits are generally characterised by a harmonic motion in the xy plane, with an uncoupled oscillation in z with a different period. The periods of these two motions (around 6 months) depend on the coupled amplitudes ($A_y \sim 3.2A_x$) in xy and A_z in z, due to the non-linear terms in the motion. For some specific combinations of large amplitudes (above 600 000 km in y) and a proper choice of the initial phase of the motions in xy and z, the in-plane and out-of-plane periods can be matched, so that a true periodic motion around L2 is generated. The family of orbits with this latter property are called 'Halo' orbits, e.g. as adopted by SOHO. Their advantage is that they pass around the Earth shadow for L2 (or some similar communications exclusion zone for L1) without special eclipse-avoidance manoeuvres. The ΔV to capture from an Ariane transfer into such a Halo orbit may be kept below 100 m s^{-1}, depending on the launch date.

In the mission analysis for FIRST/Planck and GAIA it has been demonstrated that there is another sub-class of the motions around L2 which contains orbits with large amplitudes (relative to L2 in the rotating frame) without this matching of the two periods, and which can be reached without any injection ΔV from any type of transfer orbit. This arises from the fact that the 'quasi periodic motions' around the libration points have attached both an unstable manifold (a direction in phase space along which perturbations grow exponentially) but also a stable manifold (meaning that deviations on this geometric structure attached to the orbit will decay exponentially). For the chosen sub-class of orbits, without injection ΔV, the stable manifold happens to tangentially touch e.g. the GTO perigee conditions at 200 km altitude. So with a kick manoeuvre at perigee, and with some navigational orbit corrections, a spacecraft will automatically be captured in an orbit around L2 without further major manoeuvres.

Orbits around L2 which do not satisfy the special condition on the amplitudes of a Halo orbit, seen from the Earth in the yz plane, describe a Lissajous figure, because the two oscillation periods in y and z are different (see Figure 5.2 for a corresponding case with a small amplitude). Consequently it will eventually enter into eclipse depending on the initial phase y versus z. The time span from eclipse to eclipse is about 6 years, so either the initial phase may be chosen to reach 6 years without eclipse for a space mission or, more preferably as discussed later, an eclipse avoidance manoeuvre at some time in the mission before the eclipse occurs, of about 15–20 m s^{-1}, may open another 6 year observational period without eclipse.

A large amplitude Lissajous orbit (with a 'free' transfer from Ariane launch conditions) is now the baseline for FIRST. Planck and also GAIA cannot accept the large amplitudes, because of

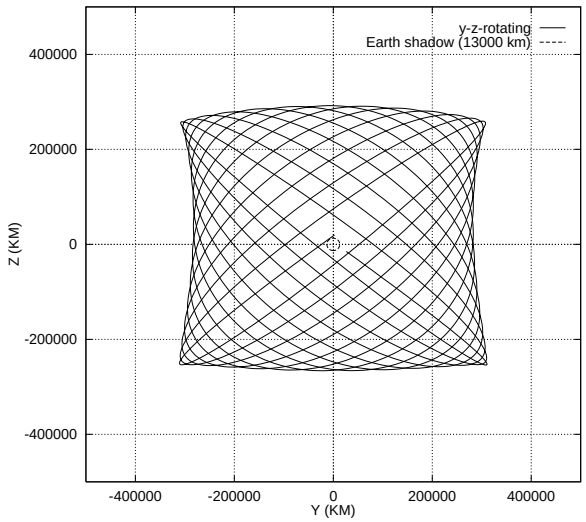


Figure 5.2: The Lissajous orbit projection into the yz plane as seen from Earth, starting tangentially at Earth shadow, employing high-order analytic propagation in the restricted circular 3-body problem for 6.8 years until the occurrence of the next eclipse.

the sky scanning strategy and the resulting constraints on the Earth to Sun viewing angle arising from thermal and communications design. A limitation of the amplitudes such that the Sun-spacecraft-Earth angle remains below 15° was accepted as a compromise of those spacecraft design requirements versus the propellant budget (injection ΔV increases with decreasing amplitudes) and the eclipse avoidance strategy. A summary of the baseline orbit is given in Table 5.2.

Table 5.2: Orbit summary.

Parameter	Property
Orbit location	Around L2 of Sun-Earth system
Transfer orbit	'Fuzzy' boundary
Transfer orbit duration	200–280 days
Total ΔV from GTO	1150 m s^{-1}
L2 orbit type	Lissajous (6-month period)
Lissajous orbit parameters	$\pm$400 000 km, $\pm$100 000 km
Eclipses	None

5.3.2 Transfer from Ariane Midnight GTO Double Launch Conditions

A special feature of the GAIA transfer (compared e.g. to FIRST/Planck) is that the GAIA spacecraft mass budget allows for sharing an Ariane 5 launch, for example with communications spacecraft which have to be delivered into a Geostationary Transfer Orbit (GTO). To accommodate typical attitude constraints for the apogee burn of such commercial spacecraft, Arianespace has fixed a standard midnight launch window for double launches: the perigee of the GTO is then pointed away from the Sun direction, with the apogee pointing towards the Sun. This is a very unfavourable initial geometry for a transfer to L2 as the apogee has to be away from the Sun, so best perigees to start such a transfer are on the Sun side somewhere near local midday. Also the fact that the transfer has to be started from a GTO apocentre placed, by definition, in the equatorial plane, has disadvantages for some launch dates (mid winter and mid summer) because the L2 region lies near the ecliptic, which for those unfavourable launch dates is far above or below the initial orbit plane.

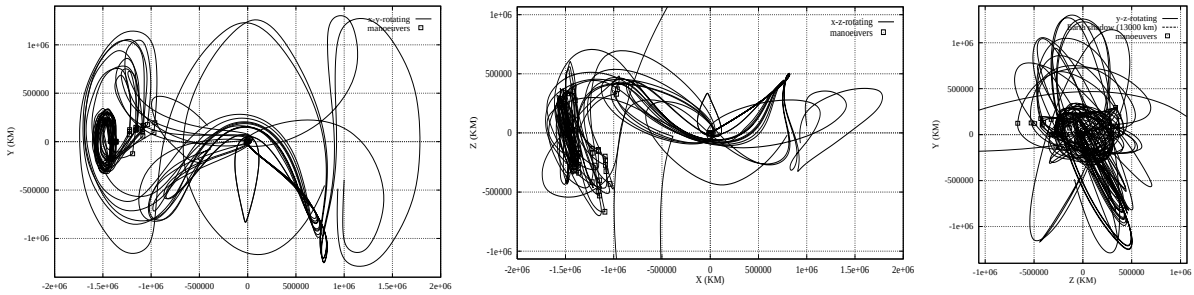


Figure 5.3: Different fuzzy boundary transfer orbits from GTO midnight launch conditions to a small Lissajous orbit around L2 (launch on 2009–01–01). The range of pericentre velocities is 5 m s^{-1} . The figure includes the amplitude reduction manoeuvres and three revolutions in the operational orbit.

Fortunately the Lissajous orbits around L2 and their stable manifolds are by definition in the 'weak stability boundary' between Earth and Sun. Near this boundary the usual two-body celestial mechanics laws sometimes can be 'overcome'. In fact in the four-body problem, also including the moon, very unusual properties of the transfer orbit from midnight GTO conditions to orbits around L2 have been observed. The L1 region and the L2 region seem to be dynamically connected, so when injecting at the GTO perigee to a transfer towards the L1 region, in a narrow band of perigee velocities a discrete set of different orbits exists which 'fall' from L1 to L2 and are captured there into a Lissajous orbit around L2. This very complex topology in the four-body problem has been known theoretically for some time. In the mission analysis, numerical studies have been performed which clearly display this 'fuzzyness', and which can be used to the advantage of the mission.

Figure 5.3 shows the different orbits (in xy, xz, and yz projections) which neither fall to the Earth nor to the Sun within 450 days (these are the orbits at the weak stability boundary being targeted), most of them within a band of 5 m s^{-1} in perigee velocity but with the same initial time (midnight launch window opening 2009–01–01) and the same initial orbit parameters except for the perigee velocity modulus. A simple bisection method has been used to construct these orbits. It should be noticed that these orbit only exist at some sharp 'resonances', with the continuum in between consisting of orbits around the Earth or Sun.

From Figure 5.3 it is evident that the first two orbits with the largest initial velocities (near 760 m s^{-1} above GTO velocity) are captured at L1, while all orbits with perigee velocities above this limit escape from the Earth-moon system. The next orbit 'does not decide' within the chosen integration time whether it goes to L1 or L2. The next case (759 m s^{-1} from GTO) reaches L2 without coming close to the Earth any more. The following 8 orbits (perigee velocities between 756 m s^{-1} and 751 m s^{-1} above GTO velocity) are captured around L2 after one or two additional perigee passages. Another two transfers (perigee velocity 746 m s^{-1} above GTO) reach L2 after three revolutions around the Earth. The minimum total ΔV (for this launch date 204 m s^{-1} for the insertion manoeuvre) is reached for the first orbit (756 m s^{-1} above GTO perigee velocity) in the group of orbits with one perigee passage (perigee radius 288 000 km). It is the orbit which has the minimum amplitude before the amplitude reduction manoeuvre.

One of the orbits in between happens to be deflected by a lunar flyby (see the xz projection). For other launch dates, over 20 different transfers to orbits around L2 have been found. For all orbits which reach the L2 region, a manoeuvre to reduce the amplitudes such that the Sun-spacecraft-Earth angle remains below 15° is performed, and the forward integration for three revolutions including eclipse avoidance manoeuvres if necessary are included in the figure. All manoeuvres are marked, also including the very small 'mathematical' corrections once per revolution which are necessary to allow such a long term numerical orbit integration of the unstable differential equations.

The strategy selected to construct the transfer is to inject at perigee to the stable manifold of a large Lissajous orbit around L2 and then, upon arrival in the L2 region, to perform one ore two orbit manoeuvres which enter from the one stable manifold to the other of a Lissajous orbit with the desired smaller amplitude. This strategy is very similar to the strategy for the FIRST/Planck carrier option (FIRST remains on the large amplitude Lissajous orbit) except for the midnight condition. The chosen strategy is not the optimum (in terms of ΔV) but it is believed (for FIRST/Planck it has been demonstrated) that it is not far from optimum.

For all these fuzzy boundary GTO to L2 solutions, for every day over one full year, Figure 5.4 shows the sum of all ΔV contributions, this means the perigee burn from GTO + the amplitude reduction manoeuvres + the eclipse avoidance for a 5 years mission. the launch time at every day has been chosen at the opening of the Ariane midnight double launch window. Some of the solutions (those marked with a +) have been removed applying a filter of additional conditions, e.g. constraining the Sun aspect angle of the insertion manoeuvres to a range of 120° to 150° and limiting the total ΔV to 1200 m s^{-1} and the total eclipse avoidance ΔV to 30 m s^{-1}, and removing transfer times above 280 days. The Earth-spacecraft-Sun angle has been verified to remain below 16° for all admitted solutions (to satisfy the 15° limit will require some fine tuning). The dashed line in Figure 5.4 connects the admitted solutions which have minimum ΔV on each day. The total allocation for an impulsive perigee ΔV

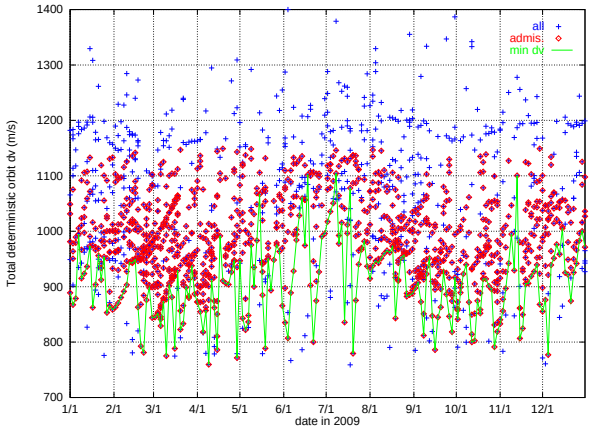


Figure 5.4: Total ΔV including eclipse avoidance for all transfers from GTO to a 15° Lissajous orbits at opening of Ariane midnight double launch window for every day in 2009.

from GTO varies between 745 m s^{-1} and 765 m s^{-1} over the year (gravity loss with Sun aspect angel constraint will be discussed below), the daily spread of the different solutions usually is in the range of 5–10 m s^{-1}. The transfer time (to Lissajous orbit insertion manoeuvre) varies from 200–280 days.

To limit the gravity loss of the apogee raising manoeuvre near perigee (around 750 m s^{-1} in total with a 400 N motor on a 3000 kg spacecraft) this manoeuvre has to be split into at least three parts. The periods of the intermediate orbits can be adapted to optimize ground station coverage. The split of the manoeuvre also allows a fine calibration of the last burn such that the required transfer conditions are exactly reached. At the same time the perigee manoeuvres usually cannot be executed at the optimum point at perigee, because a Sun aspect angle constraint (ΔV vector above 120° from Sun direction) has to be observed. For a sequence of three constant burn direction manoeuvres the gravity loss in conjunction with the penalty introduced by the Sun aspect angle constraint can be kept below 51 m s^{-1}. The rotation of the line of apsides introduced by placing the manoeuvres off perigee to satisfy the Sun aspect angle condition inside the midnight launch window, has also been taken into account.

It can be concluded that with an allocation of 1150 m s^{-1} GAIA can reach a 15° maximum Sun-spacecraft-Earth angle Lissajous orbit around L2 when left by Ariane in a GTO for any launch time within the standard midnight double launch window for any day in the year. This means that GAIA will not impose any launch window conditions.

Three typical minimum ΔV trajectories for different launch dates (transfer and 5 years in orbit) are shown: (i) a case with one revolution around the Earth during the transfer for a launch on 1 January 2009 (Figure 5.5a); (ii) a case with two revolutions around the Earth during the transfer for a launch on 3 January 2009 (Figure 5.5b); (iii) a free transfer case (no deterministic ΔV at all, also no eclipse avoidance necessary) with a launch on 9 April 2009 (Figure 5.5c). Figure 5.4 indicates that these cases happen about once per month.

The first two cases contain (mild) lunar gravity assists (the orbits of many of the cases which have a small insertion ΔV are influenced by the moon), but it should be noted that the third case (the free transfer to a small Lissajous orbit) does not come close to the moon. It can be seen (in the yz projection) that the choice of the combination of the amplitudes and the initial phase of the Lissajous orbit in y and z is part of the optimisation, the only condition imposed is the maximum Sun-spacecraft-Earth angle.

5.3.3 Orbit Determination and Maintenance

Clearly the transfer orbits have very special properties. It has not yet been studied in detail how deviations from that initially chosen minimum transfer ΔV can be removed. For FIRST/Planck this issue has been studied and it is believed that also for GAIA the transfer is feasible. The navigation (orbit determination and orbit correction manoeuvres) in the operational orbit however has been fully studied independent of the mission. Independent of the method it could be demonstrated that an allocation of not more than 1 m s^{-1} per year is necessary to maintain the orbit, provided there are no major inertial perturbations such as asymmetric wheel off-loading. A manoeuvre will be necessary at least once per month. The simple strategy of removing the unstable motion seems to perform well and has a property from which the spacecraft design benefits. All manoeuvres will be nearly aligned or opposite to a direction 28.°4 from the Sun to Earth x axis. This is a property derived from the linearised theory of the motions around L2.

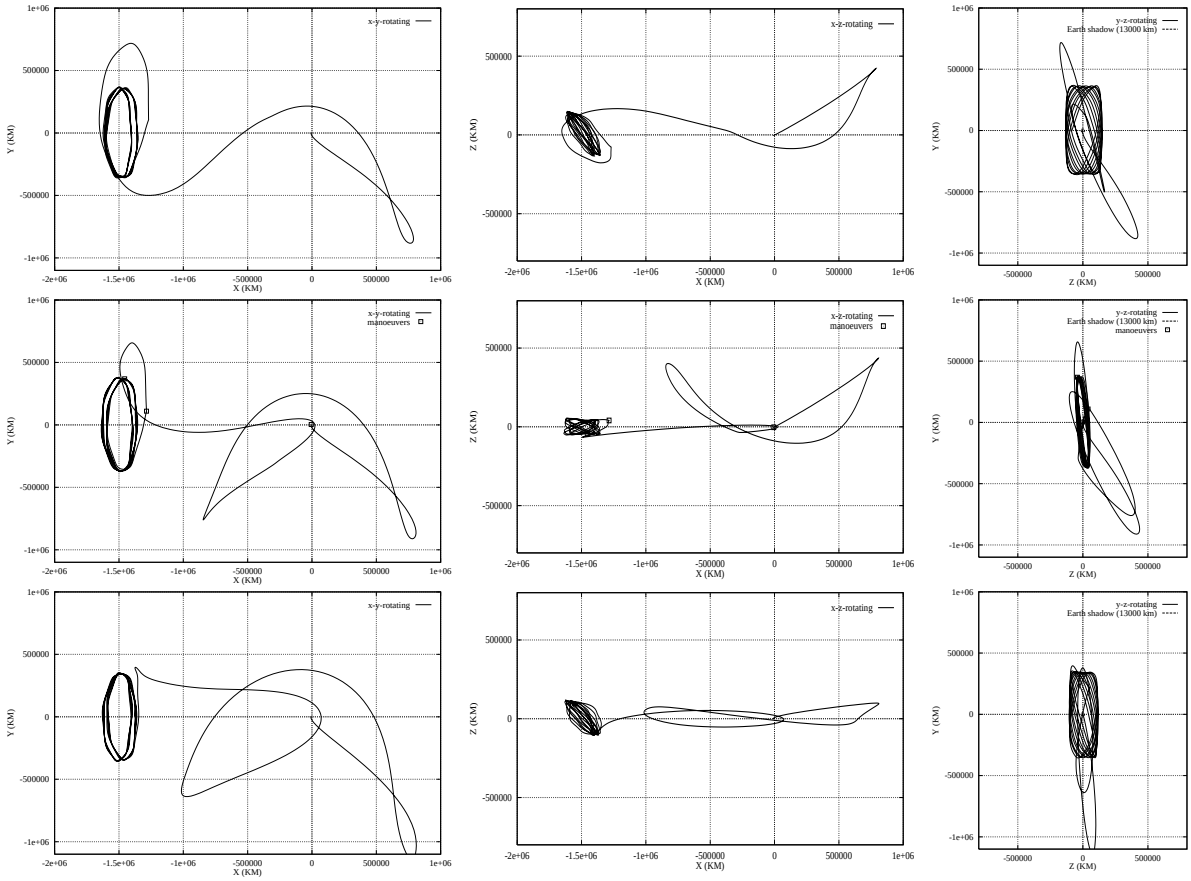


Figure 5.5: One-manoeuvre transfer and Lissajous orbit for various launch dates: (a) 2009–01–01; (b) 2009–01–03; (c) 2009–04–09.

A large number of simulations has been performed, demonstrating that with 1 mm s^{-1} Doppler and 10 m range tracking from three ground stations a position accuracy of about 1 km along track and cross track and 10 m in the Sun-Earth direction (all 1σ) can be reached (cf. Section 5.4 and Table 5.3). With one ground station and assuming Doppler only to increase the telemetry data rate, the orbit determination orthogonal to the viewing directions remains the same but the knowledge along the Sun-Earth line becomes much worse (10 km). So ranging will be necessary at least for parts of the communications time.

From the tracking the unstable deviations from the desired motion around L2 will be observed. Deviations will have to be removed without too much time delay, and the spacecraft has to be targeted back to a quasi-periodic motion around L2. Several orbit maintenance strategies to achieve this have been tested: (i) classical interplanetary navigation with a shifting target position somewhere in the future; (ii) Linear Quadratic Control, as suggested in the phase A for SOHO by Rodriguez-Canabal (1984), which minimises some sum-square deviation from a reference trajectory; (iii) removal of the velocity component along the unstable direction; (iv) re-computation of future periodic trajectory at each manoeuvre time.

5.3.4 Mission Time-Line

GAIA will be launched by Ariane 5 together with another spacecraft, as lowest passenger, into a standard Geostationary Transfer Orbit. GAIA can be launched in the 45 minutes standard midnight GTO launch window every day except for a few days in the year. After separation from the launcher, solar array deployment, spin rate and axis control will be performed autonomously by the spacecraft.

A perigee boost manoeuvre of about 800 m s^{-1} applied in 3 burns in successive orbits will inject GAIA into the transfer orbit to L2. During the transfer only stochastic orbit correction manoeuvres (30 m s^{-1} allocated) will be performed, controlling the effects of errors induced along the highly

sensitive orbit. The spacecraft will be spin stabilized during the transfer (5 rpm). Because of the midnight launch window condition the initial line of apsides orientation happens to be opposite to the direction one would normally inject to. Therefore the transfer will be along a trajectory passing through the weak stability boundary in the Earth-Sun system, passing from the Sun side of this boundary to the night side where the final orbit will be reached.

After typically 200–280 days an orbit insertion manoeuvre of up to 200 m s^{-1} (depending on the launch day) will inject the spacecraft onto the stable manifold of the operational orbit. The exact combination of the in-plane and out-of-plane amplitudes and phases in this orbit will depend on the launch day, to minimise the insertion manoeuvre. For some launch dates one more out of plane manoeuvre may be necessary within a few month from the insertion manoeuvre. From the insertion, about once per month orbit correction manoeuvres (3 m s^{-1} per year allocated) will be necessary, to counteract the random effects on the intrinsically unstable orbit.

Nominal operations will commence immediately after the insertion manoeuvre. Before that the spacecraft and the instruments will be completely checked out. The mission will be nominally completed 5 years from insertion. Consumables are allocated for 6 years.

5.4 Ephemeris Requirements

The astrometric analysis requires certain *a priori* knowledge of the position and velocity vector of the satellite with respect to the Solar System barycentre (see Figure 5.6). The observation of minor planets also places constraints on the advanced knowledge of their positions, and might be the limiting factor for the astrometric accuracy of these targets. A detailed investigation into the requirements on the position and velocity vector of the Earth, of the satellite, and of observed minor planets, has been undertaken by Mignard (1998), and is summarised in Table 5.3.

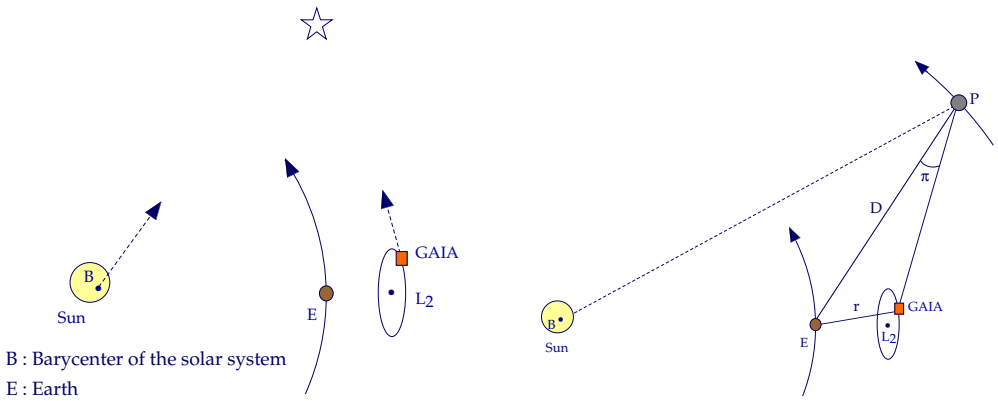


Figure 5.6: Left: the position of GAIA with respect to the Earth and the solar system barycentre. Right: The geometrical configuration during the observations of minor planets.

(a) Earth and satellite position: the observed parallactic displacement of a star depends both on the stellar parallax (classically defined as the inverse of the barycentric distance) and on the barycentric satellite distance. The latter can be decomposed into the satellite-Earth and Earth-barycentre displacements. In the most demanding case, taking $\pi = 1$ arcsec and $\sigma_\pi = 10$ microarcsec, the satellite-barycentre distance needs to be known *a priori* at the level of 10^{-6} AU = 150 km. At present, the Earth orbit is known with a relative accuracy of 10^{-8}, while a level of 10^{-10} is realistic in the near future (P. Bretagnon, private communication). The same requirement, $\sim$ 150 km, must be met for the orbital determination, i.e. to a relative accuracy of 10^{-4}.

(b) Earth and satellite velocity: astrometric observations are conventionally referred to an observer at rest with respect to the solar system barycentre by performing a Lorentz transformation, or 'correction for stellar aberration' (needed, for example, to compare observations made at different times in order to compute the parallax). Again, the satellite velocity vector is determined in two steps: firstly, with respect to the Earth by fitting the geocentric orbital parameters to (mainly) range or Doppler measurements; secondly, the geocentric velocity is transformed to the barycentric frame by adding the Earth velocity from a solar system ephemeris. The whole procedure must be carried out in a general-relativistic framework. The size of the correction is of the order of v/c, where v is

Table 5.3: Requirements on ephemerides for stellar and planetary observations. In the latter case, the positional limits refer to the case when the astrometric results are given in a geocentric frame; if the satellite frame solution is used, there are no specific constraints.

Requirements on:	Stars		Planets		
	Earth	GAIA	Earth	GAIA	Planets
Position (km)	150	150	–	0.15	30
Velocity (mm s^{-1})	1.5	1.5	1.5	1.5	–

the barycentric velocity of the observer and c the velocity of light. In order to compute the correction to within $\delta\alpha = 1$ microarcsec (5×10^{-12} rad) the satellite barycentric velocity vector is needed to within $c\delta\alpha \sim 1.5$ mm s^{-1}. Such a requirement is appropriate even for a target accuracy of 10 microarcsec at 15 mag, partly because of the guiding principle of maintaining systematic error sources well below the mean standard errors, and also because these errors are expected to be surpassed for bright objects. Today, the Earth ephemeris computed at the Bureau des Longitude in Paris is available with (more or less a random) error of 2 mm s^{-1}. This error has two main sources: truncation in the analytical development used to represent the Earth's motion, and the customary use of internationally-agreed but incorrect planetary masses (e.g., the IAU 1976 system). According to Moisson (1997) an improvement of the ephemerides to somewhat better than 0.1 mm s^{-1} is realistic within a few years. A similar requirement is needed for the satellite motion with respect to the Earth.

(c) Solar System objects: the observations of solar system objects, mainly minor planets of the asteroid main belt, is a challenging objective, requiring the parallax and aberration of each object observed. Details are given by Mignard (1998), taking into account the finite size of the bodies, and including a discussion of pre-mission observations. The conclusion is that the position of minor planets with respect to the satellite itself will have to be published, along with geocentric positions of lower accuracy.

Finally, consideration must be given to the relative orientation of the planetary ephemeris frame with respect to the extragalactic VLBI frame (ICRS), to which the observations will be tied by direct observation of radio-optical quasars. The present uncertainty of the relative orientation of the two frames is 3 mas, representing a (highly systematic) error of 0.4 mm s^{-1} in the Earth's velocity vector in the VLBI frame.

5.5 Radiation Environment

The radiation environment presents a major problem to space systems. The environment generally consists of geomagnetically trapped charged particles, solar protons, and galactic cosmic rays. It is the penetrating particles that provide the main problems, which include upsets to electronics, payload interference, damage to components and deep dielectric charging. The radiation belts encircle the Earth and contain particles trapped in its magnetic field. An inner belt contains mainly energetic protons with energies up to several hundred MeVs. This belt is reasonably stable in time and extends to a geocentric distance of 4 Earth radii. The outer belt consists primarily of energetic electrons of energies up to a few MeV. By contrast, this outer electron belt is highly dynamic, being subject to storms and injection events that follow solar-terrestrial disturbances. The electron belt extends to near the magnetopause at about 10 Earth radii. Solar protons are products of solar events, with energies in excess of several hundred MeV and peak fluxes in excess of 10^6 protons cm^{-2} s^{-1} for protons with energies greater than 10 MeV. These events, though, are relatively rare, occurring primarily during periods of solar maximum activity, which commences about 2.5 years before Sun spot maximum and lasting for seven years. The duration of individual events is usually of the order of days. The large fluxes of energetic protons and heavier ions can contribute a large dose, increase upset rates in electronics and increase radiation induced background noise in detectors.

Cosmic rays originate outside the solar system. Fluxes of these particles are low, but as they include heavy, energetic ions of elements such as iron, they cause intense ionisation as they pass through matter, and are difficult to shield. The solar wind serves to attenuate the fluxes, providing an anti-coincidence with the solar cycle. Single event upsets and latch ups are the dominant effects of cosmic rays on spacecraft electronics.

Calculating the period of solar maximum activity for solar cycle 24, during which GAIA will be launched, is subject to uncertainties being more than a solar cycle in the future. If the duration of the solar cycle 23 is 11 years, then the sunspot number for cycle 24 should peak approximately 11 years after the predicted peak for solar cycle 23, or in the year 2011. Given that the solar maximum period starts 2.5 years prior to sunspot maximum, as defined by the JPL model, a five-year mission starting in 2009 would take place during a period of solar maximum.

The geosynchronous orbit is subject to the solar proton, cosmic ray and trapped particle environments. While almost entirely out of the trapped proton belt, the trapped electron belt provides the bulk of the radiation dose for thinner shields, while Bremsstrahlung radiation from the electron belts dominates for the thicker shields. Geomagnetic shielding of the solar proton and cosmic ray particles has been ignored, as it is negligible at geosynchronous altitudes. For L2, only the solar proton and cosmic ray spectra have been considered in this analysis, for a five year mission during a period of solar maximum activity. However, the complexity of the trapped particle radiation environment will require a further analysis of the transfer phase when the ephemeris of the transfer is available.

The trapped radiation models employed were those developed in the early 1970s by Vette (1989) for NASA, i.e. AE8 and AP8 for electrons and protons respectively. These models were used to provide orbit averaged particle spectra after mapping the satellite orbit into geomagnetic co-ordinates through the Jenson & Cain (1960) geomagnetic field model and the Cain 120-term 1966 IGRF geomagnetic field model projected to 1970. The solar flare proton model used was that developed at JPL by Feynman et al. (1990). This uses a data set spanning three solar cycles, and is replacing the older King model as the standard solar flare model for spacecraft engineering. A 90 per cent confidence level was used to calculate the solar proton fluence for a period of five years, corresponding to the duration of the mission that coincides with a period of solar maximum activity. Mission average spectra are used to calculate the dose deposition in a minuscule silicon target as a function of spherical aluminium shell shielding thickness using the SHIELDOSE code (Seltzer 1980). This simplified particle transport code is well-suited for routine dose predictions in situations where the geometrical and compositional complexities of the spacecraft are not known.

The particle spectra are also used to calculate solar cell damage equivalent fluences of 1 MeV electrons as a function of cover glass thickness. These damage fluences are provided by the EQFRUX and EQFRUXGA codes, which are for silicon and gallium arsenide (GaAs) solar cells, respectively (Tada & Carter 1982; Anspaugh 1996). Infinite cell back shielding is assumed and a 10 MeV proton to 1 MeV electron equivalence ratio of 3000 is used for maximum power degradation. For body-mounted solar cells, the infinite back shielding condition applies. However, for 'wing' type solar cells, rear incidence irradiation should be considered. A lightweight panel construction, e.g. as used on HST, could result in a significant back incident radiation contribution through a lack of adequate shielding. The Cosmic Ray Effects on MicroElectronics (CREME) suite of programs provides a comprehensive set of Linear Energy Transfer (LET) spectrum, including the treatment of geomagnetic shielding and material shielding. An upset/hit rate computation is based on the path length distribution in a sensitive volume. A simplified-trapped proton induced SEU rate is also provided by the model. Calculating these upset rates, though, requires detailed information about the electronic components concerned.

Outside of possible solar flares, L2 predictions of $\sim 4-6$ cosmic ray protons cm^{-2} s^{-1}, and ~ 0.5 cosmic ray alpha particles cm^{-2} s^{-1} are derived using the well-established CREME96 models (which provide integral and differential spectra for both particle species), with some modulation from solar max to solar min conditions. Contribution from heavier ions is negligible in comparison. These fluxes are isotropic. The total count rate of particles traversing through a convex volume (such as a CCD) can be expressed as $R = (AF)/4$ where A is the volume area, and F the isotropic flux. Since the CCD thickness is very small compared with the overall area, we have $A = 174+174$ cm^2 for ASM1. Consequently the cosmic ray count rate would be of the order of 500 events per second. A rough estimate of the number of electron-hole pairs induced by the cosmic rays can be made using the CREME96 results. The mean energy of cosmic rays is very high, of the order of 1–2 GeV. At these energies they essentially penetrate the satellite regardless of the shielding. The energy loss of such a particle in the Si substrate is ~ 4 MeV cm^{-1}, or 400 eV μm^{-1}, which compared with ~ 3.6 eV required to generate an electron-hole pair, yields a the value of about 100 e$^-$ per micron. Assuming a 5 micron mean track length (depending on the exact CCD thickness) yields on average about 500 e$^-$ per pixel per cosmic ray event. The situation is somewhat complicated by the fact that a part of the energy deposit will go to high-energy knock-off electrons (delta rays), that have complex, jagged tracks within the detector and will carry off some of the energy lost locally by the primary cosmic ray. Exactly how many of those electrons end up disturbing the measurement depends on the electrical characteristics of the CCD (recombination times etc). These naturally vary depending on the CCD type, manufacturer, and batch. Results are shown in Figures 5.7–5.8. Further information concerning the space environment and its effects on spacecraft can be found in ECSS (1997).

Based on the ISO detector experience, codes are available to calculate (on a purely geometrical basis) the distribution of cosmic ray particle track lengths within a volume (CCD or pixel), and the distribution of the number of affected pixels in a CCD as a result of a large number of cosmic ray events. These, in turn, can be translated to first-approximation charge deposit histograms. Future analysis will be based on the CCD material and pixel dimensions finally selected. This code does not consider the number of secondary particles generated, both in the vicinity of the instrument and in the CCD itself. Recent work at Saclay, also based on ISO detectors, concluded that 70–80 per cent of observed glitches were due to secondaries. This is highly dependent on the instrument and satellite geometry, and on the materials chosen. If cosmic rays are the main concern, it is appropriate to avoid large amounts of heavy, high-Z materials surrounding the detector. Based on HST experience about 15–30 events s^{-1} occur for a TI CCD with 800×800 15×15 μm^2 pixels, i.e. for 1.44 cm^2. As one (faint star) ASM vertical strip (ASM1, 2 or 3)

corresponds to 174 cm^2, a number of cosmic ray events as high as 4000 per second would be expected based on an extrapolation of the HST data. However HST, at about 600 km altitude, encounters (in addition to energetic cosmic rays that come from the Solar System) magnetospheric electrons and protons, particularly at the South Atlantic Anomaly. Therefore the total HST cosmic ray count rates are higher than for GAIA, which at L2 is well out of Earth's magnetosphere.

5.6 Mission Operations Concept

ESOC will conduct the Mission Operations of the GAIA mission. Mission operation comprises (a) spacecraft operations, consisting of mission planning, spacecraft monitoring and control and all orbit and attitude determination and control; (b) scientific instrument operations, mainly consisting of the implementation of the pointing schedule and collection and data quality control of the science telemetry. ESOC will prepare a Ground Segment that comprises all facilities, hardware, software, documentation, trained staff, which are required to conduct the mission operations.

Mission Planning, Spacecraft Monitoring and Control The GAIA mission-planning concept follows the lines of the Planck mission. The scanning law will remain unmodified throughout the mission. Nominal spacecraft control during the routine mission phase will be 'off line'. Only one Ground Station (Perth) will be used. The contacts between the Mission Control Centre at ESOC and the spacecraft, except for collecting payload and housekeeping telemetry, will be primarily used for pre-programming of the autonomous operations functions on the spacecraft (up-link of master schedule), and for science data collection. Anomalies will normally be detected with a delay. The payload will also mainly be operated by ESOC. The health of the scientific instruments will be permanently monitored from the Housekeeping telemetry and necessary control actions will be taken following the same procedures as for the spacecraft subsystems. Part of the payload telemetry will be made available in near real time to the science consortium via public networks. The bulk of the science data will be distributed on some mass storage device directly from the ground station. Commands received from the scientists (e.g. calibrations etc. which require processing of the science telemetry) will be implemented.

Orbit and Attitude Determination and Control Orbit determination during all mission phases will use two-way range and coherent two-way Doppler tracking data from up to three ground stations (one station during routine operations). Orbit determination includes the calibration of all main engine manoeuvres and the calibration of thrusters. The sequence of manoeuvres from the initial geostationary transfer orbit until insertion into the Lissajous orbit will be optimised to minimise propellant consumption and taking into account all operational conditions. The orbit maintenance manoeuvres will be optimised to guarantee payload operations for up to five years with a minimum of interruption. The autonomous functions of the attitude control system onboard will be monitored and calibrated. The attitude manoeuvre schedule for the payload-provided pointing sequence will be implemented.

5.7 Ground Segment Facilities

Ground Operations Facilities consist of: (a) the ground stations and the communications network (hardware and software); (b) the Mission Control Centre (infrastructure, computer hardware); (c) the Flight Control System (data processing and flight dynamics software); (d) the spacecraft simulator. All mission and flight control facilities, except the ground stations, will be located at ESOC, including the interfaces for the provision of science telemetry to the users and processing of commands from the users.

Ground Stations and Communications Network The 35-m ESA station at Perth will be used for contact with the spacecraft during all mission phases. During the first 10 days of the mission and during critical mission phases the ESA 15-m station at Kourou will be used in addition.

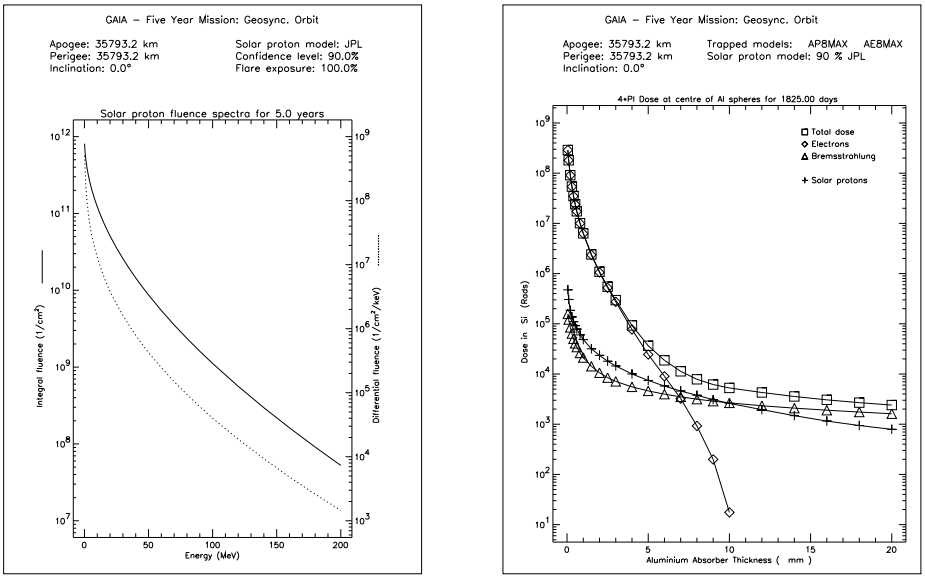


Figure 5.7: Left: Solar proton fluence spectrum for a five-year mission. Geomagnetic shielding has not been included. This spectrum is applicable to both geostationary and L2 orbits. Right: Dose in an infinitesimal silicon chip as a function of spherical shielding thickness for a five-year geosynchronous mission, starting in 2009. The solar proton contribution is also applicable to the L2 orbit.

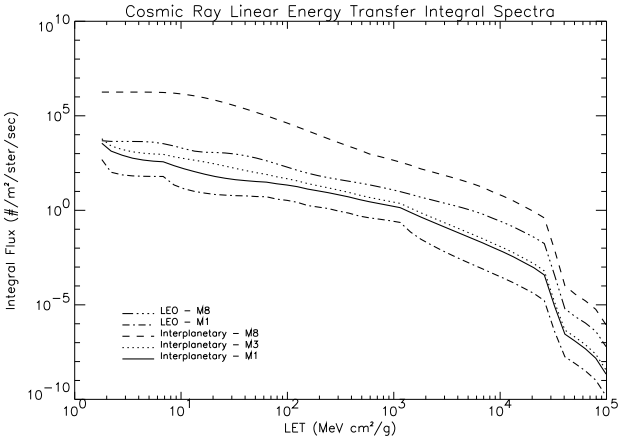


Figure 5.8: Interplanetary Linear Energy Transfer spectra for various conditions. The ECSS-E-10-04 standard recommends the use of the $M = 3$ environment for cosmic ray effects analyses. This environment is specified as the 10 per cent worst case environment with a singly ionized anomalous component. This environment, however, does not include solar particle events. For solar particle event ion effects, the $M = 8$ environment is recommended. This is the peak flux with worst-case composition from the 10 per cent worst case event; more severe environments are considered unrealistic.

The 15 m station at Villafranca will be available as backup. All ESA stations interface to the mission operations centre at ESOC in Darmstadt via the OPSNET communications network. OPSNET is a closed Wide Area Network for data (tele-command, telemetry, tracking data, station monitoring and control data) and voice.

The foreseen telemetry data rate from the spacecraft to the ground stations is 3 Mbps. Though it may be assumed that communication lines will be affordable in the mission time frame (after 2009) such that these rates can be transmitted without intermediate storage to the control centre, a concept has been adopted in which the bulk data is written to advanced bulk storage devices such as Digital Video Disks (DVDs) or Advanced Intelligent Tapes (AITs), and shipped directly to the science data processing consortium from the ground station.

Mission Control Centre Satellites operated by ESOC are controlled from the Mission Control Centre which basically consists of the Main Control Room augmented by the Flight Dynamics Room and dedicated Control Rooms and Project Support Rooms. During periods around major mission, events mainly during the Launch and Early Orbit Phase, the Main Control Room will be used for GAIA mission control. During the science operations phase when GAIA is in its final orbit around L2 and also during low activity periods in the transfer the mission control will be conducted from a Dedicated Control Room.

The control centre is equipped with workstations giving access to the different computer systems used for different tasks of operational data processing. The computer configuration used in the Mission Control Centre for the GAIA mission will be derived from existing structures. All computer systems in the control centre will be redundant with common access to data storage facilities and peripherals. They will be connected by a Local Area Network to allow transfer of data at sufficient speed and to allow joint access to data files. The external connection to the authority responsible for science data processing and the resulting command input, uses commercial/public networks.

The Flight Control System Software A Flight Control System based on infrastructure developments, using a distributed hardware and software architecture, for all spacecraft monitoring and control activities will be established including: mission Planning facilities for the preparation of command time-lines based on inputs for experiment operations; telemetry reception and analysis and distribution facilities; tele-command processing facilities; orbit and attitude determination and manoeuvre facilities.

The Spacecraft Software Simulator The simulator will provide the same telemetry/tele-command interface to the flight control system as used during the operations of the actual spacecraft. It will have to simulate realistically the behaviour of the spacecraft in reaction to its environment and to the tele-commands it receives. The GAIA simulator will be based on ESOC's Software Infrastructure for Modelling Satellites (SIMSat). SIMSat provides all the general purpose features, which are common to all simulators developed at ESOC.

Before launch, the simulator will be used to test and validate the flight control system software, to test and validate the flight operations procedures, both in nominal and in contingency situations; and to train the operations personnel. After launch, the simulator will be used to test and validate modifications to the control system software and the flight operations procedures or new procedures developed e.g. to cater for unforeseen anomaly situations; and to train new staff.

Part III
Mission Performance Assessment

6 Assessment Methodology and Basic Assumptions

6.1 Overview of Performance Assessment

One important objective of the studies coordinated by the Science Advisory Group was to assess the overall performance of GAIA in relation to its scientific goals. This must include not only a basic assessment of the astrometric, photometric and radial-velocity accuracies, but also of their implications in terms of the expected number of stars observed to given accuracy levels, the diagnostic powers of the resulting data (such as photometric indices), and limitations arising from the complexities of the real sky (such as stellar multiplicity, crowded regions, non-uniform background). This complements and extends the somewhat independent accuracy assessments made by the industrial contractors (MMS for the baseline design, and Alenia for the interferometric design given in Appendix B), which are however fully consistent with the present analysis.

A schematic representation of the performance assessment is given in Figure 6.1. As illustrated by the figure, the assessment can be seen as a mechanism for converting all the various input elements and assumptions into quantifiable statements concerning the scientific capabilities of the mission. Given, in addition, a set of scientific priorities, this mechanism could in principle be used to optimize any feature of the instruments such as the detailed sampling scheme for each CCD, or the allocation of detector area to different photometric bands.

In reality this cannot be fully achieved for a number of practical reasons, the most important being the current incomplete knowledge of what the sky looks like as seen by GAIA (which is of course what GAIA should find out). Thus, while the assessment in general follows the strategy depicted in Figure 6.1, the optimisation of the various elements is generally based on more restricted assumptions and often guided by experience or dedicated simulations.

The present section contains a summary of the technical and astronomical assumptions entering into the performance analysis, and a discussion of several other elements indicated in Figure 6.1. Details of the accuracy estimation methods are given in Section 7, while the actual results are in Section 8. The main conclusion from the performance analysis is that the original goal of 10 μas astrometric accuracy at 15 mag can be reached, and that virtually all stars and other point sources brighter than 20 mag can be observed with significant accuracies.

The accuracy assessments are presented with a considerable degree of confidence: the principles follow those successfully implemented for the Hipparcos mission, and calibration or 'noise' terms are expected to be fully calibrated by virtue of the very large number of independent observations made for each object.

6.2 Improvement with Respect to Hipparcos

It is useful to summarise qualitatively why such a dramatic advance in astrometric accuracy over Hipparcos can be achieved with the present instrument design. First, it should be recalled that the typical accuracies are not simply given by the diffraction limit of the optics: at 500 nm wavelength, and with a primary mirror diameter of about 0.3 m, the Airy radius for Hipparcos was about 400 mas. Yet the parallax accuracy was better than 1 mas for many of the stars. Final accuracies are defined by the achievable precision on the determination of the image *centroid* for a single observation, multiplied by a factor related to the number of observations (and observational geometry) throughout the measurement period.

There are three main components involved in the improved performance of GAIA compared with Hipparcos. The enlarged optics (the primary mirror aperture is 1.7 m in the scan direction, and 0.7 m across scan) provides both a smaller diffraction pattern ($\lambda/D \sim 70$ mas in the scan direction) and a significantly larger collecting area (by more than a factor of 30). The enlarged collecting area, and the improved quantum efficiency and bandwidth of the detector (CCD rather than photocathode), all contribute to improved photon statistics, significantly extending the limiting magnitude as well as the precision of the individual centroiding measurements. Finally, use of the

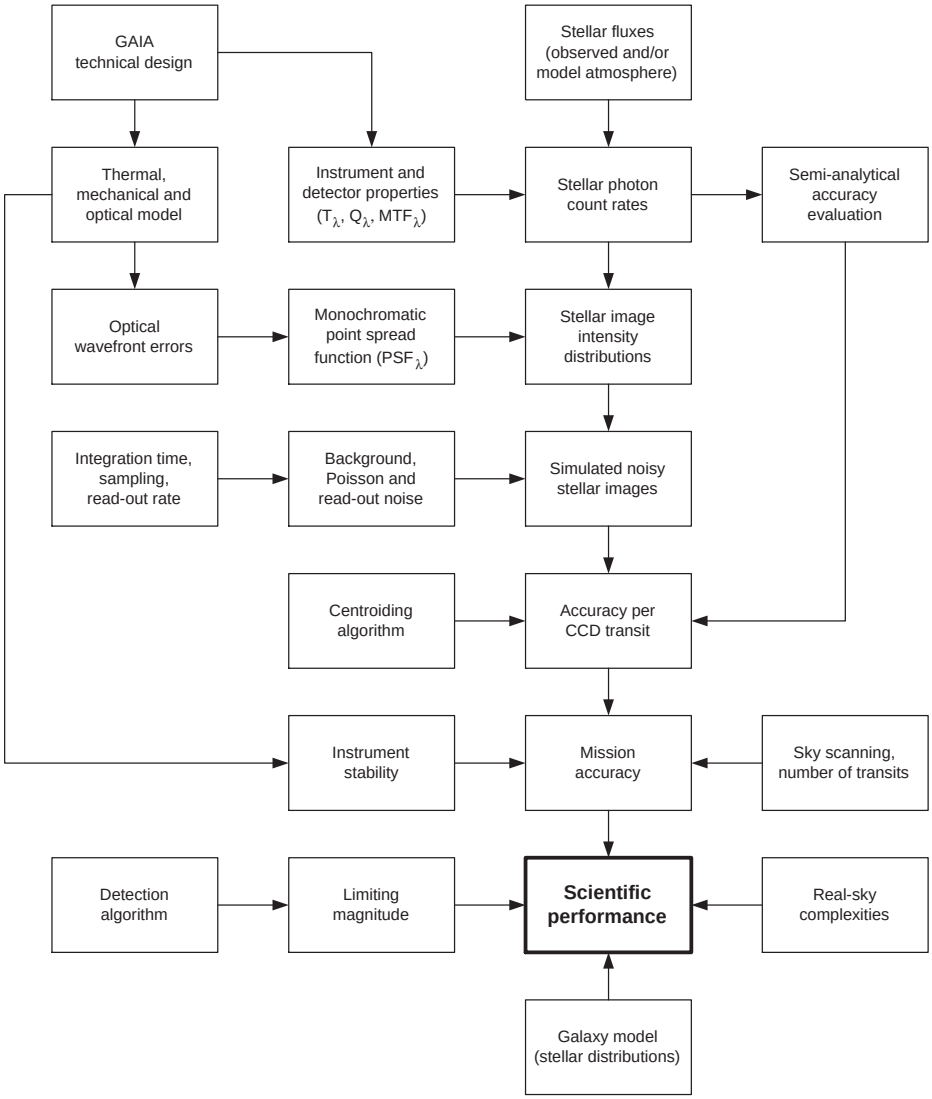


Figure 6.1: GAIA performance assessment: schematic representation of the main elements and their interrelations.

CCD detector results in important multiplexing advantages: the parallel observations of all objects crossing the focal plane results in improved photon statistics for each object (the available crossing time is not partitioned between competing stars as in the case of the sequential observations of the Hipparcos image dissector tube detector), as well as the possibility of observing *every* object down to the instrument's detection limit. A variety of other improvements complete the main aspects of the comparison summarised in Table 6.1.

6.3 Summary of Technical Assumptions

Detailed technical data for the current baseline design of GAIA are given in Part B. For convenience, the main parameters relevant for the performance assessment are summarised in Table 6.2. Specifically, these are the input data for the basic accuracy estimations described in Section 7, and yielding the results described in Section 8.

Table 6.1: The total gain in parallax accuracy compared with Hipparcos, of about $\sim$ 150, depends on the combination of many factors, which are summarised in this table. Since GAIA is optimised to observe much fainter stars than Hipparcos, stars of different magnitude (15 versus 10 mag) are compared, which gives a disadvantage for GAIA by the factor 0.1. The 'grid disadvantage' for Hipparcos refers to the intrinsic inefficiency of the grid modulation technique of Hipparcos compared with the direct imaging in GAIA (Equation (38) in Lindegren (1978)). GAIA has the additional multiplexing advantage that many ($\sim 10^4$) objects can be observed simultaneously. However, this reflects mainly in the total number of objects observed not in their individual accuracies.

Quantity	Hipparcos	GAIA	Gain on σ_π^{-1}
Magnitude compared	10 mag	15 mag	0.1
Optical resolution (λ/D)	400 mas	70 mas	6
Total collecting area	0.062 m^2	2.38 m^2	6
Wavelength bandwidth (FWHM)	200 nm	500 nm	1.5
Peak quantum efficiency	0.08	0.84	3
Detector MTF	$\simeq 1$	0.4	0.4
Optical transmittance[1]	0.25	0.90	2
Observing time per transit	4 s	14 s	2
Scan geometry ($\sin\xi$)	0.68	0.82	1.2
Relative sphere rigidity[2]	1.2	1.0	1.2
Effective mission length[3]	2.2 yr	4.0 yr	1.3
Grid disadvantage[4]	4	1	4
Total gain (rounded)			200
Actual/predicted accuracy[5]	1.5 mas	10 μas	150

[1] Includes grid bands and diffraction losses for Hipparcos
[2] Based on condition numbers estimated by Makarov (1998)
[3] Includes dead time in radiation belts for Hipparcos
[4] For the same number of detected photons and pupil size
[5] Predicted accuracy for GAIA includes 20 per cent margin

6.4 Basic Astronomical Data

6.4.1 Stellar Fluxes

Calculation of photon fluxes for objects of known magnitudes, and their subsequent conversion to detected photoelectron rates, is fundamental for the estimation of accuracies, detection probabilities and limiting magnitudes. An effort was made to obtain correct estimates of fluxes based on real stellar energy distributions, including strongly reddened spectra.

The relative energy distribution in various (real) stellar spectra were taken from a standard stellar spectrophotometric atlas (Gunn & Stryker 1983). For much of the accuracy analysis, 16 (de-reddened) spectra were selected to cover the main sequence from B1V to M8V and giants from G8III to M7III. The relative fluxes (originally expressed in energy per frequency interval) were converted to absolute fluxes (in ph m^{-2} s^{-1} nm^{-1}) for apparent magnitude $V = 0$, using the condition that the energy flux at $\lambda = 550$ nm should be 3.66×10^{-11} W m^{-2} nm^{-1} (corresponding to 1.0133×10^8 ph m^{-2} s^{-1} nm^{-1}) for un-reddened stars of spectral type A0V (Table 21 of Straižys 1992). For other spectral types and for reddened stars, the fluxes were scaled to give the same wavelength-integrated energy flux as the un-reddened A0V star, when weighted by the V band sensitivity curve in Table 2 of Bessell (1990).

Table 6.2: Main technical parameters and their assumed values in the performance assessment.

Parameter	Value (and designation in Section 7)
Mission length:	
total length	5.0 yr
effective observing time	$L = 4.0$ yr
Sky scanning:	
Sun/spin-axis angle	$\xi = 55°$
speed of spin axis relative to stars	1441 deg yr^{-1} = 0.1644 arcsec s^{-1}
inertial spin rate	$\omega = 120$ arcsec s^{-1} (3 hr period)
high-frequency pointing noise	< 7 μas rms ($f > 0.05$ Hz)
Astrometric instrument (Astro):	
rectangular pupil size	$D = 1.70$ m along scan, $H = 0.70$ m across scan
effective focal length	$F = 50$ m
detector solid angle per telescope	$\Omega = 0.305$ deg^2 (astrometric field [AF01–AF16] only)
optical transmittance	$T_\lambda = 0.9$ at all wavelengths (λ)
optical wavefront error	$W = 33.3$ nm rms (Table 3.2)
CCD quantum efficiency	Q_λ, see Table 3.15
CCD along-scan MTF	see Table 3.15
CCD pixel size	$\Delta\xi = 37.1$ mas (9 μm) along scan, $\Delta\eta = 111$ mas (27 μm) across
CCD chip size	$N_\xi = 2780$ pixels (25 mm) along scan, 2150 pixels (58 mm) across
integration time per chip	$\tau_1 = N_\xi \Delta\xi/\omega = 0.86$ s
sample sizes (pixels)	2×2 (ASM), 1×8 (AF01–AF16)
total readout noise	$r = 10.9$ e^- (ASM1), 6.0 e^- (ASM3 and AF01–AF16)
Astrometric assembly (Astro-1 & -2):	
number of telescopes	$N_i = 2$
basic angle	$\Gamma = 106°$
basic angle stability	6.4 μas rms
Spectrometric instrument (Spectro):	
rectangular pupil size	$D = 0.75$ m along scan, $H = 0.70$ m across
effective focal length	$F = 4.17$ m
optical aberrations	negligible
photometric field (SSM and MBP):	
detector solid angle	2 deg^2
optical transmittance	$T_\lambda = 0.9$ at all wavelengths (λ)
CCD quantum efficiency	Q_λ, see Table 3.15 for CCD#1B (CCD#2 also used)
CCD pixel size	10×10 μm^2 = 0.5×0.5 arcsec^2
sample size (pixels)	1×1 (SSM), 1×4 (MBP)
total readout noise	5.3 e^- (SSM), 3.0 e^- (MBP)
Radial Velocity Spectrometer (RVS):	
detector solid angle	1.0 deg^2
total efficiency	$T_\lambda Q_\lambda = 0.35$
wavelength range	850–875 nm
CCD pixel size	20×20 μm^2 = 1 arcsec $\times$ 0.075 nm
sample size (pixels)	1×1
total readout noise	3.0 e^-

For application of interstellar extinction the curve in Table 3 of Straižys (1992) was used, with linear interpolation in $1/\lambda$. For given spectrum, V magnitude and extinction, it is then possible to compute the monochromatic stellar flux f_λ expressed in ph m^{-2} s^{-1} m^{-1}.

By means of the sensitivity curves for the (Cousins) R and I bands in Bessell (1990), the R and I magnitudes of the stars could also be obtained, adopting values of $V{-}R = -0.01$ and $V{-}I = -0.02$ for the un-reddened A0V star (Table 27 of Straižys 1992). This allows, if required, to parametrise the accuracy results by colour index rather than by spectral type and extinction.

These flux calculations based on a sample of real spectra were checked against theoretical fluxes derived from model atmospheres based on the Kurucz ATLAS9 code. Differences in the broad-band responses were typically in the range ±4 per cent for most stars.

6.4.2 The GAIA Magnitude Scale

Signal-to-noise ratios, detection limits and the astrometric precision for observations in the astrometric fields basically depend on the stellar photon flux folded through the instrument transmittance T_λ and the quantum efficiency Q_λ curve for the relevant CCD type (CCD#1B; see Table 3.15):

$$e = D \times H \times \int_0^\infty f_\lambda T_\lambda Q_\lambda \, \mathrm{d}\lambda \qquad [e^- \ \mathrm{s}^{-1}] \tag{27}$$

The corresponding wavelength band $T_\lambda Q_\lambda$ is very wide, and does not correlate well with any of the standard photometric pass-bands such as B, V, R or I. The effective wavelength varies from below 500 nm for O and early B stars to more than 800 nm for late M stars and strongly reddened stars. At G0V it is about 600 nm. It is useful to introduce a special magnitude scale G defined by the GAIA (astrometric) passband, thus:

$$G = -2.5 \lg(e/e_0) \tag{28}$$

Following standard photometric practice, the zero point is fixed such that $G \simeq V$ for un-reddened A0V stars. Using the stellar flux calculations described above, and assuming the quantum efficiency of CCD#1B and a constant transmittance of 0.9, the normalization constant for the astrometric field was found to be $e_0 \simeq 4.0 \times 10^{10} \ e^- \ \mathrm{s}^{-1}$.

The approximate transformation from $(V, V-I)$ to G can be expressed in the form:

$$G = V + 0.51 - 0.50 \times \sqrt{0.6 + (V-I-0.6)^2} - 0.065 \times (V-I-0.6)^2 \tag{29}$$

which is valid (within ± 0.1 mag) at least for $-0.4 < V-I \lesssim 6$.

It can be noted than $G \simeq V$ in a rather wide colour range: for $-0.4 < V-I < 1.4$ we have, according to Equation 29, $G-V = 0.0 \pm 0.1$ mag. This is understandable since the central wavelength of the G band (i.e. the response curve for CCD#1B, Figure 2.12) is close to 550 nm, as for the V band.

6.4.3 Sky Brightness

For the basic astrometric accuracy analysis, a mean sky brightness of $V = 21.0$ mag arcsec^{-1} was assumed, with a Solar-type spectral distribution ($V-I \simeq 0.7$). The equivalent GAIA magnitude is $G = 21.1$ mag arcsec^{-1}. For most of the sky this assumption is very conservative. The sky background measured with HST at high ecliptic latitudes is $V = 23.3$ mag arcsec^{-1} and $I = 22.0$ mag arcsec^{-1} (Windhorst et al. 1994), equivalent to $G = 23.3$ mag arcsec^{-1}. In the ecliptic (but $> 60°$ from the Sun) the sky background is 1–2 mag brighter due to zodiacal light (HST/WFPC2 Instrument Handbook). The much higher brightness observed in some areas of the sky is discussed in Section 6.5.

Much of the sky is however considerably darker, and the accuracy assessment for photometry and radial velocities has been based on the more realistic (fainter) sky brightness of $V = 22.5$ mag arcsec^{-2}.

The sensitivity of both the astrometric and photometric errors to the assumed sky background is detailed in Section 8.

6.4.4 Galaxy Model

Models of the Galaxy, from the simplified star-count models to the complete evolutionary models, aim to provide a realistic representation of the Milky Way. The star-count models (Bahcall & Soneira 1980; Gilmore 1984; Chen 1997 among others) are 'static' models that provide the expected number of stars of a given magnitude in a determined direction of the sky via the integration of the fundamental equation of stellar statistics. On the other hand, the population synthesis models result in a detailed simulation of Galactic evolution, and thus luminosities and colours are

decoupled from the age-density distribution. Among others, the models of Robin & Crézé (1986); Haywood et al. (1997) and the Padova group (Ng et al. 1995) are included in this category. The Padova model contains a very detailed treatment of the stellar populations towards the Galactic centre, distinguishing them by their age, metallicity and spatial distribution. The inclusion of the kinematic properties of each population in a Galaxy model (e.g. Robin & Oblak 1987; Ratnatunga et al. 1989; Méndez & van Altena 1996; Chen 1997) leads to a prediction of the distribution of proper motions and radial velocities. The goal of GAIA in this area is to replace such plausible models by measurements.

To evaluate the number of stars to be observed by GAIA, to estimate the distribution of errors on their kinematic parameters, and to assess the overlapping, crowding and sky background due to faint stars, the Chen (1997) star-count and kinematic model was updated and applied (Torra et al. 1999). This model contains three main components: disk, thick disk and spheroid; a bulge population is not considered separately, but is included in the spheroid component. A detailed treatment of spheroid and disk white dwarfs is also included (Figueras et al. 1999). Using V magnitudes and $V-I$ colours, the photometric transformation established by Equation 29 allows prediction of the number of stars as a function of the GAIA G magnitude. The most critical points in this prediction are: (a) the extinction law used: the Arenou et al. (1992) law, the more recent model of Hakkila et al. (1997) and the Chen et al. (1999a) law, based on the COBE/IRAS all sky reddening map can be used in the model. Depending on the extinction law the estimated number of stars varies by about 10–15 per cent; (b) the poor knowledge of the faint end luminosity function (LF). The Wielen et al. (1983) single-star LF used in this model provides only lower limits for stars fainter than $M_V = 12$ mag; (c) binary stars have not been treated in the model; (d) the spatial distribution model is known to be simplistic.

6.4.5 Prediction of Star Counts

In Table 6.3 the number of stars predicted to be observable by GAIA using the Hakkila et al. (1997) and the Arenou et al. (1992) extinction laws is presented. The total of ~ 1.1 billion stars corresponds to about 1–2 per cent of the total stellar content of the Galaxy. The estimated number of stars observed by GAIA belonging to the three Galactic populations (disk, thick disk and spheroid) shows that the main scientific goal, a representative census of the Galaxy, can be achieved. More detailed predictions of the model are presented in Tables 6.4–6.6 for different stellar populations and types of stars, where the Hakkila et al. (1997) extinction law has been used.

Table 6.3: Total number of stars to $G = 20$ mag ($V \simeq 20$ mag) predicted by the Galaxy model.

	Hakkila et al. (1997) extinction law	Arenou et al. (1992) extinction law
Disk (MS + WD)	779×10^6	977×10^6
Disk (Giants)	92×10^6	115×10^6
Thick disk	97×10^6	104×10^6
Spheroid	67×10^6	68×10^6
Total	$1\,035 \times 10^6$	$1\,264 \times 10^6$

In addition, there are many rare (in the sense of low numbers) objects of high astrophysical interest for which star count predictions cannot be obtained reliably from a Galaxy model (Cepheids, RR Lyrae, planetary nebulae, AGB stars, Wolf-Rayet stars, brown dwarfs, etc.). Special consideration also has to be given to the numbers of variable stars to be detected by GAIA. A preliminary estimation, taking into account the predictions of the Galaxy models, expected photometric errors, and the detection threshold of variability for the foreseen time sampling and noise distribution as a function of magnitude, indicates that GAIA will observe of order 50 million variable stars, including some 10 million periodic variables.

Table 6.4: Star counts for populations and spectral types, predicted by the Galaxy model, as functions of limiting magnitude.

	$G \leq 15$	$G \leq 17$	$G \leq 20$
All	25×10^6	120×10^6	1035×10^6
Disk (MS + WD)	14×10^6	70×10^6	780×10^6
Disk (Giants)	6×10^6	25×10^6	91×10^6
Thick disk	4×10^6	17×10^6	97×10^6
Spheroid	1×10^6	9×10^6	67×10^6
O–B	2×10^5	5×10^5	2×10^6
A	2×10^6	6×10^6	51×10^6
F	6×10^6	24×10^6	226×10^6
G	8×10^6	47×10^6	400×10^6
K	9×10^6	41×10^6	299×10^6
M	5×10^5	3×10^6	55×10^6
White dwarfs (disk)	4×10^2	5×10^3	212×10^3
White dwarfs (halo)		1×10^1	1×10^3

6.4.6 Globular Clusters

Globular cluster fields are among the densest on the sky, and therefore provide an important indication of the limiting performances of GAIA in terms of measurable star densities. In order to estimate their observability, simulations were carried out using a grid of King models for values of the concentration $0.5 < c < 2.5$, core radius $R_C = 1$ arcmin, and $N_* = 10\,000$. The densities at the centre, at $3R_C$, and at $10R_C$, were then computed. A simulation of the real densities in each of the 140 globular clusters compiled by Harris (1996) was then computed by interpolating the grids at the true value of R_C, and by normalising according to the actual number of stars in the cluster. This number was obtained using the mass of the cluster taken from the literature or interpolated as a function of c and the central surface brightness, and by using the mass function of M92 as being applicable for all clusters. The resulting uncertainties on the number of stars and the densities (at the centre, at $3R_C$, and at $10R_C$) is probably more than a factor of 5–10 due to these simplifying assumptions.

The maximum density that the astrometric field of the satellite can handle is about 0.25 stars arcsec^{-2} (i.e. about 3 million stars per square degree). Figure 6.2a gives the star densities in the centre, at $3R_C$, and at $10R_C$ for the 140 clusters of the Harris Catalogue. One third of the clusters will be fully observed by GAIA, two thirds will be observable at $\geq 3R_C$, and only a dozen will have a density higher than 0.5 stars arcsec^{-2} at $10R_C$. The number of stars with $G \leq 20$ mag is given in Figure 6.2b, and lies between 100–100 000 stars per cluster. The number of observable stars, i.e. in areas of density smaller than 0.25 star per square arcsec, is smaller but similar. The size of the area into which the density will be too high is given in Figure 6.3. The radius of the area is smaller than 1 arcmin for the vast majority of the clusters and is larger than 3 arcmin for only 5 clusters. These values translate to a few R_C. Such regions will be observable by HST, NGST, and SIM.

6.5 Real-Sky Complexities

The assumptions outlined in Section 6.4 essentially describe the 'nominal' sky, which is the basis for accuracy predictions presented in Section 8. The 'real' sky is however vastly more complex, and it is necessary to discuss to what extent this will affect the overall mission performance. Although GAIA is designed to make a global astrometric and photometric survey of the sky to a limiting magnitude around $G = 20$ mag, this does not imply that every square arcmin of the sky needs to be covered to a uniform detection limit and accuracy. The main science goals, which depend on spatial, kinematic and photometric mappings of large stellar samples, can in general be achieved

Table 6.5: Star counts for populations and spectral types, predicted by the Galaxy model, as functions of limiting distance.

Distance limit (pc)	$r < 100$	$r < 200$	$r < 500$	$r < 1000$
All	417.5×10^3	3.0×10^6	20.5×10^6	66.0×10^6
Disk (MS + WD)	407.1×10^3	2.9×10^6	19.8×10^6	63.1×10^6
Disk (Giants)	1.2×10^3	9.9×10^3	97.6×10^3	0.5×10^6
Thick disk	8.7×10^3	65.6×10^3	542.4×10^3	2.2×10^6
Spheroid	0.5×10^3	3.7×10^3	33.9×10^3	0.2×10^6
O–B	0.1×10^3	0.5×10^3	3.7×10^3	0.02×10^6
A	2.5×10^3	16.3×10^3	121.9×10^3	0.5×10^6
F	8.9×10^3	68.5×10^3	651.2×10^3	3.1×10^6
G	15.7×10^3	126.5×10^3	1.3×10^6	6.7×10^6
K	38.7×10^3	321.1×10^3	3.4×10^6	17.7×10^6
M	341.9×10^3	2.4×10^6	14.8×10^6	37.7×10^6
White dwarfs (disk)	9.8×10^3	46.7×10^3	154.5×10^3	209.3×10^3
White dwarfs (halo)	0.06×10^3	0.2×10^3	0.7×10^3	1.0×10^3

Table 6.6: Star counts (per deg^2), predicted by the Galaxy model, as functions of Galactic latitude b and limiting G magnitude.

$\|b\|$	$G \leq 15$	$G \leq 17$	$G \leq 20$
$0 - 5°$	1.6×10^3	9.1×10^3	101.6×10^3
$5 - 10°$	1.4×10^3	7.4×10^3	79.8×10^3
$10 - 20°$	0.8×10^3	3.9×10^3	31.2×10^3
$20 - 30°$	0.5×10^3	2.1×10^3	11.4×10^3
$30 - 90°$	0.3×10^3	0.9×10^3	3.8×10^3
$0 - 90°$	0.6×10^3	2.9×10^3	25.1×10^3

with incomplete data, provided that the selection function is known, and that the relevant objects are not exclusively found in 'difficult' areas.

Despite the many expected complexities of real objects, the GAIA instruments are extremely well adapted to the accurate survey-type observations required for its science. This is a consequence of the high-angular resolution, with a nearly optimal sampling of the point spread function, coupled with the flux-limited performance for the majority of stars. To account for un-modeled complexities in the general accuracy evaluation, an error margin of 20 per cent is presently included in the astrometric and photometric accuracy evaluations. Further studies will quantify this.

6.5.1 Double and Multiple Stars

The well-known fact that many 'stars', on closer scrutiny, turn out to be double or multiple systems is perhaps the single most conspicuous deviation from the basic assumptions outlined in Section 6.4. The experience from Hipparcos was that stellar multiplicity (in which we include duplicity) was a problem for about 10 per cent of the objects, in the sense that no sensible solution would have been obtained for these objects, had their non-single nature been ignored in the data analysis. For another 10 per cent, multiplicity was noticeable in the data but could have been ignored without much impact e.g. on the parallaxes. The remaining 80 per cent were effectively seen as 'single' stars by Hipparcos.

With the much higher accuracy of GAIA it might be suspected that multiplicity will be a much more severe problem than for Hipparcos, so that perhaps a majority of the objects will require

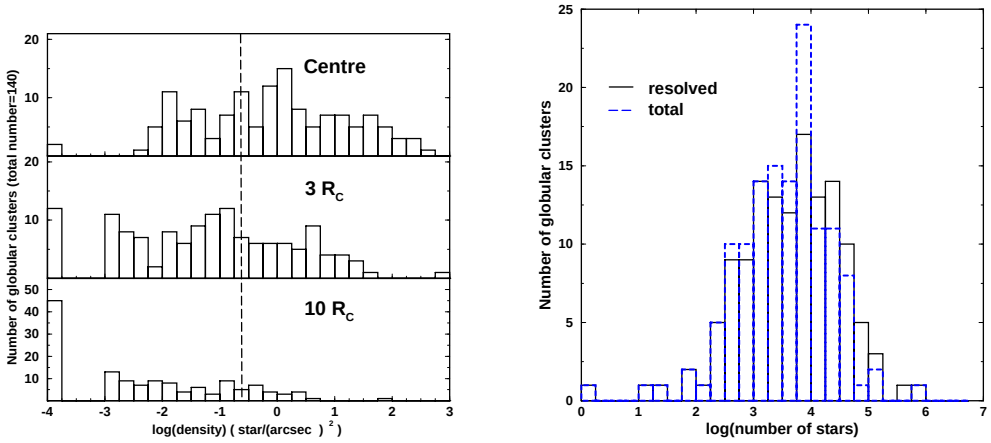


Figure 6.2: Simulation of GAIA performance in Galactic globular clusters, for the 140 clusters from the Harris Catalogue: (a) histograms of the star densities at the centre, at $3R_C$, and at $10R_C$. The dashed line marks the maximum density that the astrometric field of the satellite can handle (0.25 stars per square arcsec); (b) total number of stars and number of resolvable stars per globular cluster.

special consideration. Simulations of GAIA double-star observations (Section 1.5.1) indicate that the discovery frequency will indeed be very high for the bright and relatively nearby systems, but for the mission as a whole the fraction of stars concerned will not be radically different from the Hipparcos case. This is simply a consequence of the GAIA stars being much more distant, on the average, than the Hipparcos stars. Further study will be required in this area.

The categorization of GAIA double and multiple stars will roughly follow that for the Hipparcos double stars (see Section 1.4.2 of Vol. 1 of ESA SP–1200), appropriately scaled to the resolution and accuracy of GAIA:

- systems with very small separations ($\lesssim 5$ μas to 500 μas, depending on magnitude) are effectively single from the viewpoint of the GAIA astrometry;
- systems with somewhat larger separations, up to 10–20 mas, are still unresolved and consequently observed by their photocentres. Depending on the period, mass ratio and magnitude difference, the photocentre motion may or may not reveal the true nature of the systems. Non-linear proper motion will however be detected in a large fraction of them, and in many cases a complete orbit for the photocentre can be determined;
- for separations in the range from $\simeq 20$ mas to $\simeq 1$ arcsec, the components are resolved but have partially overlapping images. Multi-epoch, combined astrometric and photometric solutions are mandatory to disentangle the components. Since many of these systems also show appreciable orbital motion, and many have additional (resolved or unresolved) components, the complete analysis may be quite complex, but also rewarding in terms of astrophysical results (e.g. individual component masses);
- well-separated systems ($\gtrsim 1$ arcsec) with a long period can again be treated as effectively single, although they may be recognized as physical pairs from their common proper motion and parallax. Systems with shorter periods will have curved motions and therefore require common treatment similar to the previous category (for instance, the mass ratio can be derived from the curvature ratio even if the complete orbits cannot be determined);
- additional complications arise in multiple systems and in systems with variable components, e.g. variability-induced photocentre motions in unresolved pairs.

Photocentre motions from the orbits of unresolved binaries will in some cases bias the parallax estimates (for periods close to one year) and more often the proper motions (for moderately long periods). While the effects can be very significant in exceptional cases, the overall statistics are less affected. The proper motion effect, when translated to an error in the tangential velocity, is equivalent to the radial-velocity error caused by long-period orbital motion (typically a few km s^{-1}).

The basic accuracy results given in Section 8 refer to effectively single stars. A somewhat degraded performance must be expected for complex objects. However, as shown by Hipparcos, un-complicated binaries need not cause more than a marginal degradation e.g. of the parallax, provided they are correctly modeled. This will be an interesting area for future study.

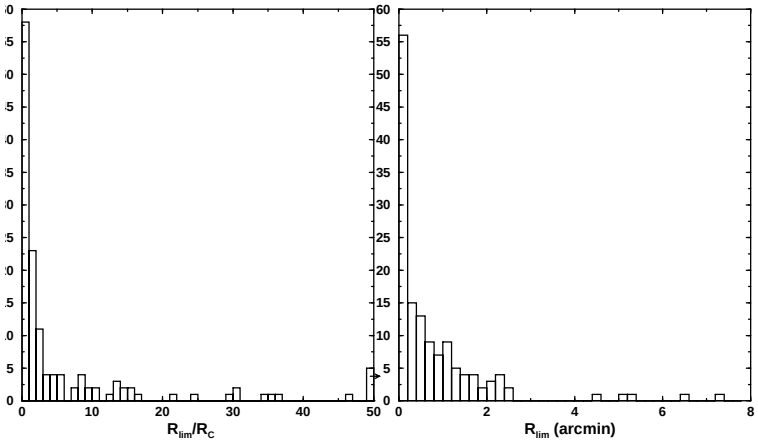


Figure 6.3: Distribution of the radii of the area over which the densities in globular clusters are larger than the critical density of 0.25 stars per square arcsec, expressed (a) as a multiple of R_C; (b) in arcmin.

6.5.2 High-Density Areas

The Galaxy model described in Section 6.4.4 predicts a smooth variation of star densities with sky coordinates, with a maximum density of $N_{20} \simeq 400\,000$ stars deg^{-2} towards the Galactic centre. (N_{20} is the density of stars with G magnitude ≤ 20.) The real sky is characterised by highly irregular density fluctuations on a wide range of angular scale lengths. Arbitrarily high densities can be found by looking at sufficiently small areas. Among the more extended regions with a very high star density are the central Bar of the Large Magellanic Cloud ($N_{20} \simeq 1 \times 10^6$ stars deg^{-2}) and Baade's Window ($l \simeq +1°$, $b \simeq -4°$; $N_{20} \simeq 2 \times 10^6$ stars deg^{-2}). At the latter density the point-spread functions in the astrometric instruments cover approximately 1 per cent of the area. This is still somewhat below the practical limit where image crowding and the combined contributions of the many more fainter stars (down to $G \simeq 25$) will have a significant impact on the performance. The on-board data handling is therefore designed to handle short intervals with very high star densities and resulting high data rates (Section 3.6). Small regions with even higher densities, as in the centres of globular clusters (Section 6.4.6), cannot be properly observed by GAIA, and will remain the preserve of HST, NGST, and SIM..

It would be useful to quantify more precisely the irregularities in the star distributions, e.g. the fraction of sky with star densities above a certain limit. Unfortunately, no existing stellar catalogue has sufficient depth, resolution and photometric precision to be of direct relevance to GAIA. Some useful statistics can however be derived from the USNO A1.0 catalogue (or the improved version A2.0). The catalogue contains blue and red magnitudes of 488 million stars. This corresponds in number to all stars with $G \lesssim 19$ mag, but the cut-off is not sharp and differs over the sky so that care must be exercised in the use of the catalogue. The catalogue is constructed from scans of Schmidt plates going somewhat fainter south of $-32.5°$ than to the north. Because of the modest resolution, it is not suited for areas around bright stars, dense clusters and bright nebulae; all such areas will appear empty in A1.0 and may be confused with empty sky patches. Also, multiple stars will often be recorded as single. The catalogue gives a more realistic picture of the star density and its variations over the sky down to $G \simeq 19$ mag than any of the available Galaxy models.

Stars in A1.0 were counted in 18.6 million bins of 4×2 arcmin^2 each. The size of the bins was chosen to match the astrometric CCDs of GAIA. In order not to underestimate the frequency of high densities, bins with no stars were assumed to have a very high density. For the entire catalogue, 1 per cent of the bins contain more than 200 stars per 8 arcmin^2 and 0.4 per cent contain more than 300 (Figure 6.4). A rough extrapolation to a total number of 10^9 stars (the expected content of the GAIA catalogue) gives the corresponding densities $N_{20} \simeq 200\,000$ and $300\,000$ stars deg^{-2}, respectively. These numbers are similar to the maximum star densities predicted by the Galaxy model, but apply to a smaller fraction of the sky than in the model. Thus it can be concluded that the Galaxy model probably does not underestimate the total area of high-density regions at the 8 arcmin^2 resolution. This is a useful constraint for dimensioning of the on-board data handling.

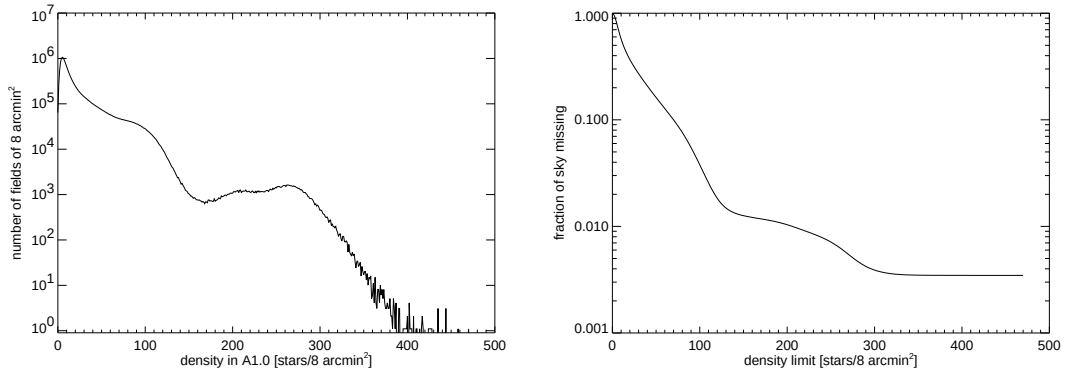


Figure 6.4: Two representations of star densities in the USNO catalogue A1.0 (containing stars with $G < 19$ mag): (a) number of fields of 8 arcmin^2 versus density; (b) fraction of sky versus star density limit.

Table 6.7: Diffuse Galactic sky background. Approximate size of the area in deg^2 in the zone $|b| < 15°$ exceeding a given surface brightness in the R band. The second column is an estimate from visual inspection of Figure 67 R of Leinert et al. (1998), the third from a comparison of Figures 66 and 67 of the same paper. The figures are based on data with a resolution of $0.25° \times 0.25°$.

Sky brightness in R (mag arcsec^{-1})	Area of sky in the zone $\|b\| < 15°$ (deg^2)	
	$278° < l < 314°$	$0° < l < 360°$
< 19.5	10	30
< 19.9	70	200
< 20.3	450	1400
> 20.3	630	9400
Total area	1080	10800

6.5.3 Sky Background and Galaxies

For the basic astrometric accuracy assessments a uniform sky brightness of $G = 21.1$ mag arcsec^{-2} was assumed (Section 6.4.3). This corresponds to $\simeq 5$ e^- per CCD sample (1×8 pixels) in the astrometric field. The associated Poisson noise is small compared with the readout noise ($\simeq 6$ e^- rms) and the Poisson noise from the star signal ($\simeq 10$ e^- rms in the centre of a star image at $G = 20$ mag).

Small-scale irregularities Higher and irregular surface brightness is found in astrophysically important objects such as galaxies and star-forming regions, and around some stars with associated nebulosity. The effects of these backgrounds on the GAIA observations vary depending on the brightness and the spatial structures on scales from the angular resolution of GAIA ($\lesssim 0.1$ arcsec) up to a few arcsec. General statements are difficult to make as each case presents a unique scenario. For the astrometric measurements (image centroiding), brightness gradients on the scale of the width of the point spread function (FWHM $\simeq 0.1$ arcsec) are the most critical. For example, a gradient of 1 count sample^{-1} in the along-scan direction (e.g. a variation from $G = 21.1$ mag arcsec^{-2} to $G = 20.7$ mag arcsec^{-2} over 0.1 arcsec) will induce a systematic displacement of 0.1 mas in the centroid for a star of $G = 18$ mag, 0.25 mas for $G = 19$ mag, and 0.6 mas for $G = 20$ mag. While the position would thus be biased, this is not necessarily the case for the proper motion and parallax, which depend on differential shifts over (usually) much shorter angular scales than the width of the point spread function.

Table 6.8: Statistics of galaxies versus integrated I magnitudes: N_c = density of galaxies brighter than the given magnitude, r_{hl} = mean half-light radius, $\mu_{I,av}$ = mean surface brightness inside the half-light radius, and R_{hl} = the fraction of the sky inside the half-light radii of all galaxies brighter than the given magnitude.

I (mag)	$N_{\rm c}$ (deg^{-2})	$r_{\rm hl}$ (arcsec)	$\mu_{I,\,\rm av}$ (mag arcsec^{-2})	$R_{\rm hl}$
14	2.1	6.19	20.0	0.000066
15	7.4	3.93	20.0	0.000096
16	27	2.49	20.0	0.00014
17	97	1.58	20.0	0.00020
18	350	1.00	20.0	0.00029
19	1 100	0.66	20.1	0.00041
20	2 900	0.44	20.2	0.00053
21	7 500	0.29	21.3	0.00067
22	17 300	0.23	21.7	0.00082
23	35 400	0.17	21.2	0.00099
24	68 800	0.14	21.6	0.0012

Diffuse Galactic sky background Table 6.7 gives the estimated areas in the Galactic plane in which the diffuse Galactic light exceeds a given level, based on data from Leinert et al. (1998) with a resolution of 0.25°. $R = 19.5$ mag arcsec^{-2} corresponds to $G \simeq 20.1$ mag arcsec^{-2}, i.e. 1 mag brighter than the sky background assumed for astrometry (Section 6.4.3). This will be noticeable in the observations of the faintest stars, but affects only a very small fraction of the sky.

Galaxies Vaccari & Høg (1999b) summarise the average properties and frequencies of E and S galaxies relevant for GAIA, including an estimation of the cumulative galaxy counts $N_{\rm c}$, the mean half-light radius $r_{\rm hl}$, the mean surface brightness inside the half-light radius $\mu_{I,\,\rm av}$ of these types of galaxies in the I band, and the cumulative fraction of the sky $R_{\rm hl}$ inside the half-light radii of these galaxies. The resulting values are given in Table 6.8. The cumulative galaxy counts in this estimation are consistent to within 30 per cent with those obtained in the red r_f band by Metcalfe et al. (1995), when a correction $r_f = I + 0.5$ mag for the different photometric band is introduced. A preliminary estimate obtained from these values and the characteristics of the detection system in the astrometric instruments indicates that galaxies of total magnitude $I = 16.5$ mag can normally be detected. Thus, the total number of galaxies potentially observed by GAIA is at least 1 million. To this should be added the fainter galaxies (including active galactic nuclei and quasars) for which a compact core with $G \lesssim 20$ mag can be detected essentially as the stars.

The last column in Table 6.8 shows that only a small fraction of the sky resides inside the central parts of galaxies where the detection and measurement of stars would be seriously disturbed. Galaxies extend far outside the half-light radius, but the disturbance of stars and background is much smaller there, taking into account that a background value is estimated for each star from samples very close to the star. The typical surface brightness e.g. at radius $3r_{\rm hl}$ is only about $I = 23.6$ mag arcsec^{-2}, much fainter than the general sky background.

6.5.4 Polarization

The faint magnitude limit of the GAIA survey implies that most of the (faint) targets will be polarized by the interstellar medium to a substantial degree. This may have effects on photometry and spectroscopy. In the 6-channel photometer used by the Copenhagen group (Knude, private communication) dependence on polarization is minimised by introduction of a Lyot-type de-polarizer. Similarly, the dispersive element in the radial velocity system must be considered if equivalent widths are to be measured; Brewster angles must be avoided in the folded beam. The effects of polarization are not taken into account in the astrometric accuracy estimation.

Table 6.9: The main scintillation parameter values for the interstellar and interplanetary media in the radio and optical ranges. $\theta_{\rm scat}$ is the rms scattering angle, m the scintillation index (a measure of the depth of modulation of the intensity fluctuations), τ the de-correlation time. The proportionality relation between those parameters and the frequency ν is indicated.

Medium	ν or λ	$\theta_{\rm scat}$ (arcsec)	m (dimensionless)	τ (sec)
Proportionality		$\propto 1/\nu^2$	$\propto 1/\nu^2$	$\propto \nu$
Interstellar	318 MHz	10^{-3}	1	10^3
	700 nm	6×10^{-16}	2×10^{-6}	1.4×10^9 ($\sim$ 45 yrs)
Interplanetary	200 MHz	10^{-3}–10^4	1	0.2
	700 nm	2×10^{-16}–2×10^{-9}	2×10^{-14}–2×10^{-7}	4×10^5 ($\sim$ 5 days)

6.5.5 Scintillation

Scintillation is, from the point of view of astrometry, a perturbing phenomenon occurring when a signal propagates through a turbulent transparent medium. Changes in the density distribution of the medium give rise to fluctuations of the index of refraction and, thus, to a possibly time-dependent phase change and scattering. Significant scintillation is observed in the atmospheric, ionospheric, interplanetary and interstellar media, all of which, though at a very different scale, show fluctuations in (gas or electron) density. Atmospheric scintillation is known to be a major limiting factor for all ground-based observations (in the optical range, the rms angular deviation is of order 1 arcsec), whereas interplanetary and interstellar scintillation have been measured and studied mainly in the radio range, through the analysis of the spectral scattering and time irregularities of pulsar signals.

GAIA, due to its orbital location, is only in principle affected by the interstellar and interplanetary media. No complete theory exists for interstellar scintillation where both the source and the observer are immersed in an extended medium, but the theory of thin-screen scattering (Lang 1971; Lang 1978) is a simple and adequate approach assuming that the scattering is produced by a thin phase-changing screen located along the line of sight between the source and the observer. The estimates of this theory closely parallel those obtained by the more complicated extended-medium analysis (Rickett 1970) and are consistent with the observed statistical measurements of pulsar radiations. Moreover, it is applicable to the interplanetary and atmospheric medium by adapting the size, wind velocity and density parameters of the thin-screen model.

Table 6.9 summarises the results of a detailed study by Vannier (1998b). At 700 nm, both the interstellar and the interplanetary scintillation have the following characteristics: a scattering angle far lower than the value relevant for GAIA astrometry ($\theta_{\rm scat} \ll 10^{-6}$ arcsec) and a very low scintillation index ($m \ll 1$). The very weak modulations of the signal intensity due to scintillation should not be measurable, whatever time scale is considered. Moreover, the worst values correspond to the case of the interplanetary scintillation in regions with high electron densities, i.e. close to the Sun. These areas will not be pointed at by GAIA. In conclusion, these typical values indicate that the interstellar and interplanetary scintillation phenomena are not relevant for microarcsec astrometry in the optical range.

6.6 Imaging Capabilities of GAIA

GAIA will provide a multi-colour, multi-epoch sampling of the entire sky, of great astrophysical interest, in spite of some rather complex characteristics. The sky sampling is *non-uniform* (some parts of the sky are scanned more frequently than others), *anisotropic* (the optical resolution and sampling density is higher in the scan direction than across the scan, and the scan direction varies between the different observations of a given point on the sky), *heterogeneous* (the sampling may be different depending on the instrument and filter bands), and *incomplete* (not all the information is sent down, but only small patches of the sky centred on each detected object, and possibly compressed information for more extended areas, such as mean background levels and other statistics which may be computed on-board).

The non-uniformity and anisotropy are direct consequences of some very basic characteristics of the mission, namely the scanning law and the optimisation of astrometric measurements along the scan. The heterogeneity and incompleteness are caused by practical constraints of the present technical design: firstly, the limited telemetry rate, which prevents sending down more than a small fraction of all the potential pixel values; secondly, the need to reduce

readout noise by on-chip binning (constructing samples out of several adjacent pixels) and window-mode operation (quickly skipping the unwanted pixels). Concerning the on-chip binning it is necessary to find a good compromise between resolution and noise, and for the window-mode operation there is similarly a compromise between the amount of useful information and the data rate. It is hardly surprising that this complex optimisation problem has not yet found a definitive solution. The current sampling scheme, summarised in Figure 3.7, is consistent with the overall science goals and current technical design (including a realistic telemetry rate), but will require further optimisation in view of several of the complications discussed in Section 6.5.

The transmitted samples permit two-dimensional mapping of the patches of sky surrounding each detected object. Subject to the complex characteristics described above, this mapping can reach a considerable sensitivity as a result of the accumulation of observing time during the mission. For instance, extended light sources can be detected and measured photometrically to a limit of $\simeq 20$ mag arcsec^{-2} (Figure 6.5), and faint stars ($G < 23$) in the vicinity of a detected object can be located and measured out to a distance of 1 arcsec (using the extended samples of AF17) or 3 arcsec (using the Spectro Sky Mapper). There are several reasons for performing this mapping:

- faint but potentially disturbing stars near the target object (including physical companions) can be detected and taken into account in a refined astrometric and photometric analysis;
- the mapping provides initial estimates of the positions and intensities of point sources, essential for the subsequent detailed analysis, which is strongly non-linear (see below);
- the mapping can identify general features such as background irregularities which are not seen in individuals scans, and thus help to select an appropriate background model;
- there are also instances where the background features themselves are astrophysically interesting, or where GAIA could trigger more detailed observations with other instruments.

Full astrometric and photometric accuracy is only achieved by very careful modeling of the individual CCD data samples in terms of specific point sources and an adequate background model. This is essentially done by a weighted least-squares (or maximum-likelihood) fitting of relatively few object and background parameters to the many observed sample values. The estimation problem is strongly non-linear in the positions of the point sources, and thus requires a good initial approximation, including prior identification of all relevant point sources. It basically requires to fit the complete assembly of condition equations similar to Equation 87 to all the samples in a given patch, collected throughout the mission, with the source positions (κ_j, μ_j) and intensities a_j expressed in terms of the unique astrometric and photometric parameters of the relevant objects.

This estimation process differs fundamentally from traditional image processing techniques (e.g. tomography or more general deconvolution algorithms), which typically estimate a large number of intensity values on a pre-defined mesh. Such techniques are applicable to the mapping of sky patches from GAIA samples, but not to the final estimation of point-source parameters.

6.7 Simulation of GAIA Observations

In the preparation of the GAIA mission several kinds of simulations play a critical role. On the one hand, existing catalogues and synthetic models of the sky are used to predict star and galaxy counts. These counts are needed for a number of satellite design considerations (maximum star density, telemetry rate, etc.), for an evaluation of the mission accuracies, and to predict the mission's expected scientific output. Synthetic models of the Galaxy are a key ingredient for this because there is presently no high angular resolution multi-colour sky survey as complete and deep as the one that will be produced by GAIA, and because many astrophysical parameters needed for the analysis of performances (i.e., the 'true' distances of the stars) cannot be obtained otherwise. On the other hand, instrument data simulations are used for developing and testing the algorithms to be used for data processing on-board and on-ground: the detection of stars entering the instrument fields, the centroiding of individual astrometric observations, the analysis of photometric data and the derivation of radial velocities from spectra are studied using CCD images obtained from computer simulations (sometimes including information from real images of the sky) or experimental test setups in laboratory. These issues are discussed in this section.

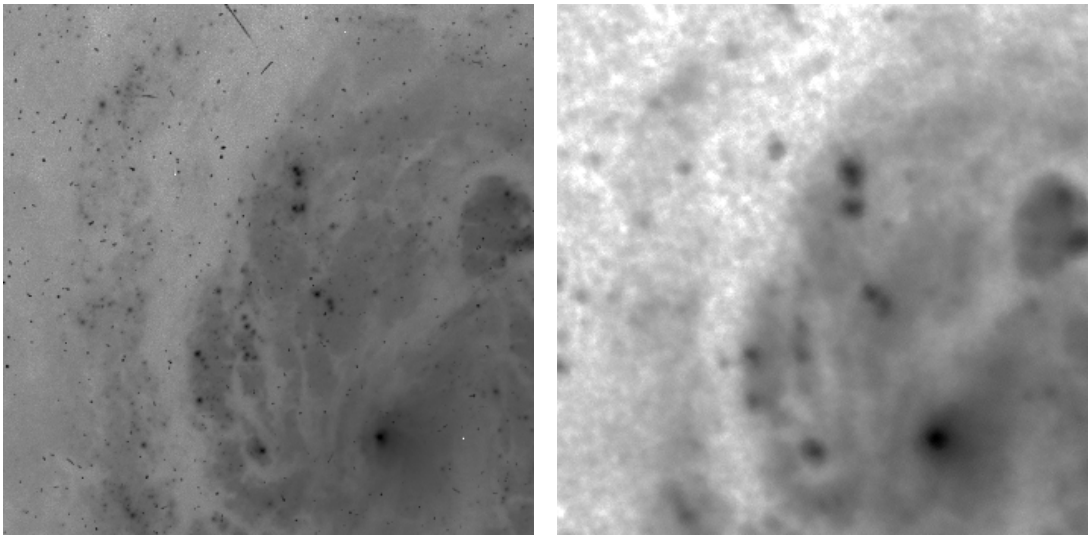


Figure 6.5: Images of the Sc galaxy M100. $V \sim 9.3$ mag, and the half-light radius given in the RC3 Catalogue is 104 arcsec. The central part of the galaxy is at the lower right, North is towards the upper left, and the field size is 16×16 arcsec^2. (a) HST image, 900 s exposure, filter F555W corresponding to V; (b) GAIA image (flux map) reconstructed from 50 simulated observations in the V band of the broad-band photometer in Astro-2, i.e. 45 s total exposure time. The stellar images have a FWHM = 350 mas, corresponding to the image quality in an excellent ground-based site during the short intervals of very good seeing. The simulated GAIA observations were based on the HST image. The conspicuous lumpiness especially in the faint part of the GAIA background is produced by features in the HST image (from Vaccari & Høg 1999a).

6.7.1 Simulation of Raw CCD Data

All GAIA instruments (astrometric telescopes, radial velocity instrument, and medium-band photometer) use CCDs as detectors. Thus, the simulation of their observations is reduced to the problem of simulating CCD images and their processing to obtain the final data. A set of tools for GAIA image simulation is being developed by teams at the *Observatoire de Paris-Meudon* and the *Universitat de Barcelona*. These tools can be accessed and used through a dedicated web server.

The tools allow the production of simulated images by adding different components: backgrounds, stellar fields, extragalactic objects and real images (using HST images obtained through the Hubble Data Archive). The combination of these elements provides a wide range of base images that can then be adapted to specific simulations by selecting values of a set of adjustable parameters: orientation of the satellite scans, image size, photometric band, etc. The generation takes into account the characteristics of the GAIA instruments, such as the point spread function, binning, exposure times, etc. An example is presented in Figure 6.6. The goal of these tools is to provide a reference set of standard CCD images for the GAIA instruments that can be used for further analysis or study. Eventually a set of images in the various CCDs computed at different epochs of a simulated mission may be used to check the reduction algorithms and to assess the expected astrometric and photometric precisions. Other applications include the assessment of the detection of binaries as a function of separation and magnitude, variable stars, star/galaxy separation, selection function resulting from the detection process, etc.

6.7.2 Detection in the Astrometric Sky Mapper

A critical aspect of the design of the astrometric instruments is the detection of objects entering the focal planes (see Section 3.3.2). Simulated images have been used for the analysis of performances of detection algorithms applied to the astrometric sky mapper (ASM) detection in the astrometric instruments. In these tests a sky field is generated using the galaxy model and used to produce CCD images. The images are analysed using the algorithm and a list of detections is produced. The list of detections is then cross-matched with the input list of simulated stars. In this way, the efficiency of the algorithm for detections is assessed.

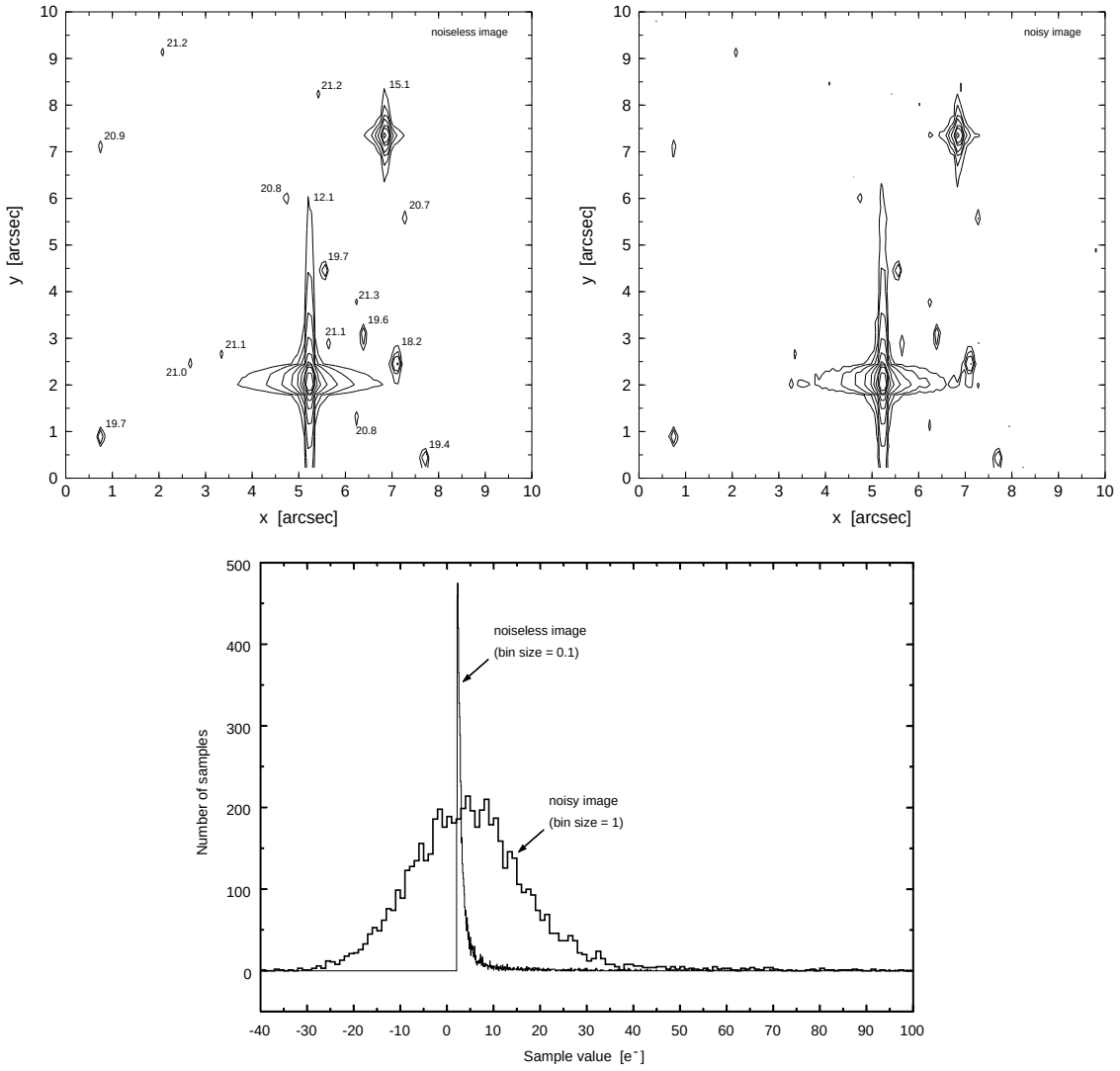


Figure 6.6: Simulation of stars in an area of 10 × 10 arcsec^2 in Baade's Window, as registered by the Astrometric Sky Mapper (ASM1). This is one of the densest areas in the Milky Way that will be observed by GAIA (outside of the centres of globular clusters). The simulation shows that, even in this area, a nearly complete detection of stars to G = 20 mag can be achieved with the 0.86 s integration time of the ASM1. Based on an extrapolation of the mean number densities found by Paczyński et al. (1994), the simulation includes stars brighter than G = 26.75 mag, at which magnitude each star contributes less than one photoelectron to the detected image. The assumed background is flat with G = 21.1 mag arcsec^{-2}. The depicted area includes the images of some 2200 stars, plus the diffraction wings of many more stars in the surrounding area. The brightest star (G = 12.1 mag) was inserted manually to show more clearly the diffraction spikes in the star images; such a bright star is otherwise not expected in this small area. The sample size is 2 × 2 pixels (0.074 × 0.223 arcsec^2). Top left: simulated image before adding the noise. The contour levels correspond to 40, 100, 251, 631, 1585, 3981, 10 000, ... electrons per sample above the dark level. The numbers show the approximate G magnitude of the stars visible by the contours. Top right: same image after adding Poisson noise and readout noise (10.9 electrons rms). The contour levels are the same as in the previous image. Bottom: histograms of the sample values in the two images. Note that the background is dominated by the readout noise: the many faint stars covering the whole area have little influence on the detection of stars at G = 20 mag, corresponding to sample values above some 100 electrons.

Several algorithms have been tested, in particular APM (Irwin 1985), Sextractor (Bertin & Arnouts 1996) and a sliding-window algorithm specifically developed for GAIA (Høg 1998g; Babusiaux & Arenou 1999). After rather extensive evaluations, both APM and the sliding window algorithm have been shown to perform well, detecting point sources down to $G = 20 - 21$ mag with high reliability and low spurious detection rates, even in complex fields. The APM algorithm has not been fully optimised to operate with the GAIA PSF, but has the merit of handling extended objects efficiently. The sliding window algorithm operates more robustly for point sources, even in the presence of strong variations in the local background although it has not been optimised for the detection of extended sources (e.g. compact galaxies). A full detection simulation, taking outputs from ASM1–3 to filter cosmic ray events, and confirm preliminary detections, has not been undertaken, but seems to present no conceptual problems.

An example of the APM detections in a star field is given in Figure 6.7, along with the detection statistics in Figure 6.8. The efficiency of APM when applied to ASM1 detections is very high, resulting in about 80 per cent completeness even at $20 \leq G < 21$ mag in the example given. Even in very crowded regions such as Baade's Window, where detection is evidently particularly difficult, 50 per cent completeness is reached at $20 \leq G < 21$ mag. These results indicate that the expected performances of the detection system give a very high completeness of GAIA observations even with a single observation transit.

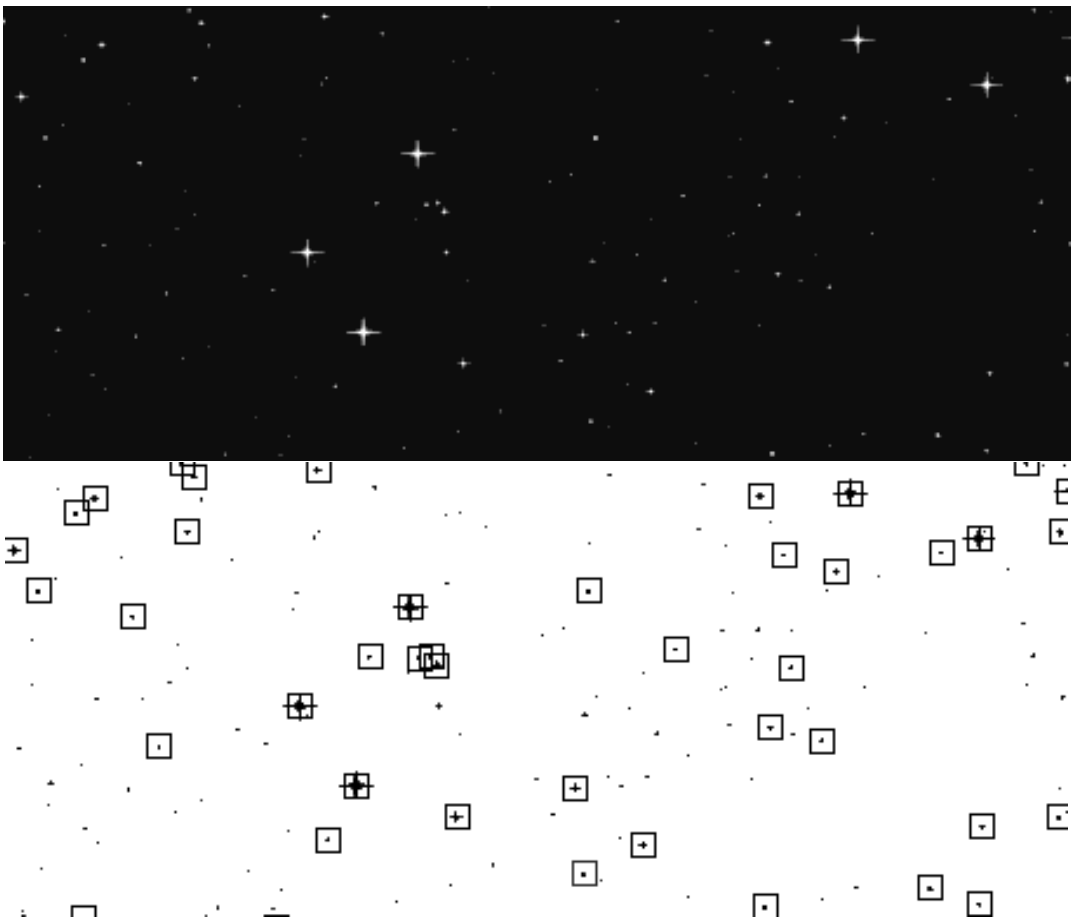


Figure 6.7: APM analysis of an ASM1 image of a typical star field at $(l = 45°, b = 5°)$. *Top: ASM1 CCD image (noiseless, subfield). Bottom: APM detections; boxes indicate the location of detections.*

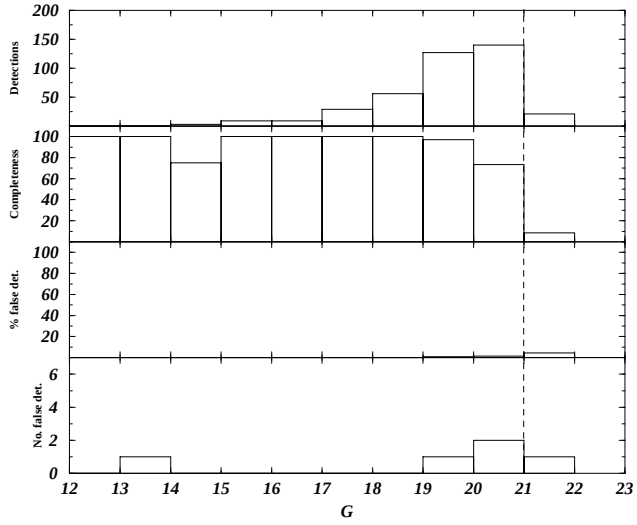


Figure 6.8: Results of the application of APM analysis of an ASM1 image of a typical star field at ($l = 45°, b = 5°$). From top to bottom: number of detections, completeness (percentage of stars in the simulated sample effectively detected by APM), percentage of false detections and absolute number of false detections.

6.7.3 Photometric Analysis

The purpose of acquiring detailed photometric measurements for each star, both in the wide bands of the Astro fields, and in the medium bands of the Spectro field, is to characterize the physical nature of the object observed through its spectral energy distribution, and its time variability. It is therefore important to understand the limitations of these goals given the complex nature of the sky: many objects (even the majority) are double or multiple, certain sky regions are densely populated, and the sky background may vary strongly on a variety of angular scales due to nebulosities, galaxies, etc. The problems associated with precision photometry in crowded regions are complex, and are under investigation for many projects. GAIA cannot solve these problems by acquiring data at arbitrarily high angular resolution: read-noise and data rate considerations prohibit this. The questions to be posed are: can accurate photometry be acquired down to the faintest limits of the GAIA survey? At what level does known sky background variations or the density of faint background objects degrade the measurements to unacceptable levels? How can reconstruction of the photometry on ground, using different scanning orientations with different along-scan and transverse angular resolutions, be optimized? These issues have been considered in a number of ongoing studies by E. Høg and colleagues at Copenhagen University Observatory, and by A.G.A. Brown at Ensenada (now ESO).

Some preliminary simulations have been carried out using the proposed sampling geometry of AF17 and the broad-band photometer of the Astro fields, detailed in Section 3, using simulated fields in the region of Draco, the Trapezium Cluster in Orion (M42), and the extremely dense region of Baade's Window, supplemented by HST Planetary Camera images to provide realistic nebular background data. The simulations comprise the generation of realistic GAIA type data, using realistic source counts, spectral and filter responses, optical configuration, and scanning geometry. Background determination and photometric analysis follow the provisional procedures proposed for GAIA on-ground analysis, but not yet the more elaborate procedures described in Section 9.4.2.

Even at $V = 19$ mag, accurate photometry is achievable in all but the densest sky regions. In the case of Baade's Window, the background structure combined with the high stellar density leads to degraded photometric results at the level of 0.05 mag even for relatively bright stars ($V = 18$ mag). However, the analysis so far has not included more robust background determination methods, and has neglected the information provided by stacking all of the photo sky mapper data (AF17) collected during the mission lifetime, which will provide details on the surroundings of each object (Section 9.4.2). The results indicate that high-quality photometry can be acquired, to the faintest GAIA magnitudes, and at least for the majority of the sky, while further optimization of the background determination, sampling strategy, and on-ground analysis will lead to further improvements.

Table 6.10: An example of results from the CCD centroiding experiment (data set 7, 15 s exposure), showing the dispersion of the reduced coordinates x', y'.

Source	1	2	3
S/N	471.95	498.35	501.25
$\sigma(x')$ [pixels]	0.00211	0.00201	0.00204
$\sigma(y')$ [pixels]	0.00224	0.00229	0.00189
$1/\sigma(x')$ [pixel fraction]	474.56	498.43	490.89
$1/\sigma(y')$ [pixel fraction]	445.47	435.87	530.13

6.8 Centroiding Experiments

To verify the basic location performance using a realistic CCD configuration, a low-cost laboratory experiment was implemented by the Observatory of Torino. The goal was to evaluate the measurement repeatability using a set of images of a simulated stellar field. For a given imaging system, in stable conditions, the expected location dispersion of a point-like image is:

$$\sigma = \alpha \frac{\lambda}{\pi D \ \mathrm{S/N}} \tag{30}$$

where D is the telescope aperture, S/N is the signal-to-noise ratio, λ is the observing wavelength, and α is a degradation factor depending on system geometry, sampling resolution, detector noise and finite pixel size, optics aberrations, and the applied location procedure (Gai et al. 1998; Lattanzi et al. 1997; Lindegren 1978).

The set-up was mounted on an optical bench (Figure 6.9a). An LED source, through a frosted glass diffuser, illuminates a target mask with several pinholes, producing an image, through an optical doublet with 2 mm aperture, over a commercial CCD camera head, which has the same 9 μm pixel size selected for the baseline option of GAIA. The chosen sampling resolution corresponds to about 7.6 pixels per Airy disk diameter, close to the GAIA value (6 pixels in the central peak). The desired 1/1000 pixel accuracy, in order to ensure photon limited performance for bright targets, translates at the focal plane level into 9 nm linear equivalent dimension.

For a given signal level, a set of 50 elementary exposures is collected, in order to provide a statistically reasonable, self-consistent sample of images under the same nominal conditions. A significant spatial noise, of about 25 counts, is observed. The target location is evaluated by means of the IDL procedure `gauss2dfit`, providing the best approximation of the data to an elliptic (i.e. asymmetric) Gaussian over a uniform plateau. The fit parameters returned are the estimates of background, peak intensity, Gaussian width and centre in both directions. Since the fit is a maximum likelihood estimate, in the least-squares sense, the limiting performance is expected to be the same as for the optimal location estimator. The mismatch between the Gaussian PSF used and the expected Bessel profile results in a performance degradation of < 10 per cent. The limited sampling resolution provides an additional dispersion increase by a similar amount.

The data corresponding to one of the data sets are shown in Figure 6.9 for three separate sources (top to bottom). Crosses and asterisks are, respectively, the x and y frame coordinates. The positions of the three sources are strictly correlated, and the raw data dispersion is of order 1–2 per cent. The dominant common mode effect is suppressed by evaluating and subtracting the average coordinate x_{cm} to derive the reduced coordinates x'_n:

$$x_{\mathrm{cm}} = \frac{x_1 + x_2 + x_3}{3}, \quad x'_n = x_n - x_{\mathrm{cm}}, \qquad n = 1, 2, 3 \tag{31}$$

The reduced coordinates are shown by the solid (x') and dotted (y') lines; their dispersion is reduced by about one order of magnitude (see Table 6.10).

Several data sets were collected, under different conditions, and in all cases a comparably large common mode disturbance was noted, with a residual position dispersion close to the inverse of the signal-to-noise ratio. From the above expression on the location dispersion, the degradation factor α and its behaviour for increasing S/N can be evaluated; the present setup limits the measurement range to $\mathrm{S/N} \leq 850$. The results are shown in Figure 6.10a; there is a rather large dispersion around the average value $\alpha \simeq 1.4 \pm 0.17$, but no evidence of a trend depending on the S/N. The contribution from unmodelled experimental effects, including spatial noise, is only 10–15 per cent. The results are therefore still far from a regime dominated by systematic errors, other than the slow common mode motion. In principle, the theoretical expectations can be extrapolated to much brighter sources, with progressively increasing accuracy. The current location performance is therefore a conservative estimate of what can be achieved with a CCD-based imaging instrument, and the limiting performance is still well beyond the regime which has been explored experimentally.

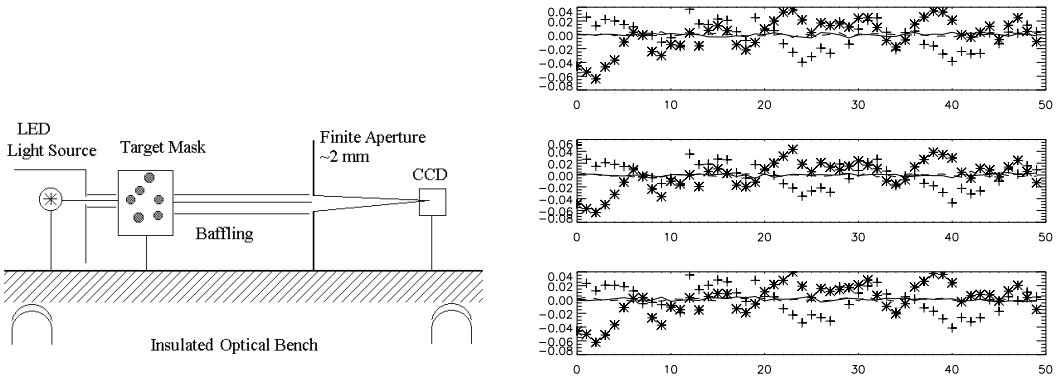


Figure 6.9: (a, left) Schematic of the experimental set-up for the CCD centroiding experiments. (b, right) Sequence 7: 15 s exposures, x (crosses) and y (asterisks) frame position dispersion; reduced coordinates (x': solid line, y': dashed line) dispersion; top to bottom: sources 1, 2 and 3.

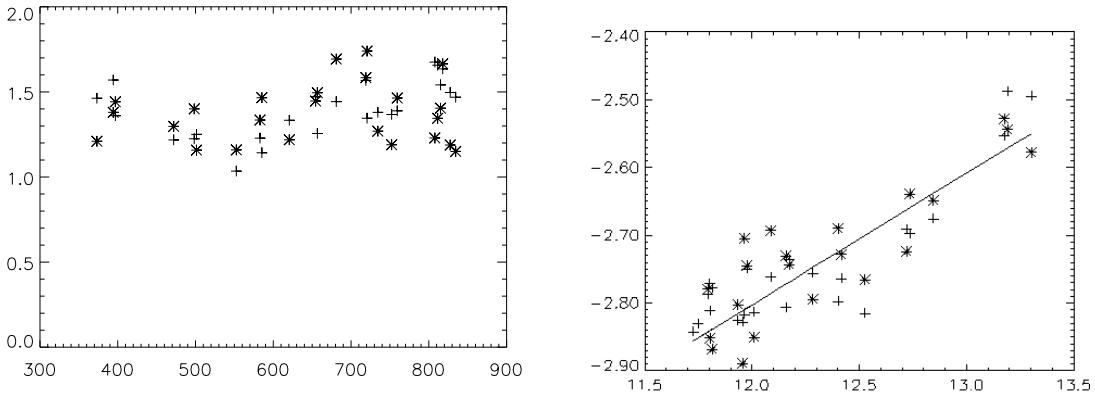


Figure 6.10: (a, left) Degradation factor (α in Equation 30) for the frame coordinates, from the whole data set (x: crosses; y: asterisks), versus the signal-to-noise ratio S/N. (b, right) Accuracy of the reduced coordinates (in pixels) versus instrumental magnitude (logarithmic plot).

The result can also be expressed by a plot of the dispersion σ (in pixels) of the reduced coordinates x', y' versus the instrumental magnitude m, the latter defined on a scale selected to match approximately the GAIA signal characteristics. The logarithmic fit of the experimental location accuracy to the instrumental magnitude provides a reasonable agreement, within ~ 10 per cent errors: in the units adopted, $\log \sigma = -5.134 + 0.194m$. The initial common mode systematics appear to be induced by the thermal control of the CCD camera head, which affects structural stability. In actual implementations, such problems can be greatly reduced by proper design, i.e. passive cooling of the detectors and high thermal stability of the overall system.

In conclusion, the limiting accuracy in the determination of the target position in CCD images follows the theoretical predictions, at least within the range explored in the present experiment, i.e. S/N $\simeq$ 850, which is comparable to the case of the bright targets of GAIA (S/N $\sim$ 2000). The intrinsic limitations of this experimental set-up provide a dispersion of the raw data of about 1/100 pixels. However, very simple corrections for the common mode motion allow suppression of the disturbances by up to an order of magnitude. Similar simple self-calibration strategies for removal of satellite attitude and payload geometry perturbations from the science data appear feasible within the GAIA operation scheme and adequate to its accuracy requirements.

7 Accuracy Estimation: Methods

The semi-analytical methods used for the basic accuracy estimation are summarised in this section. The resulting formulae permit efficient computation of the astrometric and photometric accuracies as functions of the assumed parameters. They are thus a useful tool for optimizing certain aspects of the instrument and observations. The general validity of the semi-analytical approach has been checked by extensive simulations. Its relation to the overall performance assessment is indicated by Figure 6.1. Results of the basic accuracy estimation are given in Section 8.

7.1 Notations

The following general variables are used (with units in square brackets): t is time [s] and λ the optical wavelength [m]. Linear coordinates in the telescope pupil plane are denoted (x, y) [m], with origin at the centre of the pupil, the x axis along the scan and the y axis across the scan. In the image plane, the corresponding angular coordinates are denoted (ξ, η) [rad], and the spatial frequencies (u, v) [rad^{-1}], with ξ and u along the scan. $i = \sqrt{-1}$ is the imaginary unit. $\Pi_w(x)$ denotes a centred, normalized rectangular function of width w, i.e. $\Pi_w(x) = 1/w$ for $|x| < w/2$, otherwise $\Pi_w(x) = 0$.

The celestial coordinates needed are (l, b), the Galactic longitude and latitude, and ecliptic latitude β. Notations for the main input parameters and other quantities are listed in Tables 7.1 and 7.2.

7.2 Image Formation

The telescope aperture is described by the pupil function $a(x, y)$, which is 1 inside the pupil and 0 elsewhere. For the baseline configuration with a rectangular pupil of size $D \times H$, $a(x, y) = 1$ if and only if $|x| < D/2$ and $|y| < H/2$. The equations below are however valid for arbitrary pupil shapes, including the twin pupil of a Fizeau interferometer (cf. Figure 2.1).

With $A = \iint a(x, y)\,\mathrm{d}x\mathrm{d}y$ denoting the total pupil area, the monochromatic flux distribution in the focal plane is:

$$I_\lambda(\xi, \eta) = AT_\lambda f_\lambda P_\lambda(\xi, \eta) \tag{32}$$

Here, $P_\lambda(\xi, \eta)$ is the monochromatic optical point-spread function (PSF), normalized such that $\iint P_\lambda(\xi, \eta)\,\mathrm{d}\eta\mathrm{d}\eta = 1$. For arbitrary pupil shape $a(x, y)$ and wavefront errors $w(x, y)$, the monochromatic PSF may be computed by means of the Fraunhofer diffraction formula. Defining the complex pupil function as:

$$a_\lambda(x, y) = a(x, y) \exp\left[i\frac{2\pi}{\lambda} w(x, y)\right] \tag{33}$$

the diffraction formula is:

$$P_\lambda(\xi, \eta) = \frac{1}{A\lambda^2}\left|\iint a_\lambda(x, y) \exp\left[i\frac{2\pi}{\lambda}[x\xi + y\eta]\right]\,\mathrm{d}x\mathrm{d}y\right|^2 \tag{34}$$

Since measurements are (mainly) made in the ξ direction, it is often not necessary to consider the full two-dimensional PSF, but only its marginal distribution in ξ (the line-spread function, LSF):

$$L_\lambda(\xi) = \int P_\lambda(\xi, \eta)\,\mathrm{d}\eta \tag{35}$$

For polychromatic images the total flux distribution is obtained by integrating $I_\lambda(\xi, \eta)$ over the wavelengths. In practice it is the distribution of the detection probability of the photoelectrons which is needed. Neglecting the spatial inhomogeneities of the detector (which are smoothed out by the TDI mode of observation), this distribution is given by the photoelectron-weighted polychromatic PSF and LSF:

$$P_\mathrm{O}(\xi, \eta) = \frac{\int P_\lambda(\xi, \eta) Q_\lambda T_\lambda f_\lambda\,\mathrm{d}\lambda}{\int Q_\lambda T_\lambda f_\lambda\,\mathrm{d}\lambda} \tag{36}$$

Table 7.1: Designations and units for the main parameters input to the accuracy estimation. An asterisk marks the parameters whose values are given in Table 6.2.

Designation		Meaning [unit]
A		total pupil area per astrometric instrument [m]
$a(x, y)$		astrometric pupil function (= 1 inside pupil, 0 elsewhere) [–]
b_λ		monochromatic background photon flux [ph m^{-2} s^{-1} m^{-1} rad^{-2}]
D	*	pupil dimension along scan for the astrometric instrument [m]
F	*	focal length [m]
f_λ		monochromatic stellar photon flux [ph m^{-2} s^{-1} m^{-1}]
g_0		geometric factor for determination of position [–]
g_μ		geometric factor for determination of proper motion [yr^{-1}]
g_π		geometric factor for determination of parallax [–]
H	*	pupil dimension normal to scan for the astrometric instrument [m]
k		constant (~ 0.5) used in Cramér-Rao approximation [–]
L	*	effective mission length [s]
N_i	*	number of astrometric instruments [–]
N_ξ	*	number of pixels per CCD in scan direction [–]
Q_λ	*	detector quantum efficiency [–]
r	*	detector rms readout noise [e^-]
T_λ	*	intensity transmittance [–]
W	*	rms wavefront error, including polishing errors [m]
$w(x, y)$		detailed wavefront errors (for given field position) [m]
$\Delta\eta$	*	angular pixel size normal to scan [rad]
$\Delta\xi$	*	angular pixel size along scan [rad]
$\Delta\tau$		time window for TDI (per phase) [s]
$\Delta\omega$		attitude rate error [rad s^{-1}]
ξ	*	(also) the Sun–spin axis angle of the scanning law
Ω	*	detector solid angle per astrometric instrument [rad^2]
ω	*	scan speed [rad s^{-1}]

$$L_\mathrm{O}(\xi) = \frac{\int L_\lambda(\xi) Q_\lambda T_\lambda f_\lambda \,\mathrm{d}\lambda}{\int Q_\lambda T_\lambda f_\lambda \,\mathrm{d}\lambda} \tag{37}$$

The subscript stands for 'optical', as the functions represent the instantaneous optical image as it falls on the detector.

Diffraction introduces a wavelength dependence of the PSF (and LSF) even in the absence of optical aberrations. In the presence of aberrations, the monochromatic PSF may in addition be asymmetric, resulting in wavelength-dependent centroid definitions. Equations 33–37 have been used extensively to study chromaticity, i.e. variations in the centroid position with stellar colour or spectral distribution (see Section 3.2.4). The double integral in Equation 34 is simply the two-dimensional Fourier transform of $a_\lambda(x, y)$, which permits efficient calculation of the PSF by means of the Fast Fourier Transform. The wavefront aberrations $w(x, y)$ are obtained by means of ray-tracing calculations, in which polishing and mirror alignment errors may be taken into account. It should be noted that $w(x, y)$ in general depends on the position in the field of view, i.e. on the field angles (ξ, η). This dependence is not explicitly shown in the equations above, where (ξ, η) must be interpreted as local field angles centred on the field point taken as reference for the wavefront errors — usually the chief ray, i.e. somewhere within the geometrical image. (The Fraunhofer diffraction formula is only valid to first order in the angles ξ, η.) A separate calculation of the wavefront errors, and subsequently of the PSF, is therefore required for each field point considered.

A slightly different approach is taken for the purpose of the astrometric accuracy estimation in Section 7.4. Introducing the optical transfer function (OTF):

$$O_\lambda(u, v) = \frac{1}{A} \iint a_\lambda(x, y) a^*_\lambda(x - u\lambda, y - v\lambda) \,\mathrm{d}x\mathrm{d}y \tag{38}$$

the PSF can also be computed as:

$$P_\lambda(\xi, \eta) = \iint O_\lambda(u, v) \exp\left[i2\pi(u\xi + v\eta)\right] \,\mathrm{d}u\mathrm{d}v \tag{39}$$

while the LSF becomes:

$$L_\lambda(\xi) = \int O_\lambda(u) \exp\left(i2\pi u\xi\right) \,\mathrm{d}u \tag{40}$$

with $O_\lambda(u) \equiv O_\lambda(u, 0)$.

Table 7.2: Designations and units for other important or derived quantities in the accuracy estimation.

Designation	Meaning [unit]
$a_\lambda(x,y)$	complex pupil function [–]
b	background count per pixel [e^-]
$C(\xi)$	detected stellar image after TDI [e^-]
$I_\lambda(\xi,\eta)$	two-dimensional monochromatic image [ph s^{-1} m^{-1} rad^{-2}]
$L_\lambda(\xi)$	monochromatic optical line-spread function (LSF) [rad^{-1}]
$L_\mathrm{O}(\xi)$	polychromatic optical line-spread function (LSF) [rad^{-1}]
$L(\xi)$	global line-spread function (LSF) [rad^{-1}]
$M_0(u,\lambda)$	monochromatic modulation transfer function (MTF) for aberration-free instrument [–]
$M_d(u,\lambda)$	monochromatic MTF due to charge diffusion in CCD [–]
$M_\xi(u)$	MTF due to integration over pixel width [–]
$M_\tau(u)$	MTF due to TDI time window [–]
$M_\omega(u)$	MTF due to attitude rate error [–]
N	total number of stellar photons detected in one chip crossing [–]
$O_\lambda(u,v)$	monochromatic optical transfer function (diffraction only) [–]
$O_\lambda(u)$	monochromatic optical transfer function at $v=0$ [–]
$P_\lambda(\xi,\eta)$	monochromatic optical point-spread function (PSF) [rad^{-2}]
$P_\mathrm{O}(\xi,\eta)$	polychromatic optical point-spread function (PSF) [rad^{-2}]
$P(\xi,\eta)$	global point-spread function (PSF) [rad^{-2}]
S_λ	Strehl ratio [–]
$S(\eta)$	spectral distribution across scan in radial-velocity spectrometer [rad^{-1}]
s_0, s_2	angles characterising $U(u)$ [rad]
$U(u)$	total effective MTF [–]
u_max	$= D/\lambda_\mathrm{min}$ maximum spatial frequency for non-zero $U(u)$ [rad^{-1}]
$X(\eta)$	template spectrum for radial-velocity determination
λ_max	maximum wavelength for non-zero response $Q_\lambda T_\lambda$ [m]
λ_min	minimum wavelength for non-zero response $Q_\lambda T_\lambda$ [m]
σ_cal	accuracy of astrometric or photometric calibration
σ_0	astrometric accuracy in position at mid-mission [rad]
σ_μ	astrometric accuracy in proper motion [rad s^{-1}]
σ_π	astrometric accuracy in parallax [rad]
σ_ξ	angular precision in scan direction from one CCD crossing [rad]
τ	$= L\Omega/4\pi$ total integration time on object per astrometric instrument [s]
τ_1	$= N_\xi \Delta\xi/\omega$ integration time per CCD [s]

7.3 Detector Signal

In the equations above the reference point for the optical image is stationary at $(\xi,\eta)=(0,0)$. In the absence of (asymmetric) aberrations, this is also the centre of the image. In reality the image moves with angular velocity ω across the detector, and the instantaneous intensity at the fixed point (ξ,η) on the detector should be written $I_\lambda(\xi-\omega t,\eta)$, where time t is reckoned from the passage of the reference point across the line $\xi=0$.

In the TDI mode of observation, the charge image is being integrated for a short time $\Delta\tau$ (the TDI window), before it is shifted forward along with the image. In a four-phase CCD, $\Delta\tau$ is 1/4 of the TDI period $\Delta\xi/\omega$. The optical image thus moves the distance $\omega\Delta\tau=\Delta\xi/4$ while the charge image remains stationary. The result is a smearing of the electronic image along the scan, mathematically described by a convolution of the PSF with $\Pi_{\omega\Delta\tau}(\xi)$, a rectangular function of width $\omega\Delta\tau$ in the scan (ξ) direction. The pixel size $\Delta\xi\times\Delta\eta$ similarly produces a smearing of the image, described by convolutions along the ξ and η by rectangular functions. Combining these effects gives the 'global' PSF:

$$P(\xi,\eta)=P_\mathrm{O}(\xi,\eta)*\Pi_{\omega\Delta\tau}(\xi)*\Pi_{\Delta\xi}(\xi)*\Pi_{\Delta\eta}(\eta) \tag{41}$$

where $*$ is the convolution operator. The global LSF is similarly given by:

$$L(\xi)=L_\mathrm{O}(\xi)*\Pi_{\omega\Delta\tau}(\xi)*\Pi_{\Delta\xi}(\xi) \tag{42}$$

In practical computations, these convolutions are best made by via the Fast Fourier Transform.

The 'data numbers' obtained by reading out the CCD pixels are expressed on a scale with, in principle, arbitrary zero point (bias) and scale factor (gain). In the accuracy analysis it is convenient to fix the scale such that (1) one unit of the CCD signal corresponds to a charge of one electron; (2) zero charge (no light) gives the output signal r^2 units, where r is the rms readout noise. This allows to treat the readout noise as if it were Poisson noise on the signal. With this device, the basic model for the CCD pixel values n_{jk} is a Poisson process with expectation:

$$\mathrm{E}(n_{jk}) = r^2 + b + NP(j\Delta\xi - \xi_0, k\Delta\eta - \eta_0) \tag{43}$$

Here, (ξ_0, η_0) is the image centroid, b is the expected number of electrons per pixel generated by the sky background:

$$b = A\tau_1\Delta\xi\Delta\eta \int Q_\lambda T_\lambda b_\lambda \,\mathrm{d}\lambda \tag{44}$$

and N is the total number of electrons generated by the star:

$$N = A\tau_1 \int Q_\lambda T_\lambda f_\lambda \,\mathrm{d}\lambda \tag{45}$$

If several pixels are binned to create larger samples, then b must be multiplied by the corresponding factor, and $P(\xi, \eta)$ or $L(\xi)$ modified by using the appropriate widths in Equations 41–42.

7.4 Astrometric Accuracy

7.4.1 The One-Dimensional Detector Signal

For the basic accuracy estimation it is not necessary to take the aberrations into account in a rigorous way. The main effect of (small) aberrations will be a reduction of the peak intensity of the monochromatic PSF by the Strehl ratio:

$$S_\lambda = \exp[-(2\pi W/\lambda)^2] \tag{46}$$

where W is the rms wavefront error. The theoretical (aberration-free) OTF can therefore be used, after degradation by the factor S_λ. In the absence of aberrations $O_\lambda(u)$ is real and equal to the modulation transfer function (MTF), which we write:

$$M_0(u, \lambda) = \frac{1}{A} \iint a(x, y)a(x - u\lambda, y)\,\mathrm{d}x\mathrm{d}y \tag{47}$$

For simple pupil shapes this is easily evaluated analytically. In particular, for a rectangular pupil of length D (in the x direction):

$$M_0(u, \lambda) = 1 - \left|\frac{u\lambda}{D}\right| \tag{48}$$

if $|u| < D/\lambda$, otherwise $M_0(u, \lambda) = 0$. In the present formulation the convolutions in Equations 41–42 reduce to multiplicative modifications of the MTF. Thus, the explicit forms for the global PSF derived in the previous section are not directly used in the following.

Assuming that the transverse pixel (or sample) size is sufficient to integrate all the light diffracted in the transverse (η) direction, i.e. that $\Delta\eta \gg \lambda/H$, the expected number of photoelectrons per pixel due to the star is:

$$C(\xi) = 2A\tau_1\Delta\xi \int_{\lambda_{\min}}^{\lambda_{\max}} S_\lambda Q_\lambda T_\lambda f_\lambda \int_0^{u_{\max}} M_0(u, \lambda)M_d(u, \lambda)M_\xi(u)M_\tau(u)M_\omega(u)\cos(2\pi u\xi)\,\mathrm{d}u\mathrm{d}\lambda \tag{49}$$

where $\tau_1 = N_\xi\Delta\xi/\omega$ is the CCD crossing time (or, more precisely, the integration time on the CCD before readout), $u_{\max} = D/\lambda_{\min}$ is the maximum along-scan spatial frequency for non-zero MTF, $[\lambda_{\min}, \lambda_{\max}]$ is the support of $Q_\lambda T_\lambda$, and $M_0(u, \lambda)$ is the one-dimensional diffraction MTF in the scanning direction. M_d, M_ξ, M_τ and M_ω are the modulation transfer functions associated with charge diffusion in the CCD, pixel width, TDI time window, and attitude rate errors. These functions are modeled as follows:

Charge diffusion: Diffusion of photoelectrons across the pixel boundaries (pixel cross talk) degrades the electronic modulation transfer function of the CCD. It is assumed that charge diffusion can be modeled as an exponential effect with a characteristic $(1/e)$ lateral diffusion length of d_λ [m]:

$$M_d(u,\lambda) = \frac{1}{1+(2\pi u d_\lambda/F)^2} \tag{50}$$

The wavelength dependence arises from the varying depths in the silicon at which photons are absorbed. In practice d_λ depends also on the size of the charge packet. For the present accuracy evaluation it was determined so as to reproduce the MTF values in Table 3.15. For small charge packets (low signals) this gives $d_\lambda \simeq 3.0$ μm for $\lambda \leq 450$ nm, decreasing to $d_\lambda \simeq 1.1$ μm for $\lambda \geq 950$ nm. For high signal levels the diffusion length is a factor 1.85 bigger.

Pixel width: Integration over a rectangular pixel response function of width $\Delta\xi$ (as it would be in the absence of charge diffusion) gives a sinc factor on the modulation:

$$M_\xi(u) = \mathrm{sinc}(\pi u \Delta\xi) \tag{51}$$

TDI time window: Let $\Delta\tau$ be the time interval between successive shifts of the charge image in the TDI mode. The motion of the image in this interval is $\omega\Delta\tau$. This motion produces a smearing of the image which can be represented by the spatial frequency filter:

$$M_\tau(u) = \mathrm{sinc}(\pi u \omega \Delta\tau) \tag{52}$$

A four-phase CCD is assumed, for which $\omega\Delta\tau = \Delta\xi/4$.

Rate errors: Rate errors and other time-variable attitude errors accumulated over the CCD crossing time τ_1 induce an additional smearing of the image. The effect can be represented by a filter function, which in general is the Fourier transform of the probability density function of the instantaneous attitude errors (deviations from the CCD clocking). For a constant rate error of $\Delta\omega$:

$$M_\omega(u) = \mathrm{sinc}(\pi u \tau_1 \Delta\omega) \tag{53}$$

If the rate error has a normal distribution with zero mean and standard deviation $\Delta\omega_{\mathrm{rms}}$, then:

$$M_\omega(u) = \exp\left[-\frac{1}{2}(2\pi u \tau_1 \Delta\omega_{\mathrm{rms}})^2\right] \tag{54}$$

Equation 54 is the form used in the accuracy estimation, with $\tau_1\Delta\omega_{\mathrm{rms}} = 0.3\Delta\xi$.

According to Equation 49 an effective transfer function can be defined as:

$$U(u) = M_\xi(u) M_\tau(u) M_\omega(u) \frac{\int M_0(u,\lambda) M_d(u,\lambda) S_\lambda Q_\lambda T_\lambda f_\lambda \,\mathrm{d}\lambda}{\int Q_\lambda T_\lambda f_\lambda \,\mathrm{d}\lambda} \tag{55}$$

Introducing N, the total number of stellar photons detected during the CCD crossing, from Equation 45, Equation 49 can be expressed more concisely as $C(\xi) = NL(\xi)$, with:

$$L(\xi) = 2\Delta\xi \int_0^{u_{\max}} U(u) \cos(2\pi u \xi) \,\mathrm{d}u \tag{56}$$

In analogy with Equation 43, the sequence of charge counts n_j in successive pixels is now modeled as a Poisson process with expectation:

$$\mathrm{E}(n_j) = r^2 + b + NL(j\Delta\xi - \xi_0) \tag{57}$$

where ξ_0 is the position of the image centroid at some reference time, and b is given by Equation 44 (taking into account the binning of 1×8 pixels in a sample). It can be noted that, for any ξ_0:

$$\sum_{j=-\infty}^{+\infty} L(j\Delta\xi - \xi_0) = \frac{1}{\Delta\xi} \int_{-\infty}^{+\infty} L(\xi)\,\mathrm{d}\xi = 1 \tag{58}$$

The diffraction normal to the scan is not formally included in the basic estimation method. Nominally the sample width of 8 pixels should be sufficient to include nearly all the light from a point source. However, to account for imperfect centering in the η direction, transverse motion of the image, and similar, it is assumed (conservatively) that only 90 per cent of the light falls within the sample. An extra factor 0.9 is thus introduced in Equation 45.

Stars brighter than $G \simeq 12$ will cause pixel saturation in the astrometric field and will have to be observed in a special mode permitting shorter integration time. This means that the accuracy does not improve for brighter stars but remains at a roughly constant level up to $G \simeq 6$. This effect is included in the accuracy estimation by putting an upper limit of 5×10^5 electrons on the total charge within the PSF at readout. This limit is reached in 0.86 s for a $G = 12.1$ mag star.

7.4.2 Precision of Location Estimator

The Cramér-Rao (CR) bound, which is asymptotically reached with the Maximum Likelihood estimator, can be used to evaluate the precision by which ξ_0 can be estimated from Equation 57. If the image is not too badly sampled, the CR bound is:

$$\sigma_\xi = \left[\frac{N^2}{\Delta\xi} \int_{-\infty}^{+\infty} \frac{[L'(\xi)]^2}{r^2 + b + NL(\xi)}\,\mathrm{d}\xi\right]^{-1/2} \tag{59}$$

It does not seem possible to evaluate this integral analytically in terms of the effective transfer function $U(u)$. However, since $0 \le L(\xi) \le L(0)$, one can use the inequalities:

$$\int_{-\infty}^{+\infty} \frac{[L'(\xi)]^2}{r^2 + b + NL(0)}\,\mathrm{d}\xi \le \int_{-\infty}^{+\infty} \frac{[L'(\xi)]^2}{r^2 + b + NL(\xi)}\,\mathrm{d}\xi \le \int_{-\infty}^{+\infty} \frac{[L'(\xi)]^2}{r^2 + b}\,\mathrm{d}\xi \tag{60}$$

to bracket the CR bound, viz.:

$$\sigma_\xi = \left[\frac{N^2/\Delta\xi}{r^2 + b + kNL(0)} \int_{-\infty}^{+\infty} [L'(\xi)]^2\,\mathrm{d}\xi\right]^{-1/2} \tag{61}$$

where k is some number between 0 and 1. The value of k to be used is best determined by means of Monte Carlo simulations (see below). From Equation 56:

$$L(0) = 2\Delta\xi \int_0^{u_{\max}} U(u)\,\mathrm{d}u\,, \qquad \int_{-\infty}^{+\infty} [L'(\xi)]^2\,\mathrm{d}\xi = 8\pi^2\Delta\xi^2 \int_0^{u_{\max}} U(u)^2 u^2\,\mathrm{d}u \tag{62}$$

These equations suggest that two characteristic angles s_0 and s_2 can be defined, which describe the effective transfer function $U(u)$ with respect to the astrometric content of the image:

$$s_0 = \left[2\int_0^{u_{\max}} U(u)\,\mathrm{d}u\right]^{-1}, \qquad s_2 = \left[2s_0 \int_0^{u_{\max}} U(u)^2 u^2\,\mathrm{d}u\right]^{-1/2} \tag{63}$$

In fact, s_0 measures the overall angular size of the image envelope, and consequently determines the sensitivity of the location estimator to the background level; by contrast, s_2 measures the (inverse) slope of the diffraction pattern and thus determines the astrometric accuracy for a given signal-to-noise ratio. From Equation 61:

$$\sigma_\xi = \frac{1}{2\pi}\,\frac{s_2}{\sqrt{N}}\,\left[k + \frac{(r^2 + b)s_0}{N\Delta\xi}\right]^{1/2} \tag{64}$$

It remains to determine the constant k in the last equation. Dedicated Monte Carlo simulations were performed to calibrate k under realistic assumptions for the sampling of the image. Only Poisson noise in the number of detected (stellar and background) photons, plus read-out noise, was considered. For each of 12 different spectral type/magnitude combinations, the rms location error was calculated from 60 000 experiments, in which the true location was uniformly distributed over a pixel. The location process used the readout values (n_j) from five consecutive pixels ($j = 1 \ldots 5$), where the middle one contained the actual centre of the LSF. The location (ξ_0) was found by maximising the relative log-likelihood function:

$$\mathcal{L}(\xi_0) = \sum_{j=1}^{5} [n_j \ln s_j(\xi_0) - s_j(\xi_0)] \tag{65}$$

where $s_j(\xi_0) = r^2 + b + NL(j\Delta\xi - \xi_0)$ is the expectation of n_j if the LSF is centred on ξ_0 (Equation 57). The function $\mathcal{L}(\xi_0)$ was calculated in steps of 1/8 pixel and the maximum was found by first locating it to the nearest step and then refining the estimate by means of a parabolic fit using also the values in the two adjacent steps.

This procedure gave no significant bias as function of sub-pixel position, and the precision of the location was also practically independent of the position. Using 7 pixels instead of 5 would only give a very marginal improvement ($\simeq -1$ per cent in the standard error), while 3 pixels would deteriorate the results significantly (by +13 to 26 per cent). The location process failed (i.e., no maximum was found within ± 1 pixel of the expected centre) in only 12 of the 720 000 experiments; all of the failures occurred at $G = 21$ mag.

Comparing the rms errors found from these experiments with the analytical formula for different values of k, it was found that the analytical formula could reproduce the simulation results with k ranging from 0.32 (for $G \leq 15$) to 0.25 (for $G \sim 21$). For the following assessment the analytical approximation was used, assuming $k = 0.32$. For bright stars this gives practically the same results as the Monte Carlo simulations, while for faint stars it overestimates the standard error by some 5 to 10 per cent.

7.4.3 Instrument Stability and Calibration Errors

The location measurements in pixel coordinates need to be transformed to angular field coordinates through a geometrical calibration of the focal plane, and subsequently to coordinates on the sky through calibrations of the instrument attitude and basic angle. No complete model exists for this part of the accuracy analysis. It can be assumed that the errors are negligible for faint stars, where photon, readout and background (including crowding) noise will dominate. But for bright stars they could well be a limiting factor. They are therefore represented by a constant variance $\sigma^2_{\rm cal}$ added to the variance of the individual transits (σ^2_ξ).

A rough estimation of $\sigma_{\rm cal}$ was obtained in the following way. It is assumed that the contribution from attitude and field angle calibration errors is equivalent to the total weight of location measurements between $V = 12$ and 20 obtained in 0.86 s (one CCD crossing) in a single field of view pointing to the Galactic pole. Using the numerical results in Section 8 this is found to give a constant contribution of $\simeq 25$ μas to the standard error of the location process.

To this should be added metrology errors associated with the basic angle monitoring, and various systematic effects otherwise unaccounted for. The same amount as above is allocated for each of these effects, making the total contribution equivalent to $\sigma_{\rm cal} \simeq 40$ μas at the level of individual CCD crossings. The high-frequency attitude noise (above 0.05 Hz) of up to 7 μas rms and the expected level of basic-angle fluctuations, 6.4 μas rms, are accommodated within this budget. The CCD contribution should therefore not exceed 10 μas, or 0.0024 μm with the 50 m focal length. This implies a requirement of about 0.003 μm stability over a single scan, that is on time scales between seconds and up to several days.

7.4.4 Propagation to Astrometric Accuracy

The accuracy in the astrometric parameter (parallax or annual proper motion) is finally estimated as:

$$\sigma_a = g_a \left[\frac{\tau_1}{N_i \tau p_{\rm det}(G)} (\sigma^2_\xi + \sigma^2_{\rm cal}) \right]^{1/2} \tag{66}$$

where N_i is the number of instruments, $\tau = L\Omega/4\pi$ the (average) total time available per object and instrument, and $p_{\rm det}(G)$ the detection probability as function of magnitude. It can be noted

that $N_i\tau/\tau_1$ is the (average) total number of CCD crossings of the object during the mission, so that $N_i\tau p_{\rm det}(G)/\tau_1$ is the expected number of detected CCD crossings, on which the estimation of the astrometric parameters is based.

The factor g_a relates the scanning geometry to the determination of the astrometric parameters. It is different for each of the five parameters (parallax, position at mean epoch in two coordinates, proper motion components in two coordinates) and varies as function of position in the sky, mainly as function of ecliptic latitude (β). Simulations of the scanning law are used to determine the mean values of g_a for given mission parameters, and their large-scale variations with ecliptic latitude (Table 8.3). For a revolving scanning law with solar aspect angle ξ (Section 3.1.4) the sky-averaged factor relevant for the determination of parallaxes is:

$$g_\pi = 1.64(\sin\xi)^{-1} \tag{67}$$

For the positions at the mean epoch of the mission the sky-averaged factor is practically independent of ξ, if the mean accuracy in the two coordinates is considered:

$$g_0 = 1.73 \tag{68}$$

For the proper motions, an additional factor enters depending on the length of the mission, L (years):

$$g_\mu = 6.0L^{-1} \tag{69}$$

The estimation of $p_{\rm det}$ is a complex problem, since it depends on many factors besides the brightness of the star. For the accuracy analysis an estimation of $p_{\rm det}$ as function of N, the total number of photoelectrons in the stellar image, was made by means of dedicated simulations. Stellar fields with mean densities up to 25 000 stars brighter than $G = 25$ per CCD (representative of fields in the Galactic plane) were generated and the APM algorithm was used to detect stars (Section 6.7.2). Detection probabilities did not change significantly with star density up to this density, and mean values were computed for different magnitude intervals. The resulting mean detection probabilities are given in Table 8.1 as function of G and N.

7.4.5 Error Margin

An explicit error margin of 20 per cent is added on the astrometric standard errors resulting from the preceding analysis. Results of the accuracy estimation are given in Section 8.1.

7.5 Photometric Accuracy

In general the photometric analysis cannot be separated from the astrometric analysis. The basic model for the pixel or sample values, Equation 43, or more complex variants of it, must be fitted simultaneously for the centroid coordinates ξ_0, η_0 and the photometric quantities b and N. Moreover, this fitting must be done globally, by considering together all the transits of the object throughout the mission, in which the centroid positions are constrained by the appropriate astrometric model. This procedure is referred to as 'global PSF fitting' (as opposed to 'local PSF fitting', where the centroid position is not constrained by an astrometric model).

A simpler procedure, referred to as 'aperture photometry', is used for the photometric accuracy estimation. The idea is to estimate the total intensity of the star by adding the pixel (or sample) values in a suitable area, covering most of the PSF. This sum is then corrected for the mean background estimated from other pixels located some distance away from the stellar image. Provided proper modeling of the pixel values is possible, PSF fitting is always superior to aperture photometry, both in terms of random and systematic errors. It should be emphasized that aperture photometry is assumed only for the accuracy estimation, where it gives more conservative estimates than PSF fitting. But for the real data reductions, global PSF fitting should be the baseline.

7.5.1 Single-Epoch Precision

Formulae for the photometric accuracy, in the case of aperture photometry, have been summarised by Høg et al. (1999c). The stellar intensity (N) is estimated in an 'aperture' of n_s samples. The background is estimated from n_b other samples, in which the contribution from the star is negligible. Let A be the sum of the sample values in the aperture. Then:

$$A = N + n_\mathrm{s} b + \text{noise} \tag{70}$$

where b is the background level. The noise consists of the Poisson noise from the total signal $N + n_\mathrm{s} b$, plus the readout noise from the n_s samples. The variance of A is therefore:

$$\sigma_A^2 = N + n_\mathrm{s}(b + r^2) \tag{71}$$

where r is the rms readout noise per sample.

The sky background b is estimated from the n_b background samples, e.g. by the mean or median method. The resulting estimate, $\widehat{b}$, has standard error σ_b. The estimated stellar intensity is:

$$\widehat{N} = A - n_\mathrm{s}\widehat{b} \tag{72}$$

and has the variance:

$$\sigma_N^2 = \sigma_A^2 + n_\mathrm{s}^2\sigma_b^2 \tag{73}$$

If b is estimated as the mean of the n_b background samples, and if the only noise sources are the Poisson noise from the (constant) background b and the readout noise r, then:

$$\sigma_b^2 = \frac{b + r^2}{n_\mathrm{b}} \tag{74}$$

Combining Equations 71, 73 and 74 gives:

$$\sigma_N^2 = N + n_\mathrm{s}(b + r^2)(1 + n_\mathrm{s}/n_\mathrm{b}) \tag{75}$$

The estimated stellar magnitude is computed as $\widehat{m} = \text{const} - 2.5 \lg \widehat{N}$ and has the standard error:

$$\sigma_m = (2.5 \lg e) \times \frac{\sigma_N}{N} \simeq 1.086\ \frac{\sigma_N}{N} = \frac{1.086}{\sqrt{N}} \left[1 + \frac{(r^2 + b) n_\mathrm{s}(1 + n_\mathrm{s}/n_\mathrm{b})}{N}\right]^{1/2} \tag{76}$$

If the stellar intensity is calculated as the mean of n_obs observations, then σ_m is reduced by the factor $\sqrt{n_\mathrm{obs}}$.

An explicit error margin of 20 per cent is added to the photometric standard errors resulting from the preceding analysis as a first approximation to a realistic performance. Results of the accuracy estimation are given in Section 8.2.

7.5.2 Calibration and Global Accuracy

The formal photometric *precision* reached at the end of the mission, typically by averaging $n_\mathrm{obs} \sim$ 50–100 observations, is on the level of one or a few millimag for the bright stars ($\sim$ 12 mag). A mission average photometric stability of 0.001 mag is therefore desirable. This implies a requirement of 0.01 mag stability in a single scan, that is on time scales between seconds and up to sevaral days. An interesting question is at which level calibration uncertainties will actually set in and limit the actual *accuracy* that can be achieved. Although the experience from Hipparcos shows that a continuously scanning space instrument is an excellent platform for high-accuracy photometry, the actual limitations are difficult to quantify.

An order-of-magnitude estimation of the calibration uncertainties can be made on the basis of the expected number of calibration parameters in relation to the available number and precision of observations. It is important to realize that the photometric calibrations of the CCDs is only to a very small part based on the observation of

'standard stars', i.e. stars with accurately known magnitudes and/or colour indices in some reasonably well-defined photometric system. In fact, such standards will almost exclusively serve to fix the overall zero points of the various magnitude scales defined by the space instruments, and the linearity of these scales over a large dynamic range. Most of the calibration is done by means of anonymous stars, distinguished only by their (presumed, and eventually confirmed) constancy from one transit to the next, and that they are generally well-behaved, e.g. not obviously double. Thus, literally millions of suitable calibration stars are available scattered over the whole sky. The transits of the same calibration star across two different CCDs, or along two different pixel columns on the same CCD, produces a condition equation relating the photometric responses of the two CCDs or columns.

For a concrete example, consider the calibration of the F63B band of the Broad Band Photometer (BBP) of Astro-1. How many instrument parameters are required to describe the photometric response in this band? There are a total of 21 500 pixel columns in the relevant CCDs. Although the samples are 8 columns wide, in practice every pixel column needs to be calibrated, because in a particular transit most of the light is concentrated in one or a few columns. To quantify the response of one column may require one parameter per magnitude interval. Temporal changes of the photometric response should be slow, but assume that effectively an independent calibration is required every 6 months. This gives a total of $N_{\rm par} \simeq 200\,000$ calibration parameters per magnitude interval. How accurately can these be determined?

During the 4 year effective mission length, there are on average 67 transits (observations) per star across this band. Assume that one star in ten is suitable for calibration. This gives, per interval of V magnitude, the total number of transits ($N_{\rm obs}$) indicated in the second column of Table 7.3. The subsequent columns contain the mean number of transits per parameter (the redundancy factor $N_{\rm obs}/N_{\rm par}$), the formal photometric precision per transit (σ_m, calculated for the faintest magnitude in the interval), and finally the quantity $\sigma_{\rm cal} = \sigma_m \times (N_{\rm obs}/N_{\rm par})^{-1/2}$. The redundancy factor is rather high already for the bright stars (near pixel saturation), and very high for the faint stars. $\sigma_{\rm cal}$ can be regarded as an approximate lower limit for the accuracy of the individual calibration parameter. Since the mean magnitude of a star is calculated as the mean of $n_{\rm obs} \simeq 67$ transits, and since these transits usually occur in different pixel columns and consequently have different calibration errors, the effect of the calibration errors in the mean magnitudes is reduced by a further factor. A conclusion from this elementary consideration is that nothing *a priori* prevents the achievement of sub-millimag photometric accuracy, at least in a limited magnitude range, but further more detailed evaluation is required.

Table 7.3: Order-of-magnitude estimation of the achievable photometric calibration accuracy ($\sigma_{\rm cal}$) in the F63B band of the Broad Band Photometer in Astro-1. $N_{\rm obs}$ is the available number of observations (transits of calibration stars) during the mission, $N_{\rm par} \simeq 200\,000$ the number of calibration parameters per magnitude interval, and σ_m the photometric precision in the F63B band per transit.

V (mag)	$N_{\rm obs}$	$N_{\rm obs}/N_{\rm par}$	σ_m (mmag)	$\sigma_{\rm cal}$ (mmag)
12 – 13	1.6×10^7	80	8	0.9
13 – 14	3.8×10^7	190	13	0.9
14 – 15	8.0×10^7	400	20	1.0
15 – 16	1.6×10^8	800	30	1.1
16 – 17	3.0×10^8	1500	50	1.3
17 – 18	5.6×10^8	2800	100	1.9
18 – 19	1.0×10^9	5000	200	2.8
19 – 20	1.8×10^9	9000	450	4.7

7.6 Radial-Velocity Accuracy

7.6.1 Single-Epoch Precision

The spectrum recorded from a single transit across the RVS field can be modeled in analogy with Equation 57:

$$\mathrm{E}(n_j) = r^2 + b + NS(j\Delta\eta - \eta_0) \tag{77}$$

where $S(\eta)$ is the normalized true spectral distribution (multiplied by $T_\lambda Q_\lambda$ and convolved with the pixel width $\Delta\eta$ and the distribution of transverse position errors due to attitude, optical distortion, etc) and η_0 is the required location, which is subsequently transformed to a radial velocity. The radial velocity also produces a stretching or compression of the spectrum in the direction of dispersion. This may have to be considered in the real data analysis, but for the accuracy estimation it is a second-order effect. The best precision by which η_0 can be determined

is given by the CR bound, in analogy with Equation 59:

$$\sigma_\eta = \left[\frac{N^2}{\Delta\eta} \int_{-\infty}^{+\infty} \frac{[S'(\eta)]^2}{r^2 + b + NS(\eta)} \, \mathrm{d}\eta \right]^{-1/2} \tag{78}$$

Although this expression possibly gives a useful lower bound for the precision in η_0, it cannot be used without modification in the practical accuracy analysis. The main reason is that it assumes that the true shape of $S(\eta)$ is perfectly known. For the astrometric accuracy estimation in Section 7 it could reasonably be assumed that the line-spread function is perfectly known, as a result of its simple shape and the high redundancy of data for its calibration. But there are too many varieties of stellar spectra for such an assumption to hold in the case of $S(\eta)$. The traditional remedy is to match the observations against a template spectrum $X(\eta)$ rather than $S(\eta)$, e.g. by digital cross-correlation. More generally, the analysis would minimise a goodness-of-fit statistic such as:

$$\chi^2 = \sum_j \left[n_j - b - NX(j\Delta\eta - \eta_0) \right]^2 w_j \tag{79}$$

For the Poisson noise model (where readout noise is included in b) the weights w_j are inversely proportional to $\mathrm{E}(n_j)$, which could be estimated from the fitted template, or from the raw or smoothed n_j. Assuming that the estimation of b and N are nearly orthogonal to that of η_0, the following expression is derived for the precision of η_0:

$$\sigma_{\eta_0} \simeq \frac{N^{-1/2}}{-\sum_j w_j Z_j S_j} \left[\sum_j w_j^2 Y_j^2 S_j + \frac{r^2 + b}{N} \sum_j w_j^2 Y_j^2 \right]^{1/2} \tag{80}$$

where $S_j \equiv S(j\Delta\eta - \eta_0)$, $Y_j \equiv X'(j\Delta\eta - \eta_0)$, and $Z_j \equiv X''(j\Delta\eta - \eta_0)$. This formula highlights the dependence on N, r and b in a form similar to Equation 64 for the astrometric accuracy, and Equation 76 for the photometric accuracy.

The possibility to determine η_0 by template matching depends on the negative correlation between $X''(\eta)$ and $S(\eta)$ — if no such correlation exists, σ_{η_0} is unbounded. It is therefore important to have for each observed spectrum a template with the best possible matching. This is probably best achieved within the framework of the automated spectral parameterization discussed in Section 9.4.3.

Another detail which needs consideration is what happens at the endpoints of the spectrum. The support of $S(\eta)$ is not determined by the stellar spectrum, but by $T_\lambda Q_\lambda$, which is fixed in the rest frame of the observer. The flanks of $S(\eta)$ must therefore not be allowed to influence the matching. In the accuracy estimation, this is taken into account by restricting the sums in Equation 80 to the samples (j) where $T_\lambda Q_\lambda$ is relatively flat.

7.6.2 Calibration and Global Accuracy

The two main calibrations needed for the radial-velocity measurements are: (i) calibration of template spectra for a wide range of spectral types (including different metallicities and chemical peculiarities) as function of time and the transverse coordinate η — this is equivalent to the PSF calibration in the astrometric and photometric data analysis; and (ii) calibration of the radial-velocity zero point as function of time and the transverse coordinate — this is equivalent to the geometric calibration in the astrometric data analysis. The two calibrations are not independent: the radial-velocity zero point depends on the adopted template origins. The latter should be chosen to give the same radial-velocity zero point for all templates. The template origins could be determined via the observed spectra of radial-velocity standards, but a more likely scenario for GAIA is that this can be achieved mainly via model-atmosphere spectra.

The absolute spectral shifts are obtained by comparing the observed spectra with the template spectra calculated for the known star positions. Thus, the accurate astrometry provided by the

Spectro Sky Mapper, alone or in combination with the Astro instruments, is an essential element of the processing. The positional requirements are modest compared with the astrometric accuracy (1 km s^{-1} corresponds to 0.04 pixel across scan or 40 mas on the sky) and with the mechanical stability of the instrument. The geometric zero point should therefore be completely stable and contribute negligibly to the mean radial velocity uncertainty obtained over the mission.

For very faint, individual spectra the precision will be worse than Equation 80 due to higher-order terms in N^{-1} neglected in the previous analysis. This can be circumvented by estimating the radial velocity simultaneously from all the recorded spectra of a given object, using a procedure equivalent to the 'global PSF fitting' in the photometry (Section 7.5).

8 Accuracy Assessment: Results

8.1 Astrometric Accuracy

8.1.1 Basic Accuracy Predictions

The details of the extensive accuracy assessment which has been made for the astrometric instrument is given in Section 7.4. Results for the sky averages of the accuracy in parallax and proper motions, as a function of the V and G magnitudes, are given in Tables 8.4–8.7. For the faintest stars these accuracy estimates take into account that a given star may not be observed in all transits. They include the factor $p_{\rm det}(G)^{-1/2}$, where $p_{\rm det}(G)$ is the detection probability as a function of the G magnitude (Equation 66), obtained from simulating star detection in ASM1 using the APM algorithm (Section 6.7.2). These are given in Table 8.1.

Since, by construction, the G magnitude yields a rather uniform accuracy as a function of spectral type, useful mean accuracies can be derived from a straight mean of the (un-reddened) values in Table 8.6 and 8.7. The resulting mean accuracies are given in Table 8.2. An approximate analytical fit to the tabular values for the parallax accuracy, σ_π, which also takes into account the slight colour dependence (due to the widening of the PSF at longer wavelengths) is:

$$\sigma_\pi \simeq (7 + 105z + 1.3z^2 + 6\times10^{-10}z^6)^{1/2} \times [0.96 + 0.04(V-I)] \tag{81}$$

where $z = 10^{0.4(G-15)}$. This formula is valid for the entire range of magnitudes and colours covered in Tables 8.4–8.7. For the mean position and proper motions errors, σ_0 and σ_μ, the following mean relation can be used (cf. Equations 67–69):

$$\sigma_0 = 0.87\,\sigma_\pi\,, \qquad \sigma_\mu = 0.75\,\sigma_\pi \tag{82}$$

This provides a simple accuracy formulation in combination with Equation 29, which gives a prescription for the G magnitude as a function of V and $V-I$ valid over a wide colour interval. The expected standard errors vary somewhat over the sky as a result of the scanning law. The mean variations with ecliptic latitude (β) are given in Table 8.3. The approximate ecliptic latitude can be calculated from the equatorial coordinates (α,δ) or the galactic coordinates (l,b) using:

$$\sin\beta \simeq 0.9175\sin\delta - 0.3978\cos\delta\sin\alpha \tag{83}$$

$$\simeq 0.4971\sin b + 0.8677\cos b\sin(l-6.38^\circ) \tag{84}$$

Sensitivity to Sky Background For a sky background of $V = 22.5$ mag arcsec^{-2} (1.5 mag fainter than assumed in Section 6.4.3) the standard errors improve slightly at the faintest magnitudes, by factors of between 0.996 at $G = 17$ mag, and 0.963 at $G = 21$ mag. That the improvement is so small is because the read-out noise dominates, even at a background of $V = 21$ mag arcsec^{-2} (where $b = 5.54$ in Equation 64, compared with $r^2 = 36$). Even in the limit of faint stars and zero background the improvement is only 7.4 per cent. Another small effect is that the sky background also affects the detection probabilities in the ASM, which enters through Equation 66.

8.1.2 Accuracy Folded With the Galaxy Model

The accuracy analysis provides estimates of the absolute parallax accuracy as function of magnitude, colour and position on sky. For instance, if V, $V-I$ and β (ecliptic latitude) are known, the parallax accuracy σ_π is obtained by means of Equations 29 and 81, together with Table 8.3. For all Galactic and astrophysical investigations it is however the *relative* parallax accuracy σ_π/π which is more relevant. This is easily computed if the typical distance to the objects is known.

The number of stars whose distances can be determined to a certain relative accuracy can be estimated using the Galaxy model described in Section 6.4.4. For a given apparent magnitude and direction, the model provides the distribution of stars along the line of sight (in a small solid angle), as well as their distribution in colour index at each distance. It is then simple to compute the fraction of stars having relative parallax accuracy below a certain limit R.

Table 8.1: Detection probability in a single transit versus G magnitude. N is the number of electrons produced by the star during 0.86 s integration time per CCD.

G [mag]	17.0	17.5	18.0	18.5	19.0	19.5	20.0	20.5	21.0	21.5	22.0
N [e$^-$]	5450	3440	2170	1370	864	545	344	217	86	55	34
$p_{\rm det}$	1.000	0.995	0.988	0.981	0.970	0.948	0.863	0.750	0.414	0.056	0.002

Table 8.2: Mean accuracy in parallax (σ_π), position (at mid-epoch, σ_0) and proper motion (σ_μ), versus G magnitude. The values are sky averages (see Table 8.3 for the variations across the sky).

G (mag)	10	11	12	13	14	15	16	17	18	19	20	21
σ_π (μas)	4	4	4	5	7	11	17	27	45	80	160	500
σ_0 (μas)	3	3	3	4	6	9	15	23	39	70	140	440
σ_μ (μas yr^{-1})	3	3	3	4	5	8	13	20	34	60	120	380

Table 8.3: Variation with ecliptic latitude (β) of: $n_{\rm obs}$ = number of focal-plane passages per astrometric instrument; $g_\pi(\beta)/g_\pi$ = ratio of geometric factor for parallax accuracy to sky average; $g_{\mu\lambda*}(\beta)/g_\mu$ = ratio of geometric factor for proper motion accuracy in ecliptic longitude to the mean proper motion factor; $g_{\mu\beta}(\beta)/g_\mu$ = ratio of geometric factor for proper motion accuracy in ecliptic latitude to the mean proper motion factor. The sky-averaged errors in Tables 8.2 and 8.4–8.7 should be multiplied by these factors to give the expected standard error at a particular ecliptic latitude. The geometric factor for the position at mean epoch (g_0) varies as for g_π.

$\lvert\beta\rvert$	$n_{\rm obs}$	$\frac{g_\pi(\beta)}{g_\pi}$	$\frac{g_{\mu\lambda*}(\beta)}{g_\mu}$	$\frac{g_{\mu\beta}(\beta)}{g_\mu}$	$\lvert\beta\rvert$	$n_{\rm obs}$	$\frac{g_\pi(\beta)}{g_\pi}$	$\frac{g_{\mu\lambda*}(\beta)}{g_\mu}$	$\frac{g_{\mu\beta}(\beta)}{g_\mu}$
0 – 5	54	1.17	1.21	1.02	50 – 55	64	0.84	0.91	1.01
5 – 10	56	1.16	1.19	1.01	55 – 60	60	0.82	0.95	1.02
10 – 15	57	1.14	1.15	1.02	60 – 65	58	0.81	0.98	1.03
15 – 20	60	1.12	1.10	1.00	65 – 70	56	0.79	1.01	1.04
20 – 25	65	1.10	1.03	1.01	70 – 75	54	0.78	1.02	1.04
25 – 30	74	1.06	0.94	1.01	75 – 80	53	0.78	1.04	1.05
30 – 35	96	1.00	0.77	0.99	80 – 85	53	0.78	1.04	1.06
35 – 40	95	0.93	0.70	0.97	85 – 90	53	0.77	1.05	1.05
40 – 45	76	0.89	0.80	0.99					
45 – 50	68	0.86	0.86	1.01	mean sky	67	1.00	0.99	1.01

Multiplying by the number of stars per magnitude interval and integrating up to (say) $V = 21$ gives the total star density with $\sigma_\pi/\pi < R$ (Figures 8.1–8.5). The corresponding limit on the absolute magnitude error is $\lesssim 2.17R$, assuming no or well-determined extinction. $R = 0.1$ is often considered a 'good' relative accuracy for luminosity calibration of stellar samples, and at the present time, very few stars have distances known to better than 1 per cent.

Figures 8.1–8.4 show, for selected directions, the fraction of stars with relative parallax accuracies below $R = 0.01 - 1.0$. No plots are given for $|b| < 10°$, since the computed distributions at low latitudes are very sensitive to the extinction/reddening model, which is quite uncertain. Very good distance information ($R = 0.1$ or better) will be obtained for virtually all stars brighter than $V = 15$ and for significant fractions down to much fainter magnitudes, e.g. 10–50 per cent at $V \sim 18$, depending on direction. Figure 8.5 shows the expected number densities with relative parallax accuracies below $R = 0.01$, 0.02, 0.05, 0.1 and 0.2. A magnitude limit of $V = 21$ was assumed, giving a total number of stars (irrespective of parallax accuracy) of nearly 1 billion. Table 8.8 shows the expected numbers with relative accuracy better than R, obtained by integrating the densities in Figure 8.5 over the sky. The corresponding Hipparcos numbers are also shown.

Table 8.4: Sky averages of the estimated accuracy in parallax (σ_π, in microarcsec) as a function of spectral type (or colours), reddening (A_V) and V magnitude.

				V =											
Sp	A_V	$V-R$	$V-I$	10	11	12	13	14	15	16	17	18	19	20	21
B1V	0.0	–0.12	–0.31	3.68	3.68	3.68	4.41	6.17	9.24	14.4	22.9	37.4	64.3	123.	303.
A0V	0.0	0.01	0.01	3.66	3.66	3.69	4.84	6.99	10.7	16.7	26.8	44.3	78.0	158.	604.
A3V	0.0	0.02	0.05	3.66	3.66	3.71	4.90	7.10	10.8	17.0	27.3	45.2	79.8	163.	685.
A5V	0.0	0.08	0.17	3.66	3.66	3.72	4.93	7.15	10.9	17.2	27.6	45.7	80.7	165.	715.
F2V	0.0	0.19	0.38	3.66	3.66	3.72	4.94	7.19	11.0	17.3	27.8	46.0	81.3	166.	716.
F6V	0.0	0.25	0.53	3.66	3.66	3.70	4.93	7.17	11.0	17.3	27.7	45.9	81.0	165.	677.
F8V	0.0	0.31	0.62	3.66	3.66	3.70	4.93	7.18	11.0	17.3	27.7	46.0	81.1	165.	664.
G2V	0.0	0.34	0.72	3.66	3.66	3.70	4.94	7.20	11.0	17.4	27.9	46.2	81.5	166.	669.
K3V	0.0	0.50	0.99	3.65	3.65	3.67	4.92	7.19	11.0	17.3	27.9	46.1	81.2	164.	608.
M0V	0.0	0.84	1.72	3.66	3.66	3.66	4.48	6.46	9.83	15.4	24.6	40.2	69.2	133.	324.
M8V	0.0	1.24	3.23	3.72	3.72	3.72	3.72	4.62	6.76	10.4	16.3	26.0	42.7	73.8	142.
G8III	0.0	0.48	0.97	3.65	3.65	3.69	4.95	7.24	11.1	17.5	28.1	46.5	82.1	167.	648.
K3III	0.0	0.64	1.19	3.65	3.65	3.65	4.83	7.05	10.8	17.0	27.3	45.0	78.8	158.	486.
M0III	0.0	0.80	1.72	3.67	3.67	3.67	4.52	6.53	9.95	15.6	24.9	40.8	70.4	135.	333.
M7III	0.0	1.98	4.69	3.78	3.78	3.78	3.78	3.78	4.07	5.82	8.82	13.8	21.9	35.5	59.7
B0IB	0.0	–0.11	–0.30	3.68	3.68	3.68	4.38	6.12	9.17	14.3	22.7	37.1	63.6	122.	298.
B1V	5.0	0.76	1.68	3.67	3.67	3.67	4.51	6.50	9.89	15.5	24.7	40.5	69.8	134.	328.
A0V	5.0	0.89	1.98	3.68	3.68	3.68	4.30	6.15	9.32	14.6	23.2	37.8	64.4	121.	281.
A3V	5.0	0.90	2.02	3.68	3.68	3.68	4.28	6.11	9.26	14.5	23.0	37.5	63.9	119.	277.
A5V	5.0	0.96	2.13	3.69	3.69	3.69	4.18	5.95	8.99	14.0	22.3	36.3	61.5	114.	258.
F2V	5.0	1.06	2.33	3.69	3.69	3.69	4.03	5.67	8.53	13.3	21.1	34.1	57.4	105.	228.
F6V	5.0	1.13	2.47	3.69	3.69	3.69	3.92	5.48	8.21	12.7	20.2	32.6	54.7	98.5	210.
F8V	5.0	1.18	2.56	3.70	3.70	3.70	3.84	5.35	7.99	12.4	19.6	31.6	52.8	94.5	198.
G2V	5.0	1.22	2.65	3.70	3.70	3.70	3.78	5.23	7.79	12.1	19.1	30.7	51.1	91.0	188.
K3V	5.0	1.36	2.88	3.70	3.70	3.70	3.70	4.91	7.24	11.2	17.6	28.2	46.6	81.6	162.
M0V	5.0	1.69	3.55	3.72	3.72	3.72	3.72	4.11	5.86	8.88	13.8	22.0	35.7	60.3	111.
M8V	5.0	2.27	5.13	3.78	3.78	3.78	3.78	3.78	3.78	5.03	7.49	11.6	18.3	29.4	48.7
G8III	5.0	1.34	2.88	3.71	3.71	3.71	3.71	4.93	7.27	11.2	17.7	28.4	46.9	82.2	164.
K3III	5.0	1.49	3.07	3.71	3.71	3.71	3.71	4.64	6.78	10.4	16.3	26.2	43.0	74.2	143.
M0III	5.0	1.67	3.58	3.73	3.73	3.73	3.73	4.08	5.81	8.78	13.7	21.8	35.3	59.5	109.
M7III	5.0	3.08	6.56	3.83	3.83	3.83	3.83	3.83	3.83	3.83	4.10	5.88	8.94	14.0	22.2
B0IB	5.0	0.76	1.69	3.68	3.68	3.68	4.51	6.50	9.90	15.5	24.8	40.6	69.8	134.	328.

Table 8.5: Sky averages of the estimated accuracy in proper motion (σ_μ, in microarcsec yr^{-1}) as a function of spectral type (or colours), reddening (A_V) and V magnitude.

				V =											
Sp	A_V	$V-R$	$V-I$	10	11	12	13	14	15	16	17	18	19	20	21
B1V	0.0	–0.12	–0.31	2.76	2.76	2.76	3.30	4.62	6.92	10.8	17.1	28.0	48.2	92.4	227.
A0V	0.0	0.01	0.01	2.75	2.75	2.76	3.63	5.24	7.98	12.5	20.1	33.2	58.4	118.	452.
A3V	0.0	0.02	0.05	2.74	2.74	2.78	3.67	5.32	8.12	12.7	20.5	33.9	59.8	122.	513.
A5V	0.0	0.08	0.17	2.74	2.74	2.79	3.69	5.36	8.19	12.9	20.7	34.2	60.5	124.	536.
F2V	0.0	0.19	0.38	2.74	2.74	2.79	3.70	5.39	8.24	13.0	20.8	34.5	60.9	125.	536.
F6V	0.0	0.25	0.53	2.74	2.74	2.78	3.69	5.37	8.22	12.9	20.8	34.4	60.7	124.	507.
F8V	0.0	0.31	0.62	2.74	2.74	2.77	3.69	5.38	8.23	12.9	20.8	34.4	60.7	124.	497.
G2V	0.0	0.34	0.72	2.74	2.74	2.77	3.70	5.40	8.27	13.0	20.9	34.6	61.0	124.	501.
K3V	0.0	0.50	0.99	2.74	2.74	2.75	3.69	5.39	8.26	13.0	20.9	34.5	60.8	123.	455.
M0V	0.0	0.84	1.72	2.74	2.74	2.74	3.36	4.84	7.36	11.5	18.4	30.1	51.9	99.3	243.
M8V	0.0	1.24	3.23	2.79	2.79	2.79	2.79	3.46	5.06	7.76	12.2	19.5	32.0	55.3	107.
G8III	0.0	0.48	0.97	2.74	2.74	2.76	3.71	5.43	8.33	13.1	21.1	34.9	61.5	125.	486.
K3III	0.0	0.64	1.19	2.74	2.74	2.74	3.62	5.28	8.10	12.7	20.4	33.7	59.1	118.	364.
M0III	0.0	0.80	1.72	2.75	2.75	2.75	3.39	4.89	7.46	11.7	18.7	30.6	52.7	101.	250.
M7III	0.0	1.98	4.69	2.83	2.83	2.83	2.83	2.83	3.05	4.36	6.61	10.3	16.4	26.6	44.8
B0IB	0.0	–0.11	–0.30	2.76	2.76	2.76	3.28	4.59	6.87	10.7	17.0	27.8	47.7	91.2	223.
B1V	5.0	0.76	1.68	2.75	2.75	2.75	3.38	4.87	7.41	11.6	18.5	30.4	52.3	100.	246.
A0V	5.0	0.89	1.98	2.76	2.76	2.76	3.22	4.61	6.99	10.9	17.4	28.3	48.3	90.5	211.
A3V	5.0	0.90	2.02	2.76	2.76	2.76	3.21	4.58	6.94	10.8	17.3	28.1	47.9	89.5	207.
A5V	5.0	0.96	2.13	2.76	2.76	2.76	3.14	4.46	6.74	10.5	16.7	27.2	46.1	85.3	193.
F2V	5.0	1.06	2.33	2.77	2.77	2.77	3.02	4.25	6.39	9.94	15.8	25.6	43.0	78.3	171.
F6V	5.0	1.13	2.47	2.77	2.77	2.77	2.94	4.11	6.15	9.55	15.1	24.5	41.0	73.8	157.
F8V	5.0	1.18	2.56	2.77	2.77	2.77	2.88	4.01	5.98	9.27	14.7	23.7	39.5	70.8	148.
G2V	5.0	1.22	2.65	2.77	2.77	2.77	2.83	3.92	5.84	9.03	14.3	23.0	38.3	68.2	141.
K3V	5.0	1.36	2.88	2.78	2.78	2.78	2.78	3.68	5.43	8.36	13.2	21.1	34.9	61.2	122.
M0V	5.0	1.69	3.55	2.79	2.79	2.79	2.79	3.08	4.39	6.65	10.4	16.5	26.8	45.2	82.8
M8V	5.0	2.27	5.13	2.83	2.83	2.83	2.83	2.83	2.83	3.77	5.61	8.68	13.7	22.0	36.5
G8III	5.0	1.34	2.88	2.78	2.78	2.78	2.78	3.69	5.45	8.40	13.2	21.3	35.1	61.6	123.
K3III	5.0	1.49	3.07	2.78	2.78	2.78	2.78	3.48	5.08	7.79	12.2	19.6	32.2	55.6	107.
M0III	5.0	1.67	3.58	2.80	2.80	2.80	2.80	3.05	4.35	6.58	10.3	16.3	26.4	44.5	81.3
M7III	5.0	3.08	6.56	2.87	2.87	2.87	2.87	2.87	2.87	2.87	3.07	4.41	6.70	10.5	16.6
B0IB	5.0	0.76	1.69	2.75	2.75	2.75	3.38	4.87	7.42	11.6	18.6	30.4	52.3	100.	246.

Table 8.6: Sky averages of the estimated accuracy in parallax (σ_π, in microarcsec) as a function of spectral type (or colours), reddening (A_V) and G magnitude.

				$G =$												
Sp	A_V	$V-R$	$V-I$	10	11	12	13	14	15	16	17	18	19	20	21
B1V	0.0	–0.12	–0.31	3.68	3.68	3.68	4.59	6.50	9.81	15.3	24.5	40.2	69.8	137.	354.
A0V	0.0	0.01	0.01	3.66	3.66	3.69	4.84	6.99	10.7	16.7	26.8	44.3	78.0	158.	604.
A3V	0.0	0.02	0.05	3.66	3.66	3.70	4.88	7.05	10.8	16.9	27.1	44.9	79.0	161.	648.
A5V	0.0	0.08	0.17	3.66	3.66	3.68	4.84	7.00	10.7	16.8	26.9	44.4	78.1	158.	586.
F2V	0.0	0.19	0.38	3.66	3.66	3.66	4.78	6.89	10.5	16.5	26.4	43.6	76.3	153.	489.
F6V	0.0	0.25	0.53	3.66	3.66	3.66	4.74	6.83	10.4	16.3	26.1	43.1	75.2	150.	440.
F8V	0.0	0.31	0.62	3.66	3.66	3.66	4.73	6.83	10.4	16.3	26.1	43.1	75.2	149.	431.
G2V	0.0	0.34	0.72	3.66	3.66	3.66	4.75	6.87	10.5	16.4	26.3	43.4	75.9	151.	442.
K3V	0.0	0.50	0.99	3.65	3.65	3.65	4.81	6.99	10.7	16.8	27.0	44.5	77.9	156.	479.
M0V	0.0	0.84	1.72	3.66	3.66	3.66	4.87	7.15	11.0	17.3	27.7	45.8	80.3	161.	496.
M8V	0.0	1.24	3.23	3.72	3.72	3.73	5.17	7.70	11.9	18.8	30.3	50.3	89.2	182.	674.
G8III	0.0	0.48	0.97	3.65	3.65	3.65	4.83	7.03	10.8	16.9	27.1	44.8	78.6	157.	501.
K3III	0.0	0.64	1.19	3.65	3.65	3.65	4.83	7.05	10.8	17.0	27.3	45.0	78.9	158.	487.
M0III	0.0	0.80	1.72	3.67	3.67	3.67	4.92	7.23	11.1	17.5	28.1	46.5	81.7	164.	534.
M7III	0.0	1.98	4.69	3.78	3.78	3.78	5.12	7.64	11.8	18.7	30.1	49.8	87.8	176.	504.
B0IB	0.0	–0.11	–0.30	3.68	3.68	3.68	4.56	6.45	9.72	15.2	24.2	39.7	68.9	135.	345.
B1V	5.0	0.76	1.68	3.67	3.67	3.67	4.87	7.14	11.0	17.2	27.7	45.7	80.1	160.	480.
A0V	5.0	0.89	1.98	3.68	3.68	3.68	4.93	7.25	11.1	17.6	28.2	46.6	81.9	164.	503.
A3V	5.0	0.90	2.02	3.68	3.68	3.68	4.94	7.28	11.2	17.6	28.3	46.9	82.3	165.	515.
A5V	5.0	0.96	2.13	3.69	3.69	3.69	4.94	7.28	11.2	17.6	28.3	46.9	82.3	165.	506.
F2V	5.0	1.06	2.33	3.69	3.69	3.69	4.95	7.30	11.2	17.7	28.5	47.1	82.7	166.	502.
F6V	5.0	1.13	2.47	3.69	3.69	3.69	4.97	7.33	11.3	17.8	28.6	47.3	83.2	167.	510.
F8V	5.0	1.18	2.56	3.70	3.70	3.70	4.96	7.32	11.3	17.8	28.6	47.3	83.0	166.	495.
G2V	5.0	1.22	2.65	3.70	3.70	3.70	4.97	7.35	11.3	17.9	28.7	47.5	83.5	168.	507.
K3V	5.0	1.36	2.88	3.70	3.70	3.70	4.97	7.35	11.3	17.9	28.7	47.5	83.4	167.	485.
M0V	5.0	1.69	3.55	3.72	3.72	3.72	5.01	7.43	11.5	18.1	29.1	48.2	84.6	170.	486.
M8V	5.0	2.27	5.13	3.78	3.78	3.78	5.27	7.89	12.2	19.4	31.2	51.9	92.0	188.	643.
G8III	5.0	1.34	2.88	3.71	3.71	3.71	4.98	7.36	11.4	17.9	28.8	47.6	83.6	167.	490.
K3III	5.0	1.49	3.07	3.71	3.71	3.71	4.94	7.30	11.3	17.8	28.5	47.1	82.6	165.	451.
M0III	5.0	1.67	3.58	3.73	3.73	3.73	5.00	7.42	11.5	18.1	29.1	48.1	84.4	169.	469.
M7III	5.0	3.08	6.56	3.83	3.83	3.86	5.45	8.21	12.8	20.3	32.7	54.5	97.2	200.	758.
B0IB	5.0	0.76	1.69	3.68	3.68	3.68	4.88	7.16	11.0	17.3	27.8	45.9	80.5	161.	487.

Table 8.7: Sky averages of the estimated accuracy in proper motion (σ_μ, in microarcsec yr^{-1}) as a function of spectral type (or colours), reddening (A_V) and G magnitude.

				$G =$												
Sp	A_V	$V-R$	$V-I$	10	11	12	13	14	15	16	17	18	19	20	21
B1V	0.0	–0.12	–0.31	2.76	2.76	2.76	3.44	4.87	7.35	11.5	18.3	30.1	52.3	103.	265.
A0V	0.0	0.01	0.01	2.75	2.75	2.76	3.63	5.24	7.98	12.5	20.1	33.2	58.4	118.	452.
A3V	0.0	0.02	0.05	2.74	2.74	2.77	3.65	5.29	8.06	12.7	20.3	33.6	59.2	120.	485.
A5V	0.0	0.08	0.17	2.74	2.74	2.76	3.63	5.24	8.00	12.6	20.1	33.3	58.5	118.	439.
F2V	0.0	0.19	0.38	2.74	2.74	2.74	3.58	5.17	7.87	12.3	19.8	32.6	57.1	114.	366.
F6V	0.0	0.25	0.53	2.74	2.74	2.74	3.55	5.12	7.80	12.2	19.6	32.3	56.4	112.	330.
F8V	0.0	0.31	0.62	2.74	2.74	2.74	3.55	5.12	7.80	12.2	19.6	32.3	56.4	112.	323.
G2V	0.0	0.34	0.72	2.74	2.74	2.74	3.56	5.15	7.85	12.3	19.7	32.5	56.8	113.	331.
K3V	0.0	0.50	0.99	2.74	2.74	2.74	3.60	5.24	8.02	12.6	20.2	33.3	58.4	117.	359.
M0V	0.0	0.84	1.72	2.74	2.74	2.74	3.65	5.35	8.22	12.9	20.8	34.3	60.2	121.	371.
M8V	0.0	1.24	3.23	2.79	2.79	2.80	3.87	5.77	8.92	14.1	22.7	37.7	66.8	136.	505.
G8III	0.0	0.48	0.97	2.74	2.74	2.74	3.62	5.27	8.07	12.7	20.3	33.6	58.9	118.	375.
K3III	0.0	0.64	1.19	2.74	2.74	2.74	3.62	5.28	8.10	12.7	20.4	33.7	59.1	118.	365.
M0III	0.0	0.80	1.72	2.75	2.75	2.75	3.69	5.42	8.33	13.1	21.1	34.8	61.2	123.	400.
M7III	0.0	1.98	4.69	2.83	2.83	2.83	3.84	5.72	8.86	14.0	22.5	37.3	65.8	132.	377.
B0IB	0.0	–0.11	–0.30	2.76	2.76	2.76	3.42	4.83	7.28	11.4	18.1	29.8	51.6	101.	259.
B1V	5.0	0.76	1.68	2.75	2.75	2.75	3.65	5.35	8.21	12.9	20.7	34.2	60.0	120.	360.
A0V	5.0	0.89	1.98	2.76	2.76	2.76	3.69	5.43	8.35	13.2	21.1	34.9	61.3	123.	377.
A3V	5.0	0.90	2.02	2.76	2.76	2.76	3.70	5.45	8.39	13.2	21.2	35.1	61.7	124.	385.
A5V	5.0	0.96	2.13	2.76	2.76	2.76	3.70	5.45	8.39	13.2	21.2	35.1	61.7	124.	379.
F2V	5.0	1.06	2.33	2.77	2.77	2.77	3.71	5.47	8.42	13.3	21.3	35.3	61.9	124.	376.
F6V	5.0	1.13	2.47	2.77	2.77	2.77	3.72	5.49	8.46	13.3	21.4	35.5	62.3	125.	382.
F8V	5.0	1.18	2.56	2.77	2.77	2.77	3.72	5.49	8.45	13.3	21.4	35.4	62.2	125.	371.
G2V	5.0	1.22	2.65	2.77	2.77	2.77	3.73	5.51	8.49	13.4	21.5	35.6	62.5	126.	380.
K3V	5.0	1.36	2.88	2.78	2.78	2.78	3.72	5.50	8.49	13.4	21.5	35.6	62.4	125.	364.
M0V	5.0	1.69	3.55	2.79	2.79	2.79	3.75	5.57	8.60	13.6	21.8	36.1	63.4	127.	364.
M8V	5.0	2.27	5.13	2.83	2.83	2.83	3.95	5.91	9.18	14.5	23.4	38.9	69.0	141.	481.
G8III	5.0	1.34	2.88	2.78	2.78	2.78	3.73	5.52	8.51	13.4	21.6	35.7	62.6	125.	367.
K3III	5.0	1.49	3.07	2.78	2.78	2.78	3.70	5.47	8.44	13.3	21.4	35.3	61.9	123.	338.
M0III	5.0	1.67	3.58	2.80	2.80	2.80	3.75	5.56	8.58	13.5	21.8	36.0	63.2	126.	351.
M7III	5.0	3.08	6.56	2.87	2.87	2.89	4.09	6.15	9.57	15.2	24.5	40.8	72.8	150.	568.
B0IB	5.0	0.76	1.69	2.75	2.75	2.75	3.66	5.36	8.24	13.0	20.8	34.4	60.3	121.	365.

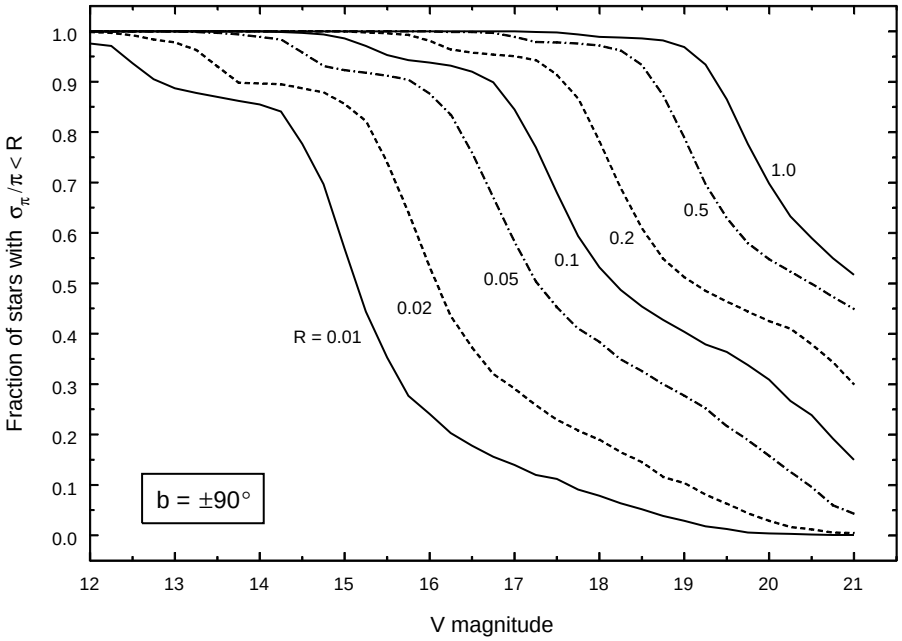


Figure 8.1: Fraction of stars at a given magnitude having a relative parallax error less than 1, 2, 5, 10, 20, 50 and 100 per cent. The curves were computed from Tables 8.4 and 8.3 for a direction towards the Galactic poles, using the Galaxy model described in Section 6.4.4.

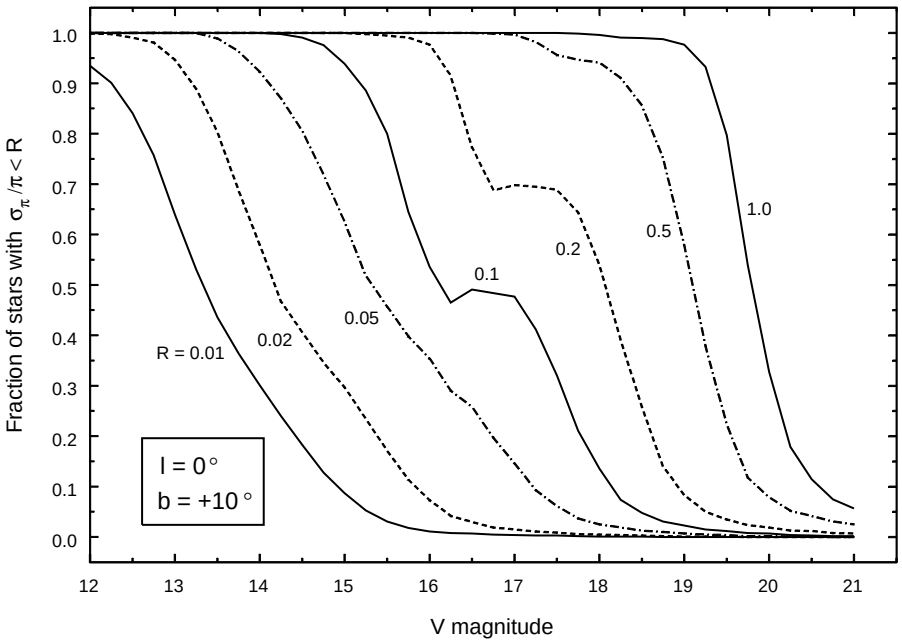


Figure 8.2: Same as Figure 8.1, but for Galactic coordinates $(l = 0°, b = +10°)$.

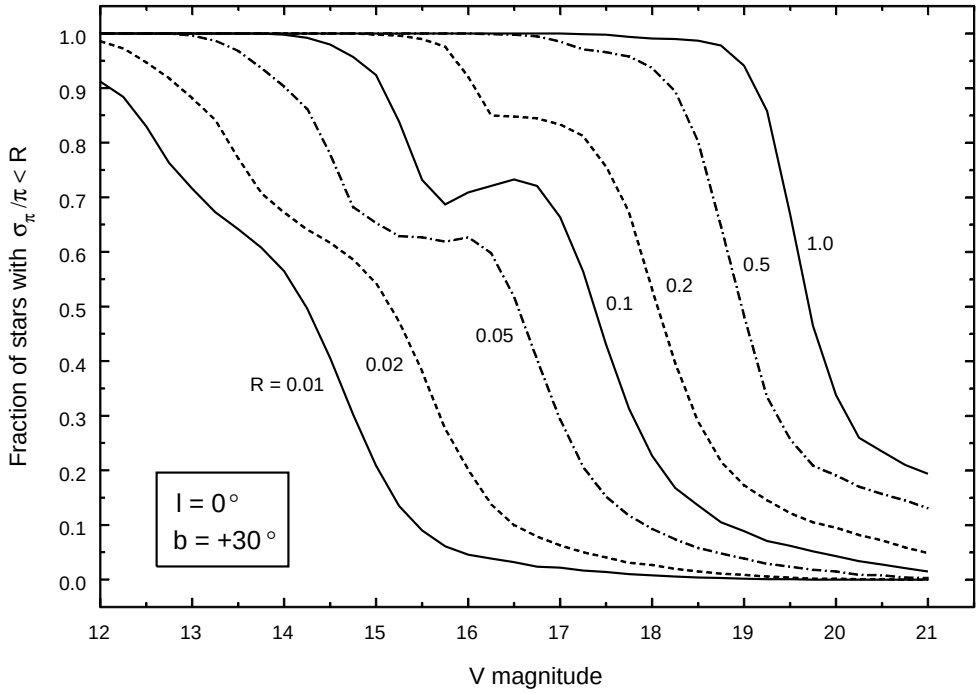


Figure 8.3: Same as Figure 8.1, but for Galactic coordinates $(l = 0°, b = +30°)$.

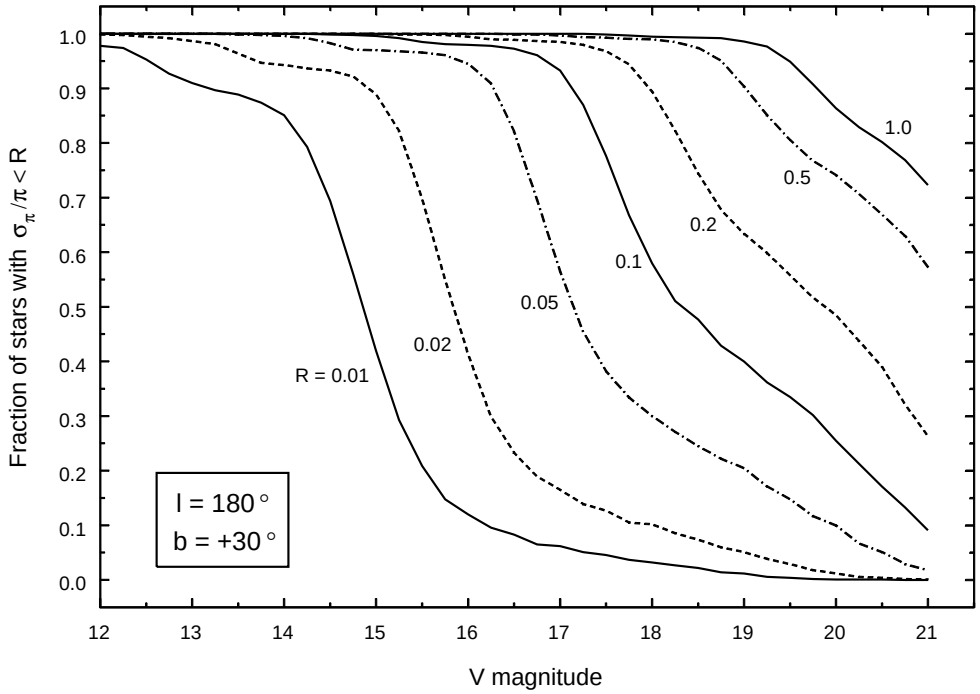


Figure 8.4: Same as Figure 8.1, but for Galactic coordinates $(l = 180°, b = +30°)$.

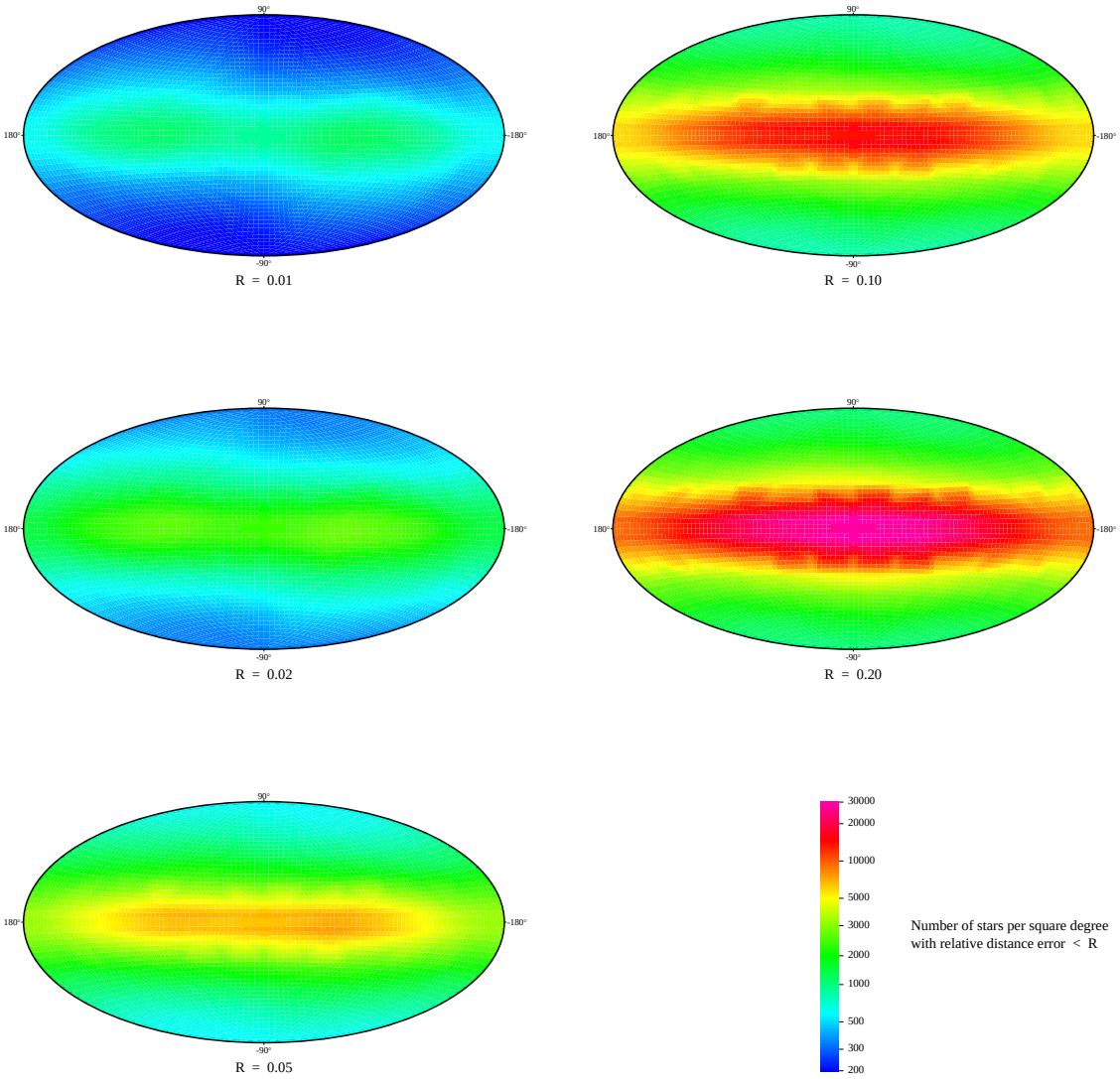


Figure 8.5: Density of stars with $V < 21$ and relative parallax error less than $R = 0.01$, 0.02, 0.05, 0.1 and 0.2. Calculations are based on the accuracy estimates in Tables 8.4 and 8.3, using the Galaxy model described in Section 6.4.4. The whole sky is shown in Aitoff's projection with the direction $l = b = 0$ at the centre of the maps.

Table 8.8: Total number of stars with relative distance error below a certain limit: $\sigma_\pi/\pi \leq R$. For GAIA the numbers are estimated from the accuracy analysis and Galaxy model described in this Chapter. For Hipparcos the numbers have been derived by means of Celestia 2000 (ESA SP–1220).

Maximum relative parallax error (R)	Number of stars	
	GAIA	Hipparcos
0.01	20×10^6	188
0.02	40×10^6	878
0.05	100×10^6	6238
0.10	180×10^6	21014
0.20	290×10^6	49545
0.50	500×10^6	90186
1.00	680×10^6	104579

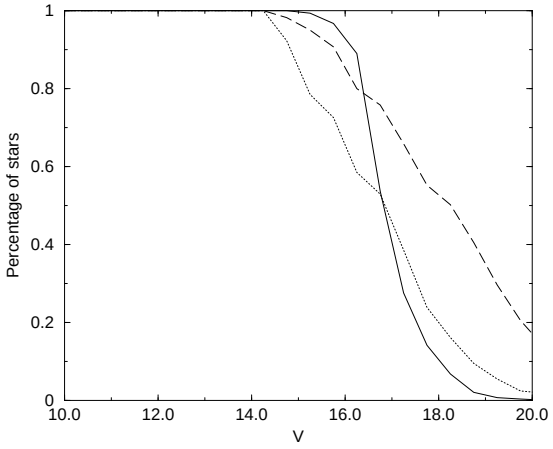


Figure 8.6: Percentage of stars with $G \leq 20$ mag with relative error in parallax smaller than 10 per cent as a function of the V magnitude for three different galactic directions. Solid line: $(l, b) = (0°, 2°)$, dotted line: $(l, b) = (0°, 30°)$, dashed line: $(l, b) = (0°, 90°)$

Using the predicted error on the astrometric parameters given in this section (depending on the G magnitude and the ecliptic coordinates) the error distribution in parallaxes and tangential velocities for the stars simulated by the Galaxy model presented in Section 6.4.4 can be derived and studied. In Figure 8.6 the percentage of stars with $\frac{\sigma_\pi}{\pi} < 10$ per cent as a function of the V magnitude is plotted for three representative galactic directions. All stars brighter than $V = 14.5$ mag have relative error in parallax smaller than 10 per cent. For fainter magnitudes the fraction of stars with this accuracy decreases, reaching at $V = 17.5$ mag the $20 - 60$ per cent, depending on the galactic latitude. Thus, at the end of the mission, parallaxes with a relative error better than 10 per cent will be available for millions of stars of all types and populations.

The error distributions for the tangential velocity are plotted in Figure 8.7 for two limiting G magnitudes. For $G \leq 15$ mag more than 85 per cent of the stars will have errors smaller than 5 km s^{-1}, and 75 per cent of them will have errors smaller than 2 km s^{-1}. These figures worsen if all the stars with $G \leq 20$ mag are considered, but even in this case, for galactic latitudes higher than $5 - 10°$, 40 per cent of the stars observed by GAIA will have tangential velocities accurate to better than 10 km s^{-1}. These figures are a challenge for the measurement of radial velocities using the radial velocity spectrometer instrument, because if one requires a similar accuracy in radial and tangential velocities for kinematic and dynamical studies, then it is desirable to acquire radial velocities with an accuracy better than 5 km s^{-1} at least up to $G = 15$ mag.

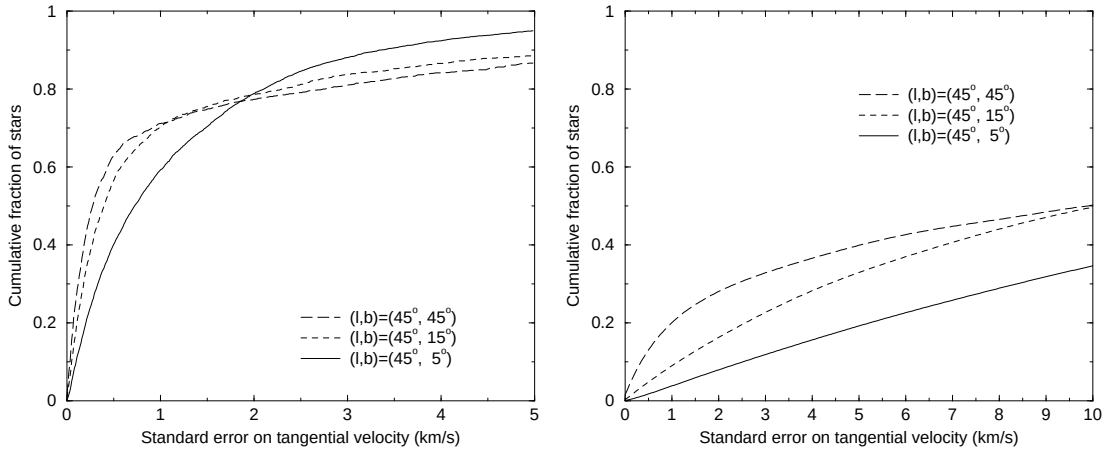


Figure 8.7: Cumulative number of stars with (a, left) $G \leq 15$ mag, and (b, right) $G \leq 20$ mag as a function of the standard error on tangential velocities.

8.2 Photometric Accuracy

8.2.1 Basic Accuracy Predictions

Photometric observations in a number of colour bands (see Section 2.3.2) will be obtained for all stars detected. Accuracy estimates for these observations have been evaluated, and a summary of the most important results is provided in this section. The assessments assume that data for 100 field transits for the (wider field) Spectro field of view and 67 transits per Astro field of view are collected during a mission of 4 years effective observing time. Each field crossing gives a number of CCD samples (see Figure 3.7) in each colour band which are transmitted to ground.

A quick photometric reduction will be carried out for each field crossing, providing rapid epoch photometry results which will be used for verification purposes, and for scientifically time-critical phenomena such as the detection of supernovae, burst, lensing, or other transient events.

A significantly improved photometric reduction can be obtained later in the mission, when accurate satellite attitude, CCD calibration data, and astrometric information for each star become available. Since no on-board processing is carried out, except for some data compression, it will be possible to reduce all photometric data for each star in an optimal way, using all available information (cf. Section 9.4.2). At this stage the highest angular resolution of multiple and complex stars is also obtained, including detection of close companions to the stars detected in the ASM which are considerably fainter than the originally detected star.

The two Astro telescopes provide photometry in a wide spectral band defined by the CCD sensitivity. This G band photometry (comparable to V-band magnitudes) is performed by the Astro sky mapper (ASM) and by the 17 CCDs covering the width of the astrometric field. Photometry at different times throughout the mission ('epoch photometry') from an average of one field crossing has a standard error of typically 0.01 mag for $G \sim 18.5$ mag. The epoch photometry is more accurate in G than in other bands, and is therefore better suited for detection of variability (for example, the signal is about 5.0 times higher than that in the F45B band for the same integration time, and the standard error per CCD is about 2.2 times smaller). Altogether, a given error in G per field crossing is obtained for stars nearly 5 mag fainter than in the F45B band.

The two Astro telescopes also provide broad-band photometry in the four bands F33B, F45B, F63B and F82B described in Section 2.3.2. The magnitudes resulting from averaging 67 observations obtained during a mission time of 4 years will have a precision of about 0.02 mag in the F63B band at $V \sim 20.0$ mag for all spectral types (Figure 8.8).

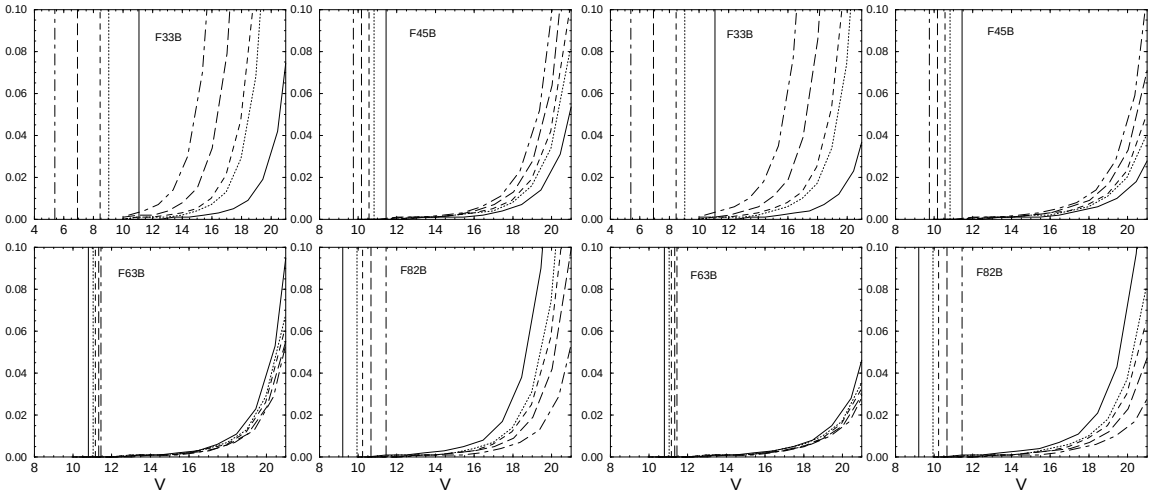


Figure 8.8: Photometric accuracy (in mag) for 67 observations of a star in the two Astro (left, Astro-1, right, Astro-2) telescopes in each of the four broad photometric bands. The accuracy has been computed, for an unreddened star, assuming (for Astro-1) the use of 6 samples for the determination of the star flux and of 6 samples for the measurement of the background, resulting in a total read noise of 6.2 e^-. For the Astro-2 telescope the number of samples is 2 and 4 (star and background), resulting in a total read noise of 5.0 e^-. Different line types indicate different stellar types: from left to right in the left upper panel M0III, K0III, G2V, F0V, B1V. The saturation limit for each spectral type is also indicated, corresponding to 500 000 e^- per star.

The Spectro telescope will collect photometry in the 11 bands described in Section 2.3.2. Each star will be observed on average over about 100 field crossings. The resulting average magnitudes will have a precision of 0.01 mag in the F57 band at $V = 19$ mag, and 0.02 mag in the F33 band for an unreddened G2V star (Figure 8.9) The corresponding accuracies for the colour indices (from which the astrophysical quantities such as $T_{\rm eff}$, etc., are derived) are shown in Figure 8.10.

The estimated standard errors illustrated in the plots contain photon noise and read noise from the star and background observations (assumed to be at the level of the zodiacal light observed in HST observations, of about $V = 22.5$ mag arcsec^{-2}, see Section 6.4.3). Lossless compression of the transmitted data has been assumed. Other error sources exist, e.g. cosmic ray events, calibration errors, disturbance from stars or galaxies, non-ideal performance of the CCDs, all of which have been tentatively taken into account by multiplying the resulting standard errors by a factor of 1.2.

A further asymptotic standard error of about 0.01 mag should possibly be added to the plots to obtain realistic values for the absolute accuracy of the photometry. This will mainly affect stars brighter than $V \simeq 18$ mag. Additional errors will be encountered for small-separation double stars and stars in dense regions which, although accounting for only a small fraction of all stars, may be astrophysically important.

The standard error of a single 'epoch photometry' measurement in the Astro field is $\simeq 8$ times the error shown in Figure 8.8, or about 0.08 mag for the F63B band for $V \simeq 19$ mag. When individual observations are considered for the detection of variability, realistic standard errors for isolated, sufficiently bright, single stars may well be smaller than this asymptotic error of 0.01 mag, and indeed sub-millimag photometric accuracy is a realistic possibility, according to Section 7.5.2.

The sensitivity of the resulting accuracies to the assumed value of the sky background is illustrated in Table 8.9.

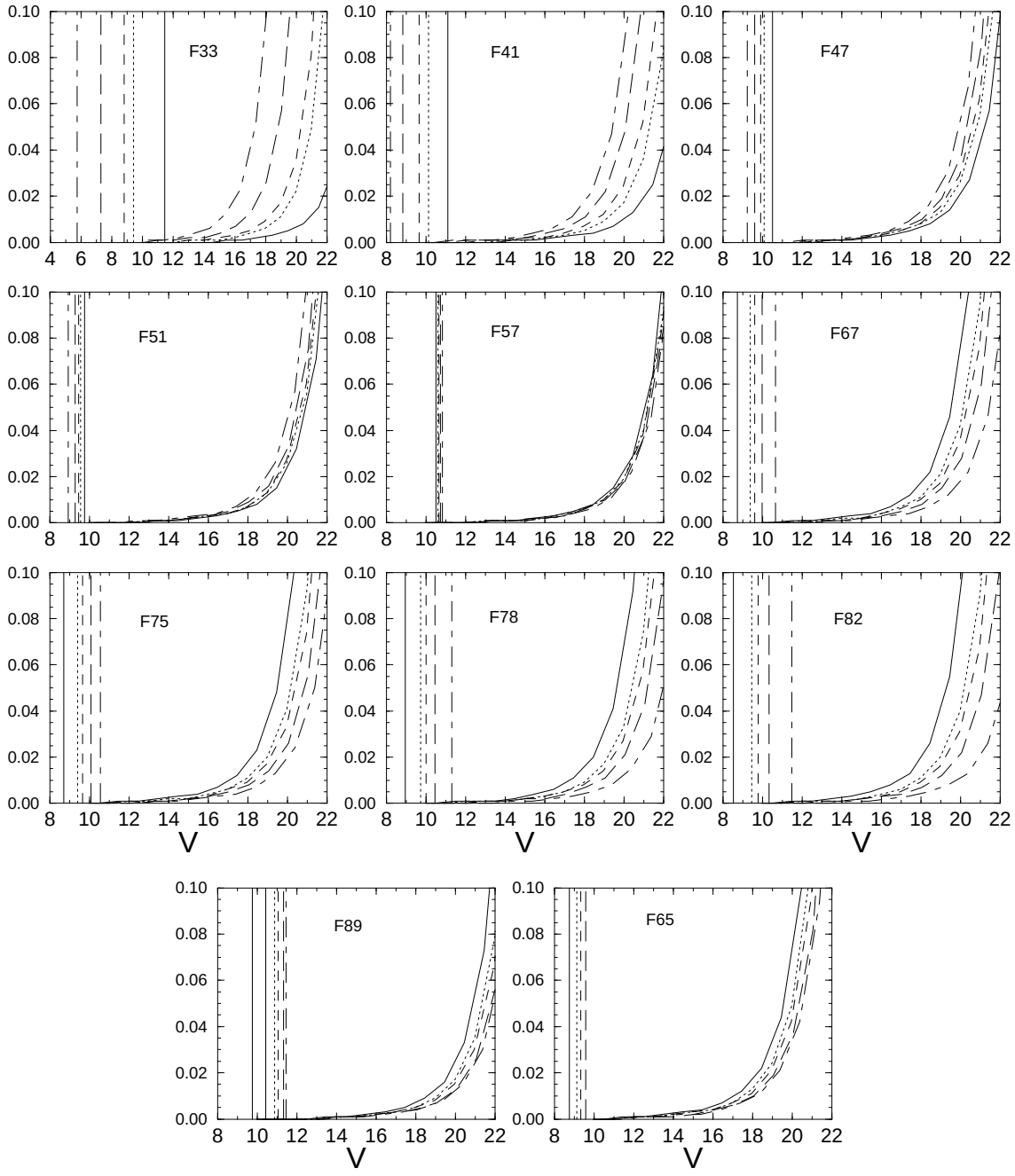


Figure 8.9: Photometric accuracy (in mag) for 100 observations of a star in the Spectro telescope in each of the 11 medium photometric bands. The accuracy has been computed, for an unreddened star, using 3 samples for the determination of the star flux and 6 samples for the measurement of the background, resulting in a total read noise of 3.0 e^-. Different line types indicate different stellar types: from left to right in the left upper panel M0III, K0III, G2V, F0V, B1V. The saturation limit for each spectral type is also indicated, corresponding to 500 000 e^- per star.

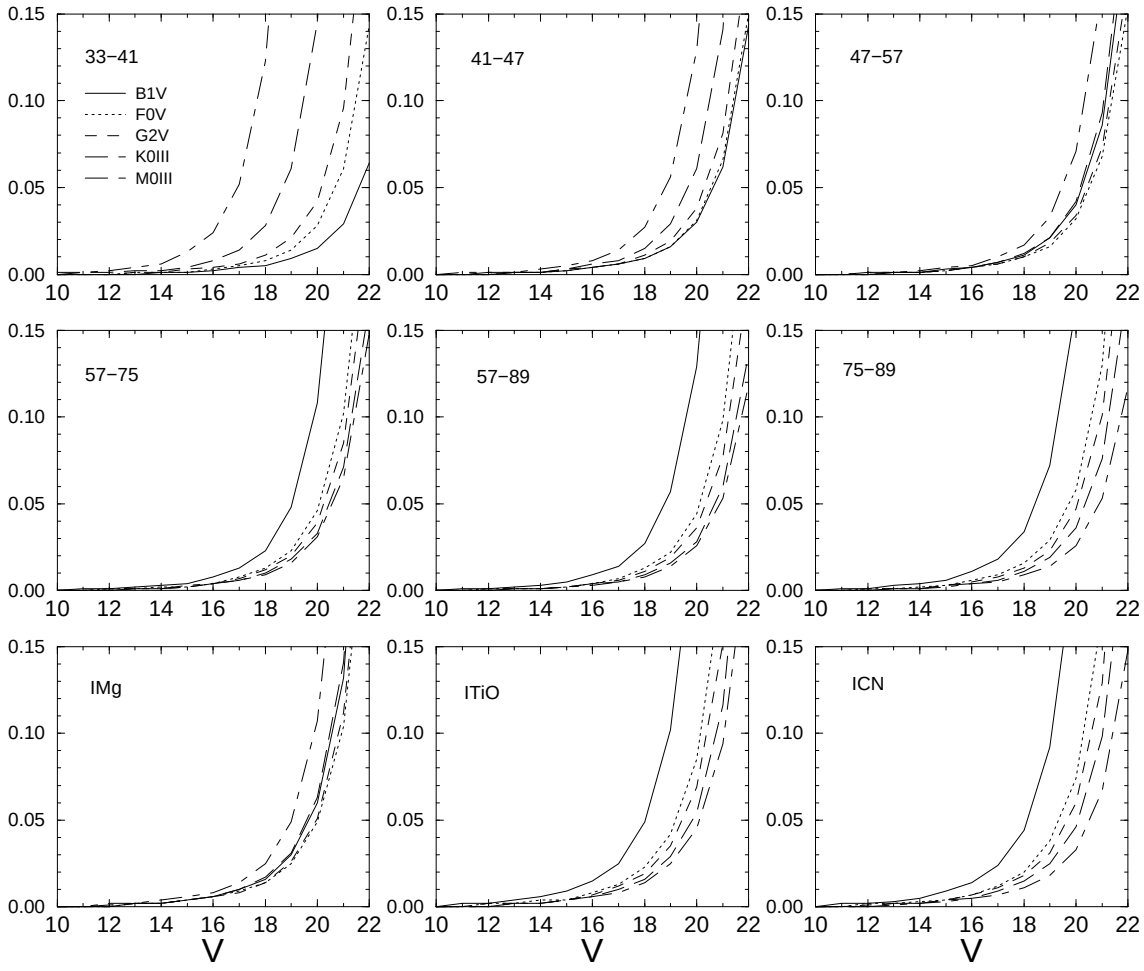


Figure 8.10: Photometric accuracy (in mag) for 100 observations of a star in the Spectro telescopes in each of the relevant colour indices derived from the 11 medium photometric bands. The accuracy has been computed for an unreddened star. Different line types indicate different stellar types: from left to right in the left upper panel M0III, K0III, G2V, F0V, B1V.

Table 8.9: Sensitivity of the photometric accuracies to the assumed value of the sky background. Nominal photometric accuracies have been computed assuming a sky background of $V = 22.5$ mag arcsec^{-2} (Section 6.4.3). This table indicates the degradation of the photometric standard errors for the medium band F57 that would be obtained in the presence of a brighter sky background of $V = 21$ mag arcsec^{-2}, as a function of spectral type and G magnitude.

Spectral Type	< 17	18	19	20	21
B1 V	1.00	0.89	0.83	0.72	0.68
F0 V	1.00	1.00	0.83	0.76	0.70
G2 V	1.00	1.00	0.83	0.75	0.70
K0 III	1.00	1.00	0.83	0.75	0.70
M0 III	1.00	0.86	0.79	0.76	0.70

8.2.2 Accuracy of Some Astrophysical Diagnostics

Using the photometric indices discussed in Section 2.3.2 the photometric accuracies for the colour indices shown in the previous section can be converted in accuracies on the astrophysical parameters. These accuracy estimates are computed for the case of a perfectly known background structure and of a perfect detector. In practice, the accuracy at the faint end will again be limited by the imperfect knowledge of the background, while at the bright end detector systematics will likely prevent the photometry from reaching the sub-millimag accuracies implied by the photon statistics and by the effect of known noise sources.

As an example, Figure 8.11 shows the uncertainty in effective temperature for G- and M-type dwarfs ($T_{\rm eff} = 5750$ and 3500 K respectively) as a function of magnitude, for single transit and mission-average photometry. Figure 8.13 (left) shows the uncertainty in [Ti/H] for M dwarfs, while Figure 8.13 (right) shows the uncertainty in [M/H] for Population I K giants. Again, these plots show the effect of photon noise (plus read-out noise, a uniform sky background, and a factor of 1.2), so that in practice an effective floor to the accuracy, at the bright end, will come from both the intrinsic cosmic spread as well as from the uncertainty in the model atmospheres used to convert from observed colours to astrophysical parameters. Based on current state of the art, such systematic uncertainties will be at the level of $\simeq 50$ K in effective temperature and $\simeq 0.05$ dex both in [Ti/H] and in [M/H].

Figure 8.12 (left) shows a typical reddening-free diagram, i.e. the location of main sequence and giant stars spanning spectral types from O to M in a plane in which the effect of reddening is (for the earlier spectral types) effectively negligible. The error bars are computed for a single transit of a $G \simeq 15$ mag star. The early-type (O to A2) stars form the ascending branch of the diagram, and the very near coincidence of the reddened and unreddened sequences, together with the small error bars, show the feasibility of deriving the spectral parameters (and thus the reddening) from the photometry without any *a priori* knowledge of the reddening or of the spectral type. The Balmer discontinuity causes the change of behavior at spectral type about A2; the overlapping of the various sequences (reddened and unreddened, giant and main sequence) at redder colours together with the large error bars shows that for F and later type stars the reddening-free approach is not feasible. The effect of metallicity is shown in Figure 8.12 (right).

The small resulting uncertainties on the astrophysical parameters derived from the photometry, across the whole magnitude range accessible to GAIA, allow the characterization of the target populations to the precision level required for the derivation of accurate distribution functions in both abundance and effective temperature (and thus, through the mass-luminosity raltion, the mass) for the key kinematic tracers for GAIA, satisfying the main astrophysical requirements for the GAIA photometry. The small uncertainties on the key astrophysical parameters (limited by model systematics) will also allow detailed and accurate characterization of individual targets.

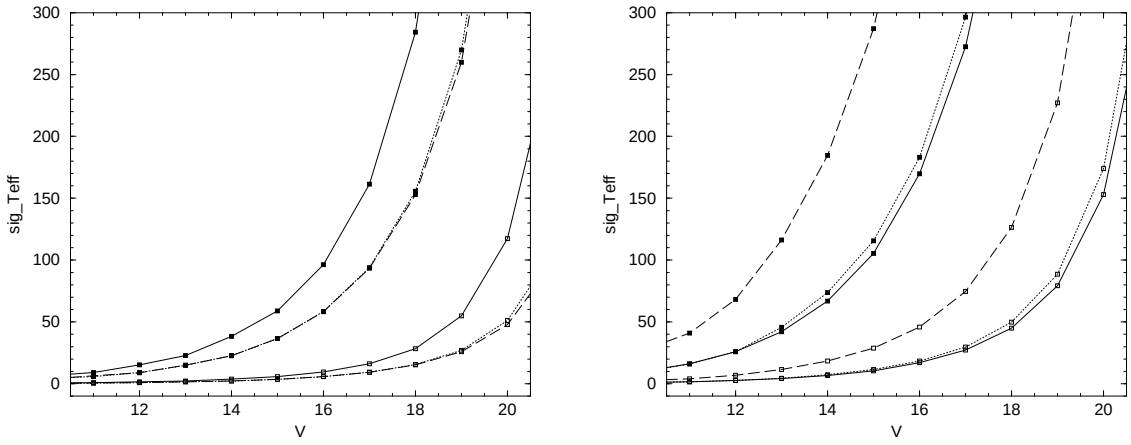


Figure 8.11: Left: the errors on $T_{\rm eff}$ when derived from C47–57 (continuous line), C57–75 (dotted line) and C75–89 (dashed line) indices, for M dwarfs with $T_{\rm eff}$ = 3500 K, as a function of the V magnitude. Solid symbols correspond to the errors for a single observation, open symbols to the mission-average measurements (100 single observations). Right: the errors on $T_{\rm eff}$ when derived form C47–57 (continuous line), C57–75 (dotted line) and C75–89 (dashed line) indices, for G dwarfs with $T_{\rm eff}$ = 5750 K, as a function of the V magnitude. Solid symbols correspond to the errors for a single observation, open symbols to the mission-average measurements (100 single observations).

8.3 Radial Velocity Accuracy

The performance of the radial velocity spectrograph has been studied through both simulations and through real observations of the spectral region under analysis for a variety of stars. Simulations were performed by producing synthetic spectra of stars with different atmospheric parameters in the Ca II spectral region (using the Kurucz model atmospheres and spectral synthesis code, i.e. the ATLAS9 and SYNTHE 1996 packages), degrading it to the resolution and signal-to-noise ratio of a single GAIA observation and then determining the radial velocity by the cross-correlation method. The intrinsic spectral resolution of the simulated spectra is $R = 30\,000$, and they were sampled at 0.075 nm per pixel, covering the region of the Ca II triplet, i.e. from 849–874 nm. The photon budget was computed from the design shown in Section 3.5.2, and in addition to Poisson noise both the total read-out noise (3 e^-) and the sky background (normalized to a solar spectrum with $V = 22.5$ arcsec^{-2}) were considered.

Both individual (single observation) spectra and mission-average spectra were considered. As each individual spectrum is sampled differently, depending on the position of the star with respect to the spectrograph, mission-average spectra were obtained by producing a set of 100 individual spectra with random sampling, which was then linearly re-sampled at 1/6 of pixel resolution and then added together. The value of the radial velocity was determined by cross-correlating the simulated observed spectrum with an infinite S/N template and determining the position of the cross-correlation peak.

Figure 8.14 shows the expected accuracy of the radial velocity determination for hot ($T_{\rm eff}$ = 12 500 K) and cool ($T_{\rm eff}$ = 5500 K) stars under different conditions. The top left panel shows the accuracy obtained for individual observations as well as from the average of all the observations obtained throughout the 5 yr mission (mission averages). For cool stars, the mission-average velocity accuracy is $\sigma_v \simeq 5$ km s^{-1} at $V = 18$, while for hot stars the performance is limited to $\sigma_v \simeq 10$ km s^{-1} at $V = 16$ mag.

The dependence of the radial velocity performance on the stellar parameters has been explored by considering stars of different metal abundances and surface gravities, as well as rotational velocities. For cool stars decreasing metallicity will (as expected) degrade the radial velocity accuracy. The

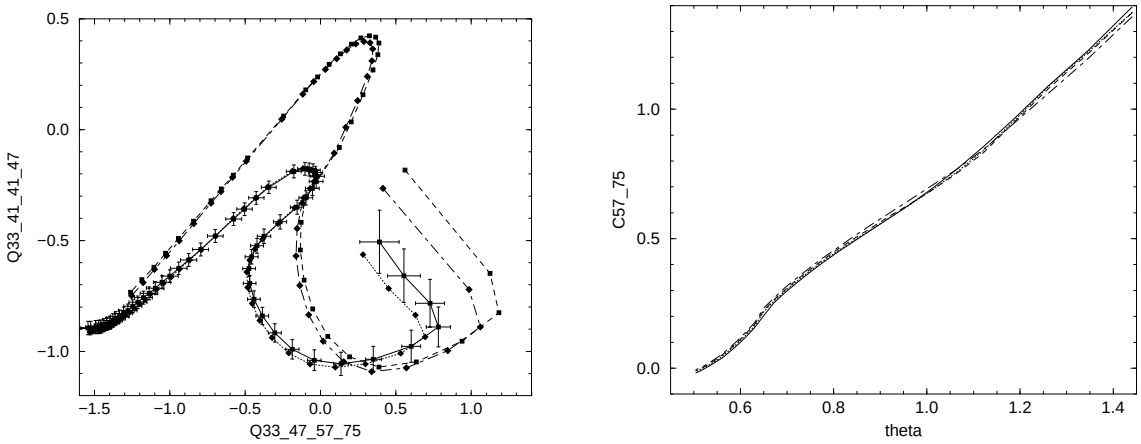


Figure 8.12: Left: a reddening-free diagram employing the $Q_{33-41-41-47}$ versus $Q_{33-47-57-75}$ indices. The lines show the position of unreddened and reddened main sequence stars (lower curves, log g = 4.5, $A_v = 0.0$ and 1.0), and giants (upper curves, log g = 2.5, $A_v = 0.0$ and 1.0). Error bars correspond to a $G = 15$ mag star. The ascending branch is the position of OBA stars, where the effect of reddening in such a diagram is effectively negligible. The overlap of the curves, together with the larger error bars, show that the reddening-free approach cannot be used for later-type stars. Right: the relationship between C57–75 and θ for main sequence stars (log g = 4.5) with [M/H] = 0.0, −0.5, −1.0 and −2.0 in the $T_{\rm eff}$ range 10 000 to 3500 K. Note the absence of significant residuals as a function of [M/H].

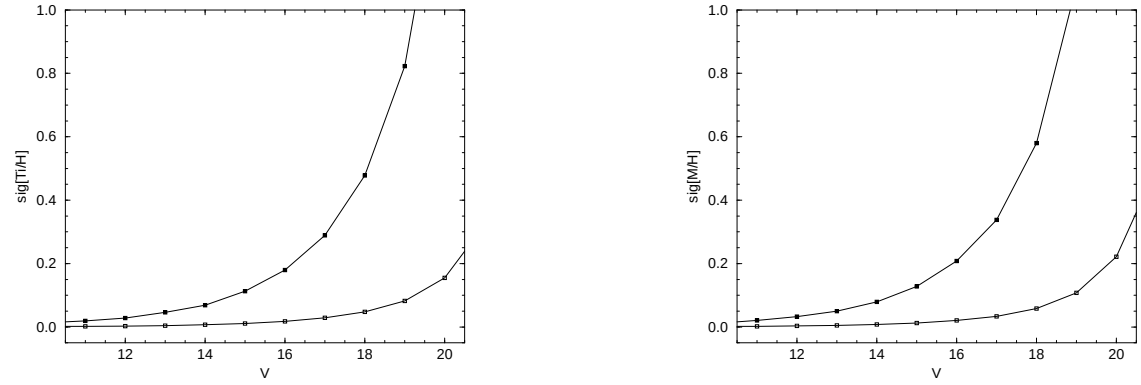


Figure 8.13: Left: the errors on [Ti/H] when derived from the TiO index, for $T_{\rm eff}$ = 3500 K M dwarfs, as a function of the V magnitude, for single observations (filled symbols) and for mission average-measurements (open symbols). Right: the errors on [M/H] when derived from the metallicity index, I_Z, for unreddened Population I K giants ($T_{\rm eff}$ = 4500 K), as a function of the V magnitude, for single observations (filled symbols) and for mission average-measurements (open symbols). For Population II giants the corresponding error is about twice as large.

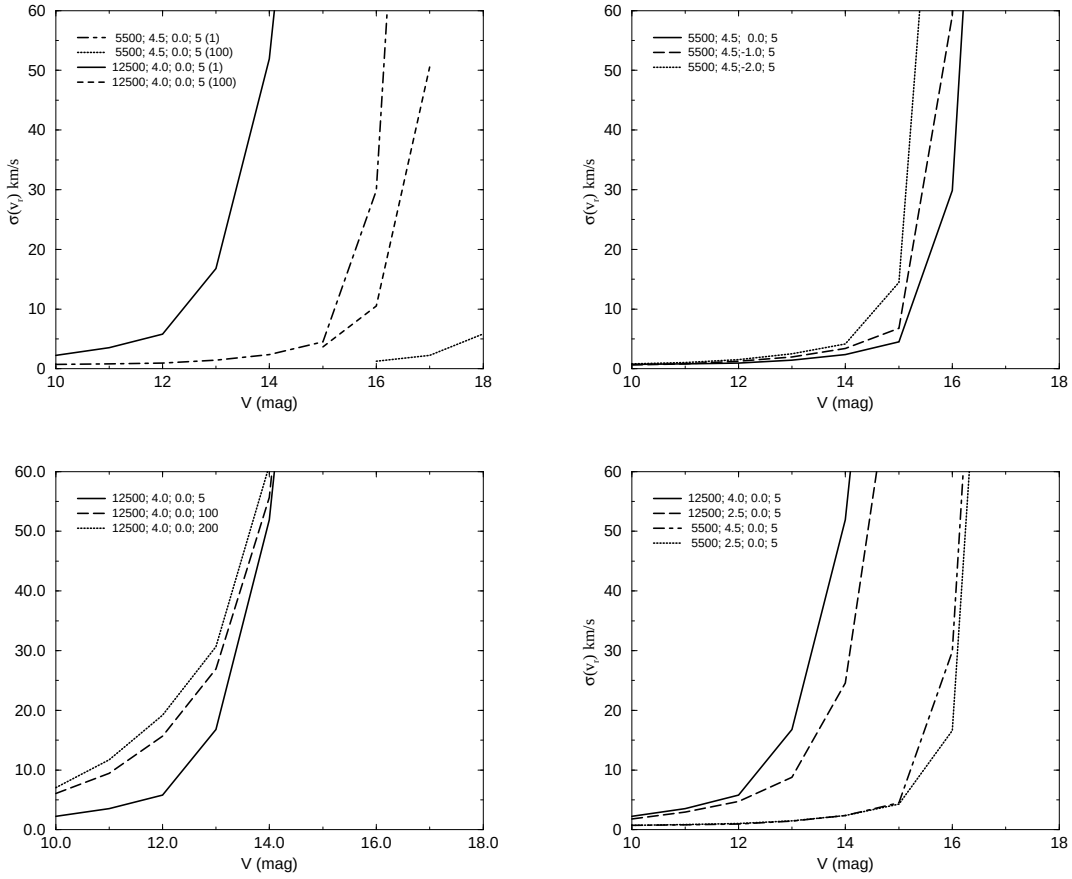


Figure 8.14: Top left: radial velocity accuracy versus magnitude for individual observations and systemic velocities, for hot and cool stars. Curves are labelled with the corresponding values of $T_{\rm eff}$ (K), log g (cgs units), [Fe/H], and rotational velocity (km/s). The number in parentheses is the number of individual observations: the two highest lines are the individual accuracies for the hot and cool stars respectively; the systematic accuracies are the corresponding lines to the right. Top right: the effect of metallicity on accuracy for cool stars (more metal rich spectra yield more accurate radial velocities). Bottom left: the effect of rotational velocity for hot stars (fast rotation decreases the accuracy). Bottom right: the effect of surface gravity for cool stars.

effect is however small (see top right panel of Figure 8.14), being effectively negligible for the range of disk metallicities, and implying a loss of ≤ 1 mag in limiting magnitude for halo metallicities. Gravity (bottom right panel of Figure 8.14) is irrelevant for cool stars, while it is important for the hot stars, where low-gravity (giant stars) result in more accurate radial velocity determination (due to the narrowing of the Paschen H lines). Finally, rotation significantly decreases the accuracy (bottom left panel of Figure 8.14) for fast rotating hot stars.

One additional potential (systematic) error source is the mismatch between the observed star and the template spectrum. In the performance assessments shown in Figure 8.14 a $S/N = \infty$ with the same atmospheric parameters as the simulated observed star has been used. If a template spectrum with different atmospheric parameters is used, a systematic bias in the radial velocity may result. This effect has been studied by using different templates for the source spectrum. For sufficiently cool stars (i.e. $T_{\rm eff} \leq 10\,000$ K) the final accuracy is rather insensitive to the possible mismatch between observed and template spectrum (with the additional uncertainty < 1 km s^{-1} for F stars, becoming effectively zero for K stars). For hotter stars the effect will be larger (up to $\simeq 30$ km s^{-1} for early B stars), due to the changing blend between the Ca II and the Paschen H lines.

However, given the high accuracy with which the GAIA photometry will determine the atmospheric parameters for the target stars, this will not be a limiting factor. In practice, especially for cool stars, astrophysical effects will introduce additional systematic differences between the physical center of mass velocity of the star being observed and its Doppler-determined radial velocity, which can reach up to $\simeq 2$ km/s. These include the effect of stellar surface turbulence (specially for the later spectral types) and the gravitational redshift.

One essential trade-off element in the design of the radial velocity spectrograph is the dispersion: in principle higher dispersion will yield spectra containing more information, at the price of a shallower limiting magnitude and a higher confusion problem at high stellar densities. However, the Ca II triplet lines which carry the radial velocity information have broad wings (0.2–0.3 nm), and the spectral sampling used (0.075 nm per pixel) properly samples them. Over-sampling does not lead to further increases in the radial velocity accuracy as a function of magnitude, also because the further spreading of the signal on the CCD results in an increase in the read-out noise per spectral line. The dependance of the performance on the spectral sampling has been veryfied by repeating the simulations with different spectral dispersions. As long as the Ca II lines are properly sampled the resulting performance as a function of magnitude is only very weakly dependant on the sampling, and values between $\simeq 0.075$ and $\simeq 0.15$ nm per pixel yield very similar performance.

Source confusion, i.e. overlap between spectra of the different objects will affect the spectra obtained in more crowded regions. 'Severe' or non-recoverable overlap will occur whenever two spectra overlap in the dispersion direction, i.e., when they are within 1 pixel perpendicular to the dispersion direction and within 1 length of the spectrum in the dispersion direction. In this case the two spectra (partly) overlap on the same pixel so that part of the information is lost, even if it may be recoverable *a posteriori* by only using the non-overlapped part of the spectra. 'Mild' overlap will occur when the spectra fall at a distance between 1 and 2 pixels perpendicular to the dispersion direction. In this case the wings of each spectrum contaminate the neighboring spectrum (and compromise sky subtraction); this can be corrected for at the expense of degraded S/N. Simulations of the sky however show that while spectral overlap is an issue in the more crowded regions (i.e. bulge or cores of globular clusters), but will not be a serious concern in most of the sky.

A further issue is the determination of the velocity zero-point. Given the stability of the GAIA instruments, it is possible to reconstruct the zero point from 'first principles' using the precise astrometry available from the main mission and the knowledge of the instrument's geometry. This would be calibrated in absolute terms through a network of reference stars, including for example the ones for which planetary companions are being searched, all of which have radial velocities known to excellent accuracy.

Part IV
Data Analysis and Mission Management

9 Data Analysis

9.1 Overview of the Data Analysis Task

The total amount of (compressed) science data generated in the course of the five-year mission is about 2×10^{13} bytes (20 TB). Most of this consists of CCD raw or binned pixel values with associated identification tags. The data analysis aims to 'explain' these values in terms of astronomical objects and their characteristics (Figure 9.1). Since the observations depend on conditions quantified by the attitude data and various instrument calibrations, these too must be determined as part of the data analysis. In principle the analysis is done by adjusting the object, attitude and instrument models until a satisfactory agreement is found between predicted and observed data (dashed lines in Figure 9.1). Although in practice this general scheme will not be strictly followed, it is useful to keep in mind as an idealized prototype for the GAIA data analysis.

Successful implementation of the GAIA data analysis task will require expert knowledge from several different fields of astronomy, mathematics and computer science to be merged in a single, highly efficient system. The following are discussed in this section:

- (a) accurate physical modeling of the observations in terms of detectors, optics, satellite attitude and the astrometric and photometric characteristics of the objects (Section 9.2), including a fully general-relativistic treatment consistent at least to the 1 μas level (Section 9.3);
- (b) statistically efficient estimation methods to utilize the information optimally (Section 9.4);
- (c) accurate calibration of the instruments, both geometrically and photometrically, including the celestial orientation (attitude) of the instrument axes (Section 9.5);
- (d) efficient procedures for generating and maintaining software, and for the management, processing and dissemination of data (Section 9.6).

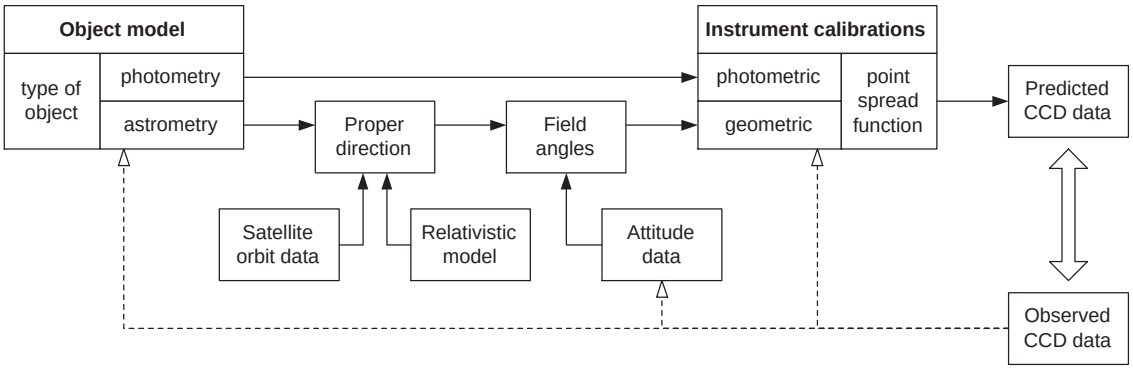


Figure 9.1: Model of CCD data interrelations for an astronomical object. In principle, the data analysis aims to provide the 'best' representation of the observed data in terms of the object model, satellite attitude and instrument calibration. Certain data and models can, from the viewpoint of the data analysis, be regarded as 'given'; in the figure these are represented by the satellite orbit (in the barycentric reference system) and the relativistic model used to compute celestial directions. Other model data are adjusted to fit the observations (dashed lines).

9.2 Modeling the Observations

The data reductions build upon a consideration of the various astronomical and instrumental reference systems, from a consideration of the type of data generated by the on-board detectors, from the attitude perturbations experienced by the satellite, and from the astronomical source model that the observations are finally fitted too. These various steps are described in this section.

9.2.1 Coordinate Systems

We briefly review the main coordinate systems relevant for the data analysis.

Barycentric reference system The motions of celestial bodies, including the GAIA satellite itself, are described in the barycentric reference system $(T, \mathbf{X})$, where T is the barycentric coordinate time and $\mathbf{X}$ are space-like coordinates with origin at the Solar System barycentre (Brumberg 1991). The choice of coordinates is in principle arbitrary, but in practice set to a first approximation by the coordinate system used for Solar System ephemerides. The orientation of the axes is set by the International Celestial Reference System, ICRS (Feissel & Mignard 1998) which is by definition non-rotating with respect to distant extragalactic objects.

Satellite reference system Observed directions and their relations with the satellite axes are more conveniently expressed in a local inertial reference system in the vicinity of the point of observation, the satellite reference system $(t, \mathbf{S})$. Here t is the proper time of the observer and the axes of the space-like coordinate system $\mathbf{S}$ are chosen by asymptotic matching to the barycentric reference system (i.e., for $c \to \infty$ the transformation from $\mathbf{X}$ to $\mathbf{S}$ reduces to a shift of origin corresponding to the parallax). $\mathbf{S}$ is also known as the proper frame of the observer (Murray 1983). The significance of this frame can be seen by considering the directions to two objects, $\mathbf{p}_1$ and $\mathbf{p}_2$. If these are unit vectors expressed by their Cartesian components relative to $\mathbf{S}$, then the *measurable* angle ψ_{12} between the objects is given by the Euclidean expression:

$$\cos\psi_{12} = \mathbf{p}_1 \cdot \mathbf{p}_2 \tag{85}$$

The relation between the barycentric and satellite reference systems is discussed in detail by Brumberg (1991). See also Section 9.3 for an alternative approach.

Instrument reference system GAIA contains three instruments which we may distinguish by index $i = 0$ for the spectrometric instrument and $i = \pm 1$ for the two astrometric instruments. By projection onto the sky of the designated field centre, each instrument defines a viewing direction $\mathbf{f}_i$. The spatial axes of the instrument reference system are:

$$\mathbf{x} = \frac{\mathbf{f}_{-1} + \mathbf{f}_{+1}}{|\mathbf{f}_{-1} + \mathbf{f}_{+1}|}\,, \qquad \mathbf{z} = \frac{\mathbf{f}_{-1} \times \mathbf{f}_{+1}}{|\mathbf{f}_{-1} \times \mathbf{f}_{+1}|}\,, \qquad \mathbf{y} = \mathbf{z} \times \mathbf{x} \tag{86}$$

i.e., $\mathbf{x}$ bisects the basic angle between the two astrometric viewing directions, while $\mathbf{z}$ is normal to the plane containing the basic angle.

The spatial coordinates (x, y, z) are related to the satellite reference system $\mathbf{S}$ by a pure rotation, which defines the instrument attitude (Section 9.2.4). The time coordinate for the instrument system (at the centre of mass of the satellite) is the same as in the satellite reference system, i.e. the satellite proper time t. Strictly speaking, the instrument reference system is rotating, and thus non-inertial. At the present time it is not clear if this has a measurable effect on the observations.

Electronic reference system All the science data produced by GAIA consist of CCD pixel values read out in the TDI (time delay and integration) mode. A particular pixel value may be identified by means of the three integers k, m and n. Here, k is a sequential numbering of the TDI periods, and thus corresponds to the along-scan location of the sample within the data stream. m is the pixel column number, counted from the output node on the CCD chip, and thus represents the across-scan coordinate of the sample. Finally, n identifies the output node (or video channel) through which the data are taken.

The centroid of an electronic image may be represented by the channel number n and the pixel coordinates (κ, μ), which are continuous variables obtained by interpolation in the (k, m) mesh. The sequencing of the CCDs is governed by a central clock which defines the on-board time scale τ. The on-board datation (determination of the time of a given event) allows an on-board time coordinate τ_{kn} to be associated with each sample. Relations between the electronic reference system (τ, κ, μ)

and the instrument reference system (t, x, y, z) are part of the instrument calibrations (Section 9.5). The relation between τ and t is accurately known from the on-ground orbit determination and datation of telemetry data.

9.2.2 CCD Sample Model

The sample values obtained by reading the CCDs are usually the results of on-chip binning of several pixels according to the CCD sampling mode. For instance, in the astrometric field the sample contains 8 serially binned pixels and the elementary observation normally consist of 6 consecutive such samples centred on the object detected in the astrometric sky mapper.

For a given CCD output node (n) the samples actually extracted and transmitted to ground can be regarded as an excerpt from the infinite stream of sample values s_{kmn}, where k and m are pixel coordinates defined above (referring, for instance, to the first pixel in the sample). Together with the (known) sampling mode, (k, m) specifies the location of the sample in the data stream. In principle, the (decoded) sample values s_{kmn} must be modeled in terms of a large number of parameters describing the observed objects and sky background, the instrument characteristics (both geometric and photometric), and the attitude of the instrument (Figure 9.1).

The basic CCD sample model, for point-like sources, is a local background plus a linear superposition of point-spread functions:

$$s_{kmn} = b + \sum_j a_j P_n(k - \kappa_j, m - \mu_j) + \text{noise} \tag{87}$$

where b is the background, a_j the amplitude of the jth source, P_n the (continuous) point-spread function and (κ_j, μ_j) the centroid position of the source expressed in pixels and fractions thereof. In most cases there is only one source to be included in the sum. The noise term is modeled in terms of the known noise sources (Poisson, read-out, quantization and decoding noise), which essentially means that the total noise variance is known, or can be calibrated, as a function of signal level. Non-linearities from CTE effects and pixel saturation are not considered here, but will most likely require quasi-empirical correction of the raw data.

9.2.3 Instrument Model

The instrument model includes both a photometric and a geometric description. The photometric part describes how the point-spread function P_n varies with time, transverse pixel coordinate (μ), and the spectral characteristics of the object such as colour index. (In this case the point-spread function is not necessarily normalized to unit volume.) The geometric part specifies the relation between the electronic reference system (τ, κ, μ) and the instrument reference system (t, x, y, z).

It should be noted that the description of P_n, together with the actual centroiding algorithm, implicitly define what is meant by the centroid of the electronic image — a non-trivial issue if the image is asymmetric. The photometric and geometric models are therefore not independent of each other. This is especially clear for the chromaticity (Section 3.2.4), which can be seen either as a shift of the point-spread function centroid relative to a fixed geometric reference point (such as defined by the chief ray), or as a colour-dependent geometrical distortion (if the centroid is taken as origin of the point-spread function), or as a mixture of both cases. This illustrates the importance of always treating the photometric and geometric characteristics of the GAIA instruments in parallel.

For a given CCD channel (n) we may imagine a fiducial line across the CCD, set by the location of the image centroid at the time $\tau_{\kappa-K,n}$. Here κ is the along-scan pixel coordinate of the centroid and K is the nominal distance (in pixels) of the fiducial line from the output serial register ($K \simeq$ half the number of pixels in the along-scan direction). The corresponding time in the satellite reference system, $t_{\kappa-K,n}$, is also known by means of the calibrated relation $t(\tau)$.

Let (ξ_i, η_i) be field angles in instrument i, with origin at $\mathbf{f}_i$ and with η_i parallel to the instrument $\mathbf{z}$ axis. The mapping from the across-scan field angle to the across-scan pixel coordinate is:

$$\mu = M_{in}(\eta_i) \tag{88}$$

where M_{in} may be a low-order polynomial. Similarly, the location of the fiducial line is given by:

$$\xi_i = \Xi_{in}(\eta_i) + \Delta\xi_{imn} \tag{89}$$

where Ξ_{in} again may be a low-order polynomial, while $\Delta\xi_{imn}$ is the displacement of pixel column $m = \mathrm{int}[M_{in}(\eta_i)]$ relative to the mean position. The geometrical model of the instrument includes three more parameters. One is the basic angle Γ, given by:

$$\cos\Gamma = \mathbf{f}_{-1} \cdot \mathbf{f}_{+1} \tag{90}$$

The other two define the location of the Spectro field centre, $\mathbf{f}_0$, in the instrument frame, e.g. by its components along $\mathbf{y}$ and $\mathbf{z}$, which are nominally zero.

The instrument model is thus described by a considerable number of parameters in P_n, M_{in}, Ξ_{in}, $\Delta\xi_{imn}$, Γ and $\mathbf{f}_0$. Many of them will change significantly over the mission. However, a very important assumption for the data analysis is that the instrument changes are slow enough that the parameters can, if necessary, be calibrated as functions of time.

9.2.4 Attitude Model

The attitude specifies, as a function of time, the orientation of the instrument axes $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ in the satellite reference system $(t, \mathbf{S})$. At any instant the orientation may be given by three angles, and since the satellite spin is assumed to be smooth, these angles could be modeled as continuous functions of time, e.g. by means of polynomials, trigonometric functions or splines. In practice the satellite will be controlled to follow a completely smooth nominal scanning law to within a few arcmin. It is then probably convenient to model the true attitude differentially with respect to the nominal law, using three Eulerian angles that are continuous functions of time. Using spline functions, for instance, the choice of knot sequence must take into account the power spectrum of the pointing error. Based on Hipparcos experience, it is estimated that about one knot per minute of time will give an optimum compromise between modeling errors and estimation noise.

9.2.5 Astrometric Model

The motions of Solar System objects are directly described in the barycentric reference system $(T, \mathbf{X})$. In particular, the satellite ephemeris $\mathbf{X}_{\mathrm{G}}(T)$ can be assumed known from the orbit determination (Section 5.3.3). The track of a photon from an arbitrary source within the Solar System to GAIA can be computed in the barycentric system using perturbation methods described for instance in Brumberg (1991), or by the non-perturbative approach outlined in Section 9.3. The incident coordinate direction in the barycentric system is transformed to the satellite reference system, which gives the (measurable) proper direction to the object.

The motions of sources outside the Solar System, i.e. stars and extragalactic objects, are modeled with respect to the barycentric date

$$T_{\mathrm{B}} = T_{\mathrm{G}} + \frac{1}{c}\,\mathbf{R} \cdot \mathbf{X}_{\mathrm{G}}(T_{\mathrm{G}}) - \frac{\Delta\gamma}{c} \tag{91}$$

Here T_{G} is the barycentric coordinate time of light reception at GAIA, $\mathbf{R} = \mathbf{X}/X$ is the barycentric coordinate direction to the object, and $\Delta\gamma$ is the relativistic delay (Shapiro effect) between the source and GAIA. Thus, for a single, unperturbed star the barycentric motion is modeled by

$$\mathbf{X}(T_{\mathrm{B}}) = \mathbf{X}(T_0) + (T_{\mathrm{B}} - T_0)\mathbf{V} \tag{92}$$

where T_0 is the chosen reference epoch for the astrometric data analysis and $\mathbf{V}$ is the barycentric coordinate velocity of the star with respect to the barycentric date. ($\mathbf{V}$ differs from the barycentric space velocity of the star by the Doppler factor $\mathrm{d}T_{\mathrm{S}}/\mathrm{d}T_{\mathrm{B}}$, where T_{S} is the time of light emission at the source. The variation of this factor is here neglected.) $\mathbf{X}(T_0)$ and $\mathbf{V}$ are parameterized by the standard five astrometric parameters (α_0, δ_0, π, $\mu_{\alpha*}$, μ_δ) plus the radial velocity (V_{R}).

9.3 Relativistic Formulation

Besides its fundamental role in unraveling some of the most intriguing problems of modern astrophysics, GAIA will also contribute to some outstanding topics in fundamental physics (see Section 1.9). The satellite may even be the first in-orbit experiment capable of astrometric detection of gravitational waves. Further studies must be devoted to both a rigorous formulation of the effects and its actual observability within the context of GAIA: clearly, global astrometry with GAIA must be formulated in a fully general relativistic framework. Correction terms of order $M/R \sim v^2/c^2$ were already included in the Hipparcos reductions; the expected increase of observational accuracy with GAIA will significantly increase the degree of higher approximations that must be included. Recent attempts at modeling astrometric variables in a relativistic framework have been performed by Soffel (1989); Brumberg (1991); and Klioner & Kopeikin (1992). Soffel (1989) concentrated on a PPN formulation, while Brumberg (1991) used a first post-Newtonian formulation of general relativity. The work of Klioner & Kopeikin (1992) was undertaken to model the measurements of the POINTS satellite and followed Brumberg's approch for integration of relativistic light-ray equations, extending the analysis to post-post-Newtonian effects of the spherical Sun at the 1 μas level as well as including corrections due to oblateness and angular momentum of the solar system bodies (Sun, Jupiter, Saturn, etc.). Very recently, Kopeikin & Schäfer (1999) have succeeded in developing the Lorentz covariant theory of light propagation in the gravitational fields of arbitrary moving bodies. The theory has been used for precise calculation of the light deflection caused by bodies in motion.

9.3.1 The Non-Perturbative Approach for a Spherical Sun

The objective of relativistic astrometry is to determine the position of a star and its proper motion, relative to a selected reference frame, within a general relativistic framework. The development requires fixing the following data: (i) the sources of the gravitational potential, namely the background geometry; (ii) a reference frame which allows for a coordinate representation of the space-time which is generated by the source (or sources) and is compatible with the prescription of the ICRS reference frame; (iii) the rest-frame of the observer (the orbiting telescope) in which the measurements are performed.

The method is that of solving the inverse problem in ray tracing, namely to find the position of a star from the measurements, made in the observer's rest-frame, of the angle between any two incoming photon trajectories. The non-perturbative general relativistic approach to global astrometry requires that the background geometry is described by a space-time metric which is an exact solution of Einstein's equations rather than a series expansion about a Newtonian term, truncated according to the chosen level of accuracy.

In the case of a satellite orbiting the Sun, the latter being assumed spherical and non-rotating, this simply implies the use of the Schwarzschild metric as background geometry. Once the orbit of the satellite is fixed and the satellite's rest frame is identified in terms of a suitable tetrad, it is relatively simple to solve the inverse problem of deducing the Schwarzschild coordinate positions and proper motions of distant stars from direct measurements made in the rest frame of the orbiting satellite and referred to a Sun-centred frame. This has been done in de Felice et al. (1998) limited to the static sphere, and work is in progress to extend this to a dynamical case for proper motions and parallaxes. In both cases, standard ray-tracing methods would enable one to deduce analytically the observation equations and integrate them with no approximations apart from those imposed by the accuracy of the numerical code. The results show that, regardless of the intrinsic complications of the relativistic model, repeated measurements of arcs between stars, each good to ~ 500 μas (as expected for $V \simeq 15$ mag stars) can be modeled to yield ~ 10 μas errors on the final relativistic astrometric parameters after only one year of continuous observations.

9.3.2 The More Realistic Sun–Earth Case

Global general relativistic astrometry may be hard to tackle in a non-perturbative approach if one considers the more realistic case of a background geometry given by the Sun–Earth system with possible inclusion of the other planets and of the Sun's rotation. In this case there are no exact solutions available which provide a metric for an exact ray-tracing analysis. However if

the satellite is orbiting the Sun at the Lagrangian point L2 of the Sun–Earth system, then the Newtonian gravitational potential felt by the satellite is that of an effective non-rotating, axially symmetric mass $\mathcal{M}$ centred at the Sun–Earth barycentre and endowed with a constant quadrupole moment which is expressed in terms of the Earth–Sun mass ratio. Using standard Newtonian physics we have at L2:

$$V_{\rm L2} = -\frac{G(\mathcal{M}_\odot + \mathcal{M}_\oplus)}{r}\left\{1 + \frac{\mu}{(1+\mu)^2}\left(\frac{d}{r}\right)^2 + O\left[\left(\frac{d}{r}\right)^3\right]\right\} \tag{93}$$

where $\mathcal{M}_\odot$ and $\mathcal{M}_\oplus$ are the masses of the Sun and the Earth respectively; $\mu \equiv \mathcal{M}_\oplus/\mathcal{M}_\odot$, and d and r are the Sun–Earth and the Sun–L2 distances respectively. Comparison with a standard multiple expansion shows that the Sun–Earth system, as viewed from L2, is equivalent to a static source with mass $\mathcal{M} = \mathcal{M}_\odot + \mathcal{M}_\oplus$ and with a quadrupole fixed by a coefficient $J_2 = \mu/(1+\mu)^2$.

In general, the potential felt at any point p at a distance r from the Sun–Earth barycentre and latitude $\pi/2 - \psi$, is given by:

$$V_p(r,\psi) = -\frac{G(\mathcal{M}_\odot + \mathcal{M}_\oplus)}{r}\left\{1 + \frac{\mu}{(1+\mu)^2}P_2(\cos\psi)\left(\frac{d}{r}\right)^2 + O\left[\left(\frac{d}{r}\right)^3\right]\right\} \tag{94}$$

where $P_2(\cos\psi) = (1/2)(3\cos^2\psi - 1)$ and $(d/r) \ll 1$. Hence, at L2, the satellite motion and the measurement conditions (namely fixing the local rest frame) can be described in terms of the Erez-Rosen metric (Erez & Rosen 1959). The latter is an exact solution of Einstein's equations which describes, to lowest order in (d/r), the gravitational field of a non-rotating spherically symmetric source with a quadrupole component. The parameter which carries the information about the quadrupole term is denoted by q and comparison with the GAIA situation yields:

$$q = \left(\frac{d}{2m_\odot}\right)^2 \frac{\mu}{(1+\mu)^4} \tag{95}$$

where $m_\odot = G\mathcal{M}_\odot/c^2$, G being the gravitational constant and c the velocity of light. The Erez-Rosen metric is available in analytic form to order $(m/r)^4$, r being the distance of the satellite (or of the light ray) from the source centre (Quevedo & Parkes 1989).

The contribution of Jupiter as well as that of any other type of source which perturbs the gravitational field of the Sun, can be described within the same metric in terms of an effective mass and an effective quadrupole moment which, however, are no longer constant but can be expressed in terms of the satellite's proper time. Despite the analytical complication of this approach, the mathematical formulation is the same as that used in the simplified spherical Sun case. Evidently a numerical code is needed which can account for the more complicated structure of the background metric. This work is now in progress within the GAIA framework.

9.3.3 Reference Frames

The non-perturbative general relativistic algorithms that are now being considered for the astrometric reduction of the GAIA data, require the use of two reference frames: a 'main frame' which allows the coordinate representation of space-time and with respect to which one fixes the coordinates of a star, and a 'local rest-frame' for the given observer, with respect to which the basic astrometric measurements are fixed. The latter are angles between light-ray directions coming from any pair of nearby stars. These angles are given by:

$$\cos\psi_{12} = \frac{h_{\alpha\beta}k_1^\alpha k_2^\beta}{(h_{\iota\pi}k_1^\iota k_1^\pi)^{1/2}(h_{\rho\sigma}k_2^\rho k_2^\sigma)^{1/2}} \tag{96}$$

where k_1^α and k_2^β are the components of the tangents to the null geodesics of the photons emitted by the two stars respectively, and $h_{\alpha\beta} = g_{\alpha\beta} + u_\alpha u_\beta$ is a tensor operator which projects in the

rest frame of the given observer whose four-velocity is given by the vector components u^α. In Equation 96 all components are relative to the main frame while the angle which is so defined is a measurement made by the local observer. It is then crucial to be able to determine its trajectory, namely the four-velocity u, in the given space-time together with the coordinate representation of the light rays, namely the four-vectors k's. The local frame can be further specialized in terms of spatial directions: the most obvious ones are the direction L2–Earth–Sun, the direction of motion, and the direction orthogonal to these two. Since the satellite oscillates about the point L2, then a local frame must be identified more rigorously in terms of mean directions.

9.4 Estimation Methods

It is clear that the processing of vast amounts of data will require numerical methods that can be highly automated and thus are robust, yet efficient in terms of information extraction. While this obviously applies to all levels of the data processing, it is particularly critical for the image centroiding (or PSF fitting) of the elementary astrometric and photometric observation in the astrometric instruments, and the corresponding analysis of spectral data in the spectrometric instrument. In this section we consider two complementary approaches to this kind of problem: one is the application of classical estimation techniques to the image centroiding and the photometric data analysis; the other is the use of neural networks for spectral parameterization.

9.4.1 Image Centroiding

Accurate and efficient estimation of the centroid coordinate κ_j in Equation 87, based on the noisy CCD samples s_{kmn}, is crucial for the astrometric performance. Although many effects such as the charge transfer inefficiency (Section 3.7.6) will contribute to the error budget, the Poisson statistics of the photoelectrically and thermally generated charges, together with the readout noise (including quantization and coding errors), will remain the dominant source of random errors, with the possible exception of very bright and very faint stars.

The left panel of Figure 9.2 illustrates the situation for the observation of an unreddened K3III star of magnitude $G = 18$ ($V \simeq 18$). The curve is the expected variation of the number of electrons as a function of pixel coordinate for the integration time of one CCD crossing (0.86 s). The total number of electrons produced by the star is 2340, of which 2110 fall within the sample width of 8 pixels in the across-scan direction. The centre of the point-spread function is at the pixel coordinate $\kappa = +0.3$. The background level is about 6 electrons per sample (for a sky brightness of 21 mag arcsec^{-2}), hardly visible at the scale of the figure. The circles show a single simulation of the sample data, including Poisson noise and readout noise with an rms value of $r = 5.6$ e$^-$. In reality only 6 samples approximately centred on the peak would be read out from the CCD. The centroiding, as well as the magnitude estimation, must be based on these six values.

The statistical efficiency of any centroiding algorithm can be assessed by comparison with the theoretical limit known as the Cramér–Rao bound. In the case of an isolated star, and considering only the along-scan coordinate, Equation 87 can be simplified to:

$$s_k = b + aP(k - \kappa) + \text{noise} \tag{97}$$

If s_k is expressed in electrons, with $s_k = 0$ corresponding to zero charge, then the noise variance due to Poisson statistics is simply $b + aP(k - \kappa)$. To this must be added the readout noise variance, r^2. This gives the following well-known expression for the Cramér–Rao bound:

$$\sigma_\kappa \geq \left[\int_{-\infty}^{+\infty} \frac{[aP'(u)]^2}{r^2 + b + aP(t)} \, \mathrm{d}t \right]^{-1/2} \tag{98}$$

where σ_κ is the standard error of the estimated centroid, $\hat{\kappa}$, expressed in pixels. Equation 98 assumes that the point-spread function is well sampled and entirely covered by the available data.

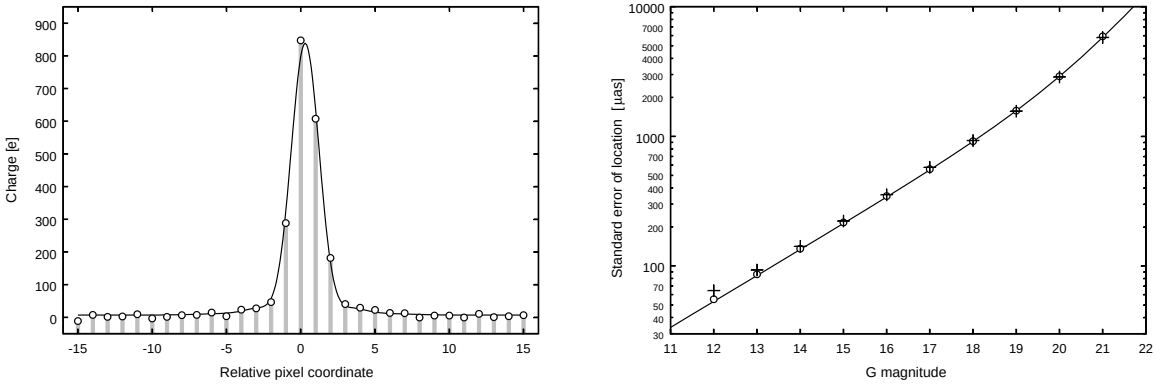


Figure 9.2: Illustration of image centroiding in the astrometric field, based on noisy CCD sample values. Left panel: the solid curve is the expected variation of signal (background plus scaled point-spread function) as a function of the continuous pixel coordinate, for a star of 18 mag. The circles show a simulation of the CCD sample values including Poisson and readout noise. Right panel: the curve is the theoretical (Cramér–Rao) lower bound for the centroiding error (Equation 98); the circles and crosses show the rms errors actually obtained in Monte Carlo experiments. In these experiments the centroiding was made with a maximum-likelihood estimator, using the five central sample values. For the circles, the background and star intensity (b and a) were assumed to be perfectly known; for the crosses they were crudely estimated from the sample values (Equation 99). It is seen that the theoretical bound is practically reached in nearly all cases.

Moreover, it assumes that b, a and P are perfectly known. Finally, it should be recalled that there is no guarantee that any existing estimator can actually reach this theoretical bound, even when these conditions are met.

In reality the sampling of the point-spread function is rather sparse, the available data are limited to six samples around the maximum, and at least b and a are also subject to estimation errors. (On the other hand, the function P can be calibrated quite accurately because it is common to a very large number of observations.) It is important to study how a real estimator performs under these non-ideal circumstances. Results of a large number of Monte Carlo experiments are summarised in the right panel of Figure 9.2. The assumed point-spread function, background, and readout noise were the same as in the left panel of the figure, but with the true centroid randomly placed within ± 0.5 pixel. The solid curve gives the theoretical bound according to Equation 98, while the circles and crosses show the results of the Monte Carlo experiments.

The centroiding algorithm used was the maximum-likelihood estimator described in Section 7.4.2. Five consecutive samples were used, viz. the highest sample and the two adjacent ones on either side. In this algorithm the background and stellar intensity, b and a in Equation 97, are supplied as input to the centroiding. A set of experiments were first made in which the *true* values of b and a were used for the centroiding. The results, shown as circles in the right diagram, are seen to follow the theoretical bound to within a few per cent. Next, the experiments were repeated with the following very crude estimates for the photometric parameters:

$$\begin{aligned} \hat{b} &= \max\left(1,\, s_{\min} - 0.08\, s_{\max}\right) \\ \hat{a} &= (s_{\max} - \hat{b})/0.95\, P(0) \end{aligned} \tag{99}$$

Here, $s_{\min}$ and $s_{\max}$ are the extreme values among the five samples. Resulting statistics are shown as crosses in the diagram. Equation 99 is certainly not the recommended way to estimate a and b; the point is rather to show that even with such crude photometric estimates the centroiding is almost perfect, at least for the fainter stars. In reality a and b can be estimated quite accurately by means of the sky mapper data (for the background) and from the combination of the 17 successive transits in the astrometric field.

The conclusion from these experiments is that a rather simple maximum-likelihood algorithm performs extremely well under these idealized conditions, and that six samples is sufficient to determine the centroid accurately. Much work remains to extend the analysis to more complex cases, including in particular overlapping stellar images.

9.4.2 Photometric Data Analysis

The photometric data analysis is carried out on the ground. A preliminary analysis can be made of the individual patches of a star, each containing the samples from a single field crossing. The corresponding sky mapper observation and preliminary geometric and photometric calibrations of the CCDs are required, but not any other observations or data; e.g. the accurate absolute satellite attitude is not needed at this stage. A local PSF fitting is carried out on each of the patches, using the position of the star given by the sky mapper observation (see Høg et al. 1999c, Section 7.3.3). Such preliminary photometric data can be used for verification purposes and for scientific studies, e.g. of supernovae and other sudden events.

The final photometric results should not be obtained simply as a mission average of the preliminary epoch photometry for reasons mentioned below (see also Section 7.5). All observations of a star must be treated together and the accurate absolute satellite attitude must be available.

The first step is to detect and estimate the positions of all stars (point sources) brighter than $G = 23.0$ out to a certain distance from the main star. This is possible by means of sky mapper observations (without filter) to a distance of 1 arcsec for the broad-band photometer and 3 arcsec for the medium-band photometer (cf. Section 6.6). The subsequent photometric analysis shall provide the photometric parameters of the stars and background, in the sky mapper and for each colour band. This procedure may be called 'global PSF fitting', as opposed to the 'local PSF fitting' described above, where the centroid position is not constrained by an astrometric model.

Let there be $n_{\rm star}$ stars in the area and $n_{\rm obs}$ observations per filter. The stars may at first be assumed to be of constant magnitude so that one amplitude value per star shall be determined. The background may also at first be assumed to be constant over the area, but it will be different for each patch, e.g. due to zodiacal light. The simplest observation model thus contains $n_{\rm star}$ amplitudes and $n_{\rm obs}$ background values which shall be determined from the samples in each band. The currently assumed scheme of sampling is shown in Figure 3.7. It has been developed in a number of technical reports from the Copenhagen University Observatory (Høg 1997d, Høg 1997e, Høg 1997b, Høg & Sørensen 1997, Høg 1997c, Høg 1997f, Høg 1997a, Høg 1998b, Høg 1998c, Høg 1998d, Høg et al. 1998a, Fabricius & Høg 1998c, Fabricius & Høg 1998a, Høg 1998a, Fabricius & Høg 1998b, Høg et al. 1998b, Høg 1998g, Høg & Knude 1998, Høg 1998e, Høg et al. 1999c, Høg et al. 1999b, Høg & Knude 1999, Høg et al. 1999e, Høg 1999).

We shall henceforth consider the medium-band photometry in some detail according to Høg et al. (1999e). This case can easily be converted to that of broad-band photometry. There are 8 samples per band along scan in a patch of medium-band photometry; the four outer ones are wider than the four inner and they sample the background and disturbing stars. We assume for the final photometry that all geometric calibration values, satellite attitudes, CCD sensitivities, and PSF values are available. The derivation of star amplitudes and background values in each band can then be formulated as a linear least-squares problem.

A simple consideration shows that the problem is usually well determined. Consider for example a high star density of 200 000 stars deg^{-2} brighter than $G = 20$ mag. Such a density or higher is found in less than 1 per cent of the sky (cf. Section 6.5.2). There are then on average 1.8 additional (disturbing) stars brighter than $G = 23$ mag in the 28 arcsec^2 area within 3 arcsec of the main star. Thus, $n_{\rm star} = 1 + 1.8 \simeq 3$ amplitude values and $n_{\rm obs} = 100$ background values are to be estimated. The number of equations for a single photometric band (with 8 samples) is $n_{\rm eq} = 8n_{\rm obs} = 800$. The system is thus overdetermined with 697 degrees of freedom. That the method is feasible in principle has so far been demonstrated on simulations of double stars (Høg & Makarov 1998, Høg & Fabricius 1998).

After a least-squares solution the residuals may be analysed in order to find variable stars and variation of the background over the field, e.g. due to a galaxy. This should first be carried out in the sky mapper data because the count rate without filter is much higher than in any of the bands of the medium-band photometer.

On-board astrometric and photometric analysis The possibility of estimating, on-board the satellite, the astrometric and the two photometric parameters (a and b in Equation 97) from each CCD observation has been considered. The advantage would be that only these three numbers and their standard errors would have to be transmitted to the ground which could lower the telemetry rate. This appears to be unacceptable for a number of reasons (Høg 1999). It would not be possible to return to the original observations in order to take information into account which becomes available only when all observations of a star are treated together. This is information on duplicity or multiplicity of the star, or the variability, or the disturbance from neighbouring stars or nebulosities. Even for isolated single stars the method would be unacceptable for the photometry of faint stars because the magnitude estimate could not approach the Cramér–Rao limit. Another point is that a magnitude estimation is more accurate if it is based on a knowledge of the position of the star, therefore the (nearly) final astrometry and geometric calibrations of CCDs and satellite attitude must be available.

Variables The number of new variable stars with high-precision multi-colour observations will be numbered in the tens or hundreds of millions. The general problem of determining (pulsation or orbital) periods for tens of millions of short-period objects will require special effort focussed on fully automated analysis procedures (for the Hipparcos analysis, 'manual' methods were applied to check and fine-tune the periods given by standard Fourier methods, but even a 10-second inspection at 10 trial periods for each of 50 million light-curves would correspond to about 1000 years of effort).

9.4.3 Automated Spectral Parameterization

The high-resolution (radial velocity) spectrometer will produce spectra for about a hundred million stars, and multi-epoch, multi-band photometry will be obtained for about one billion stars. The analysis of such large numbers of spectra and photometric measurements needs to be performed in a fully automated fashion, with no manual intervention. Automatic determination of (at least) the surface temperature $T_{\rm eff}$, the metallicity [M/H], and the relative α element abundance [α/Fe] is necessary. As discussed in Section 2.3.1 the determination of $\log g$ is, given the availability of parallaxes for most stars, of lesser importance. The size of the data set rules out any 'traditional' approach to the determination of astrophysical parameters from individual spectra and magnitudes. The possibilities for a fully automated system for the derivation of astrophysical parameters from the large number of spectra and magnitudes collected by GAIA, using all the available information for each star (rather than the limited amount of information contained e.g. in a colour-colour diagram) have been studied, with the results showing the feasibility of an approach based on the use of neural networks.

In the classification system foreseen for GAIA spectra will be extracted, reduced and converted to calibrated spectral energy distributions, and then sent to an 'Initial Classifier' software package. Photometric measurements will be treated similarly. The initial task of this package will be to sort objects into stellar and non-stellar (galaxies, quasars, asteroids etc.), possibly using neural networks trained on pre-identified data sets. The details of this subsystem (which would however be based on the same principles as the rest of the package) will not be described here. Stellar data sets are then passed to the 'Automated Stellar Parameterization' sub-package.

The proposed system will parameterize observed magnitudes and spectra in terms of physical parameters rather than MK classifications. The primary disadvantage of the MK system for the GAIA project is that MK lacks a well-defined metallicity parameter. This, combined with the fact that the MK system is based on nearby bright stars, makes it unable to discriminate between, and hence correctly classify members of, the various stellar populations which will be observed by GAIA. Additional problems with the MK system include its arbitrarily defined spectral type parameter and its non-numeric definitions of 'peculiar' stars. Moreover, it is the physical parameters of stars which are really of interest. Therefore the proposed system aims to derive physical parameters directly from a stellar spectrum. If desired, MK classifications can later be assigned by those interested in doing so.

The Automated Stellar Parameterization system uses a feed-forward neural network operating on the entire set of spectral and photometric measurements (rather than some transformation of it). Details of such a network can be found in Bailer-Jones et al. (1997). For the purpose of the present study the neural network has been trained using synthetic stellar spectra with known physical

parameters, generated using stellar structure and atmosphere models. Once trained, the network is used to deduce the physical parameters of the real observed stars. The main parameters of interest are $T_{\rm eff}$, $\log g$, [M/H], $\{a_i\}$ (the relative abundance ratios of the elements whose signatures are present in the spectrum under analysis, mostly [α/Fe]) and possibly ξ, the micro-turbulence velocity. Additional parameters could be used to quantify peculiarities.

The derived values for the stellar parameters are thus naturally linked to the models used to train the network. Given the extreme rapidity of neural networks, if and when models are improved, re-classification of the entire data set can be done extremely quickly: an archive of 10^8 spectra or photometric measurements could be reclassified in about a day with the present-day computing power of a scientific workstation.

Neural networks A neural network is a device which performs a non-linear parametrized mapping between an input vector, x, and an output vector, y. The input vector is passed through a layer of 'hidden nodes', h, which form a non-linear weighted combination of the inputs:

$$h_j = \tanh\left(\sum_i w_{i,j} x_i\right) \tag{100}$$

Each output from the network is then given by:

$$y_k = \sum_j w_{j,k} h_j \tag{101}$$

The weights, w, are the free parameters of the network. In order for the network to give the required input–output mapping, the network must be trained on a set of representative data pairs, that is inputs (e.g. stellar spectra) for which the true outputs (e.g. stellar parameters) are known. This training can be achieved by a numerical least-squares minimisation: A set of spectra are fed through the network and the actual outputs compared with the desired outputs (targets). The gradient of this error with respect to each weight in the network is then used to iteratively perturb the weights towards a minimum of the error function. Thus the training is a minimisation problem in an N-dimensional space, where N is the number of weights. Once the network has been trained, the weights are fixed and the model used to obtain parameters for a new set of inputs. For the present study a network with one hidden layer consisting of five hidden nodes has been used. The complexity of the network is determined by the number of hidden layers and the number of nodes in each hidden layer. Weight decay regularisation has been used to avoid overtraining.

Any classification system is a compromise between retaining the full information in the spectrum and the need for a compact summary of it. The optimal 'summary' is given by the physical parameters, and the current work has focused on parameterization in terms of the fundamental parameters $T_{\rm eff}$, $\log g$ and [M/H]. Additional parameters will ultimately be required to describe the various 'peculiarities' in the spectra (such as emission lines, abundance peculiarities, etc.). As GAIA will obtain spectra and photometry of hundreds of millions of stars, it defines an entirely new stellar parameterization system. Considerable effort in collaboration with a wide community of astronomers will be required to ensure that the best, most useful and long-lasting system can be developed.

As the GAIA spectrograph will be slit-less, some spectra will inevitably overlap. In many cases, complete (un-overlapped) spectra can be reconstructed from the individual spectra taken at each epoch at different positions angles (a consequence result of the satellite's scanning law). However, rather than throwing away those remaining contaminated spectra, the surviving parts should still be used to obtain physical parameters. This will require some modification of the parameterization approach to make it robust to missing data.

For the present study a large grid of synthetic spectra have been generated using Kurucz models (Kurucz 1992) and the synthetic spectral generation program of Gray (Gray & Corbally 1994). The parameter grid consists of 36 $T_{\rm eff}$ values between 4000 K and 30 000 K (step sizes of between 250 K and 5000 K), 7 values of $\log g$ between 2.0 and 5.0 (in 0.5 steps) and 15 values of [M/H] between -3.0 and $+1.0$ (step sizes of between 0.1 and 0.5). The microturbulent velocity was fixed at 2.0 km s^{-1}. Spectra were calculated between 300 and 1000 nm with a line list of over 900 000 atomic and molecular lines. The spectra were then convolved with Gaussians to degrade the resolution, r, to 2.5, 5.0, 10, 20 and 40 nm (two pixels) and combined with the transmission curves of the F filter system (as well as with some variations on it) to produce a set of multi-filter fluxes ($r \simeq 800$). Poisson noise was added to all data sets to simulate signal-to-noise ratios of 5, 10, 20, 50 and 1000 (S/N per resolution element). The result is 30 sets of 3537 absolute spectral energy distributions at each combination of resolution (or filter system) and S/N. The absolute flux information is retained, as GAIA parallaxes will enable calculation of absolute fluxes.

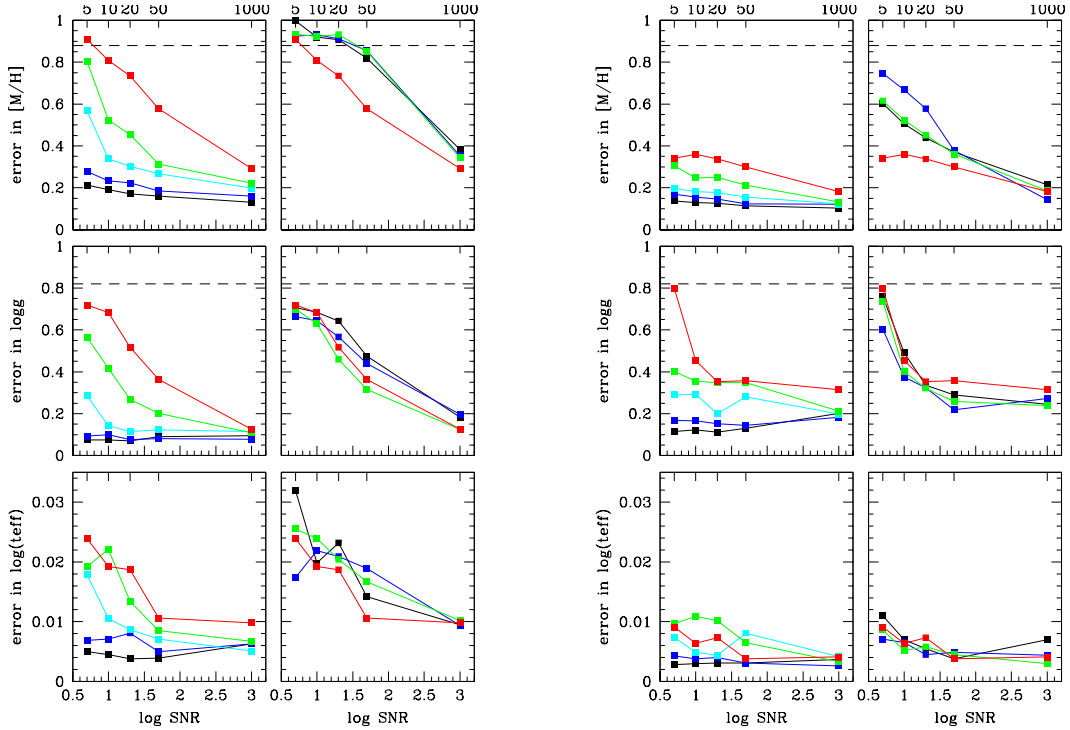


Figure 9.3: Neural network parameterization of stellar spectra and photometric measurements at various resolutions and signal-to-noise ratios. The left two columns show the performance of spectroscopy (leftmost) and filters (right) on hot stars ($T > 10\,000$ K), the right two columns the same performance on cool stars ($T < 10\,000$ K). The lines for spectroscopy are arranged in order of spectral resolution (with higher resolutions yielding more precise parameters) while the lines for the photometry are different variations on the baseline GAIA photometry system. The horizontal dashed lines on the $\log g$ *and [M/H] plots are the performances of random (untrained) networks. The corresponding line for the* $T_{\rm eff}$ *plot is at* $\epsilon = 0.16$.

Kurucz models do not produce highly accurate spectra, particularly at low $T_{\rm eff}$, as they exclude water opacity (and no H_2O lines are present in the line lists); this sets the lower limits of 4000 K. Furthermore, the models lack chromospheres and will not reproduce features such as emission in the cores of the Ca II IR triplet and H and K lines). For the present investigation, however, it is not necessary to have highly accurate individual spectra, but rather that the spectra reflect differences of the appropriate scale and complexity between spectra with different parameters.

Network training is a relatively CPU-intensive operation, while classification of individual spectra is very rapid. For the purpose of the present study each network was trained for 10 000 iterations. Longer training did not reduce the error further. The longest training time (for the largest input vector, the r=2.5 nm spectra) was about one day on a Sun SPARC Enterprise.

Figure 9.3 summarizes the results (separately for hot stars ($T > 10\,000$ K) and for cool stars ($T < 10\,000$ K) showing the feasibility of classifying large photometric and spectroscopic data sets in a fully automated way, and show that the GAIA filter system can be used in such a way, yielding results (at least for the temperature determination) comparable to 5–10 nm resolution spectroscopy. In these simulations the performance of the system is actually limited by the spacing of models in the grid. The difference in performance between hot stars (for which little metallicity information is available in the spectrum) and cool stars (which, due to the large number of metallic lines, allow good estimation of metallicity even on relatively low S/N spectra) is a consequence of the physics of the stars.

The work presented here is only a first assessment of the parameterization possibilities with the GAIA data. The development of a complete classification system for GAIA will need to address

the following questions in detail: (i) the approach for the Initial Classifier: Principal Components Analysis has been investigated by Bailer-Jones et al. (1998); (ii) the approach must be extended to include (at least) determination of the interstellar extinction and abundance ratio variations; (iii) GAIA will observe many (very) low-mass stars (including brown dwarfs), and model spectra of these objects need to be included in the training and test data sets; (iv) most stars are in multiple systems, and in some cases will be physically or spectroscopically resolvable. The networks need to be trained to take this into account, possibly by having multiple sets of output parameters; (v) fully simultaneous use of photometric measurements and spectra should be implemented. Another issue which needs to be addressed is how the neural networks are to be trained for the GAIA mission, i.e. whether on entirely synthetic spectra and then applying to real data, or on a small fraction of real GAIA data which are calibrated using synthetic spectra.

Status of automatic stellar spectral classification Most existing work on automating classification of stars has attempted to automate the MK system (Morgan et al. 1943) starting from low-resolution spectroscopy, through various approaches. Recently, much of this work has focused on the use of artificial neural networks, details of which can be found in, for example, Hertz et al. (1991); Bishop (1995); Lahav et al. (1996) and Bailer-Jones (1996). Reviews of the applications of neural networks in astronomy more generally are given by Storrie-Lombardi & Lahav (1994) and Miller (1993). Kurtz (1982) classified intermediate resolution (1.4 nm) spectra using cross-correlation with standard spectra and achieved a mean classification error of 2.2 spectral subtypes for stars in the range B0 to M2, however with gave poor luminosity classification capability. LaSala (1994) used the related technique of minimum distance classification to classify a set of 350 B-star spectra, and achieved a mean error of 1.14 spectral subtypes.

One of the first applications of neural networks to stellar spectral classification was made by von Hippel et al. (1994), who achieved an rms classification error of 1.7 spectral subtypes for spectra in the range B3 to M4. Gulati et al. (1994) reported a classification error of 2 subtypes over the range O–M, using however a very complex neural network with over 18 000 free parameters (network weights), with no justification for this complexity (the determination of these weights was likely to be poorly constrained by the small amount of training data used). Bailer-Jones et al. (1997) extended the work of von Hippel et al. (1994) to a much larger database (> 5000 spectra) and to luminosity classification, achieving accurate spectral type classifications ($\sigma_{\rm rms} = 1.09$ subtypes) across the spectral range B2–M7, with correct luminosity class determination at high confidence for over 95 per cent of both dwarfs and giants. They also investigated the application of Principal Components Analysis to the optimal compression of spectra, finding that it would compress the spectra by a factor of over 30 while retaining essentially all of the useful information in the data set. This work also demonstrated how such a compression optimally removes noise and can be used to identify unusual spectra.

All the above results concentrated (for historical reasons) on the 'classical' (i.e. blue-violet) spectral region. Recently, there have been attempts to classify stellar spectra beyond this range. Weaver & Torres-Dodgen (1995) used neural networks to classify red-infrared spectra (580–890 nm) of A stars at 1.5 nm, and achieved spectral type and luminosity class classification precisions of 0.4 subtypes and 0.15 luminosity classes respectively. More recently they have achieved good results in the infrared for a wide range of spectral types (O–M) and luminosity classes (I–V) (Weaver & Torres-Dodgen 1997). Vieira & Pons (1995) used a neural network trained on a set of 64 IUE ultraviolet spectra (15–320 nm) in the range O3 to G5, and reported a classification error of 1.1 spectral subtypes, although this possibly suffered from an underdetermined network with 110 000 weights. Whitney (1983) examined the use of Principal Components Analysis for spectral classification of a set of 53 A and F stars. The data consisted of 47 photoelectric measurements of spectra over the wavelength range 350–400 nm. He applied Principal Components Analysis to his data set and then performed a regression on the three most significant components, achieving an average classification error of 1.6 spectral subtypes.

The use of the MK spectral system (with its missing 'third parameter', i.e. the metallicity) is clearly not adequate for the classification and the determination of the stellar parameters of a large number of stars belonging to different populations (as required for GAIA). Bailer-Jones et al. (1998) trained neural networks on synthetic spectra to determine physical parameters for spectra directly, i.e. without using the MK system; the proposed approach to the analysis of the GAIA spectroscopic data is based on this work.

9.5 Calibrations and Attitude Determination

The photometric and geometric calibrations of the GAIA instruments, and the attitude determination (which is also a form of calibration), constitute an integral and very central part of the data analysis task. It is important to realize that these calibrations must be derived from normal observations, i.e. from exactly the same data as are used to derive the astrometric and photometric characteristics of the objects.

The data analysis task would be impossible without certain regularizing assumptions. Concerning the astronomical objects it is necessary to assume that a substantial fraction of them follow a very simple model, viz. (apparently) single stars with little or no photometric variability, whose motions can be described by the standard five astrometric parameters (α, δ, π, $\mu_{\alpha*}$, μ_δ). For the satellite attitude and instrument characteristics it must be assumed that sudden changes are rare, so that time-averaging and smoothing are effective in reducing observational noise. Without these assumptions the problem would simply have too many degrees of freedom.

While such regularity conditions must be valid in a broad sense, it is clear that they cannot be guaranteed to hold in any particular situation or for a specific object. The data analysis must be able to filter out the cases where the conditions do not apply, and divert them to a separate branch of the analysis. The efficiency of the filtering process depends critically on the quality of the instrument calibrations and attitude determination, which initially is quite low. Thus an iterative process is needed in which the selection of objects and observations is successively improved along with the calibrations and attitude determination.

We shall consider in some detail this process as it applies to the simultaneous determination of astrometric, attitude and (geometric) instrument parameters for a subset of well-behaved stars. It represents the core of the GAIA astrometric data analysis in the sense that once completed, the resulting instrument calibration and attitude data can be used to analyze the data for arbitrary objects, or sky patches, effectively decoupled from each other.

9.5.1 Core Processing of Astrometric Data

A possible procedure for the core processing of astrometric data is outlined below. Its purpose is to establish a globally consistent reference frame for the positions and proper motions, and a corresponding set of data for the satellite attitude and geometric calibration of the instrument. The photometric and spectrometric (radial-velocity) data need to be processed according to very similar principles in order to establish a globally consistent system for the photometry and radial velocities. Details of the spectrophotometric core processing are not given here, but it is envisaged that it could (and should) be done in parallel with the astrometric core processing, since basically the same data sets will be used and the data access patterns are practically the same.

Details of the modeling of the CCD data, instrument, attitude and astronomical objects have been given in Section 9.2. In the present outline the data analysis is considered in more general terms. Thus it it convenient to divide the unknown parameters in three separate groups: (i) the star parameters, which describe the individual object; (ii) the local parameters, which describe the observatory in a limited time interval $\mathcal{T}$, of the order of several hours to a few days; and (iii) global parameters, which affect the observations of all objects throughout the mission. The local parameters include attitude data and the variable part of the instrument parameters; the global parameters include fixed instrument parameters and data which describe, for instance, the reference system (such as the PPN parameter γ, if that is part of the data analysis). The following notations will be used:

- stars are numbered $p = 1, 2, \ldots P$ and their astrometric parameters are represented by the data arrays $\mathbf{a}_p$;
- time intervals for the local parameters are numbered $q = 1, 2, \ldots Q$ and the local parameters applicable in the interval $\mathcal{T}_q$ are represented by the array $\mathbf{l}_q$;
- elementary observations, such as the centroid pixel coordinates (κ, μ), are numbered $r = 1, 2, \ldots R$. A single observation is represented by the data array $\mathbf{o}_r$.

The whole collection of star parameters is represented by the array $\mathbf{a} = (\mathbf{a}_1, \ldots \mathbf{a}_P)$, and the collection of local parameters and elementary observations are similarly represented by the arrays $\mathbf{l}$ and $\mathbf{o}$. The array of global parameters is denoted $\mathbf{g}$. There is also an array of auxiliary data $\mathbf{t}$ which is not subject to improvement by the observations. This includes, for instance, timing information and the barycentric ephemeris of the satellite.

Given all the model parameters and auxiliary data listed above, and using the models outlined in Section 9.2, it is possible to compute the expected result of each and every observation. Specifically, for observation r, this computation can be summarised by means of the function $\mathbf{c}_r(\mathbf{a}_{p(r)}, \mathbf{l}_{q(r)}, \mathbf{g}|\mathbf{t})$. In this expression, $p(r)$ and $q(r)$ are, respectively, the star index and the time interval index associated with the observation index r. The collection of all the computed data is represented by the array $\mathbf{c}$, which is thus a function of all the model parameters.

The core processing task can now be formulated as the minimisation problem:

$$\min_{\mathbf{a},\mathbf{l},\mathbf{g}} \|\mathbf{o} - \mathbf{c}(\mathbf{a}, \mathbf{l}, \mathbf{g}|\mathbf{t})\|_M \tag{102}$$

where the norm is calculated in a metric M defined by the statistics of the data, which in the general, non-linear case need not be Gaussian. Let us assume, however, that the parameterization has been made in such a way that the problem can be solved by linearisation, e.g. using the weighted least-squares method (with appropriate precautions for robustness).

The structure of the linearised minimisation problem, with its division into stellar, local and global parameters, is very similar to the structures encountered in the great-circle reductions and the sphere solution of Hipparcos (see Chapters 9 and 11 of Vol. 3 of ESA SP–1200, and especially Figure 11.2 in that volume), and indeed for the Hipparcos astrometric solution as a whole. It could therefore, in principle, be handled by the direct method described in the quoted chapters. That is, after sorting the data either chronologically (by the q index) or systematically (by the p index), the corresponding unknowns ($\mathbf{l}_q$ or $\mathbf{a}_p$) may be eliminated, resulting in a rather dense system of normal equations for the remaining parameters. For Hipparcos the dimensions of $\mathbf{a}$ and $\mathbf{l}$ were, respectively, about $n \simeq 370\,000$ (the astrometric parameters for the primary reference stars) and $\sim 2\,000\,000$ (the number of spline coefficients and free instrument parameters in the great-circle reductions). If the local parameters were successively eliminated in Hipparcos, the direct solution of the remaining system of n astrometric and global parameters would have required some $n^3/3 \sim 10^{16}$ floating-point operations, and the administration of some $n^2/2 \sim 6 \times 10^{10}$ double-precision reals ($\simeq 500$ Gigabyte): a non-trivial task even for supercomputers and parallel processing. It was such considerations that led to the idea of the 'three-step' decomposition proposed for Hipparcos in 1976. However, the practicality of that method was gained at the expense of approximations which should be avoided with GAIA. For GAIA, the minimum dimension of $\mathbf{a}$ or $\mathbf{l}$ is many times greater than it was for Hipparcos, making a direct solution unfeasible even with the much more powerful computers available today.

9.5.2 Global Iterative Solution

Apart from the 'three-step' method, the only alternative to the direct solution proposed to date is an iterative solution. The basic idea dates back at least to 1977, when Pierre Lacroute advocated the use of intermittent guiding of the satellite and the use of 'dynamical smoothing' in the quiet intervals. In his introductory talk at the 'Colloquium on European Satellite Astrometry', held in Padova in June 1978, the idea was formulated as followings: *... it is possible to represent the attitude motion during the periods of free motion by using the coordinates of the stars and all their transit times. With the help of mechanical laws the computed attitudes should be very accurate and by using them along with the transit times we could obtain better evaluations of the coordinates.*

To iterate this procedure is an obvious possibility. The resulting method, which may be referred to as the 'global iterative solution', was subsequently proposed and studied by a group at the Istituto di Topografia, Fotogrammetria e Geofisica, Milano (Sansò et al. 1989). In the present framework it can be described as follows. Let $\mathbf{o}_p$ be the array of all the observations of a particular star p, and similarly let $\mathbf{o}_q$ be the array of all the observations made in the time interval $\mathcal{T}_q$. $\{\mathbf{o}_p\}_{p=1}^{P}$ and $\{\mathbf{o}_q\}_{q=1}^{Q}$ are thus different partitionings of the total observation array $\mathbf{o}$. In practice they could be obtained by sorting the observations according to star index (p) or time index (q), respectively, although this may not be necessary depending on the administration of the equations.

If the astrometric parameters $\mathbf{a}$ and the global parameters $\mathbf{g}$ are regarded as known, or rather as 'given', it is a simple matter to solve, for each time interval $\mathcal{T}_q$, the minimisation problem:

$$\min_{\mathbf{l}_q} \|\mathbf{o}_r - \mathbf{c}_r(\mathbf{a}_{p(r)}, \mathbf{l}_{q(r)}, \mathbf{g}|\mathbf{t})\|_M \tag{103}$$

involving only the observations $\mathbf{o}_q$ and resulting in a linearised system of equations with $\dim(\mathbf{l}_q)$ unknowns, i.e., typically a few thousand. The solution to this problem could formally be written

as the function $\hat{\mathbf{l}}_q(\mathbf{o}_q|\mathbf{t},\mathbf{a},\mathbf{g})$. This problem is somewhat analogous to the attitude reconstruction problem discussed in Chapter 7 of Vol. 3 of ESA SP–1200.

Conversely, by regarding $\mathbf{l}$ and $\mathbf{g}$ as given, the astrometric parameters of each star are obtained by solving the problem:

$$\min_{\mathbf{a}_p} \|\mathbf{o}_r - \mathbf{c}_r(\mathbf{a}_{p(r)}, \mathbf{l}_{q(r)}, \mathbf{g}|\mathbf{t})\|_M \tag{104}$$

involving only the observations $\mathbf{o}_p$ and resulting in a linearised system of equations with $\dim(\mathbf{a}_p)$ unknowns (typically 5). The solution to this problem, analogous to the astrometric parameter determination discussed in Chapter 11 of Vol. 3 of ESA SP-1200, may be written as the function $\hat{\mathbf{a}}_p(\mathbf{o}_p|\mathbf{t},\mathbf{l},\mathbf{g})$.

Finally, if both the local and astrometric parameters are regarded as given, the global parameters may be obtained as the solution to the problem:

$$\min_{\mathbf{g}} \|\mathbf{o}_r - \mathbf{c}_r(\mathbf{a}_{p(r)}, \mathbf{l}_{q(r)}, \mathbf{g}|\mathbf{t})\|_M \tag{105}$$

and denoted $\hat{\mathbf{g}}(\mathbf{o}|\mathbf{t},\mathbf{a},\mathbf{l})$. This problem involves all the observations, but still results in a relatively small system of equations with $\dim(\mathbf{g})$ unknowns.

The global iterative solution is a straightforward sequential application of the above (partial) solutions. The optimal sequence of the three estimators $\hat{\mathbf{a}}_p$, $\hat{\mathbf{l}}_q$, $\hat{\mathbf{g}}$ is not obvious, but the following order seems intuitively natural:

$$\left.\begin{array}{lcll} \mathbf{a}^{(0)} & = & \text{initial catalogue} & \\ \mathbf{g}^{(0)} & = & \mathbf{0} & \\ \mathbf{l}_q^{(m)} & = & \hat{\mathbf{l}}_q(\mathbf{o}_q|\mathbf{t},\mathbf{a}^{(m-1)},\mathbf{g}^{(m-1)}), & q=1,\ldots Q \\ \mathbf{a}_p^{(m)} & = & \hat{\mathbf{a}}_p(\mathbf{o}_p|\mathbf{t},\mathbf{l}^{(m)},\mathbf{g}^{(m-1)}), & p=1,\ldots P \\ \mathbf{g}^{(m)} & = & \hat{\mathbf{g}}(\mathbf{o}|\mathbf{t},\mathbf{a}^{(m)},\mathbf{l}^{(m)}) & \end{array}\right\} \quad m=1,2,\ldots \tag{106}$$

If the iterations converge, the end result is evidently equivalent to a direct solution of the global minimisation problem, Equation 102.

Concerning the convergence properties, it can be noted that the linearised form of the procedure, written in the form of normal equations, is equivalent to the Gauss-Seidel iteration method for the solution of the linear system of equations. It is well known that this method converges for any symmetric and positive definite matrix. Due to the (theoretical) rank deficiency of the problem, this condition is in principle not satisfied. However, it can be argued that the particular degeneracy due to the undefined reference frame is of no practical consequence for the iterative solution, since each of the partial minimisation problems (Equations 103–105) does not suffer from this degeneracy. The tentative conclusion is therefore that the method does converge, namely to the particular solution closest, in some sense, to the initial estimate $\mathbf{a}^{(0)}$, $\mathbf{g}^{(0)}$.

Intuitively, the global iterative solution is expected to converge as a consequence of the geometrical structure of the problem, namely that in a given interval $\mathcal{T}_q$ many different stars contribute to the determination of $\mathbf{l}_q$, while, conversely, many different intervals contribute to the determination of a given star. Thus, an initial error in the coordinate of one star gives only a much smaller error in the attitude parameters of the affected intervals, and these errors in turn are diffused, in the next iteration, to a large number of stars, and, in rather few iterations, to the whole set of stars. It can be noted that this diffusion is strengthened by the superposition of the two fields of view, by the incommensurability of the basic angle to 360°, and by the diversity of scan directions across any point on the sky; i.e., by the very properties that make the reference frame internally 'stiff' (as with Hipparcos). It is a very likely hypothesis that the convergence properties are closely linked with the stiffness of the resulting reference frame: a well-designed space astrometry project should ensure good convergence of the global iterations.

A simplified version of the global iterated solution for Hipparcos, using 2000 stars, was in fact implemented by Sansò et al. (1989) and was found to converge in only two iterations. The block iteration method used for the Hipparcos sphere solution by the FAST Consortium (Chapter 11 of

Vol. 3 of ESA SP–1200) follows the same general numerical principle (although the detailed equations are different), demonstrating its feasibility for a similar problem with $\sim 370\,000$ unknowns. Descriptions of an attempted full global iterative solution based on the Hipparcos data are given in Chapter 23 of Vol. 3 of ESA SP–1200.

The global iterative solution thus appears to be a both practically feasible and intuitively natural method for solving the general space astrometry problem. One possible disadvantage of the method is that it seems to be difficult to estimate reliably the uncertainties of the astrometric parameters. The curvature matrix associated with the restricted problem in Equation 104 gives only a lower bound to the covariance matrix of $\mathbf{a}$, by neglecting the uncertainties in $\mathbf{l}$ and $\mathbf{g}$. This aspect of the global iterative solution requires additional study.

9.6 Practical Implementation

9.6.1 Computational Complexity

The computational complexity of the GAIA data analysis arises not just from the amount of data to be processed, but even more from the intricate relationships between the different pieces of information gathered by the various instruments throughout the mission, and from the complexities of the astronomical objects themselves at the required resolution and sensitivity. The following three points illustrate the complexities: (i) no input catalogue is used — each of the three main instruments (Astro-1/2 and Spectro) has its own real-time detection of sources, with all object identification and data collation done *a posteriori*; (ii) no sufficiently accurate catalogue of reference positions will be available to define the local reference frame for the observations — on the contrary, a global reference frame must be built from the observations themselves; (iii) no dedicated calibration exposures will be available from which the photometric response or geometric distortion of the detectors will be known to sufficient accuracy — practically all instrument calibrations must be derived directly from the science data. This implies that the data analysis cannot be designed as a simple processing pipeline (data input | calibrations applied | models fitted | output), at least not until calibrations have been provided through the core processing.

The total amount of data to be handled is of the order of 100 TB. Problems of data storage, data access (where different applications have distinctly different, but predictable, access patterns), data integrity, and the implementation of flexible and transparent user interfaces are probably best handled by means of an object-oriented data base system (Section 9.6.2).

It is very difficult to assess the magnitude of the data analysis problem in terms of processing requirements. Certain basic algorithms that have to be applied to large data sets can be translated into a minimum required number of floating-point operations, but the associated overhead remains very uncertain. For instance, an elementary process that will certainly be needed is the estimation of the location and amplitude of a stellar image from about eight successive CCD samples. A fairly straight-forward maximum-likelihood algorithm for this purpose has been used in Monte Carlo experiments designed to assess the precision of the estimates. From these experiments it appears that some 3000 floating-point operations are required for each estimation. For 10^9 objects, some 3×10^{12} such estimations will be needed, requiring $\sim 10^{16}$ floating-point operations. Since this is only a small part of the analysis, the entire effort can be estimated to be at least of order 10^{18} floating-point operations.

A different approach to assess the processing requirements is to compare with other projects of similar relative complexity and apply a sensible scaling factor. The data analysis for Hipparcos (10^5 stars) is estimated to have required $\sim 10^{15}$ floating-point operations. The relative complexity of that task for GAIA is perhaps not too different from that of Hipparcos, but the number of objects is 10^4 times larger. This leads to an estimated 10^{19} floating-point operations for the GAIA data analysis task. While these estimates are probably naive in the context of the practical management of the data, they do indicate that very serious attention must be given to the implementation of the data analysis, and that this effort must start very early in the project.

9.6.2 Implementation Using an Object Oriented Approach

The processing of the GAIA raw data into fully calibrated and consistent sets of astrophysical data will be an extremely challenging task. Observations of each object are distributed throughout the mission, so that calibrations and analysis must be feasible both in the time-domain (for CCD and geometric calibrations, photometric response evolution, etc.) and in the object domain. Such demands create very specific requirements which are not necessarily present in other large astronomical data base projects. A highly automated, yet sophisticated data processing system will be required to take care of the bulk reductions. At the same time a great deal of flexibility and interaction is needed to cope with special objects or astrophysical investigations, many of which cannot be foreseen at the software design stage. At the same time, the delicate calibration of instruments and celestial pointings, necessary to interpret the data in terms of absolute astrometric and photometric quantities, must be protected from unintentional modification. The system must provide access to data at different levels for software testing, algorithm testing, and data validation as well as actually processing the data. It must allow effective participation by a large number of active scientists working on the data analysis, but through strictly controlled interfaces.

Object Oriented (OO) methodologies for data modeling, storage and processing are ideal for meeting the challenges faced by GAIA. The recent emergence of the Unified Modeling Language (UML, Eriksson & Penker 1998), as a fusion of the best parts of many OO design approaches, has finally provided a harmonized approach to software design across the computing industry. Furthermore tools such as Rational Rose provide code generation from UML diagrams making it possible to construct the outline of the entire system on the screen and then produce the outline code.

Most new software is now designed in an OO manner because of the inherent scalability and abstraction of the OO design methodologies which allow for extendable frameworks to be built for software which will change in the future. An example of an astronomical framework available in C++ is the Hierarchical Triangular Mesh (HTM) from the Sloan Digital Sky Survey Project for partitioning data according to position on the sky (Szalay & Brunner 1998). Logically software designed in an OO manner should be implemented using an OO language. Here the choice is wide: although the most used language at present is C++, more development is being done in the Java language. Java is proving to be portable, quick to code, and far less error prone than other OO languages. The only current drawback to Java is speed but, at this point in the GAIA project, ability to test designs and algorithms is considered to be more crucial.

Data storage is correspondingly most easily done in an OO database management system. The transition from 'objects' to storage is seamless for OO databases and languages, whereas in a traditional system data types always had to be converted between the database and the program accessing it. Additionally such systems have excellent scalability and flexibility in the storage of data over their relational counterparts. Many companies such as Motorola and Siemens, as well as scientific institutes such as CERN and STScI, are now using OO databases.

Prototype Experiments The feasibility of the OO approach has been demonstrated by a short prototyping exercise carried out during the present study phase. 'Rational Rose' was used to build an object model for the Hipparcos Intermediate Astrometry Data. Algorithms for three processes (reasonably well related to the types of processing foreseen for GAIA) were provided and incorporated into the OO model, underlining one important feature of OO design: the ability to have complex data structures and operations described in a single model. Java code was generated from the model and the algorithms implemented. The algorithms were to produce a mission chromaticity matrix, reference great circle harmonics, and astrometry updates. The interdependence of the three sets of results is similar to the type of dependencies which will occur in the GAIA data. A framework for the distribution of the processing was also produced allowing the processing to be performed in parallel on several machines. This was easily supported by the commercial database system 'Objectivity' which was used to store the initial data and the derived results of the reduction. The prototype was highly successful and reinforced confidence in the OO approach for treating the GAIA data. Further details are given in O'Mullane & Lindegren (1999).

As an example of the work underway to create simulations of the GAIA data, Figure 9.4 shows two 'snapshots' of the sky generated as HTM formats for different time intervals. An observing list is generated for objects contained within the time interval specified, and increasingly realistic sky models can be incorporated into the simulation chain.

Application to GAIA Work has commenced on the defining of the GAIA object model. The UML design shown in Figure 9.5 summarises some of the major packages of the model namely: GAIA, RawObject, Telemetry and util.

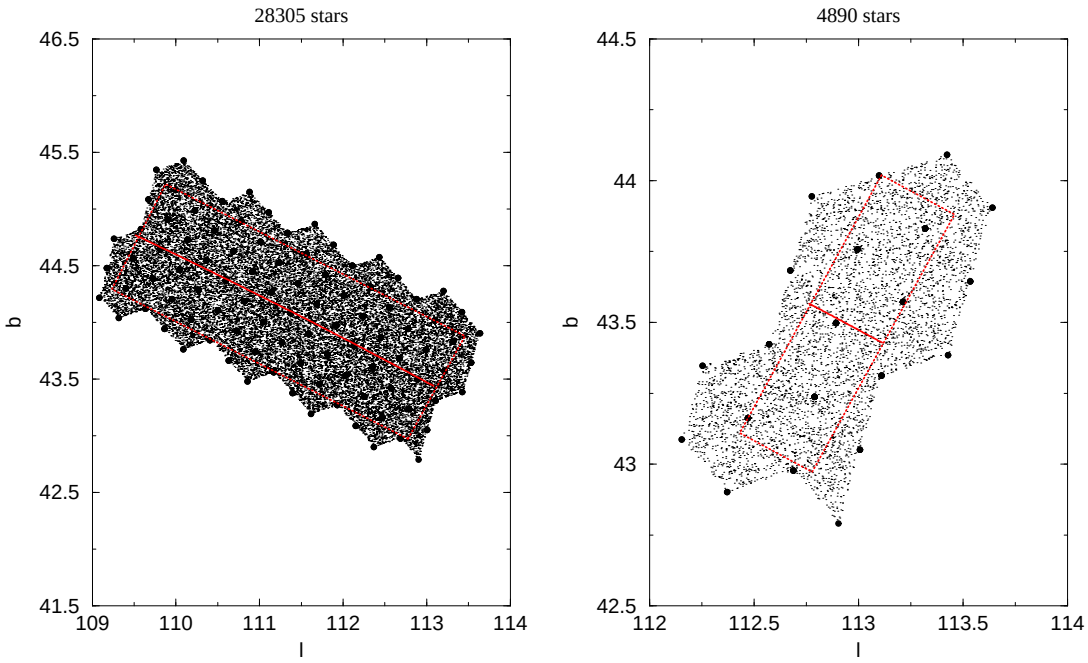


Figure 9.4: The GAIA sky simulator generates 'snapshots' of the sky over a specified time interval of the satellite motion (two such examples are shown here). An observing list is generated for objects contained within the dotted rectangle. The HTM nodes are shown as large circles, stars to $G = 25$ mag are shown as small dots, and the scanning law is shown as the solid line.

Each package contains a number of classes. The grouping of classes into packages is the highest level of abstraction in OO design and serves as a method for the designer (and user) to concentrate on one area of the system without being crowded by classes from other packages. Later this also becomes the partitioning for the software as it is developed. Each box in the diagram represents a Class, each of which has attributes and methods (not shown), and similarly each class becomes a class in the OO language used to implement the system. The lines between the classes represent different types of relationships. A line denotes a relationship between objects in the system: for example it can be seen that Photometry is related to many RawPhotometric objects. An arrow indicates that the relationship will usually be traversed in this direction. A diamond on the end of the line denotes 'aggregation', which generally may be taken to mean that the other object cannot exist without the one with the diamond on it. Thus if a GAIAobject is deleted then the Photometry, Spectroscopy and Astrometry Objects associated with it will also be deleted.

The triangle represents 'inheritance' which is one of the most powerful parts of OO design and OO languages. Inheritance and Polymorphism are directly supported in OO languages allowing essentially different objects to be treated in the same manner in code. In Figure 9.5 GAIAobject is the super class of Single and Complex. In turn these are further sub-classed to different individual classes of entity. This, for example, allows a complex binary object to be treated simply as a GAIAobject, while a WellBehaved Single star may be treated in the same manner. Another advantage of this particular hierarchy is that an object which is initially detected can be categorized as a Single but during processing it may be discovered it has multiple parts. In turn it will become complex, and in the process will be related to new single star components which is contains. The ability to treat all objects in a similar manner at the same time, and still allow them to inherently be quite different is very beneficial for GAIA.

As evident in Figure 9.5 it is possible, through relationships, to reach each observation of a particular GAIAobject back to the Telemetry from which it came, and it is important that this type of traceability be kept. An OO database will prove very useful for this as it supports the different types of relationships between objects directly.

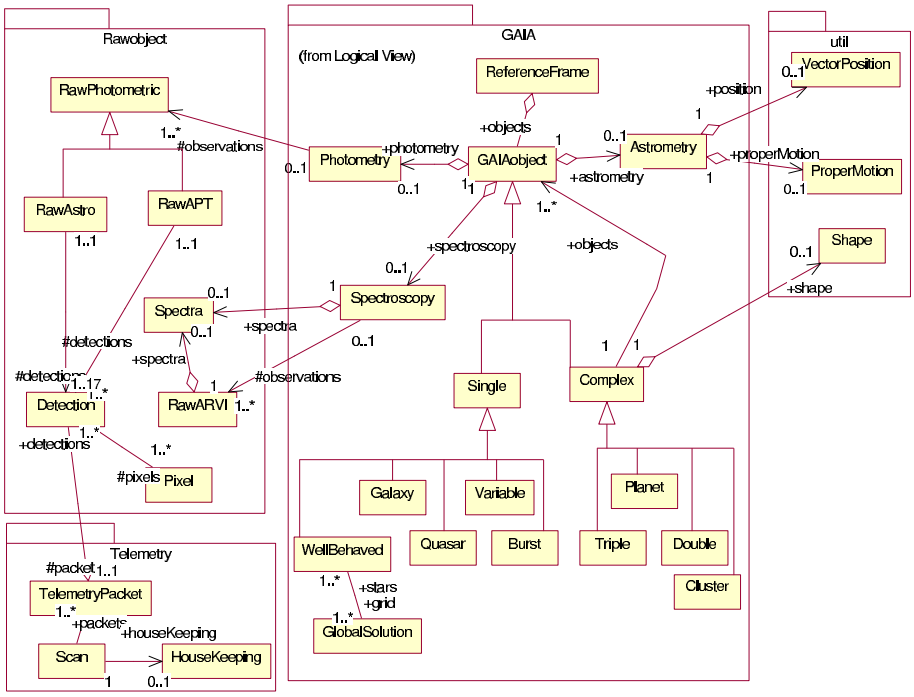


Figure 9.5: Overview of some of the GAIA data analysis packages, in the form of a diagram using the Unified Modeling Language (UML). Once the interrelations between the different classes (represented by the small boxes) have been designed, outline code can be produced directly from the UML diagram.

Distribution The reduction process for the GAIA data is inherently distributed. Some processes act on an individual star and all of its observations while others will operate on groups of observations in a time ordered manner. Hence multiple processors may be engaged to perform reductions in parallel. The obvious advantage of this approach is that it is scalable. Early on in the mission lifetime one or more multiprocessor machines may be purchased for data processing and as the mission continues and the volume of data increases additional machines and processors may be purchased. In this scenario we will also benefit from Moore's Law, which predicts the doubling of transistor density on chips (hence doubling of CPU power) every eighteen months. At this rate the physical limit for current microprocessor design would be reached around 2017. By 2010, the projected processing speed would be sufficient to process the data within acceptable time scales. Hence the budgeting of any hardware purchase for GAIA should be phased accordingly; the later it can be bought the more powerful and cheaper it will be. Here again portability and flexibility of any software developed will be essential to make use of newly available hardware. Budget wise, it will be better to suffer with inadequate machines in the early years of the GAIA project to reap the benefits of far more powerful machines which will be available more cheaply during the mission.

The distribution of the data in multiple databases can be done using positional information. A framework such as the Hierarchical Triangular Mesh (Szalay & Brunner 1998) can be used to achieve balanced partitioning of the data into individual databases. A level six HTM would probably provide sufficient distribution for GAIA, and this would yield $8 \times 4^6 = 32768$ databases, with each containing a little more than the data of one degree on the sky. Within each database partitioning can be in a time-ordered manner according to the circle the observations belong to (Figure 9.6). In this way access will be quick for positional information and also reasonable for time-ordered data. It will still mean opening most databases to get the information for a full circle but it will not require too much overhead in pointers, which was the case for the current (1999) prototype. Commercial databases such as Objectivity allow for this kind of flexibility: it is possible to have thousands of databases and to interrogate them individually or as a whole.

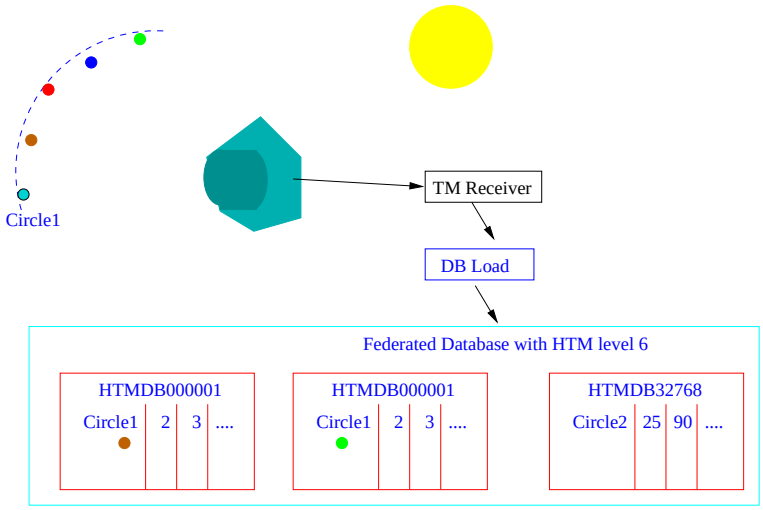


Figure 9.6: An example of the distribution of GAIA observations into a federated database. The positional information is used to partition the data into 32 768 data bases, corresponding to a partitioning of the sky according to a Hierarchical Triangular Mesh (HTM) with six levels. Each database represents a little more than one square degree of the sky, which is probably close to optimal for the field of view sizes of the GAIA instruments. In each database the observations are time-ordered according to the successive circles of scanning.

Having split the data in multiple databases these become ideal processing units for a distributed processing such as demonstrated in the prototype with the Hipparcos intermediate data. Again OO assists in this providing excellent frameworks such as Remote Method Invocation (RMI) and the Common Object Request Broker Architecture (CORBA) for coordinating distributed processes.

Possible Architecture An initial setup for GAIA might contain several machines and a tape robot. The robot would be for backups and for temporary off-line data storage. In Figure 9.7 the machines are drawn as separate boxes, although it is also possible to purchase multiple machines in a single box which may or may not be more cost effective. The approach taken by BeoWulf (http://cesdis.gsfc.nasa.gov/linux/beowulf/beowulf.html) and indeed by Sloan is to buy reasonably powerful Pentium PC's and tie them together with a high-speed network. An architecture such as this will be ideal for the GAIA processing and will result in the benefits of more powerful processors and cheaper disks as the mission progresses.

Ridge et al. (1997) describe a 1.25 Gigaflop 'Beowulf' machine costing $50 000; this was several years ago and a similar set of processors now would achieve 3 Gigaflops. Extrapolating this to a launch date of 2009 for GAIA and using Moore's law (a doubling of processing power every 18 months) a further increase of 2^6 should be achieved, suggesting that 192 Gigaflops should be available for the same price in 2009, i.e. roughly $260 per Gigaflop. Taking 10^{19} flops as the required computational load for GAIA, and imposing a requirement of 3 years on the overall computing time would mean that the machine would cost $87 000 at the start of the mission. After a further 3 years the price of a Gigaflop should fall to about $65, meaning that the same machine would cost under $25 000. Hence with an 18 months phased buying plan far more computing power may be purchased for half the price of buying a complete machine at the beginning of the mission.

9.6.3 Other Large Data Base Efforts

Many international efforts to develop efficient and effective techniques to handle large scientific databases are underway internationally, driven by particle physics, earth resources, and similar requirements, as well as by survey astronomy and cosmology, including the ESA Planck mission. We are aware that such developments are underway, and will ensure that whichever consortia of countries eventually develop the GAIA databases, and the analysis methodologies, they will include and build upon such expertise, rather than acting in isolation.

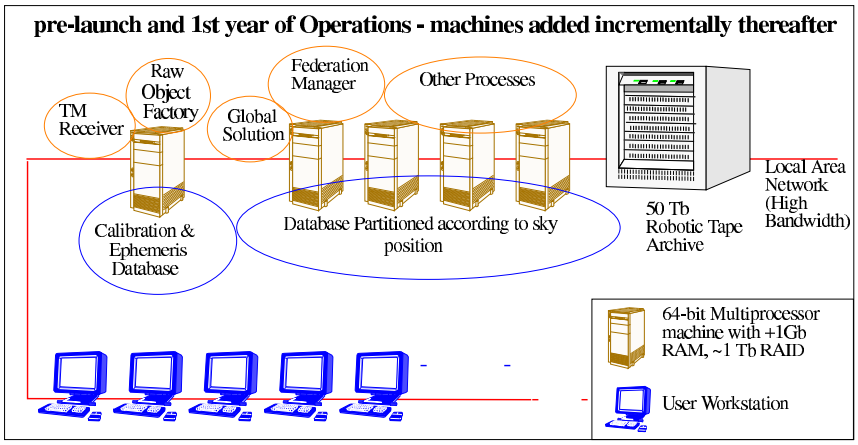


Figure 9.7: Possible initial GAIA data processing setup using multiple databases and processing units. The architecture allows the setup to be easily extended and upgraded during the mission.

Contact concerning data processing and storage details has been made with several institutes, and some of their experiences are listed here:

- Sloan Digital Sky Survey (SDSS): SDSS aim to make a digital map of the northern sky. An estimated 20 TB of raw data will be collected, and the final catalogue will be about 1 TB. All raw data will be kept on tape at Fermilab (where the initial reduction is carried out) and then calibrated quantities are stored in the science archive. The science archive is built using Objectivity and will use the HTM as a structuring mechanism.
- Planck: ESA's Planck mission will deal with maps of the entire sky at different frequencies. It is a global scanning mission with some similarity to GAIA, although the expected products are somewhat different. Typically the maps will be of the order of 2 GB, and the HEALPIX scheme has been developed as a pixelisation mechanism which may be used in conjunction with HTM to fragment maps into more manageable chunks. There is a tendency to adopt Java at least at the presentation level of the system and the initial 'coordinator' pipeline has been written in Java. Work is under way to define the OO model for the mission data which will be processed at four separate locations.
- VLT: The ESO Very Large Telescope (VLT) will deliver a science archive of astronomical observations well exceeding 80 TB already within its first six years of operations. ESO is undertaking the design and development of both on-line and off-line archive facilities. Details are given at http://archive/VLT-SciArch.
- STScI Guide Star Catalog 2 (GSC2): the GSC2 database is around 6 TB as of the end of 1999. This is using Objectivity on NT client/server network, Dell Server, 4 cpu, with 5 client workstations. The main purpose of the database is to have an online catalogue which provides efficient storage and retrieval methods such that global astronomical calibrations can be performed to (a) remove systematic errors; (b) multiple plate operations for computation of proper motions and colours for all-sky; (c) cross referencing between catalogue. Disk storage is currently on a 500 MB Dec Raid system but an HSM (Hierarchical Storage Management) system, OTG DiskXtender, and 4 TB DLT (Digital Linear Tape) jukebox is being incorporated. HTM is used as a partitioning mechanism for the database.
- CERN: there are a number of large projects under development or study at CERN. There is a tendency towards OO methodologies and databases. In general CERN deals with more data than GAIA will produce, with volumes for the Large Hadron Collider (LHC) being in the PetaByte Range. The following projects use (or plan to use) Objectivity in CERN: (a) NA45: an existing experiment that has used Objectivity since 1996 for a number of purposes and stored data in the tens of GB range. This may increase to tens of TB; (b) COMPASS: an experiment still in preparation that plans to use Objectivity to store $>$ 300 TB of data per year; (c) CHORUS: an existing experiment that has been using Objectivity as part of their production chain; (d) LHC experiments (ATLAS, CMS, ALICE, LHCb).
- SLAC: the BaBar detector at the Stanford Linear Accelerator Center studies the millions of B mesons produced by the PEP-II storage ring. Some 25 million positrons/electron collisions are generated per second, with volumes of more than 1 TB of data per day. Data is stored in an Objectivity database on an HPSS system, and the analysis involves data sets of hundreds of TB.
- Alta Vista: Alta Vista runs on a farm of 64-bit Alpha workstations running DEC Unix. There are currently $>$ 36 Alphas each with 1 GB of RAM. This is a very dynamic setup with more machines and disks being added as required. Alta Vista is a proprietary database built by DEC. It is designed for free text retrieval on documents and hence is not suited to the needs of data processing, yet the dynamic distributed architecture is similar to that outlined above.

- Human Genome: established at Johns Hopkins University in Baltimore, Maryland, USA in 1990, the Genome Database is the official central repository for genomic mapping data resulting from the Human Genome Initiative, a worldwide research effort to analyze the structure of human DNA and determine the location and sequence of the estimated 100 000 human genes. In support of this project, the database stores and curates data generated worldwide by those researchers engaged in the mapping effort of the Human Genome Project. The task of supervising the content of this encyclopedia and maintaining its correctness and currency is enormous. Members of the scientific community participate by submitting their data, adding annotations to existing data, and adding links to related objects in other databases. At this time the GDB database has somewhat greater than 25 GB of data. The object-oriented structure utilizes Sybase tables under an object broker and OPM. A detailed description is given at http://www.gdb.org.

9.6.4 Management Tools for Very Large Catalogues at CDS

The developments of new detectors, and of powerful data reduction techniques have given a new impulse to the production of sky surveys, or of deep wide field observations, resulting in the production of catalogues in the billion object range: USNO-A, DeNIS and 2MASS, Sloan DSS, GSC-2, are the first of these new generation surveys to become (at least partially) available. Access to the results of these sky surveys is becoming critical for many studies: new or rare categories of objects, statistical studies over complete samples of objects, spatial distribution, references for new kinds of observations, etc.

The Strasbourg Astronomical Data Center (CDS), in close coordination with the data producers has started to develop specific tools for: (1) managing and making accessible published data from sky surveys, deep fields, or very large catalogues at any wavelength; (2) helping cross-identification, and validation of data from new surveys, through cross-matching tools. Specifically, original dedicated management techniques have been developed for compressed coding, indexing, and fast coordinate access to very large catalogues such as GSC (of size around 20 million objects), USNO A (500 million objects), or DeNIS.

The VizieR catalogue service at CDS contains many reference catalogues and surveys, of smaller size, for example, currently: Hipparcos, Tycho, IRAS, VLA FIRST, etc., and allow respectively sampling (complex queries over any catalogue column) and full retrieval of the catalogues over the network. While VizieR generally uses a commercial relational database management system, catalogues of size larger than 20 million objects (currently: GSC, USNO A, DeNIS, and 2MASS sampler) are managed through customized tools for which the performance has been optimized as much as possible by data compression and efficient indexing.

The Aladin interactive sky atlas provides, in complement, access to reference images from the large Schmidt surveys (DSS I, and DSS 2 produced by Space Telescope Science Institute), and high resolution MAMA scans), together with a set of tools to help source diagnostic and identification, specifically with very large catalogues and surveys. It is currently possible to overlay USNO-A sources, or DeNIS infrared point sources, simultaneously with Hipparcos data or SIMBAD information, on top of the optical digitized image.

The SIMBAD database already contains a few millions of cross-identifications of sources for objects referenced in the literature. Impact of an increase of the size of the data base by one (or several) order(s) of magnitude is currently being assessed. Solutions for a very large database management system, fully object-oriented, potentially storing several kinds of large catalogues together, with fast multi-criterion access, are being explored: object-oriented commercial packages are being tested in the context of astronomical specifications. These studies are being done in close contact with groups having similar needs: Sloan Digitized Survey, Space Telescope Science Institute, European Southern Observatory, European Space Agency, etc.

In the context of the VLT scientific environment, data mining tools including an automatic cross-referencing engine for catalogues is being investigated in the framework of a collaboration between ESO and CDS. While queries have been, up to now, mainly optimized for access by coordinates, new methods are being investigated to allow optimized access to queries based on all or part of the catalogue columns (object type, colours, distances, etc.), using multivariate analysis technique for the optimized indexing. All these projects, which will be interfaced in the future, will allow

CDS to provide convenient access and data mining tools for large survey catalogues such as DeNIS, 2MASS, Sloan DSS, TERAPIX, OmegaCam, XMM, Planck Surveyor, or, later on, GAIA.

The experience with existing catalogues shows that it is already possible to support efficiently on-line queries to very large catalogues. Although the GAIA catalogue will be much more complex (especially with a large diversity of information collected), it can be reasonably anticipated that the resources available at the time of GAIA data production will make possible the support of on-line access to the core catalogue in the major astronomical data centres.

9.7 Scientific Work Packages

Figure 9.8 illustrates some of the main scientific work packages likely to be associated with the large-scale processing of the GAIA data. In this figure the emphasis has been placed on the scientific elements of the data processing, rather than on a rigorous breakdown of all functional aspects, in order to illustrate the range of specialist scientific activities associated with the data processing task. Interested groups would commit to undertaking specific analysis tools, preparing the theoretical, analytical and computational tools necessary for each specific aspect. Software would then be incorporated and verified within the main processing centre and, as the mission processing gets underway, the responsible teams would have access to early results for verification and optimization. A similar approach has been adopted by the Sloan Digital Sky Survey team.

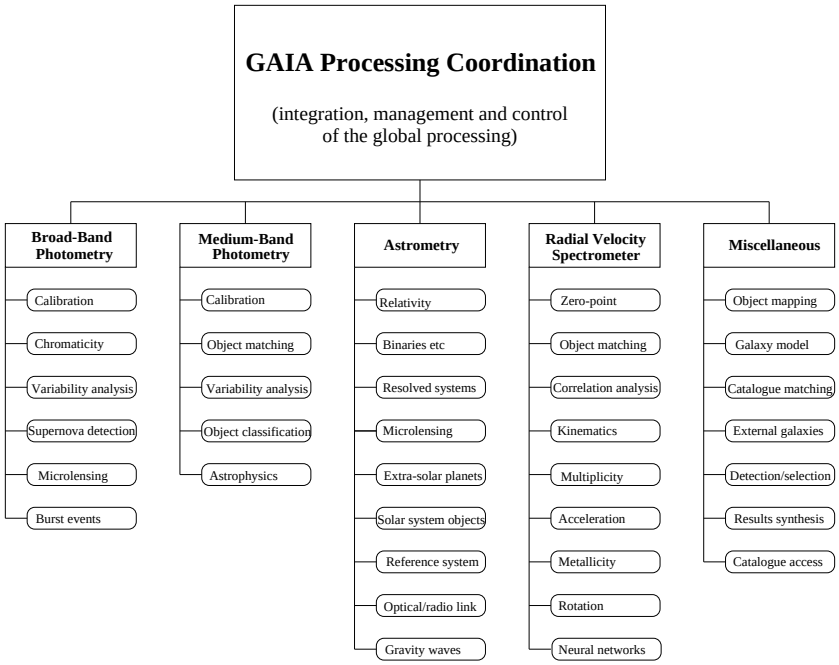


Figure 9.8: A partial breakdown of the scientific work packages associated with the GAIA data processing. Many elements can be considered in isolation, but one of the major tasks will be the consideration of data coming from all instruments—the astrometry, radial velocity, and photometry—to ensure global (astrophysical) consistency in the final catalogue.

10 Management, Science Operations, and Public Outreach

Project Management for the GAIA mission will of course be decided by the ESA Executive and Advisory Committees. The following paragraphs outline a possible management approach for the mission implementation. Although the outline below recommends continued involvement of the Concept and Technology Study prime contractor, selection of teams to implement all development should be by open, competitive tender. It is also recommended that implementation of the Core Technology programme should be the responsibility of the ESA Science Directorate, working closely with the GAIA Science Advisory Group, to ensure that technology developments remain closely relevant to the needs of the scientific programme.

10.1 Technology Developments and Project Contractor Selection

As a part of the Concept and Technology Study, several technology activities which are required to demonstrate the detailed feasibility of the proposed satellite concept were identified (see Section 10.2). Several of these required technological developments are of relevance to other proposed missions, and are of general utility to the ESA and national scientific programmes.

Based on the importance and implications of each of these for system design, an indicative technology development plan has been defined (Figure 10.1), assuming that the GAIA mission is scheduled for launch in 2009. For an initial period of about 3 years, technology developments would most efficiently be performed with the overall management and co-ordination of the D-SCI Future Study Office, with the scientific support of the Science Advisory Group and with the technical assistance of two industrial contractors, to be selected by an open call for tender for the technology technical assistance and for the Definition Phase of the GAIA Mission. Towards the end of this 3-year period, a short wrap-up study should confirm the feasibility of the overall concept and complete the technology activities. During the Definition Phase Study, the GAIA Project Team should be formed to select, at the end of the Definition Phase, a single industrial contractor for the Phase B/CD, and to manage the GAIA satellite activities.

10.2 Areas of Technological Studies and Development

Several technological activities are foreseen prior to the construction of GAIA, in order to ensure that the spacecraft development plan and cost are reliable. A comprehensive list of technological developments has been established during the study. Identified development is foreseen for:

- three side buttable, small pixel, high performance CCDs for the astrometric instruments (design and development pre-qualification);
- large-area, highly integrated CCD based focal plane assembly for the astrometric instruments (preliminary design and bread-boarding activities);
- large size, high performance CCDs for the spectrometer (design and development pre-qualification);
- large size, high performance CCD for the photometer (design and development pre-qualification);
- highly integrated, high speed, low noise detection chains for the astrometric focal planes (design, bread-boarding, and development of ASIC hybrids);
- large area SiC mirrors (study and validation of the manufacturing and testing process);
- ultra-stable large size SiC structure and application to the payload primary structure (study and validation of the assembly and testing process);
- large size deployable solar array sunshield assembly (preliminary design and bread-boarding);
- phased array antenna for high data rates and L2 orbits (study, with preliminary design and bread-boarding);
- complementary qualification of a FEEP based reaction control subsystem (delta-qualification for the adequate range of thrust and lifetime);
- optimised on-board compression algorithm for the science data (study, preliminary design, and validation);
- payload data handling electronics architecture optimisation;
- study of ground verification calibration approach and required facilities;
- inch-worm actuators for the refocusing mechanism (breadboard and testing);
- data base architecture, including storage, archive and processing of the satellite data.

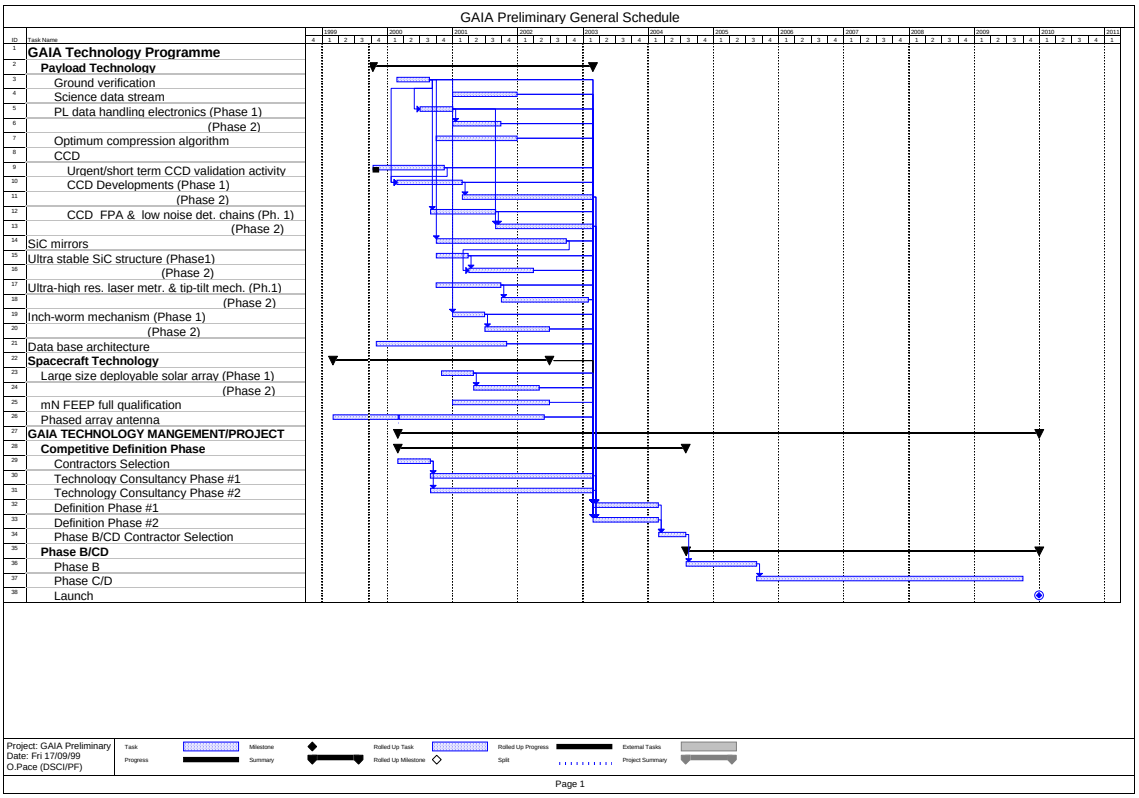


Figure 10.1: Technology development schedule for launch in 2009.

10.3 Procurement Philosophy

The management model outlined here assumes that the satellite, including payload and operations, is fully under ESA responsibility. The substantial scientific activities are assumed to be the responsibility of the ESA scientific community, and include activities related to instrumental and satellite design and optimisation, scientific performance assessment, and preparation and eventual implementation of the data reduction tasks leading to the production of the final mission science products. This approach, which is driven by the complexity and 'system nature' of the payload, and also by the effort demanded of the scientific community with respect to the data analysis, is a generalisation of that adopted for Hipparcos, and provides the basis for the mission schedule and cost assessment. Thus the cost presented corresponds to an all-inclusive 'cost at completion', covering satellite, payload, integration, testing, launch, and spacecraft and science operations.

10.3.1 Involvement of Scientific Institutes

The involvement of scientific institutes in the hardware development is anticipated and expected. If desired by participating national agencies, elements of the payload could be funded and supplied by (consortia of) scientific institutes. The proposed management model assumes, in that case, that specification, integration and testing would remain under the responsibility of the industrial Prime Contractor in order to ensure overall system integrity, and thus maximise the probability of achieving the stated mission goals. Although full modularity of the payload, with well-separated and stand-alone instruments, is not realistic, an integrated design approach is proposed. This is based on one common optical bench, which accomodates the various optical elements and focal plane systems. This minimises the design complexity, the overall mass and size, and the integration and test activities. This approach is reflected in the product tree (Figure 10.2).

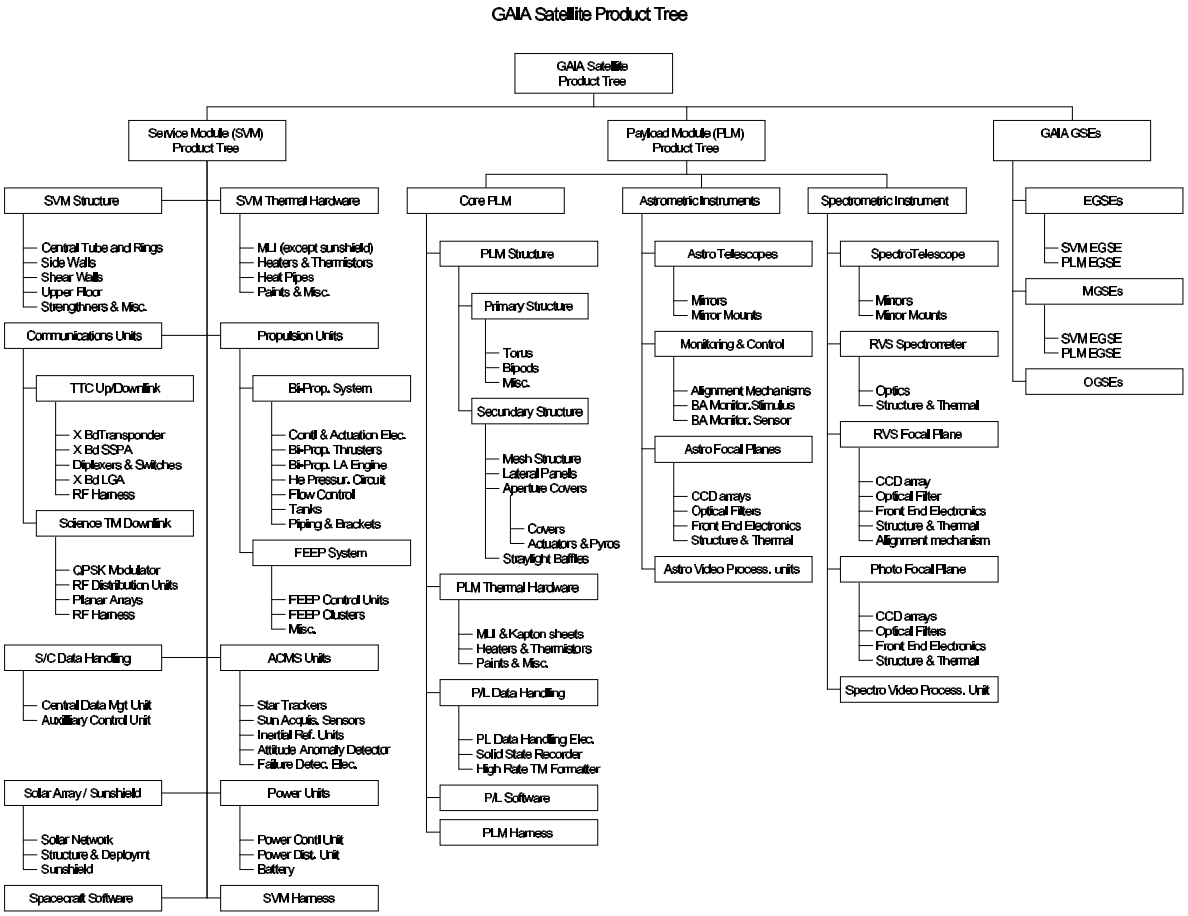


Figure 10.2: Product tree.

In order to maximise opportunities for open involvement and tendering, the satellite design has been made modular as far as possible, with well-separated payload and service modules, and with well-defined interfaces. Consequently, assembly, integration and testing at payload and service module level may be largely decoupled, providing additional flexibility in the implementation approach. As part of the GAIA study, and in order to initiate discussions of community involvement, an initial identification of studies suitable for specific laboratory development of flight hardware in European scientific institutes has been made. The cost of these tasks has been estimated. This list should be considered together with the list of technological developments as part of the ESA technology programme noted above. It includes:

- support for the astrometric instrument development (cost being dependent on activity definition) specifically: detection chain performance validation; star detection/discrimination algorithm optimisation and validation; data compression algorithm validation; refocusing mechanism: design consolidation, development and verification; and basic angle monitoring device: design consolidation, development and verification;
- radial velocity spectrometer development: spectrometer optics (imager, disperser, collimator, structure); spectrometer focal plane (CCDs, front end electronics, actuator); and spectrometer video processing unit (sequencing, acquisition, digitization). Activities would be associated with Phase B+C/D, i.e., related to the detailed design, manufacturing, testing, qualification, support to instrument level integration and testing, and in-orbit calibration;
- medium-band photometer development: dedicated optics, if any; focal plane with filters; video processing unit. Activities would be associated with Phase B+C/D, i.e., related to the detailed design, manufacturing, testing, qualification, support to instrument level integration and testing, and in-orbit calibration;
- combination of the spectrometer and medium-band photometer functions: as described above.

10.3.2 Instrument Selection

After the approval of the mission, it is assumed the Agency would issue an Announcement of Opportunity calling for proposals for the science data centre or centres, and for the supply of selected components of the payload instruments, as described in Section 10.3.1. The Announcement of Opportunity would be explicit in terms of the spacecraft mission and resources, technical interfaces, schedule, deliverable items, responsibilities of the parties, etc; would contain a description of the payload; and would request proposals for the supply of the complete packages, as provisionally defined in Section 10.3.1; and for the scientific operations concept. These proposals should describe in detail the design and development of the envisaged implementation/instrumentation and its associated software, describe the management structure within the collaboration, and show how the responsibilities for the scientific, technical, operational and analysis aspects are discharged.

Selection of principal investigators would take place via the normal procedure, which includes a technical evaluation by the ESA Executive, with the concurrence of the selected industrial Contractor, a scientific evaluation by ESA's scientific advisory bodies and approval by the Science Programme Committee. The selected principal investigators would be responsible for obtaining the necessary funding from the appropriate national authorities. Following selection, refined resource allocations and instrument interfaces would be negotiated with the principal investigators prior to the start of the spacecraft Phase B. These resources and interfaces would be frozen at the end of Phase B. Special items of mechanical ground support equipment and electrical check-out equipment might also have to be supplied if appropriate.

During the development phase, the project team would conduct a preliminary design review, a critical design review, and a flight model review. The Science Advisory Group, comprising collaborating scientists, mission scientists as appropriate, and the ESA project scientist, would be established to support the project. One possibility is that ESA would give responsibility for managing the provision of the payload to the Phase B-C/D industrial Contractor, who would assist in the management of all supplied items during Phases B and C/D of the project. Further details of the proposed science organization and management are given in Section 10.4.

10.3.3 Industrial Management

It is proposed that the spacecraft procurement is carried out by an industrial prime contractor. At the end of the technology development programme, an open Invitation to Tender would be issued to select two industrial contractors for a competitive Phase A, of a few months duration. At the end of Phase A, the two contractors would supply independent satellite development proposals, including industrial costs. After review, a single prime contractor would be selected to carry out Phases B and C/D. The industrial contract would be funded and placed by ESA. Should this be considered appropriate at the time, the prime contractor might in principle be delegated responsibility for control and monitoring of the liaison between partners, sub-contractors and principal investigator groups, and/or be in charge of the launcher procurement.

10.3.4 Development Plan Overview and Schedule

The suggested development plan aims to minimise the spacecraft development cost at reasonable risk. It is based on a minimum number of module and spacecraft models and a maximum utilization of each model. Because of the relatively simple, although challenging, opto-mechanical and thermal architecture of the payload module (one single primary structure with a set of mirrors and focal planes directly attached to it), it is possible to propose a reduced model philosophy for the payload module development. Efforts will concentrate on the thermo-elastic validation of the opto-mechanical assembly and on the validation of the focal planes and video chains design and performance.

The payload models are: (i) a payload optical structural and thermal model, for the mechanical qualification of the payload structure, and the validation of the thermal design. The structure

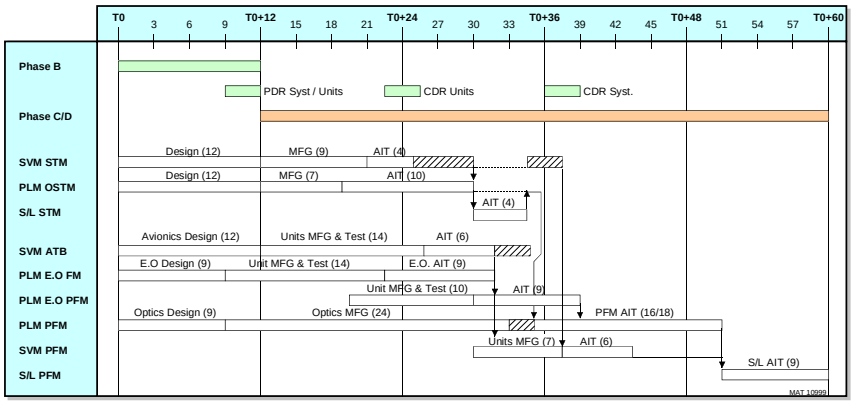


Figure 10.3: Spacecraft development schedule. Abbreviations used: SVM: service module; PLM: payload module; STM: structural and thermal model; OSTM: optical structural and thermal model; PFM: flight model.

would be built directly to flight standard and would be re-used for the protoflight model; (ii) a payload electro-optical engineering model, which will be used to validate the design of the focal planes, video processing units and digital handling units, to check the compatibility between these units, and to carry out several functional and performance tests at ambient and in vacuum; (iii) a payload proto-flight model.

A reduced model philosophy is proposed for the development of the service module, because of its relatively simple and classical design. Two service module models and an intermediate avionics test bench are foreseen: (i) a structural and thermal model, for the mechanical qualification of the payload module and launcher interfaces and for structural and thermal verification. Equipped with dummy units (solar array sunshield assembly and the payload module structural and thermal model) it will also undergo thermal tests in vacuum; (ii) an avionics test bench, made of engineering models of the non-recurring avionics electronics, actuators or sensors, and flight models of 'off-the-shelf' units. It will be used for the verification of the avionics hardware and software compatibility and the verification of the functional performance; (iii) a protoflight model, built with the refurbished structural and thermal model structure. Two spacecraft models are built from the payload and service modules; (iii) the spacecraft structural and thermal model, used for the mechanical qualification of the spacecraft and for the validation of the thermal design; (iv) a protoflight model, used to complete the thermal qualification and acceptance process and submitted to an acceptance test programme. All payload and service module models participate in the qualification and acceptance process, but flight acceptance is obtained only on the spacecraft protoflight model.

A model philosophy at assembly and unit level has been derived from this model philosophy defined at spacecraft and module level, and follows the same objective of cost minimisation at reasonable risk. It is also based on minimising the number of models, and takes into account the number of models to be built at module/spacecraft level (and their respective use) as well as the projected development and qualification status of the equipment unit at the time of the project development. All assemblies/units have been identified as a result of the present Concept and Technology Study. The proposed schedule of the overall programme is given in Figure 10.3. The schedule is based on the assumptions that the technology programme would be completed in 2003, Phase B completed mid-2005, and the launch in mid-2009. In this case, time T_0 would be mid-2004; time T_0+12 (the start of Phase C/D) would be the first quarter of 2005; and the end of the Phase C/D would be the end of the first quarter of 2009, with launch in late 2009. This launch date could be met, or even brought forward, if advanced Phase C/D activities were initiated in parallel with Phase B.

10.4 Scientific Programme and Observatory Concept

A mission which extends to such faint and complete magnitude levels as GAIA, i.e., $V \sim 20$ mag, may be regarded at one extreme as a survey experiment to which the principles of an 'observatory' mission do not apply, or as the 'ultimate' observatory in which every interested scientist can acquire any or all of the astronomical targets of interest. Procedures should maintain as much flexibility as possible for scientists to acquire data from GAIA as rapidly as possible, whilst ensuring that the data are calibrated, verified, and properly documented before becoming public.

A consideration of the time-scales involved illustrates the issues to be faced. Assuming a launch in 2009, a 6-month coasting phase, and nominal operations starting early in 2010, implies that the baseline mission duration of 5 years will extend until mid-2015. If the full analysis is completed within 3 years after the end-of-life (as achieved by the Hipparcos teams) and documented and distributed 1 year later, the final results will be available in 2019.

Science operations pre-launch will be relatively simple and inexpensive: no observing proposals need to be solicited in order to define the observing programme, and no specific scheduling or optimisation of the observation programme is necessary, or indeed possible.

The processing of the satellite data required for the derivation of the highest accuracy parallaxes, proper motions, radial velocities and photometry is, in contrast, a huge undertaking, and will require the coordinated efforts of one or more dedicated teams working on the entire database.

Community access to the data can take a number of forms. One option is to simply release the entire GAIA Catalogue after the data analysis teams have completed the processing. Another option is to solicit proposals from scientists shortly in advance of the completion of the data reduction, and then grant them priority access for one year, after which all data becomes public. This could be done at separate stages of the project, so that early access could be provided to subsets of the data. The second option has the advantage that the dedicated efforts of the teams can be rewarded with early access, but at the expense of a slight delay in full public access.

Certain photometric results should be released rapidly, close to real-time where feasible, to allow the follow-up of phenomena such as bursts, supernovae, and gravitational lensing events.

Although not an ESA policy, it should be noted that effective exploitation of the data will be helped significantly by awarding funds for personnel and computer hardware, based on proposals to use the GAIA data. This approach has been successful on NASA missions, where funds are awarded to US scientists who win observing time or who write an excellent science proposal based on data already in the archive. The Hubble Deep Field provides a particularly interesting example. The observations were planned, carried out, and reduced by a team, after which they were released immediately. The possibility to submit 'archival' proposals for science programmes based on the HDF helped involve a large part of the US science community very swiftly, and with great effect.

10.5 Requirements on Ground-Based Observations

Since the only information that will be available for many objects observed by GAIA will be the measurements returned by the satellite itself, considerable effort will be devoted to the further optimisation of the photometric measurements on-board necessary to furnish relevant astrophysical information on each target object. Specific preparatory programmes will undoubtedly be foreseen for many applications—for example, acquiring radial velocities for cluster members with higher precision than will be obtained by the radial velocity spectrograph—but for the most part, GAIA will provide its own data base without placing strong requirements on advanced ground-based observations. Follow-up work will include high-resolution spectroscopy to obtain radial velocities and element abundances for physical groups identified in the GAIA database, and studies of peculiar objects (variables, supernovae, quasars, etc.). The study of fields with high star density and complex backgrounds could benefit, for example, from high-resolution, multi-colour photometry to support the medium-band photometry observations. There may be requirements on obtaining spectra for the brighter quasars discovered, which will make up the reference frame, in order to optimise chromaticity corrections, and studies with VLTI and NGST, to quantify any microlensing effects on photometric stability. Such targets will also be of interest for many other studies.

10.6 Data Analysis

The data processing of GAIA will represent a significant challenge, for the reasons detailed in Section 9.6.1: not only is the data volume considerable (1 Mbps or some 20 TB of raw data over the mission lifetime, and a total data volume of perhaps 100 TB), but a number of other key aspects contribute to its complex nature: (i) the fact that observations of each object, or each area of sky, are distributed throughout the mission observations: a simple re-ordering of the data, by object, is inapplicable, since much of the data processing will be conducted in the 'time domain' (CCD response, photometric calibration, etc); (ii) many different pieces of mission data will need to be considered together as part of the object model: the photometry, the radial velocity, and the astrometric data; (iii) the control of systematics will be paramount; (iv) many different groups, with enormously varied expertise, will be required to provide algorithms which will all contribute to the overall reduction system; (v) the solution will be an iterative one.

A 'serial' reduction approach was adopted for Hipparcos, where the first stage of the analysis was conducted at one institute, the next stage at another institute, etc., with the final results being assembled or inter-compared at others. This approach is considered to be intractable for GAIA, considering the huge data volumes, the complexity of the processing, and the necessity of implementing an iterative approach at a number of processing levels. On the other hand, the parallel reduction of the Hipparcos data (a similar approach is being followed with Planck) provided one of the great strengths of the Hipparcos data analysis: that the results of the various complex steps of the processing could be verified, controlled, and rectified, at a number of different levels.

A plausible solution could be the formation of one or more processing centres, which would physically receive and be responsible for all steps of the multiple level processing foreseen for the total ensemble of the satellite data, including photometry, astrometry, radial velocities, etc. These groups could concentrate on the physical processing of the data, incorporating algorithms supplied by participating scientists elsewhere in Europe.

This single control centre approach is presented explicitly as one of the possible choices: the Hipparcos experience (and COBE practise) showed this was a powerful way of ensuring internal consistency. However, Planck is setting another model. In principle, every national community could have its own set of the data, if sufficient national support were available, but that would pre-suppose a coordinated science team, doing the comparison and contrast checks. The primary requirement would be to avoid lots of independent data reductions, without a coordinated overview.

To take a specific example: one or more groups interested in taking responsibility for the reduction algorithms for the radial velocity determinations would supply their algorithms to the core processing team for implementation, would test their algorithms on simulated data pre-launch (established as a separate task for each element of the GAIA payload data), and would receive all requested diagnostic or reduced data from the core processing team as the data analysis proceeded. More than one interested algorithmic group for each major analysis task or topic (e.g. planet detection, gravity wave detection, supernovae alerts, etc) would permit the same kind of algorithmic validation as realised for the Hipparcos reductions, without the duplication of the total mission data, or the transport of huge data volumes (with the associated problems of data validation) between the various institutes. Such an organization could also support the concept of a 'fast data analysis' group, which would deliver preliminary (but highly accurate) results to the community at an early stage in the mission lifetime. Evidently, if two processing groups were interested in taking on the core processing tasks, this could be considered, and may even be advantageous during the start-up phases, when the futures of such groups may be less certain.

10.7 Science Management

In view of its considerable complexity, GAIA should be undertaken with the same attention given to coordination and management of scientific activities as ESA has traditionally devoted to the spacecraft activities. Major parts of satellite development now follow relatively well-developed engineering practices, with mature industrial teams taking direct responsibility for many of the subsystems. Instruments, data volumes, science operations and processing intricacies are reaching higher levels of complexity, for which similar professional management is appropriate, and related to which a significantly enhanced level of service to the scientific community is demanded. This will require adequate resources to: (i) permit proper and efficient communication flow and in-

formation control and distribution throughout the science teams; (ii) maintain a structured and up-to-date repository of documentation and information on the mission; (iii) allow the assignment of manpower to urgent or high-priority activities to maintain the schedule of the scientific and satellite activities; (iv) provide expert support and advice in areas of the payload instrumentation, accuracy assessment, user interface, data processing, and all science-related tasks; (v) respond to demands on the web site, carry out pro-active public relations and community outreach efforts, and helpdesk-type support to the external community (see Section 10.8). Such resources have been identified in the projected mission costs, and will be essential if the mission is to fully capitalize on its inherent possibilities.

10.8 Public Relations, Education, and Outreach

Space science missions, and in particular astronomical studies of the visible Universe, have an extraordinary interest for the public at all levels. In addition to the primary ESA goals of basic research and applied technological developments, this means that related educational goals, such as helping to ensure that a continuing supply of scientists, engineers, and technologists will be available to meet the needs of the twenty-first century, can build on GAIA. The images, scientific discoveries and new appreciation of the scale and diversity of nature provided by astronomy captivate people's imaginations, inform teachers, and excite students and the public about science and exploration. To realize and enhance this educational and public outreach capability, the ESA countries (and NASA) are increasingly working to make education at all levels, and the enhanced public understanding of science, integral parts of the Space Science missions and research programmes.

GAIA is exceptionally well-suited for this educational, awareness and technical training requirement. GAIA will provide opportunities and challenges at all levels, from the evolution of the Galaxy and the search for extra-solar planets, through applied gravitation, to the technical challenges in accessing large data sets. Every one of these is of direct and topical interest, and produces knowledge of very wide and continuing general applicability.

Among many examples of GAIA science which are directly appropriate for general educational opportunities, GAIA will provide the first detailed knowledge of the content and evolution of our own Milky Way, an aspect of astronomy with immediate resonance for all people at all levels. Thus the basic level of interest will be high. Building on this, GAIA will provide detailed knowledge of kinematics, allowing a natural forum for explanation of Newtonian and General Relativity gravitational theory, chaos theory, and orbits. This can be provided naturally at the pictorial level – a movie of the sky – through to the highly technical – metric mapping, gravitational distortions of space-time – appropriate to all ages and interests, and all levels of educational requirements. At a wider level, by providing a precise measure of the distribution of dark matter near the Sun, and throughout the Galaxy, GAIA will set the boundaries of our understanding of the nature of matter, luminous and dark. The direct links with particle physics and fundamental physics are well known, and of wide general appeal.

It is worth emphasizing the natural scientific complementarity of GAIA, which uniquely addresses the nearby Universe, to the ESA Horizons 2000+ missions, especially Planck for the very early Universe, XMM and Integral for high-energy physics, Cassini/Huygens, MARS Surveyor and Rosetta for Solar System interest, NGST for high redshifts, and FIRST for star formation and continuing formation of new objects. This complementarity, which includes the major NASA missions, extends to public outreach. While these other missions are in general of somewhat more specialised interest for the public, their complementarity in providing a full understanding of the Universe over many wavelengths and all times, will become a feature of the programme. The scale of the GAIA dataset and its resulting data processing requirements will also provide outstanding opportunities for computational research.

A key feature of such future outreach opportunities is use of the interactive capabilities available on wide-bandwidth internet networks. Already national educational systems are installing internet access facilities in schools and public libraries. Private access is rapidly expanding. Thus, one will

have the capability to provide a full outreach programme with affordable resource commitments in distributed locations: a major funded single facility is not required. The only specifically national requirements are for language translation, which are modest, especially now that automated translation programmes are being perfected.

Given this opportunity, an affordable outreach programme will provide educational opportunities at every level, from starting school through to University research training, with complementary personal home learning. Introductory self-teaching interactive modules for internet learning will be the norm, with many trial developments appropriate to this scale already in operation. These educational opportunities will naturally involve astronomy, physics, technology, and information technology, and involve access by schools, universities, libraries, museums, and private individuals.

ESA projects now include a specific budget line to be dedicated to public relations and outreach programmes, and appropriate structures must be put in place to ensure an effective use of such resources not available to previous ESA space science missions. Examples could include educational modules (at a variety of levels) related to basic principles of angular measurements, Earth motion, and stellar distances and motions; planetarium programmes focusing on stellar distances and motions; and observational astronomy programmes for amateur and classroom groups related to the monitoring of variable stars. The effectiveness of these programmes has already been demonstrated by the interest of variable star astronomers in the Hipparcos measurements (see, for example, *'Mining Hipparcos's Buried Treasure'* by R.W. Sinnott in the June 1999 issue of Sky & Telescope, and *'The Educational Potential of the Hipparcos Data Base'* by the President of the IAU Commission for the Teaching of Astronomy, J.R. Percy, in ESA SP-402, 739, 1997). And the Hipparcos www site, which includes stereo images, 3-d simulations, access to the Hipparcos and Tycho Catalogues, a series of other animations, as well as links to published scientific papers, has already featured in David Ratledge's book *'Software and Data for Practical Astronomers: the Best of the Internet'*. The Haydn Planetarium in New York, for example, now includes the 'Digital Sky Project', comprising a space travel to Orion at several thousands km/s, entirely constructed from the Hipparcos Catalogue.

Complementary to these formal educational capabilities, GAIA addresses science of vast general appeal. The general public is genuinely fascinated by astronomical discoveries, and their important general implications for people's understanding of their place in the Universe. Thus activities ranging from the motivation of the creative arts, through galleries and the public media, to informed political debate, will naturally follow from public understanding of the GAIA mission goals, and the GAIA scientific return.

To take one specific example, the discovery of Near Earth Asteroids has enormous potential in terms of public outreach and communication. GAIA's contribution would represent an important opportunity for educational and outreach purposes, and would respond to the repeated invitations of the United Nations and the Council of Europe for a better assessment of the impact risk.

Part V
Appendices

A Related Ground and Space Projects

Interest in the astrophysical capabilities of astrometry has led to a variety of recent proposals for space missions dedicated to astrometry. These missions are divided into two categories: those that aim for Hipparcos-type accuracy on a comparable number of stars (i.e., some one hundred thousand objects); and those that aim for a significantly higher astrometric accuracy. In the former category were missions proposed over the past 10 years from scientists within the former Soviet Union (Lomonosov, Regatta-Astron, AIST, and STRUVE). Their scientific importance was based on maintaining the reference frame established by the Hipparcos mission throughout the coming decades, such that a mission in 10–20 years from now, with individual accuracies in the range 1–2 milliarcsec would still have a useful and important, if not fundamentally new, scientific value. Little progress has been reported on these studies in recent years.

The second category of missions includes two 'super-Hipparcos' type instruments. They were both inspired by the GAIA design, and each can be considered conceptually as a 'mini-GAIA': scanning instruments, with two viewing directions, and the same global measurement and reduction approach. DIVA is a German national programme, still at the proposal stage, employing a small Fizeau interferometer (and dispersed fringes) to reach milliarcsec level accuracies for at least 2.5 million stars down to $V = 12$ mag, with guest observer programmes down to about $V = 15$ mag (Röser et al. 1997). FAME was approved at the end of 1999, and is based on an instrument design originally formulated by the U.S. Naval Observatory, and now incorporating design features from the Harvard (POINTS/NEWCOMB) group. Its accuracy targets are similar to those of DIVA. A Japanese Astrometric Satellite study has also been mentioned (Yoshizawa et al. 1997).

The goal of JPL's OSI and SONATA concepts (Shao 1993; Buscher et al. 1995 and the POINTS mission (Reasenberg et al. 1994a), was to achieve significantly higher astrometric accuracy on specific objects. These programmes aimed to reach accuracies at the level of a few microarcsec or better, per observation, with some or all of these *pointed* instruments being able to benefit from repeated measurements throughout the mission. The limiting magnitude of POINTS was between 14–18 mag, OSI's accuracy target was 5 microarcsec on 20 mag stars, while SONATA targeted 0.5 microarcsec for 14 mag objects after 4 hours. A lower-cost alternative to POINTS was NEWCOMB (Reasenberg et al. 1994b), which aimed to reach an accuracy of about 100 microarcsec in a single measurement of about 4 min.

There is a continuing effort to develop ground-based optical interferometers, building on the success of the Mk III optical interferometer (see, e.g., Hummell et al. 1994). These instruments should ultimately provide very high relative astrometric accuracy over small fields (for example, on bright double or multiple systems) at the 10 microarcsec level (Shao & Colavita 1992; Shao 1995). Within ESO's VLTI programme, PRIMA ('Phase-Referenced Imaging and Microarcsecond Astrometry') aims to use the four 8.2-m VLT unit telescopes to achieve initially 50 microarcsec astrometry (10 microarcsec finally) for a variety of narrow-field applications. None of these instruments are considered as tools for the measurement of large numbers of objects, mandatory for many Galactic studies, nor for large-angle (positional) astrometry, mandatory for the large-scale determination of parallaxes and proper motions.

A.1 SIM

SIM (Space Interferometry Mission, Boden et al. 1997) is being developed by the Jet Propulsion Laboratory under contract with NASA and in close collaboration with two industry partners. It is planned for a launch in 2007 in an Earth-trailing solar orbit. In this orbit the spacecraft will slowly drift away from the Earth at a rate of approximately 0.1 AU per year, reaching a maximum communication distance of about 95×10^6 km after 5 years.

SIM is a space-based Michelson interferometer with a 10 m baseline capable of providing wide angle astrometry at 4 microarcsec accuracy, with parallaxes to comparable accuracies down to about 20 mag (Table A.1). Narrow-angle astrometry over a field of 1° will allow measurements of

Table A.1: SIM instrument and mission parameters.

Feature	Information
Baseline	10 m
Wavelength range	0.4–0.9 μm
Telescope aperture	0.3 m
Wide angle field of view	15°
Narrow angle field of view	1°
Imaging field of view	0.3 arcsec
Orbit	Heliocentric Earth trailing
Mission lifetime	5 years
Wide-angle astrometric accuracy	4 microarcsec
Narrow-angle astrometric accuracy	0.5 microarcsec (single measurement)
Imaging resolution	10 milliarcsec
Imaging Sensitivity	20 mag in 1 hour
Number of stars in the grid	$\approx$ 30 000
Number of tiles in the grid	$\approx$ 300

the motion of brighter stars ($V < 16$ mag) with a single measurement accuracy of 0.5 μarcsec. Thus SIM will be able to search for planetary companions to nearby stars by detecting their astrometric reflex motion.

During the technical studies two designs were considered: 'SIM Classic' with three separate baselines and a pointing subsystem using seven siderostats distributed along the spacecraft structure; and 'Son of SIM' with a single 10 m baseline common to the guide and science interferometers and a hexapod pointing system for acquisition of targets. SIM Classic was selected in October 1999.

SIM will observe relative positions of individual stars over small areas of the sky (referred to as 'tiles') with the spacecraft inertially pointed. These observations will be combined later in two dimensions over the whole sky into a self consistent astrometric catalog (the astrometric grid). The grid will comprise mainly bright objects, free of astrometric jitter, and about hundred quasi-stellar objects, some radio loud, to provide an extragalactic anchor for the astrometric grid. Science targets will be observed within a tile and linked to at least four grid stars at a time.

Simulations have shown that for a 4 microarcsec, wide-angle accuracy over a 5 year mission, the instrument will need to make a single astrometric measurement with a systematic error less than 7.5 microarcsec. This yields an integration time of the order or less than 1 s for stars brighter than 10 mag, 200 s at 15 mag, 30 min for 18 mag and more than 4 hours for the faintest targets at 20 mag. Adding the dead time for pointing and calibrations, this puts a severe constraint on the content of the science programme which will include a relatively small target observing list of several thousand pre-selected programme stars. SIM has enormous capabilities, but due to its observing time requirements it will only be able to observe a very small fraction of stars down to any given magnitude limit.

The possibility of a degraded mission has been studied with lower capabilities but still acceptable in terms of scientific performances, with accuracy figures of 10 μarcsec (compared with 4 μarcsec) for the overall mission accuracy in wide-angle astrometry for stars brighter than $V = 16$ mag (compared with 20 mag). For narrow-angle astrometry the mission is considered as non-viable if the accuracy deteriorates above 3 μarcsec for the singe measurement precision at $V = 13$ mag.

SIM's high astrometric accuracy makes it suitable for specific scientific goals, such as the detection of low mass planetary companions around selected nearby stars; however, the relatively few objects which can be observed means that its objectives do not significantly overlap with the majority of the scientific goals attainable by GAIA: as stressed in the science case, GAIA will not only yield information on huge numbers of objects (which may be used individually or statistically) but will do so without requiring *a priori* knowledge of the candidates to be observed.

Table A.2: FAME instrument and mission parameters.

Feature	Information
Astrometric accuracy $V < 9$	50 microarcsec
Astrometric accuracy $V = 15$	300 microarcsec
Wavelength range	0.4–0.9 μm
Limiting magnitude	15
Telescope aperture	0.5×0.25 m
Number of fields of view	2
Field of view	1.1°
Basic angle	65°
Scanning period	20 min
Precession period	10 days
Orbit	Geosynchronous
Mission lifetime	2.5 years
Sun aspect angle	45°
Number of CCDs	24 2k×4k
Image scale	0.0275 arcsec/μm
Number of targets	40 million

A.2 FAME

FAME (Full-Sky Astrometric Mapping Explorer) was selected by NASA in the MIDEX competition in 1999. It is a small scanning satellite, comparable to Hipparcos in its overall design and intermediate between Hipparcos and GAIA for its accuracy performances (Table A.2). It aims to build an astrometric catalogue (positions, proper motions and parallaxes) for about 40 million stars with a precision of 50 microarcsec for all the bright stars ($5 < V < 9$) and better than 300 microarcsec for fainter stars ($9 < V < 15$) with a mission lifetime of 2.5 years. Photometric data will be obtained in four Sloan bands. As far as science is concerned (there are additional goals in navigation to fulfill the needs of the Department of Defense), FAME looks like a mini-GAIA. However its design is much more comparable to Hipparcos, with a beam combiner, but using CCD technology instead of an Image Dissector Tube. The telescope has two fields of view separated by a 65° angle, and the spacecraft will spin with a short period of 20 min. The spin axis is maintained at a constant angle of 45° from the Sun, and the solar radiation pressure on the solar shield will result in a precession about the Sun-spacecraft line with a 10 day period, requiring rare thruster firings. Unlike GAIA, on-board object detection will not be used.

A.3 DIVA

Diva (Deutsches Interferometer für Vielkanalphotometrie und Astrometrie; Double Interferometer for Visual Astrometry) is a small German astronomy satellite planned for launch in 2003, but not yet approved for funding. It is aimed at the measurement of positions, proper motions and parallaxes for about 40 million stars down to $V = 15$ mag with an accuracy in the range of milliarcsec, and somewhat better at the bright end (Table A.3). It is somewhat less accurate than FAME. Broad-band photometry and spectrophotometry for stars brighter than 13.5 mag are included in the proposal. The design of DIVA is built around two small Fizeau interferometers (one per field of view) with a 15 cm baseline. The fringes are dispersed on the focal plane and acquired in time delay integration mode on CCDs. The DIVA satellite has the outer form of a cylinder with a diameter of 1.8 m, a height of 1.5 m and a total mass of 410 kg. An average data rate of 550 kbps scientific data is expected.

Table A.3: DIVA instrument and mission parameters.

Feature	Information
Astrometric accuracy V = 9	0.2 mas
Astrometric accuracy V = 12	0.6 mas
Astrometric accuracy V = 14	2.0 mas
Wavelength range	0.4–1.0 μm
Limiting magnitude	16.2 (K0 stars)
Entrance pupils	7.5×7.5 cm
Baseline	15 cm
Dispersion	200 nm/mm
Number of fields of view	2
Field of view	0.°5
Basic angle	100°
Scanning period	2 hours
Precession period	56 days
Orbit	elliptical, 24 hr period
Mission lifetime	2.0 years
Sun aspect angle	45°
Number of CCDs	60 1k×1k
Image scale	0.172 arcsec/mm

A.4 Scientific Comparison

The measurement capabilities of the various planned missions are summarised in Table A.4. The target accuracies and numbers of objects for a variety of relevant catalogues are illustrated in Figure A.1 for positions/parallaxes and proper motions respectively.

A.4.1 GAIA and SIM

The scientific capabilities, and mission goals, of GAIA and SIM are nearly orthogonal. SIM is an ideal mission for precise measurements of a small number of carefully pre-selected targets of specific scientific interest. GAIA is a survey mission, and is ideal for statistical analysis of the unknown, with broad applications to the Solar System, galaxies, large-scale structure, and primarily galactic structure and evolution. The SIM target selection is yet to happen, but one may be confident it will be optimised for studies to which SIM can contribute uniquely. These include detailed searches for low-mass planets around a few nearby stars, calibration of the distance scale, determination of the fundamental properties of stars, especially in short-lived evolutionary stages, and detailed studies of known micro-lensing events. In all these studies SIM will revolutionize knowledge, based on the few thousand targets it will be able to observe.

The SIM science case is entirely complementary to the GAIA science case. SIM will search for a few candidate nearby planetary systems, for later study in detail; GAIA will search 300 000 stars for the planetary system distribution function, allowing understanding of planetary system formation and evolutionary processes. GAIA will also survey our Solar System for minor bodies, extending the census of the outer Solar System to new mass limits. SIM will calibrate the luminosities of important (e.g. Cepheids) and rare (e.g. black hole binary, supergiants) stages of stellar evolution, advancing stellar physics. GAIA will search the Galaxy for such objects, defining their distribution, lifetimes, and kinematics. GAIA will also discover any as yet unidentified short-lived phases of evolution. It will quantify and calibrate the distribution functions of stars as they depend on the astrophysically important parameters, age, metallicity, environment, and so on. SIM will define the luminosities of some known young and pre-main sequence stars. GAIA will provide a complete local census of the numbers of such stars, their locations, and their relationship to their environs. SIM will determine distances to galaxies in the Local Group directly. GAIA will map the 2-dimensional kinematics in these galaxies, quantifying deviations from simple planar dynamics. SIM will measure

Table A.4: Summary of the capabilities of Hipparcos and GAIA, along with those of other proposed astrometric space missions. Numbers of stars are indicative; in the case of SIM they are distributed amongst grid stars and more general scientific targets. Typical accuracies are given according to magnitude where appropriate.

Mission	Agency	Launch	No. of stars	Mag limit	Accuracy (mas)	Accuracy (mag)
Hipparcos	ESA	1989	120 000	12	1	10
DIVA	Germany	2003	40 million	15	0.2 5	9 15
FAME	USNO/NASA	2004	40 million	15	0.050 0.300	9 15
SIM	NASA	2007	10 000	20	0.003	20
GAIA	ESA	2009	1 billion	20	0.003 0.010 0.200	12 15 20

the stellar dynamics in the cores of globular clusters, and their global space motions, improving mass models of the outer Galaxy. GAIA will discover many thousands of new outer halo giants, vastly extending the reliability of determination of dark matter in the Galaxy. GAIA will map the phase-space substructure of the Galaxy, quantifying models of galaxy merger histories and dark matter structures. GAIA will allow determination of the star formation and chemical evolutionary history of the Milky Way, providing for the first time a robust picture of the structure and history of a large galaxy. That is, SIM will calibrate the known, to unrivaled precision, from a few thousand observations of pre-selected targets. GAIA will quantify the Milky Way, from billions of observations of every visible object, providing fundamentally new and unbiased knowledge of our Galaxy.

A.4.2 GAIA, FAME and DIVA

The difference between GAIA and these two missions is primarily one of scale, although GAIA also includes a comprehensive package of crucial photometric and radial velocity measurements. FAME and DIVA are essentially successors to Hipparcos, with an extension of limiting sensitivity, sample size and accuracy (statistical weight) by a factor of order 100 in each. DIVA and FAME will both provide an excellent reference frame, substantially improve calibration of the distance scale and the main phases of stellar evolutionary astrophysics, and map the local Galaxy to much improved precision. GAIA exceeds these two missions in scale by a further factor of order 100.

The implications of this additional factor are substantial. GAIA will extend studies of the lowest luminosity stars through a sample which is complete to 50 pc. FAME will be complete only to 3 pc, a volume which probably does not contain any such star. GAIA will map the Solar System out to the Kuiper belt, whose brightest members are below the FAME limit. GAIA will map the kinematics of the Local Group, and the nearby Universe. FAME and DIVA will reach only the three nearest dwarf satellites. Most significantly, the core science goal, understanding the Milky Way Galaxy, require that one surveys a large part of our Galaxy. Of the proposed missions, only GAIA can achieve that scientific goal.

A.4.3 Scientific Benefits From Multiple Missions

The four missions, GAIA, SIM, FAME and DIVA, which utilise astrometry as a tool to attain their scientific goals, are very different in their capabilities, and are complementary rather than competitive in their science cases. Fortunately, a significant scientific synergy, additional to the primary science cases, but nonetheless important for all of them, will arise from the combination of their data. This results directly from the increased time over which targets will be monitored, and the correspondingly increased sampling for variability and multiplicity. We note in passing that this synergy is an argument against merging any of the missions.

Detailed studies over many years have shown that multiplicity is a common feature of stellar systems. Available results remain preliminary, but it is clear that a fraction of order unity of all stars are multiple: that is, (almost) all stars have a stellar, sub-stellar or planetary companion. The distribution of separations of these systems extends to an upper limit set by the Galactic tidal field, and corresponds to orbital periods very much longer than the duration of a single mission. Available radial velocity studies (e.g. Duquennoy & Mayor 1991) suggest that the orbital periods of known binary stars follow a log-normal distribution, with median near 200 years, and logarithmic dispersion of about 2.

For a non-single stellar object, the centre of light does not, in general, correspond with the centre of mass. Therefore the photocentric motion of a non-single object will deviate from simple linear motion of an isolated single stars. Depending on the parameters of the stellar system, and the measurement accuracy, true non-linear photocentric motion will be seen as: (a) motion on a projected keplerian ellipse; (b) motion on a slightly curved path; or (c) apparent linear motion which differs from the linear motion which would be measured at other times (e.g. Wielen 1995; Odenkirchen & Brosche 1999). That is, measurement at any one time, as in a single mission, will be precise but systematically in error. Wielen (1997) has combined Astrographic Catalogue and Hipparcos data to identify examples of such phenomena, exploiting the long time base between the data sets.

An even more complex phenomenon, of spurious apparent motion induced by time-dependent asymmetry in the surface brightness of the target star ('variability induced movement') is also expected. Such spurious motions will be generated by, for example, large starspots on a rotating surface. The combination of both phenomena, for example in a mass-transfer system with an accretion disk, will of course also occur. Systems with either very long periods, or periods near one year, will of course be particularly affected.

All these phenomena induce systematic errors in derived astrometric parameters. While in some specific cases (e.g. Wielen et al. 1999) combination of all data may allow detailed study of individual systems, in general statistical distribution functions are all that may be derived. It is worth emphasising that such distribution functions are not only essential, to quantify the true wings of the apparent distribution of measurements for analyses of such quantities as phase-space substructures and mass distributions, but are also of intrinsic interest, for studies such as the true multiplicity function, with implications for angular momentum transport during star formation. Similar conclusions apply to photometric variability, stellar rotation and surface asymmetry, and so on.

It is apparent that combination of high-precision astrometric and photometric data over the longest possible time baseline is imperative for discussion of these astrophysical phenomena, and for quantification of the true accuracy of an arbitrarily precise measurement. Detailed simulations of the necessity for, and benefits from, combining data from the various space missions with older ground-based data are underway (Section 1.5.2). It is already clear that combination of the results of GAIA, FAME and DIVA will allow quantification of the distribution functions from tens of millions of stars, while combinations of GAIA, FAME and SIM will provide very precise assessment of the effects, and consequently the reliability of the SIM reference grid and astrometric parameters, from a few thousand very accurate measurements.

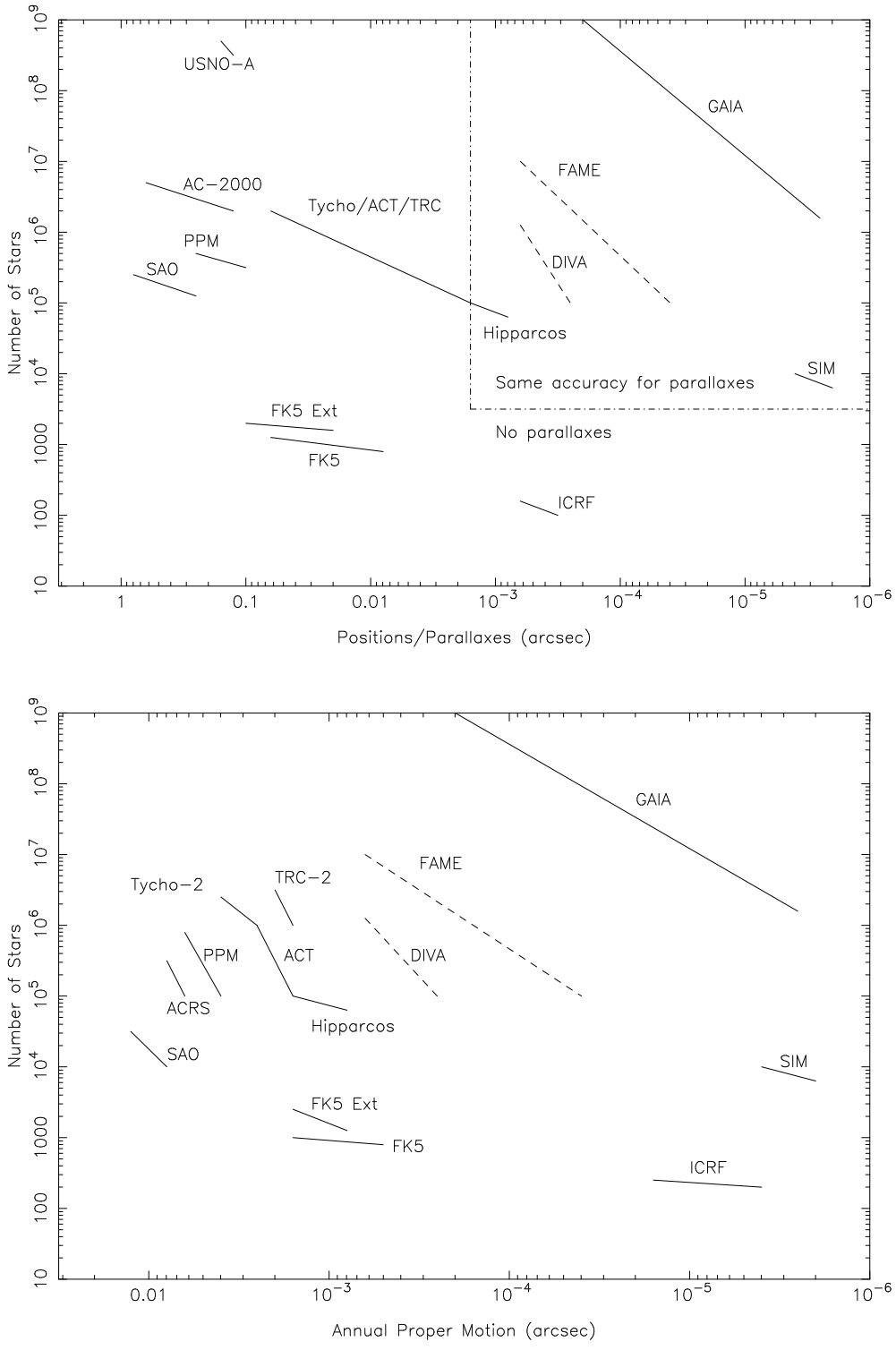


Figure A.1: A comparison of (top) the position/parallax accuracies, and (bottom) the proper motion accuracies of various ground and (planned) space programmes (courtesy J. Kovalevsky). (AC = Astrographic Catalogue; ICRF = International Celestial Reference Frame; PPM = Positions and Proper Motions Catalogue; SAO = Smithsonian Astrophysical Observatory; TRC = Tycho Reference Catalogue; USNO = US Naval Observatory)

B Interferometric Option

B.1 Introduction

The original proposal for the GAIA satellite at the outset of the present study consisted of three stacked interferometers; in the course of the baseline study it was shown that the key scientific goals of the GAIA mission (i.e. 10 μas astrometry at $V \sim 15$) could be achieved with a non-interferometric telescope. A detailed study of an interferometric mission concept was carried out in parallel, in order to establish more clearly the advantages and disadvantages of such an approach, the technological implications, and the possibilities of longer-baseline astrometric instruments in the future. The study, summarised in this appendix, was carried out as part of the ESA Technology Research Programme by a consortium led by Alenia Spazio, and which included Aerospatiale (now Alcatel), Matra Marconi Space, ETEL, and the scientific institutes Osservatorio Astronomico di Torino, Istituto Metrologico 'G. Colonnetti-CNR' (Torino) and EICAS Automazione (Torino).

This parallel study looked at a number of alternative subsystem concepts for the implementation of the GAIA mission, which can be considered either together as an unified alternative design concept, or can be considered individually as elements of possible alternative designs. The main features were: (a) an interferometric optical design, (b) an actively controlled optical system, (c) the use of a beam combiner and (d) a geostationary orbit.

The general mission profile of the Alenia study is in many important respects similar to the baseline mission profile, i.e., it consists of a scanning satellite simultaneously observing two sky regions placed at a large angular separation (the basic angle) along great circles, progressively covering the sphere by a composition of the rotation, precession and orbital motions. Each elementary observation is performed by imaging the scanned sky strip over CCD detectors operating in TDI mode. Sources to be observed are detected on-board using the data of the leading strips of CCDs from the focal plane, and selective read-out of the CCD chips is used to measure the position of each star as it drifts across the focal plane. The trailing CCD chips of the focal plane are used to perform broad-band photometry, while spectroscopic measurements and medium-band photometry are performed by a dedicated instrument, for which the same concept as described in the main report can be adopted. The main difference is that interferometric configuration produces fringes in the focal plane (rather than the elliptical point-spread function of the baseline optical configuration), which must be fully sampled by focal plane CCD detectors to derive the positional information for each star.

B.2 The Interferometric Optical Concept

In the course of the Alenia-led study the original GAIA concept was also modified significantly, whilst retaining the basic interferometric configuration. The most important changes were in the telescope design, in the number of interferometers, and in the presence of a beam combiner. The original concept consisted of three stacked, independent, three-mirror interferometers, while the present interferometric concept consists of two interferometers combined by means of a beam-combiner, thus simultaneously imaging two sky regions on the same focal plane. The third viewing direction, originally considered for reasons of stability of the global reference system, was discarded for reasons discussed in Section 2.1.1.

The advantage of the pair of interferometers with beam combiner compared with the three originally proposed is a much greater compactness, implying much better structural and thermal rigidity, and also a lower satellite mass. The advantages and disadvantages of two versus three viewing directions are discussed in Section 2.1.1. To allow for direct fringe detection the focal length of the interferometer had to be increased significantly, while maintaining a large field of view. Finally a folded Korsch design (a three-mirror telescope with two folding mirrors for compactness) was selected, which allows for a $0\overset{\circ}{.}8$ (along scan) $\times$ $1\overset{\circ}{.}4$ (across scan) field of view with fringe pattern visibility better than 77 per cent, using a red band (central wavelength $\lambda_0 = 750$ nm, a passband $\Delta\lambda \simeq 200$ nm), and an effective focal length of $\simeq 40$ m. The resulting well-corrected field of view

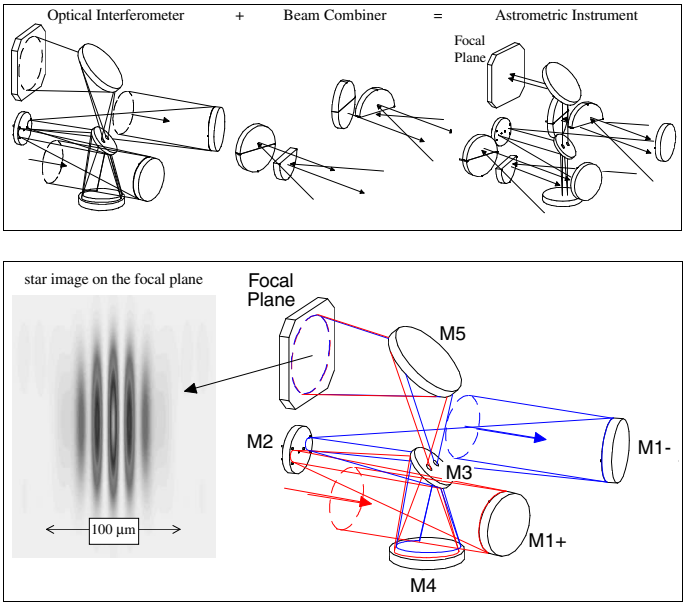


Figure B.1: The layout of the interferometric optical configuration from the Alenia-led study.

is significantly larger than for the baseline optical configuration. An overview of the optical system is shown in Figure B.1.

Each primary mirror is circular with a diameter $D = 0.65$ m, so that the collecting area of the individual interferometer primary mirror sub-apertures is 0.33 m^2, with an interferometric baseline $B = 2.45$ m. The resulting focal plane scale is 5.16 arcsec mm^{-1}, so that the physical size of the astrometric field of view is approximately 0.56×0.98 m^2. The field of view has two obscured regions with diameters of 0°.31.

B.3 Actively-Controlled Optical Systems

The stability of the basic angle is a key contributor to the error budget for bright sources, and will to a large extent determine the maximum achievable accuracy for the brightest stars. The beam combiner can provide a significant enhancement in the materialisation of the rigidity of the basic angle, by virtue of its being a physical realization of this quantity. While in the baseline study a purely passive optical system was chosen, in which the stability is achieved through accurate thermal control (and monitored through a laser device, described in Section 4.2.3), the Alenia-led study examined the possibility of achieving the same level of basic-angle stability through active control. The study included the design of adequate tip-tilt control mechanisms for the optical elements, of an interferometric measurement system (based on laser gauges) and of an appropriate control system.

This active-control system has different goals, including the initial phasing of the interferometer from a potentially distorted post-launch configuration, the maintenance of the phasing condition and the basic angle to the high accuracy required by the final accuracy of better than 10 μas. The use of the beam combiner allows these tasks to be decoupled, concentrating the control of the basic angle on the beam combiner itself, while the phasing is done by controlling the primary mirrors.

The phasing of the interferometer requires separate control of at least one optical element in each arm, since they must be phased with respect to each other, and placed in a position adequate for the alignment and focus of the whole instrument. The chosen actuation system is a tip-tilt device, operating on the three most critical degrees of freedom (one linear, i.e. displacement along the

optical axis, and two angular, corresponding to inclinations about it) out of the six of the controlled optical components, considered as a rigid body; the other movements are passively constrained by the mounting setup. The beam combiner requires active control of three out of its four flat mirrors, in order to preserve the stability of the base angle between the two viewing directions for both arms of the interferometer. Again, a tip-tilt device appears to be adequate to the requirements of optimal control. Such a system can be implemented with a two-stage control system, using low-voltage piezoelectric actuators (commonly used in space applications) for the fine control, and stepper motors for the coarse long-stroke control. This ensures achieving fine accuracy together with implementing specific long strokes (for example, to compensate for post-launch distortions).

Analysis of the optical configuration (performed through ray-tracing) shows that the control of the basic angle requires the most critical linear degrees of freedom of the beam combiner to be controlled to $\simeq$ 5 pm, while the less critical ones have a requirement of $\simeq$ 100 pm. For the interferometer mirrors the requirements on the linear displacements range from 100 pm and 10 nm.

Measurement of distances down to $\simeq$ 5 pm over $\simeq$ 1 m (the size of the beam combiner assembly) is challenging, and a specific development program was carried out in order to verify its feasibility. This work was led by the metrology institute 'G. Colonnetti' of the Italian National Research Council, CNR, whose researchers had achieved 10^{-11} accuracy in linear distance measurement and control (i.e. the level required by the GAIA requirements) using laser gauging, but on distance scales of a few centimeters. The present line of work consisted of the demonstration of the feasibility of laser gauging over much larger distances, and with components suitable for space applications.

The chosen concept relies on null metrology using Fabry-Pérot cavities. Such a cavity is obtained with a pair of high-reflectivity spherical mirrors placed at a fixed distance; it resonates for wavelengths which are an exact sub-multiple of its optical length, yielding a 'nulling signal' for such lengths. The signal is expressed in units of wavelengths, thus a very stable laser source is required to achieve the required accuracies. This can be obtained through a reference cavity of known, and above all stable, optical length, and locking a laser source to one of its modes. In this way it is possible to achieve a high stability of the laser frequency (or wavelength), which can then be distributed as a reference 'gauge' for measuring other unknown cavities. The reference cavity (of a few cm in size) is the only component in the whole system which needs to be thermally controlled to $\simeq$ mK accuracy.

This technique can reach a very high sensitivity by locking on the sharp individual resonances produced by the high finesse cavity with high reflectivity mirrors. This is achieved at the expense of the dynamic range, because basically no signal is obtained except from a very small interval (e.g. $\pm$1 nm) close to the resonance point. Therefore, the implementation of this kind of laser gauging naturally involves an active control, because the cavity length can be measured with high accuracy only close to the resonance length. The selected wavelength for the Nd:YAG laser used in the system, $\lambda = 1319$ nm, has been chosen beyond the $\lambda_{\rm Si} \simeq 1100$ nm cut-off of the silicon, so that possible stray-light from the metrology system would not affect the astronomical measurements.

The experiment implemented to demonstrate the feasibility of active control for the GAIA interferometers, named the COSI ('Control Optics Structure Interaction') testbed, consisted of two plates of diameter 0.5 m mounted on a supporting frame at a distance of 0.5 m, by means of piezoelectric actuators. The size, masses and distances of the two plates constitute a reasonable approximation to the actual GAIA mirrors in the interferometric design. The distance between these two plates was measured through three Fabry-Pérot cavities, and a closed-loop control system commanded the piezoelectric actuators. The whole assembly was contained in a vacuum chamber. The actual laboratory setup is shown in Figure B.2.

The objective of the COSI experiment was the simultaneous operation of the three cavities, with stabilization of the plates to within 100 pm, a typical, if not the most challenging, requirement for the interferometric configuration design described above. Through several test sessions the system successfully showed an internal measurement dispersion of the order of only 3–5 pm for uninterrupted operations of $\simeq$ 20 min, showing the validity of the overall proposed concept. The actual experimental setup prevented at this stage continued operation for the 3 hr which represents the GAIA rotation period (and thus the time-scale of the closure condition on the basic angle); also, full verification of the cavity stability through comparison against laboratory ultra-stable laser sources was deferred to a future phase. While a full assessment of the limiting performance of the experiment will thus have to wait for an upgrade of the existing setup, the COSI testbed has been fully successful in demonstrating the feasibility of the proposed metrology and control approach for the GAIA interferometric optical design.

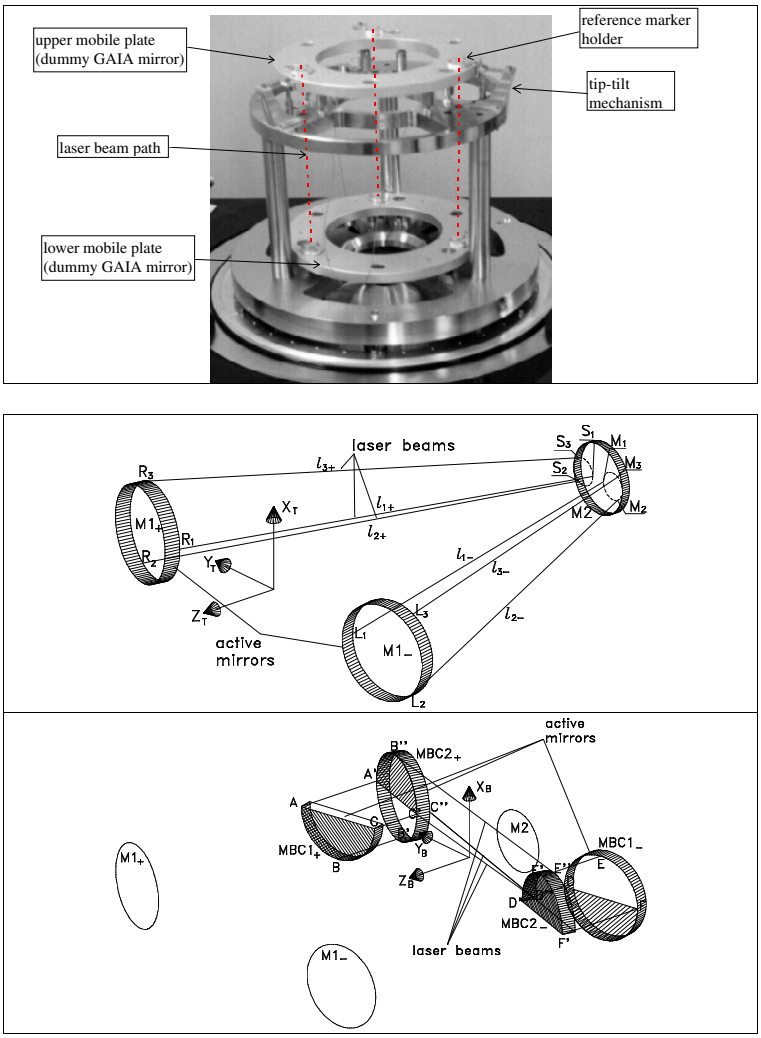


Figure B.2: Top panel: a photograph of the experimental setup (the 'COSI testbed') used to demonstrate $\simeq 5$ pm short-term control accuracy using laser gauges and piezoelectric actuators. The setup is a mechanically representative, scaled down model of the beam combiner used in the GAIA interferometric configuration. The upper and lower mobile metal plates represent the two mirrors, with the three laser beams positioned at their edges. In actual operation the assembly is positioned under vacuum. Bottom panels: a schematic view of the position of the laser gauges in the GAIA interferometric optics. The position of the three plus three gauges for the control of the position of the two (actively controlled) primary mirrors M1 with respect to the secondary M2 is shown above, while the position of the nine gauges for controlling the relative positions of the four beam-combiner mirrors is shown below.

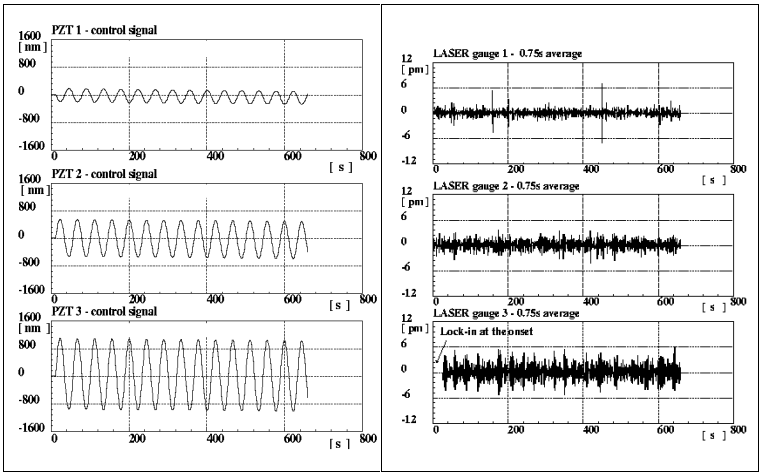


Figure B.3: Results from the COSI testbed. The left panels show the signals applied to the three piezoelectric actuators of the lower mobile plate (imposing a sinusoidally modulated deformation of the relative alignment of the two plates). This deformation is continually compensated by the control system using the error signal from the three laser gauges and the actuators on the upper plate. The error signal averaged on 0.75 s time bins (i.e. the variations from the null position of the laser gauges) is shown in the right panels. The residuals have an rms of few pm for all three lines, showing that the control system is able to fully compensate deviations of several hundreds of pm.

B.4 The Beam Combiner

To achieve observation of two separate directions in the sky with a single interferometer, use is made of a beam combiner, following a principle similar to the one adopted for the Hipparcos mission. To match the interferometer's entrance pupils the beam combiner is achieved through diluted optics, i.e. using four individual flat mirrors separately mounted on the optical bench. The overall arrangement corresponds to an X-shaped array of 4 flat mirrors.

The disadvantages of the adoption of a beam combiner have been summarized in the main report. At the same time, the main advantage of a compact beam combiner is the localization, in a well-defined and relatively compact assembly, of all the most critical degrees of freedom for the whole optical assembly, i.e. the ones influencing the basic angle. Therefore the pm-level metrology and control issues are not spread through the whole interferometers but concentrated on the four beam combiner mirrors. In fact, the beam combiner is a physical realization of the basic angle, allowing, with respect to the original concept of stacked instruments, a more convenient mechanical framework for the implementation of a reliable set-up with the required level of stability. The tolerance requirements are such that the maximum allowed linear displacement is, for the most critical degrees of freedom, of $\simeq 10$ pm.

Since the beam combiner produces overlapped images of two widely separated pointing directions it is necessary to develop a strategy allowing for clear identification of the field of view to which individual targets belong (This was not an issue for the Hipparcos mission, since the stars were observed accordingly to an input catalogue). A detailed analysis shows that this can be achieved by measuring the across-scan velocity and correlating it with the individual transit data. Targets in the two fields of view will in fact have across-scan velocities of opposite signs; also, targets belonging to the leading field of view will appear (after about half an hour) in the trailing field of view. These two criteria allow for unambiguous identification of the field of view to which each target belongs.

The multiplexing onto the same detector of the beams from the two lines of sight also helps to minimize the effects of perturbations in the optical configuration, because many of them will introduce correlated effects onto both fields of view. Such common mode disturbances are relevant only with respect to degradations of the image quality, thus increasing the random components of the measurement error, but their effect on the basic angle (a source of systematic error) appears at higher order. This residual differential systematic error is basically due to the lack of symmetry of some degrees of freedom with respect to the incoming beams; in a completely symmetrical system, the perturbation would only provide common mode terms. A possible implementation of a fully symmetric concept would require half-reflecting, full aperture mirrors in the beam combiner, but refractive optics are not appealing because of the chromatic errors introduced.

The use of a beam combiner thus succeeds in significantly relaxing the control requirements for the mirrors composing the interferometer. To preserve the stability of the basic angle between the two directions for both arms of the interferometer, the beam combiner requires the control of three out of its four flat mirrors. A tip-tilt device (analogous to the one implemented in the COSI testbed described above) appears to be adequate to the requirements of optimal control.

B.5 Focal Plane Issues

The fringe period of each stellar image ($\lambda/B = 63$ mas) must be covered by at least 3 CCD pixels, in order to ensure reasonable sampling. Each CCD pixel should therefore be $\simeq 21$ mas in size, which, given the 5.16 arcsec mm^{-1} focal plane scale, results in a ~ 4 μm physical size in the high resolution direction. As in the baseline design, pixels can be significantly larger in the across-scan directions, i.e. they can match the Airy disk, corresponding to $\simeq 20$ μm.

The focal plane is covered by a mosaic of 700 individual CCD devices, for a total of 1.4×10^5 pixels. CCD devices of different sizes are used, to follow the residual optical distortion (< 0.3 per cent over the field). The individual exposure times range from 0.75 to 1 s.

The requirements on the CCD detectors required for the interferometric optical configuration are challenging: although the specification on each relevant parameter is consistent with existing state-of-the-art components, no device currently hosts all of the desired features. The desired read-out noise level, RON = 3 e^-, at the maximum pixel rate specified in the current design, i.e. 100 kpixel s^{-1} is currently achievable, and the full-well capacity, scaled to the desired pixel geometry from the specifications of current components, is sufficient to provide non-saturated, uniform measurements over the main astrometric field up to magnitude $V \simeq 8$ mag. Anti-blooming would be needed to avoid problems from the brighter stars in the field of view.

Some Philips experimental devices (Peek et al. 1996) are available with smaller pixel size than the required 4 μm, but they feature rather poor quantum efficiency (peak $QE \simeq 20$ per cent at $\simeq 550$ nm), compared with the quantum efficiency values up to 80–90 per cent which have been achieved for backside illuminated, thinned CCDs with pixel size of 9–15 μm. A quantum efficiency of $\simeq 60$ per cent has been assumed in the performance assessment of the interferometric configuration, with the caveat that no device exists with the required high quantum efficiency, small pixel size, etc.

B.6 Telemetry Requirements

To properly sample the fringe pattern produced by the interferometer for each individual star a larger number of pixels need to be sent to ground than in the case of the baseline optical configuration. The telemetry rate produced by the interferometric configuration is thus correspondingly higher. A further increase is produced by the larger number of individual measurements produced during each field crossing (due to the larger field of view), so that the telemetry requirements of the interferometric configuration are much higher than for the baseline configuration. The raw (uncompressed) average data rate required to observe stars down to $V \simeq 18$ mag is $\simeq 15$ Mbps, increasing to $\simeq 60$ Mbps on the Galactic plane.

In a geostationary orbit the same type of phased array antenna as used in the baseline design can reach the required high bandwidth, and thus allow for transmission of the high data rate to ground. Also, the continuous visibility from a single ground station allows for the data to be continuously downloaded. From an L2 orbit such a high data rate would impose serious problems, requiring the use of a directional dish antenna, which would need to be mounted on a frictionless bearing mechanism. Also, while the high data rate appears to be achievable without significant problems from the geostationary orbit, the higher bit rate would impose a significant additional load on the data analysis system.

B.7 Geostationary Orbit

The Alenia-led study examined the implications of adopting a geostationary orbit for the GAIA spacecraft (rather than the L2 orbit examined in the baseline study). The main disadvantages of the geostationary orbit is a more disturbed thermal environment (Earth radiation and eclipses), larger tidal and magnetic Earth disturbances, and a larger fuel mass required for orbit injection. The advantages include much less stringent data communication requirements (higher bit rate possible, continuous visibility).

The study showed that the main problem in geostationary orbit is the presence of eclipses (a standard feature of geostationary orbits). These are however not a problem in terms of thermal control if the astronomical observations are performed also during the eclipse: the main driver in the thermal control is in fact not the external solar radiation, but rather the internal power dissipated by the electronics. Thus, if sufficiently large onboard batteries are included observations can be performed into eclipse. Thus the study showed that either orbit is suitable for GAIA, if necessary.

B.8 Performance of the Interferometric Configuration

The final performance of the interferometric configuration was assessed, using a detailed error budget for the various contributions of error sources such as the attitude and control system, the stability of the basic angle, the data reduction, etc. The noise contribution coming from the photon statistics was assessed assuming a quantum efficiency of $\simeq 60$ per cent for the CCD detectors (which, as explained in Section B.5 may involve significant development in this area). The accuracy figures relative to the contribution of the photon noise and of the optical quality are shown (for a K3III-type star) in Table B.1 for two different assumptions regarding the total noise for small-pixel CCD chips (i.e. 3 and 5 e^-). Individual transit and final mission accuracies are given. The systematic errors (to be added in quadrature to these numbers) are $\simeq 3$ μas with the active control strategy described above. Stars of $V \sim 7$ mag or brighter would saturate the CCD chips. A similar strategy as for the baseline configuration can be implemented for them.

The astrometric performance of the interferometric configuration for bright stars is thus very similar to that of the baseline configuration. It does not however allow to reach as faint as the baseline configuration, for two reasons: the smaller collecting area deriving from the use of a split pupil coupled with a beam combiner results in a lower photon budget, while the spreading of star light into fringes makes it much more sensitive to the effects of read-out noise.

B.9 Technical Conclusions

The results of the study summarized in the present appendix show that the nominal observational goal of GAIA (i.e. the determination of the five fundamental astrometric parameters to an accuracy of 10 μas for a star of $V \simeq 15$ mag) can be achieved both with an interferometric optical design as well as with the baseline design. In addition, the issues related to the use of a beam combiner, to the use of an active control system for the optics and to the placement of the mission in a geostationary orbit have been investigated in detail. These issues are to a large extent 'orthogonal', i.e. the active control concept could be equally well applied to the baseline configuration (and an

Table B.1: Transit-level and final accuracy for the interferometric configuration on a K3III star.

V magnitude	Accuracy [μarcsec]			
	RON= 3 e^-		RON= 5 e^-	
	Transit	Final	Transit	Final
8	2.2	0.44	2.3	0.46
9	3.6	0.72	4.3	0.86
10	5.4	1.08	6.3	1.26
11	8.4	1.68	8.6	1.72
12	14.8	2.96	13.1	2.62
13	22.1	4.42	25.7	5.14
14	37.0	7.40	30.8	6.16
15	55.1	11.02	68.8	13.76
16	86.8	17.36	115.5	23.10
17	139.2	27.84	187.8	37.56
18	265.4	53.08	418.3	83.66
19	532.9	106.58	718.6	143.72

interferometer could be implemented with a passive control concept). The interferometer mission can be implemented either in the geostationary orbit or at L2. The former is advantageous mainly for the data transmission, the latter for its quieter environment and small required Δv, and thus likely to be preferred for a mission which can cope with the data rate limitations.

The main disadvantages of an interferometric configuration (independently from its being actively or passively controlled) are the much higher data rates generated (which follows directly from the need to properly sample several fringes, as opposed to the need to sample a single Airy disk) and the need to use CCD chips with characteristics which are not currently available, thus requiring a dedicated and probably highly challenging detector development program. In terms of the science produced, the adoption of an interferometric configuration does not produce any additional science; on the contrary, it is limited to intrinsically brighter stars because of the smaller pupil and larger detector area used per star.

The adoption of an active control system (i.e. using metrology and closed loop controls) would allow the mission concept (either with a monolithic or an interferometric optical design) to be implemented even if the very accurate and stable thermal characteristics described in Section 4.3 for the baseline GAIA design would not be achievable. If the feasibility of a purely passive approach is fully demonstrated by the planned development program, it is clear that this will have the great benefit of simplicity and robustness, and thus be preferable. Should the requirements on dimensional stability turn out to be more demanding that currently foreseen, the metrology approach described above can be used to either simply monitor the evolution of the structure or, with the addition of the closed-loop controls, to actively control it.

B.10 Scientific Implications of the Interferometer Option

The primary performance difference between the baseline GAIA implementation and the interferometer option is an effective loss of one magnitude in performance near $V = 17$ mag. The scientific implications of this loss are significant, affecting most of the high-priority scientific GAIA goals. The most obvious is a reduction of the effective volume searched, by roughly a factor of ten. This seriously reduces the performance for studies of the luminosity function of low-luminosity stars, both low-mass and white dwarfs, with substantial impact on GAIA's ability to determine the metallicity-dependance of the local luminosity function. The loss in performance effectively removes the Kuiper belt objects, in the outer Solar System, from detection. The change in imaging performance substantially degrades the scientific interest in studies of galaxies and large scale structure, and removes study of the kinematics of the Local Group beyond the Milky Way from serious analysis. Perhaps most fundamentally, studies of the central regions and the outer parts of the Milky Way will be degraded, limiting scientific analysis of halo phase-space substructures, the mass distribution in the outer galaxy, and the star formation history of the inner galaxy. That is, GAIA's primary scientific goal, study of the structure and evolution of the Milky Way Galaxy, will be significantly degraded.

C Acronyms

ASIC Application specific integrated circuit
ASM Astrometric sky mapper
BBP Broad-band photometer
CDM Cold dark matter
CDS Centre de Données astronomiques de Strasbourg
CFRP Carbon-fibre reinforced plastic
DeNIS Deep Near Infrared Survey of the Southern Sky
DIVA Double Interferometer for Visual Astrometry
DSS Digitized Sky Survey
EIRP Earth-instantaneous radiated power
EM Engineering model
EQM Engineering and qualification model
FAME Full-Sky Astrometric Mapping Explorer
FEEP Field emission electric propulsion
FEM Finite element model
FM Flight model
FPA Focal plane assembly
HST Hubble Space Telescope
GTO Geostationary transfer orbit
ICRF International Celestial Reference Frame
ICRS International Celestial Reference System
IMF Initial mass function
KBO Kuiper Belt object
kbps kilobits per sec
GSC Guide Star Catalog (digitized Schmidt surveys for HST)
LAE Liquid apogee engine
LEOP Launch and early orbit phase
LMC Large Magellanic Cloud
MBP Medium-band photometer
mbps megabits per sec
MD Mechanical dummy
MLI Multi-layer insulation
MMS Matra Marconi Space
MTD Mechanical thermal dummy
MTF Modulation transfer function
OGSE Optical ground support equipment
OO Object oriented (programming)
OSR Optical solar reflector
PFM Protoflight model
PSF Point spread function
QE Quantum efficiency
ROM Rough order of magnitude
RF Radio frequency
RVS Radial velocity spectrograph
SAS Sun acquisition sensor
SEU Single event upset
SFR Star formation rate
SIM Space Interferometry Mission
SIMBAD Set of Identifications, Measurements and Bibliography for Astronomical Data (Object database at CDS)
SMA Shape memory alloy
SMC Small Magellanic Cloud
STM Structural thermal model
TDI Time-delayed integration
TERAPIX Pipeline processing of MegaPrime deep field images (Canada-France-Hawaii Telescope)
TTC Telemetry and telecommand
UML Unified Modelling Language
USNO A US Naval Observatory catalogue (all-sky digitization of photographic Schmidt surveys)
VLA Very Large Array (radiotelescope)
VLT Very Large Telescope (ESO)
VPU Video processing unit
WFE Wave front error
2MASS 2 Micron All Sky Survey (Univ. Massachusetts)

D Contributors and Acknowledgments

Science Advisory Group Members

(Chair: Michael Perryman, ESA)

K.S. de Boer
Sternwarte Univ Bonn,
Auf dem Hugel 71,
Bonn 53121,
Germany
Tel: +49 228 733656
Fax: +49 228 733672
deboer@astro.uni-bonn.de

G. Gilmore
University of Cambridge,
Institute of Astronomy,
Madingley Road,
Cambridge CB3 0HA,
United Kingdom
Tel: +44 1223 337506
Fax: +44 1223 337523
gil@ast.cam.ac.uk

E. Høg
Copenhagen University Observatory,
Juliane Maries Vej 30,
Copenhagen OE,
DK-2100,
Denmark
Tel: +45 35 325975
Fax: +45 35 325989
erik@astro.ku.dk

M.G. Lattanzi
Osservatorio Astronomico di Torino,
Strada Osservatorio 20,
Pino Torinese (TO),
I-10025 Italy
Tel: +39 011 8101923
Fax: +39 011 8101930
lattanzi@to.astro.it

L. Lindegren
Lund Observatory,
Box 43,
Lund 22100,
Sweden
Tel: +46 46 2227309
Fax: +46 46 2224614
lennart@astro.lu.se

X. Luri
Universitat de Barcelona,
Departament d'Astronomia i Meteorologia,
Avda Diagonal 647,
Barcelona 08028,
Spain
Tel: +34 93 4021128
Fax: +34 93 4021133
xluri@mizar.am.ub.es

F. Mignard
Observatoire de la Côte d'Azur,
CERGA,
Avenue Copernic,
Grasse 06130,
France
Tel: +33 493 405382
Fax: +33 493 405333
mignard@obs-azur.fr

P.T. de Zeeuw
Sterrewacht,
Jan Hendrik Oort Building,
Postbus 9513,
Leiden 2300 RA,
The Netherlands
Tel: +31 71 5275879
Fax: +31 71 5275743
tim@strw.leidenuniv.nl

ESA Science Advisory Group Members

O. Pace [Study Manager]
ESTEC (SCI-PF),
Postbus 299,
Keplerlaan 1,
Noordwijk 2200 AG,
The Netherlands
Tel: +31 71 5653458
Fax: +31 71 5655417
opace@estec.esa.nl

M.A.C. Perryman [Study Scientist]
ESTEC (SCI-SA),
Postbus 299,
Keplerlaan 1,
Noordwijk 2200 AG,
The Netherlands
Tel: +31 71 5653615
Fax: +31 71 5654690
mperryma@astro.estec.esa.nl

M. Hechler [ESOC Study Manager]
ESOC (TOS-OFA),
Robert-Bosch-Strasse 5,
Darmstadt, D-64293
Germany
Tel: +49 6151 2291
Fax: +49 6151 90495
mhechler@esoc.esa.de

S. Volonte
European Space Agency,
8-10, rue Mario-Nikis,
Paris Cedex 15,
75738,
France
Tel: +33 1 5369 7103
Fax: +33 1 5369 7690
svolonte@hq.esa.fr

F. Favata
ESTEC (SCI-SA),
Postbus 299,
Keplerlaan 1,
Noordwijk 2200 AG,
The Netherlands
Tel: +31 71 5654665
Fax: +31 71 5654690
ffavata@astro.estec.esa.nl

Other ESA Involvement:

Astrophysics Division: K.S. O'Flaherty, W. O'Mullane, M. Vannier, A. Colorado-McEvoy

The following D-TOS engineers have given support to GAIA study:

- J. Doutreleau (data processing)
- H. Evans (radiation)
- C. Gomez (thermal control)
- J. Gavira (structure)
- A. Martin Polegre (antenna)
- A. Mohammadzadeh (CCD CTE)
- R. Morgan-Owens (communications)
- P. Nieminen (radiation)
- C. Philippe (attitude and orbit control system)
- A. Popescu (basic angle monitoring)
- I. Zayer (CCD operations)

Science Working Group

(Co-Chairs: Tim de Zeeuw and Gerry Gilmore)

L. Aguilar (Ensenada), F. D'Antona (Rome), J. Bland-Hawthorn (AAO), A.G.A. Brown (ESO), C. Chiosi (ESO), C. Cutler (Potsdam), N.W. Evans (Oxford), K.C. Freeman (Stromlo), R. Kudritzki (Munich), H. Lamers (Utrecht), J. Laskar (Bureau des Longitudes), Y. Lebreton (Meudon), J. Luu (Leiden), S. Söderhjelm (Lund), M. Soffel (Dresden), C. Turon (Meudon), C. Waelkens (Leuven), V. Zappala (Torino)

Photometry Working Group

(Chair: Fabio Favata)

C. Bailer-Jones (Heidelberg), A. Bressan (Padova), M. David (Antwerp), P. Dubath (Geneva), D. Evans (Cambridge), A. Gimenez (Madrid), A. Gómez (Meudon), M. Grenon (Geneva), P. Jorden (EEV), J. Knude (Copenhagen), U. Munari (Padova), L. Pasquini (ESO), M. Romaniello (ESO), M. Spite (Meudon), C. Sterken (Brussels), V. Straizys (Vilnius), V. Vansevicius (Vilnius), W. Verschueren (Antwerp)

Instrument Working Group Members

(Chair: Lennart Lindegren)

F. Arenou (Meudon), U. Bastian (Heidelberg), A. Bijaoui (Nice), S. Casertano (STScI), M. Crézé (Vannes), M. Gai (Torino), F. van Leeuwen (Cambridge), C.D. Mackay (Cambridge), J. Pel (Groningen), R.S. Le Poole (Leiden), E. Schilbach (Potsdam), H. Schrijver (Utrecht), J. Torra (Barcelona), S. Wagner (Heidelberg), A. Wicenec (ESO), S. Worswick (Cambridge)

Members at Large

J. Andersen (Copenhagen), A. Baglin (Meudon), D. Barthès (Barcelona), S. Beaulieu (Cambridge), P. Bernacca (Padova), P. Bertelli (Padova), O. Bienaymé (Strasbourg), J. de Bruijne (Leiden), F. Crifo (Meudon), W. Dehnen (Heidelberg), S. van Eck (ESO), D. Egret (Strasbourg), L. Eyer (Leuven), C. Fabricius (Copenhagen), F. de Felice (Padova), F. Figueras (Barcelona), C. Flynn (Tuorla), M. Froeschlé (OCA-CERGA), G. Giampieri (London), J. Halbwachs (Strasbourg), A. Helmi (Leiden), G. Hensler (Kiel), D. Hestroffer (Torino), J. Holmberg (Lund), R. Hoogerwerf (Leiden), M. Irwin (Cambridge), C. Jordi (Barcelona), S. Kopeikin (Columbia), J. Kovalevsky (OCA-CERGA), S. Loiseau (JPL), J. Lub (Leiden), V. Makarov (Copenhagen), J. Mattei (AAVSO), M. Mennessier (Monpellier), G. Micela (Palermo), A. Milsztajn (Saclay), R. Neuhaeuser (Munich), I. Novikov (Copenhagen), P. Paillou (Bordeaux), F. Quist (Lund), I. Ribas (Barcelona), N. Robichon (Paris), S. Röser (Heidelberg), I.W. Roxburgh (London), R. Scholz (Potsdam), H. Sørensen (Copenhagen), C. de Vegt (Hamburg), W. Weiss (Vienna), P. Whitelock (SAAO), R. Wielen (Heidelberg), H. Zhao (Leiden), H. Zinnecker (Potsdam)

Coordinators and Contributors:

The section on the scientific goals of GAIA has been coordinated by P.T. de Zeeuw, G. Gilmore and M.A.C. Perryman based on inputs from:

- E. Høg, C. Fabricius, C. Flynn, J. Knude and V.V. Makarov (kinematic tracers);
- A. Vallenari, P. Bertelli, A. Bressan and C. Chiosi (stellar populations);
- R. Drimmel, M.G. Lattanzi, R. Smart and A. Spagna (spiral arms);
- N. Robichon (open clusters and orbits);
- W. Dehnen (kinematics of the Galactic disk, and age-velocity dispersion);
- K. Freeman (age-metallicity-velocity relation for the disk);
- F. Figueras and J. Torra (small-scale structure of the disk);
- R. Hoogerwerf and J. de Bruijne (OB associations and moving groups);
- G. Micela, F. D'Antona and F. Favata (Gould Belt);
- G. Gilmore (warps);
- K.S. de Boer (extinction and the interstellar medium);
- G. Gilmore (thick disk);
- A. Helmi (halo);
- C. Flynn and J. Holmberg (outer halo);
- G. Gilmore (tidal disruption of globular clusters);
- A. Helmi, H. Zhao and P.T. de Zeeuw (detecting halo streams);
- G. Gilmore, A. Helmi and P.T. de Zeeuw (Sagittarius dwarf);
- G. Gilmore (mass and extent of the Galaxy, and dark matter);
- N.W. Evans and M.I. Wilkinson (total mass of the Milky Way);
- P.T. de Zeeuw (wide binaries and local escape velocity);
- H. Zhao (tidal streams);
- J. Bland-Hawthorn (inter-stellar medium);
- G. Gilmore (star formation history of the Milky Way, and the local luminosity function);
- G. Micela, F. D'Antona and F. Favata (birth rate of stars in the solar neighbourhood);
- A.G.A. Brown (initial mass function in clusters and associations);
- M.A.C. Perryman, A. Baglin, Y. Lebreton and C. Turon (stellar astrophysics);
- F. Favata and M.A.C. Perryman (seismology);
- F. Favata (variability);
- D. Barthès (long-period variables);
- Y. Lebreton and A. Baglin (physics of stellar interiors);
- I. Mazzitelli (convection);
- H. Lamers (Wolf-Rayet stars);
- Y. Lebreton, A. Baglin and N. Robichon (stellar ages);
- C. Turon and N. Robichon (globular cluster distances);
- N.W. Evans and H. Zhao (microlensing);
- T.M. Eubanks ('forward' microlensing);
- S. Söderhjelm (binaries);
- F. Favata (eclipsing and contact binaries);
- M. van Kerkwijk (Be/X-ray and Be/pulsar binaries);
- G. Gilmore (low-mass stars);
- M. Haywood (isolated brown dwarfs);
- J. Isern, E. García-Berro and M. Hernanz (white dwarfs);
- S. Söderhjelm (brown dwarfs in binaries);
- M. Lattanzi, M.A.C. Perryman and J. Laskar (extra-solar planets);
- V. Zappala (minor planets);
- N.W. Evans (Trojans);
- J. Luu (Kuiper Belt objects);
- J. García Sánchez, R.A. Preston and P.R. Weissman (perturbations of the Oort Cloud);
- G. Gilmore (Magellanic Clouds and other satellites);
- J. Kovalevsky (reference frame);
- G. Gilmore (galaxies);
- E. Høg with C. Fabricius, V.V. Makarov, G. Tammann, A.D. Dolgov and B.E.J. Pagel (supernovae);
- S. Wagner (jets of active galactic nuclei);
- J. Kovalevsky, L. Lindegren and D.P. Schneider (quasars);
- T.M. Eubanks (quasar source stability);
- J.-F. Claeskens and J. Surdej (quasar microlensing and cosmology);
- U. Bastian (galactocentric acceleration);
- L. Lindegren and J. Kovalevsky (radio/optical reference frame);
- F. Mignard and M.A.C. Perryman (fundamental physics);
- T.M. Eubanks and F. Mignard (perihelion precession of solar system bodies);
- T.M. Eubanks (cosmological vorticity);
- C. Cutler (gravitational waves);
- A.G.A. Brown (curvature of stellar orbits and local density effects).

The section on the technical design has been compiled by L. Lindegren and M.A.C. Perryman (astrometry); U. Munari, F. Favata and M.A.C. Perryman (radial velocity); F. Favata, M. Grenon, E. Høg, C. Jordi, U. Munari, V. Straizys, and P.T. de Zeeuw (photometry). Sections on the payload and spacecraft have been compiled by M.A.C. Perryman and O. Pace on the basis of the MMS industrial study and final report. Details on the orbit, operations, and ground segment have been compiled by M. Hechler.

The mission performance assessment has been compiled by L. Lindegren (technical assumptions, basic astronomical data); F. Figueras (Galaxy model); N. Robichon (globular clusters); X. Luri and E. Høg (real-sky complexities); L. Lindegren (simulations); F. Arenou (image simulations); M.G. Lattanzi and M. Gai (centroiding experiments). Accuracies have been compiled by L. Lindegren (methodology and astrometric accuracy); L. Lindegren and X. Luri (folded with Galaxy model); C. Jordi and E. Høg (photometric accuracy); D. Katz, A. Gómez and F. Favata (radial velocity simulations); U. Munari (radial velocity assessment through observations acquired at Asiago).

Details of the data analysis have been compiled by L. Lindegren (modeling and estimation methods); F. de Felice and M.G. Lattanzi (relativistic formulation); C. Bailer-Jones (automated spectral parameterization); W. O'Mullane (implementation using OO methodology); D. Egret (management tools for very large catalogues). The section on management, science operations, and public outreach has been compiled by O. Pace, M.A.C. Perryman, G. Gilmore, and P.T. de Zeeuw.

The appendix on related ground and space projects, and their scientific complementarities, was compiled by F. Mignard and G. Gilmore. The appendix on the interferometric option was compiled by F. Favata and O. Pace on the basis of the Alenia industrial study and final report.

Many other members of the Working Groups provided detailed contributions and comments on various parts of the report. In addition to the specific credits above, the SAG would particularly like to thank C. Turon, K. O'Flaherty, H. Schrijver, F. Figueras, C. Jordi, I. Ribas, A. Gimenez, F. D'Antona.

Industrial Involvement

MMS Study Team:

Matra Marconi Space (F):
P. Mérat (Study Manager/System engineer)
F. Safa (Payload Manager)
J. Boyadjian (Optics)
O. Saint-Pé (Focal Plane)
B. Calvel (AIV)
L. Brouard (Mechanical/Thermal)
C. Page (Drawing Office)
F. Bouffard (Pointing and Dynamics)
A. Defendini (Basic Angle Monitoring)
G. Lahure (Programmatics)

EEV Ltd (UK):
D. Morris (CCD Technology)
P. Jerram (CCD Technology)

Alcatel Space (F):
G. Caille (Antenna Technology)

Alenia Study Team:

Alenia Aerospazio, Space Division:
S. Cesare (Study Manager)
G. Battistoni (Focal Plane)
W. Bogo (Spacecraft Design)
M. Cecconi (Optics)
F. Cometto (Attitude Control)
M. Fosson (Thermo-elastic Analysis)
P. Lanza (Telecommunications)
F. Pepe, L. Bonino (Active Optics Control System)

Istituto di Metrologia 'G. Colonnetti':
F. Bertinetto (Laboratory Tests)
M. Bisi (Active Optics Control System Prototype)
G. Mana (Metrology)

EICAS Automazione:
F. Donati (Astrometric Error Budget)
E. Canuto (Active Optics Control System Controller Design and Test)

Osservatorio Astronomico di Torino:
M. Gai, M.G. Lattanzi (Astrometric Instrument Design and Analysis)

Matra Marconi Space-F:
A. Defendini (DYNOPTIX Release 2.0)

Alcatel Space:
R. Krawczyk, E. Thomas (Optical Interferometer Design and Trade-off)
D. Dantes (CCD Technology Investigation)

Other Industrial Studies:

CCD CTE study: G. Hopkinson (Sira, Chislehurst)

Basic angle monitoring: B. Braam, B. Snijders, P. Kappelhof, R. Vink (TNO-TPD, Delft)

Part VI
References

References

Abbott D C 1982 *The Theory of Radiatively Driven Stellar Winds. II - The Line Acceleration Astrophys J* **259** 282–301

Acker A, Fresneau A, Pottasch S R & Jasnewicz G 1998 *A Sample of Planetary Nebulae Observed by Hipparcos Astron Astrophys* **337** 253–260

Alcock C, Akerlof C W, Allsman R A et al. 1993 *Possible Gravitational Microlensing of a Star in the Large Magellanic Cloud Nature* **365** 621–623

Alcock C, Allsman R A, Alves D et al. 1997 *The MACHO Project Large Magellanic Cloud Microlensing Results from the First Two Years and the Nature of the Galactic Dark Halo Astrophys J* **486** 697–726

Alcock C, Allsman R A, Axelrod T S et al. 1996 *The Macho Project LMC Variable Star Inventory. II. LMC RR Lyrae Stars Pulsational Characteristics and Indications of a Global Youth of the LMC Astron J* **111** 1146–1155

Allen B 1997 *The Stochastic Gravity-Wave Background: Sources and Detection* in J A Marck & J P Lasota, eds., *Relativistic Gravitation and Gravitational Radiation* p. 373 Cambridge University Press

Allen C & Santillán A 1991 *An Improved Model of the Galactic Mass Distribution for Orbit Computations Rev Mex Astron Astrof* **22** 255–263

Allen C W 1973 *Astrophysical Quantities* Athlone Press

Althaus L G & Benvenuto O G 1998 *Mon Not R Astr Soc* **296** 206–216

Altmann M & de Boer K S 199 *Kinematical Trends Among the Field Horizontal Branch Stars Astron Astrophys* **353** 135–146

Anspaugh B E 1996 *GaAs Solar Cell Radiation Handbook* Technical Report 96-9 JPL

Aparicio A, Bertelli G, Chiosi C & García-Pelayo J M 1990 *CCD UBVR Photometry of the Old Rich Open Cluster King 2: Comparison with Theoretical Models Astron Astrophys* **240** 262–288

Aparicio A & Gallart C 1995 *The Stellar Content of the Pegasus Dwarf Irregular Galaxy Astron J* **110** 2105–2119

Arenou F, Grenon M & Gómez A 1992 *A Tridimensional Model of the Galactic Interstellar Extinction Astron Astrophys* **258** 104–111

Asiain R, Figueras F, Torra J & Chen B 1999 *Detection of Moving Groups Among Early-Type Stars Astron Astrophys* **341** 427–436

Aubourg E, Bareyre P, Bréhin S et al. 1993 *Evidence for Gravitational Microlensing by Dark Objects in the Galactic Halo Nature* **365** 623–625

Babadzhanyants M & Belokon E T 1986 *Optical Manifestations of Superluminal Expansion of Components Belonging to the Millisecond Radio Structure in the Quasar 3C345 Astrophysics* **23** 639

Babusiaux C & Arenou F 1999 *A Possible Detection Algorithm for GAIA* Technical Report IWG-OPM-002 Observatoire de Paris-Meudon-Nancay

Bahcall J N, Flynn C & Gould A 1992 *Local Dark Matter from a Carefully Selected Sample Astrophys J* **389** 234–250

Bahcall J N & Soneira R M 1980 *Star Counts as an Indicator of Galactic Structure and Quasar Evolution Astrophys J* **238** L17–L20

Bailer-Jones C A L 1996 *Neural Network Classification of Stellar Spectra* Ph.D. thesis University of Cambridge

Bailer-Jones C A L, Irwin M, Gilmore G & von Hippel T 1997 *Physical Parametrization of Stellar Spectra - The Neural Network Approach Mon Not R Astr Soc* **292** 157–166

Bailer-Jones C A L, Irwin M & von Hippel T 1998 *Automated Classification of Stellar Spectra - II. Two-Dimensional Classification with Neural Networks and Principal Components Analysis Mon Not R Astr Soc* **298** 361–377

Bange J 1998 *An Estimation of the Mass of Asteroid 20-Massalia Derived from the Hipparcos Minor Planets Data Astron Astrophys* **340** L1–L4

Baraffe I, Chabrier G, Allard F & Hauschildt P H 1995 *New Evolutionary Tracks for Very Low Mass Stars Astrophys J* **446** L35

— 1998 *Evolutionary Models for Solar Metallicity Low-Mass Stars: Mass-Magnitude Relationships and Color-Magnitude Diagrams Astron Astrophys* **337** 403–412

Barbuy B, Renzini A, Ortolani S, Bica E & Guarnieri M D 1999 *High-Resolution Abundance Analysis of Two Individual Stars of the Bulge Globular Cluster NGC 6553 Astron Astrophys* **341** 539–546

Barnes J 1996 *Memory in Mergers Major and Minor* in H Morrison & A Sarajedini, eds., *The Formation of the Galactic Halo, Inside and Out* ASP Conf. Ser. 92 p. 415 Astronomical Society of the Pacific, San Francisco

Barthès D & Luri X 1999 *A Few Remarks on AGB Variable Stars and the GAIA Mission Baltic Astronomy*

8 285–290
Barthès D, Luri X, Alvarez R & Mennessier M O 1999 *Period-Luminosity-Colour Distribution and Classification of Galactic Oxygen-Rich LPVs. I. Luminosity Calibrations Astron Astrophys Suppl* **140** 55–67
Bastian U 1995 *100 Million Radial Velocities: How to Get Them* in M A C Perryman & F van Leeuwen, eds., *Future Possibilities for Astrometry in Space* ESA SP-379 pp. 165–168 ESA, Noordwijk
Batten A H 1985 *Radial Velocity Standards* in A G Davis Philip & D W Latham, eds., *Stellar Radial Velocities* IAU Coll. 88 p. 325 Davis Press, Schenectedy, New York
Battrick B 1994 in B Battrick, ed., *Horizon 2000 Plus: European Space Science in the 21st Century* ESA SP-1180 ESA, Noordwijk
Beaulieu J P, Grison P, Tobin W et al. 1995 *EROS Variable Stars: Fundamental-Mode and First-Overtone Cepheids in the Bar of the Large Magellanic Cloud Astron Astrophys* **303** 137–154
Bedding T R, Kjeldsen H & Christensen-Dalsgaard J 1998 *Hipparcos Parallaxes for η Boo and κ^1 Boo: Two Successes for Asteroseismology* in R A Donahue & J A Bookbinder, eds., *10th Cambridge Workshop on Cool Stars, Stellar Systems and the Sun* ASP Conf. Ser. 154 p. 741 Astronomical Society of the Pacific, San Francisco
Beers T, Rossi S, Norris J, Ryan S & Shefler T 1999 *Estimation of Stellar Metal Abundance. II. A Recalibration of the CA II K Technique, and the Autocorrelation Function Method Astron J* **117** 981–1009
Beers T C, Preston G W & Shectman S A 1992 *A Search for Stars of Very Low Metal Abundance. II Astron J* **103** 1987–2034
Bell E F & de Jong R S 2000 *The Stellar Populations of Spiral Galaxies Mon Not R Astr Soc* **in press**
Bergeat J, Knapik A & Rutily B 1998 *The PL Relation of Galactic Carbon LPVs. The Distance Modulus to LMC Astron Astrophys* **332** L53–L56
Bersier D, Alcock C, Allsmann R D et al. 1998 *LMC Cepheids in the MACHO Database: Constraints on the Star Formation Rate and on Stellar Models* in P A Bradley & J A Guzik, eds., *A Half Century of Stellar Pulsation Interpretation* ASP Conf. Ser. 135 p. 24 Astronomical Society of the Pacific, San Francisco
Bertelli G, Betto R, Chiosi C, Bressan A & Nasi E 1990 *Theoretical Isochrones with Convective Overshoot Astron Astrophys Suppl* **85** 845–853
Bertelli G, Bressan A, Chiosi C, Ng Y K & Ortolani S 1995 *The Galactic Structure Towards the Galactic Centre. II. A S1tudy of the Fields Near the Clusters NGC 6603, Lynga 7 and Terzan 1 Astron Astrophys* **301** 381–395
Bertin E & Arnouts S 1996 *SExtractor: Software for Source Extraction Astron Astrophys* **117** 393–404
Bertolami O, Mourão J M & Pérez-Mercader J 1993 *Phys Lett B* **311** 27
Bertotti B & Giampieri G 1992 *Relativistic Effects for Doppler Measurements near Solar Conjunction Class Quantum Grav* **9** 777–793
— 1998 *Solar Coronal Plasma in Doppler Measurements Solar Phys* **178** 85–107
Bessell M S 1990 *UBVRI Passbands Publ Astr Soc Pac* **102** 1181–1199
Bienaymé O 1999 *The Local Stellar Velocity Distribution of the Galaxy. Galactic Structure and Potential Astron Astrophys* **341** 86–97
Bienaymé O, Robin A C & Crézé M 1987 *The Mass Density in our Galaxy Astron Astrophys* **180** 94–110
Binney J J 1992 *Warps Ann Rev Astron Astrophys* **30** 51–74
Binney J J, Dehnen W, Houk N, Murray C A & Penston M J 1997a *The Kinematics of Main Sequence Stars from Hipparcos Data* in M A C Perryman & P L Bernacca, eds., *Hipparcos Venice 97* ESA SP-402 pp. 473–478 ESA, Noordwijk
Binney J J, Gerhard O & Spergel D 1997b *The Photometric Structure of the Inner Galaxy Mon Not R Astr Soc* **288** 365–374
Binney J J & Tremaine S 1987 *Galactic Dynamics* Princeton University Press
Bishop C M 1995 *Neural Networks for Pattern Recognition* Oxford University Press
Bissantz N, Englmaier P, Binney J J & Gerhard O 1997 *The Microlensing Optical Depth of the COBE Bulge Mon Not R Astr Soc* **289** 651–659
Blaauw A 1961 *On the Origin of the O and B-type Stars with High Velocities (the 'Runaway'), and Some Related Problems BAN* **15** 265–290
Black D C 1997 *Possible Observational Criteria for Distinguishing Brown Dwarfs from Planets Astrophys J* **490** L171–L174
Bland-Hawthorn J & Maloney P R 1999 *The Escape of Ionizing Photons from the Galaxy Astrophys J* **510** L33–36
Bland-Hawthorn J, Veilleux S, Cecil G N, Putman M E, Gibson B K & Maloney P R 1998 *The Smith Cloud: HI Associated with the SGR Dwarf? Mon Not R Astr Soc* **299** 611–624
Blinnikov S I & Dunina-Barkovskaya N V 1994 *The Cooling of Hot White Dwarfs - a Theory with Non-*

Standard Weak Interactions and a Comparison with Observations Mon Not R Astr Soc **266** 289–304
Boden A, Unwin S & Shao M 1997 *Global Astrometry with the Space Interferometry Mission* in M A C Perryman & P L Bernacca, eds., *Hipparcos Venice 97* ESA SP-402 pp. 789–794 ESA, Noordwijk
Boden A F, Shao M & van Buren D 1998 *Astrometric Observations of MACHO Gravitational Microlensing Astrophys J* **502** 538–549
Bois E & Girard J F 1998 *Impact of the Quadrupole Moment of the Sun on the Dynamics of the Earth-Moon system* in J Henrard, S Ferraz-Mello & J J Binney, eds., *Impact of Modern Dynamics in Astronomy* IAU Coll. 172
Bonnell I A & Davies M B 1998 *Mass Segregation in Young Stellar Clusters Mon Not R Astr Soc* **295** 691–698
Boss A P 1998a *Astrometric Signatures of Giant-Planet Formation Nature* **393** 141–143
— 1998b *Twin Planetary Systems in Embryo Nature* **395** 320–321
Bouvier J 1994 *The Rotational Evolution of Low-Mass Pre-Main Sequence Stars* in J P Caillault, ed., *8th Cambridge Workshop on Cool Stars, Stellar Systems and the Sun* ASP Conf. Ser. 64 p. 151 Astronomical Society of the Pacific, San Francisco
Bouvier J, Stauffer J R, Martin E L, Barrado y Navascues D, Wallace B & Bejar V J S 1998 *Brown Dwarfs and Very Low-Mass Stars in the Pleiades Cluster: a Deep Wide-Field Imaging Survey Astron Astrophys* **336** 490–502
Braginsky V B & Grishchuk L P 1985 *Soviet Physics JETP* **62** 427
Braginsky V B, Kardashev N S, Polnarev A G & Novikov I D 1990 *Nuovo Cim B* **105** 1141
Brans C & Dicke R H 1961 *Phys Rev* **124** 925
Bressan A, Fagotto F, Bertelli G & Chiosi C 1993 *Evolutionary Sequences of Stellar Models with New Radiative Opacities. II. Z = 0.02 Astron Astrophys Suppl* **100** 647–664
Brown A G A, Dekker G & de Zeeuw P T 1997 *Kinematic Ages of OB Associations Mon Not R Astr Soc* **285** 479–492
Brown A G A, Perryman M, Kovalevsky J et al. 1997 *The Hyades: Distance, Structure and Dynamics* in M A C Perryman & P L Bernacca, eds., *Hipparcos Venice 97* ESA SP-402 pp. 681–686 ESA, Noordwijk
Brumberg V A 1991 *Essential Relativistic Celestial Mechanics* Adam Hilger
Burton W B 1998 *The Structure of our Galaxy Derived from Observations of Neutral Hydrogen* in G L Verschuur & K I Kellermann, eds., *Galactic and Extragalactic Radio Astronomy* pp. 295–354 Springer-Verlag, Berlin
Buscher D et al. 1995 *Appl Opt* **34** 1081
Butler R P, Marcy G W, Fischer D A et al. 1999 *Evidence for Multiple Companions to υ Andromedae Astrophys J* **526** 916–927
Cabrera B, Clarke R M, Colling P et al. 1998 *Appl Phys Let* **73(6)** 735
Camenzind M & Krockenberger M 1992 *The Lighthouse Effect of Relativistic Jets in Blazars - A Geometric Origin of Intraday Variability Astron Astrophys* **255** 59–62
Canuto V M 1992 *Turbulent Convection with Overshooting - Reynolds Stress Approach Astrophys J* **392** 218–232
Canuto V M & Mazzitelli I 1991 *Stellar Turbulent Convection - A New Model and Applications Astrophys J* **370** 295–311
Carney B, Latham D & Laird J 1989 *A Survey of Proper-Motion Stars. VIII - On the Galaxy's Third Population Astron J* **97** 423–430
Carney B, Latham D, Laird J & Aguilar L 1994 *A Survey of Proper Motion Stars. 12: An Expanded Sample Astron J* **107** 2240–2289
Carney B W, Laird J B & Latham D W 1988 *A Survey of Proper Motion Stars V. Extreme-Velocity Stars and the Local Galactic Escape Velocity Astron Astrophys* **96** 560–566
Carney B W, Laird J B, Latham D W & Kurucz R L 1987 *A Survey of Proper-Motion Stars. II. Extracting Metallicities from High-Resolution, Low S/N Spectra Astron J* **94**(4) 1066–1076
Carollo C M 1999 *The Centers of Early- to Intermediate-Type Spiral Galaxies: A Structural Analysis Astrophys J* **523** 566–574
Carr B 1994 *Baryonic Dark Matter Ann Rev Astron Astrophys* **32** 531–590
Castor J I, Abbott D C & Klein R I 1975 *Radiation-Driven Winds in Of Stars Astrophys J* **195** 157–174
Cayrel R, Castelli F & C. V V 1999 in M Froeschlé & F Mignard, eds., *Atelier GAIA: l'Astrométrie pour l'Astrophysique et l'Astrodinamique* p. 73 Observatoire de la Côte d'Azure
Cayrel R, Lebreton Y, Perrin M N & Turon C 1997 *The HR Diagram in the Plane* $\log(T_{\rm eff})$*,* $M_{\rm bol}$ *of Population II Stars with Hipparcos Parallaxes* in M A C Perryman & P L Bernacca, eds., *Hipparcos Venice 97* ESA SP-402 pp. 219–224 ESA, Noordwijk
Chaboyer B 1999 *Globular Cluster Distance Determinations* in A Heck & F Caputo, eds., *Post-Hipparcos*

Cosmic Candles Astrophys. Sp. Sci. 237 p. 111 Kluwer, Dordrecht
Chaboyer B, Demarque P, Kernan P J & Krauss L M 1998 *The Age of Globular Clusters in Light of Hipparcos: Resolving the Age Problem? Astrophys J* **494** 96–110
Chaboyer B, Demarque P & Sarajedini A 1996 *Globular Cluster Ages and the Formation of the Galactic Halo Astrophys J* **459** 558–569
Chaboyer B & Kim Y C 1995 *The OPAL Equation of State and Low-Metallicity Isochrones Astrophys J* **454** 767–773
Chabrier G & Baraffe I 1995 *CM Draconis and YY Geminorum: Agreement Between Theory and Observation Astrophys J* **451** L29–L32
— 1997 *Structure and Evolution of Low-Mass Stars Astrophys J* **327** 1039–1053
Charbonneau D, Brown T M, Latham D W & Mayor M 2000 *Detection of Planetary Transits Across a Sun-Like Star Astrophys J* **529** L45–L48
Charbonneau D, Noyes R W, Korzennik S G, Nisenson P, Jha S, Vogt S S & Kibrick R I 1999 *An Upper Limit on the Reflected Light from the Planet Orbiting the Star τ Bootis Astrophys J* **522** L145–L148
Chen B 1997 *Comparisons of a Galactic Kinematic Model with Two Proper-Motion Surveys in the Vicinity of the North Galactic Pole Astrophys J* **491** 181–190
Chen B, Figueras F, Torra J, Jordi C, Luri X & Galadí-Enríquez D 1999a *Constraining Galactic Structure Parameters from a New Extinction Model and Four Star Count Samples Astron Astrophys* **352** 459–468
Chen B, Vergely J L, Egret D, Torra J, Figueras F & Jordi C 1999b *The Spatial Resolution of the Extinction Structure from GAIA Baltic Astronomy* **8** 195–201
Chereul E, Crézé M & Bienaymé O 1999 *The Distribution of Nearby Stars in Phase Space Mapped by Hipparcos. Clustering and Streaming Among A-F Type Stars Astron Astrophys* **135** 5–28
Chieffi A, Straniero O & Solaris M 1995 *Calibration of Stellar Models Astrophys J* **445** L39–L42
Chiosi C, Bertelli G, Meylan G & Ortolani S 1989 *Globular Clusters in the Large Magellanic Cloud - NGC 1866, a Test for Convective Overshoot Astron Astrophys* **219** 167–191
Christensen-Dalsgaard J, Bedding T R & Kjeldsen H 1995 *Modeling Solar-Like Oscillation in η Bootis Astrophys J* **443** L29–L32
Cochran W D & Hatzes A P 1994 *A High Precision Radial-Velocity Survey for Other Planetary Systems Astrophys Space Sci* **212** 281–291
Colavita M M & Shao M 1994 *Indirect Planet Detection with Ground-Based Long-Baseline Interferometry Astrophys Space Sci* **212** 385–390
Collier Cameron A, Horne K, Penny A & James D 1999 *Probable Detection of Starlight Reflected from the Giant Planet Orbiting τ Bootis Nature* **402** 751–755
Colorado McEvoy A 1999 *How Many Planets Will GAIA Detect?* Technical Report SAG-AC-001 ESTEC
Comerón F 1999 *Vertical Motion and Expansion of the Gould Belt Astron Astrophys* **351** 506–518
Comerón F, Torra J & Figueras F 1998a *Kinematic Signatures of Violent Formation of Galactic OB A1ssociations from Hipparcos Measurements Astron Astrophys* **330** 975–989
Comerón F, Torra J & Gómez A E 1998b *Kinematic Signatures of Violent Formation of Galactic OB Associations from Hipparcos Measurements Astron Astrophys* **330** 975–989
Crézé M, Chereul E, Bienaymé O & Pichon C 1998 *The Distribution of Nearby Stars in Phase Space Mapped by Hipparcos. I. The Potential Well and Local Dynamical Mass Astron Astrophys* **329** 920–936
Dalcanton J, Spergel D & Summers F 1997 *The Formation of Disk Galaxies Astrophys J* **482** 659–676
Dale C et al. 1993 *Displacement Damage Effects in Mixed Particle Environments for Shielded Spacecraft IEEE Trans Nuc Sci* **40(6)** 1628
Damour T & Esposito-Farèse G 1998 *Phys Rev D* **58** 42001
Damour T, Gibbons G W & Taylor J H 1988 *Limits on the Variability of G Using Binary-Pulsar Data Phys Rev Lett* **61** 1151–1154
Damour T & Nordtvedt K 1993a *General Relativity as a Cosmological Attractor of Tensor-Scalar Theories Phys Rev Lett* **70(15)** 2217–2219
— 1993b *Tensor-Scalar Cosmological Models and their Relaxation Toward General Relativity Phys Rev D* **48** 3436–3450
Danner R & Unwin S 1999 *SIM: Taking the Measure of the Universe* NASA/JPL
D'Antona F 1998 *The Conversion from Luminosity Functions to Mass Functions* in G F Gilmore & D Howell, eds., *The Stellar Initial Mass Function* ASP Conf. Ser. 142 p. 157 Astronomical Society of the Pacific, San Francisco
D'Antona F, Caloi V & Mazzitelli I 1997 *The Universe and Globular Clusters: an Age Conflict? Astrophys J* **477** 519
D'Antona F & Mazzitelli I 1997 *Evolution of Low Mass Stars* in G Micela, R Pallavicini & S Sciortino, eds., *Cool Stars in Clusters and Associations: Magnetic Activity and Age Indicators* volume 68 of *Mem.*

Soc. Astron. Ital. p. 807
Dar A 1992 *Nuc Phys B* **28A** 321
de Bruijne J H J 1999 *A Refurbished Convergent-Point Method for Finding Moving Groups in the Hipparcos Catalogue Mon Not R Astr Soc* **306** 381–393
de Bruijne J H J & Hoogerwerf R 2000 *Astron Astrophys* **in press**
de Felice F, Lattanzi M, Vecchiato A & Bernacca P L 1998 *General Relativistic Satellite Astrometry. I. A Non-Perturbative Approach to Data Reduction Astron Astrophys* **332** 1133–1141
De Marchi G & Paresce F 1995 *Low Mass Stars in Globular Clusters. II. The Mass Function of M15 Astron Astrophys* **304** 202–210
de Zeeuw P T 1999 *Conference Summary* in B K Gibson, T S Axelrod & M E Putman, eds., *The Third Stromlo Symposium: The Galactic Halo: Bright Stars and Dark Matter* ASP Conf. Ser. 165 pp. 515–526 Astronomical Society of the Pacific, San Francisco
de Zeeuw P T, Hoogerwerf R, de Bruijne J H J, Brown A G A & Blaauw A 1999 *A Hipparcos Census of the Nearby OB Associations Astron J* **117** 354–399
Dehnen W 1998 *The Distribution of Nearby Stars in Velocity Space Inferred from Hipparcos Data Astron J* **115** 2384–2396
— 1999 *The Pattern Speed of the Galactic Bar Astrophys J* **524** L35–L38
Dehnen W & Binney J J 1998 *Local Stellar Kinematics from Hipparcos Data Mon Not R Astr Soc* **298** 387–394
Demarque P, Krauss L M, Guenther D B & Nydam D 1994 *The Sun as a Probe of Varying G Astrophys J* **437** 870–878
Dimeo T, Valls-Gabaud D & Kerins E J 1997 *The TYCHO Database as a Control Microlensing Experiment* in R Ferlet, J P Maillard & B Raban, eds., *Proc. 12th IAP Astrophysics Coll. 1996: Variable Stars and Astrophysical Returns from Microlensing Surveys* pp. 69–73 Editions Frontières
Dirac P A M 1938 *Proc R Soc London Ser A* **165** 199
Dravins D, Lindegren L & Madsen S 1999 *Astrometric Radial Velocities: I. Non-Spectroscopic Methods for Measuring Stellar Radial Velocity Astron Astrophys* **348** 1040–1051
Drimmel R, Smart R L & Lattanzi M G 1997 *A New Tool for Studying Galactic Structure* in M A C Perryman & P L Bernacca, eds., *Hipparcos Venice 97* ESA SP-402 pp. 607–610 ESA, Noordwijk
Drinkwater M J, Webb J K, Barrow J D & Flambaum V V 1998 *New Limits on the Possible Variation of Physical Constants Mon Not R Astr Soc* **295** 457–462
Duquennoy A & Mayor M 1991 *Multiplicity Among Solar Type Stars in the Solar Neighbourhood. II Distribution of the Orbital Elements in an Unbiased Sample Astron Astrophys* **248** 485–524
Durrell P R & Harris W E 1993 *A Color-Magnitude Study of the Globular Cluster M15 Astron J* **105** 1420–1440
Durrer R 1994 *Light Deflection in Perturbed Friedmann Universes Phys Rev Lett* **72** 3301–3304
Dwek E, Arendt R G, Hauser M G, Kelsall T, Lisse C M, Moseley S H, Silverberg R F, Sodroski T J & Weiland J L 1995 *Morphology, Near-Infrared Luminosity, and Mass of the Galactic Bulge from COBE DIRBE Observations Astrophys J* **445** 716–730
Dziembowski W A & Pamyatnykh A A 1991 *A Potential Asteroseismological Test for Convective Overshooting Theories Astron Astrophys* **248** L11–L14
ECSS 1997 *System Engineering: Space Environment* Technical Report ECSS-E-10-04 European Cooperation for Space Standardisation (ECSS)
Eddington A S 1914 *Stellar Movements and the Structure of the Universe* Macmillan, London
Edvardsson B, Andersen J, Gustafsson B, Lambert D L, Nissen P E & Tomkin J 1993 *The Chemical Evolution of the Galactic Disk: I. Analysis and Results Astron Astrophys* **275** 101–152
Efremov Y N 1988 *Astrophysics and Space Physics Reviews* **7** 105
— 1993 in J Franco et al., eds., *Star Formation, Galaxies and Interstellar Medium* p. 61 Cambridge University Press
— 1994 *Star Complexes and the Structure of the Galactic Disk* in I R King, ed., *Physics of the Gaseous and Stellar Disks of the Galaxy* ASP Conf. Ser. 66 p. 167 Astronomical Society of the Pacific, San Francisco
— 1995 *Star Complexes and Associations: Fundamental and Elementary Cells of Star Formation Astron J* **110** 2757–2770
Efremov Y N & Chermin A D 1995 *Star Complexes and their Rotation Astron Astrophys* **293** 69–74
Eggen O J 1965 in A Blaauw & M Schmidt, eds., *Stars and Stellar Systems, Vol. 5, Galactic Structure* p. 111 University of Chicago Press
— 1994 in L V Morrison & G Gilmore, eds., *Galactic and Solar System Optical Astronomy* p. 191 Cambridge University Press

Eggen O J, Lynden-Bell D & Sandage A R 1962 *Evidence from the Motions of Old Stars that the Galaxy Collapsed Astrophys J* **136** 748–766

Egret D & Heck A 1999 *Harmonizing Cosmic Distance Scales in a Post-Hipparcos Era* ASP Conf. Ser. 167 Astronomical Society of the Pacific, San Francisco

Elmegreen B G 1999 *The Stellar Initial Mass Function from Random Sampling in Hierarchical Clouds. II. Statistical Fluctuations and a Mass Dependence for Starbirth Positions and Times Astrophys J* **515** 323–336

Elmegreen B G & Elmegreen D M 1983 *Regular Sof HII Regions and Superclouds in Spiral Galaxies - Clues to the Origin of Cloudy Structure Mon Not R Astr Soc* **203** 31–45

Elson R, Gilmore G, Santiago B & Casertano S 1996 *HST Observations of the Stellar Population of the Globular Cluster Omega Cen Astron J* **110** 682–692

Elson R, Tanvir N, Gilmore G, Johnson R & Beaulieu S 1998 in Y H Chu, N B Suntzeff, J E Hesser & D A Bohlender, eds., *New Views of the Magellanic Clouds* IAU Symp. 190

Erez G & Rosen N 1959 *Bull Res Council of Israel* **8F** 47

Eriksson H & Penker M 1998 *UML Toolkit* Technical report Wiley

Evans N W & Tabachnik S 2000 **TBC** TBC

Eyer L & Cuypers J 2000 *Predictions on the Number of Variable Stars for the GAIA Space Mission* in L Szabados & D W Kurtz, eds., *The Impact of Large-Scale Surveys on Pulsating Star Research* ASP Conf. Ser. TBC Astronomical Society of the Pacific, San Francisco

Faber S M, Tremaine S, Ajhar E A et al. 1997 *The Centers of Early-Type Galaxies with HST. IV. Central Parameters Astron J* **114** 1771–1796

Fabricius C & Høg E 1998a *Precision of FSM Photometry* Technical Report SAG-CUO-034 Copenhagen University Observatory

— 1998b *Precision of FSM Photometry II* Technical Report SAG-CUO-037 Copenhagen University Observatory

— 1998c *Precision of Photometry in all GAIA Telescopes* Technical Report SAG-CUO-033 Copenhagen University Observatory

Fagotto F, Bressan A, Bertelli G & Chiosi C 1994 *Evolutionary Sequences of Stellar Models with New Radiative Opacities. IV. Z = 0.004 and Z = 0.008 Astron Astrophys Suppl* **105** 29–38

Fakir R 1994a *Phys Rev D* **50** 3795

— 1994b *Gravity Wave Watching Astrophys J* **426** 74–78

Fall S M & Efstathiou G 1980 *Formation and Rotation of Disc Galaxies with Haloes Mon Not R Astr Soc* **193** 189–206

Fall S M & Rees M J 1985 *A Theory for the Origin of Globular Clusters Astrophys J* **298** 18–26

Fan X, Strauss M A, Annis J et al. 1999 *Spectroscopy of Quasar Candidates from SDSS Commissioning Data* in S Holt & E Smith, eds., *After the Dark Ages: When Galaxies were Young (the Universe at* $2 < z < 5$*)* 9th Annual October Astrophysics Conference in Maryland p. 282 American Institute of Physics Press

Favata F 1998 *Spectrophotometry with GAIA* Technical Report SAG-PWG-006 ESTEC

Favata F, Barbera M, Micela G & Sciortino S 1993 *A Search for Yellow Young Disk Population Stars among EMSS Stellar X-Ray Sources by Means of Lithium Abundance Determination Astron Astrophys* **277** 428–438

Favata F, Micela G, Sciortino S & D'Antona F 1998 *The Evolutionary Status of Activity-Selected Solar-Type Stars and of T-Tauri Stars as Derived from Hipparcos Parallaxes: Evidence for a Long-Lived T-Tauri Disk Astron Astrophys* **335** 218–226

Favata F & Perryman M A C 1995 *Parallel Acquisition of Radial Velocities and Metallicities for a GAIA-Type Mission* in M A C Perryman & F van Leeuwen, eds., *Future Possibilities for Astrometry in Space* ESA SP-379 pp. 153–164 ESA, Noordwijk

Favata F, Sciortino S, Rosner R & Vaiana G S 1988 *The Stellar Composition of X-ray Surveys from the Einstein Observatory Astrophys J* **324** 1010–1015

Feast M W & Catchpole R M 1997 *The Cepheid Period-Luminosity Zero-Point from Hipparcos Trigonometrical Parallaxes Mon Not R Astr Soc* **286** L1–L5

Feissel M & Mignard F 1998 *The Adoption of ICRS on 1 January 1998: Meaning and Consequences Astron Astrophys* **331** L33–L36

Fernandes J, Lebreton Y, Baglin A & Morel P 1998 *Fundamental Stellar Parameters for Nearby Visual Binary Stars: η Cas, XI Boo, 70 OPH and 85 Peg. Helium Abundance, Age and Mixing Length Parameter for Low-Mass Stars Astron Astrophys* **338** 455–464

Fernley J, Barnes T G, Skillen I et al. 1998 *The Absolute Magnitudes of RR Lyraes from Hipparcos Parallaxes and Proper Motions Astron Astrophys* **330** 515–520

Feynman J, Armstrong T P, L. D G & Silverman S 1990 *New Interplanetary Proton Fluence Model J Spacecraft* **27** 403–410

Figueras F, García-Berro E, Torra J, Jordi C, Luri X, Torres S & Chen B 1999 *GAIA Science Output: White Dwarfs Baltic Astronomy* **8** 291–300

Figueras F, Gómez A E, Asiain R et al. 1997 *Identification of Moving Groups in a Sample of Early-Type Main-Sequence Stars* in M A C Perryman & P L Bernacca, eds., *Hipparcos Venice 97* ESA SP-402 pp. 519–524 ESA, Noordwijk

Fleming T A 1998 *A New Sample of Nearby M Dwarfs Discovered by ROSAT Astrophys J* **504** 461–467

Flynn C & Morrison H L 1990 *An Automated Technique for Locating Metal-Deficient Red Giants from Objective-Prism Spectra Astron J* **100** 1181–1190

Freedman W L, Madore B F, Mould J R, Ferrarese L, Hill R, Kennicutt R. C. J, Saha A, Stetson P B, Graham J A, Ford H, Hoessel J G, Huchra J, Hughes S M & Illingworth G D 1994 *Distance to the Virgo Cluster Galaxy M100 from Hubble Space Telescope Observations of Cepheids Nature* **371** 757

Freeman K C 1993 *The Dynamical History of the Galactic Disk* in S Majewksi, ed., *Galaxy Evolution: The Milky Way Perspective* ASP Conf. Ser. 49 p. 125 Astronomical Society of the Pacific, San Francisco

Frémat Y, Houziaux L & Andrillat Y 1996 *Higher Paschen Lines in the Spectra of Early-Type Stars Mon Not R Astr Soc* **279** 25–31

Freudenreich H T 1998 *A COBE Model of the Galactic Bar and Disk Astrophys J* **492** 495–519

Fry A M & Carney B W 1997 *Chemical Abundances of Galactic Cepheid Variables that Calibrate the P-L Relation Astron J* **113** 1073–1087

Fuchs B, Jahreiß H & Wielen R 1998 *Kinematics of Nearby Subdwarfs* in M Spite & F Crifo, eds., *Galaxy Evolution: Connecting the Distant Universe with the Local Fossil Record* p. 20 Kluwer, Dordrecht

Fukugita M & Turner E L 1991 *Gravitational Lensing Frequencies: Galaxy Cross-Sections and Selection Effects Mon Not R Astr Soc* **253** 99–106

Gabler R, Gabler A, Kudritzki R P, Puls J & Pauldrach A W A 1989 *Unified NLTE Model Atmospheres Including Spherical Extension and Stellar Winds - Method and First Results Astron Astrophys* **226** 162–182

Gai M, Casertano S, Carollo D & Lattanzi M G 1998 *Location Estimators for Interferometric Fringes Publ Astr Soc Pac* **110** 848–862

García-Berro E, Hernanz M, Isern J & Mochkovitch R 1995 *The Rate of Change of the Gravitational Constant and the Cooling of White Dwarfs Mon Not R Astr Soc* **277** 801–810

García-Berro E, Torres S, Isern J & Burkert A 1999 *Monte Carlo Simulations of the Disc White Dwarf Population Mon Not R Astr Soc* **302** 173–188

García-Sánchez J, Preston R A, Jones D L, Weissman P R, Lestrade J F, Latham D W & Stefanik R P 1999 *Stellar Encounters with the Oort Cloud Based on Hipparcos Data Astron J* **117** 1042–1055

Gatewood G D 1987 *The Multichannel Astrometric Photometer and Atmospheric Limitations in the Measurement of Relative Positions Astron J* **94** 213–214

Gendreau K C, Prigozhin G Y, Huang R K & Bautz M W 1995 *A Technique to Measure Trap Characteristics in CCDs Using X-Rays IEEE Trans Electronic Devices* **42(11)** 1912

Genzel R 1998 *The Dark Mass in the Center of the Milky Way Bull Amer Astron Soc* **193** 6201

Georgelin Y M & Georgelin Y P 1976 *The Spiral Structure of our Galaxy Determined from HII Regions Astron Astrophys* **49** 57–79

Gerhard O E 1996 in L Blitz & P Teuben, eds., *Unsolved Problems of the Milky Way, IAU Symp. 169* p. 79 Kluwer, Dordrecht

Ghez A M, Klein B L, Morris M & Becklin E E 1998 *High Proper-Motion Stars in the Vicinity of Sagittarius A*: Evidence for a Supermassive Black Hole at the Center of Our Galaxy Astrophys J* **509** 678–686

Giampieri G 1996 *Relativity Experiments in the Solar System* in *Proc. 11th Italian Conference on General Relativity and Gravitational Physics* World Scientific

Gilmore G 1984 *New Light on Faint Stars. VI - Structure and Evolution of the Galactic Spheroid Mon Not R Astr Soc* **207** 223–240

Gilmore G & Reid I N 1983 *New Light on Faint Stars. III - Galactic Structure Towards the South Pole and the Galactic Thick Disc Mon Not R Astr Soc* **202** 1025–1047

Gilmore G & Wyse R F G 1985 *The Abundance Distribution in the Inner Spheroid Astron J* **90** 2015–2026

— 1998 *Element Ratios and the Formation of the Stellar Halo Astron J* **116** 748–753

Gilmore G, Wyse R F G & Jones J B 1995 *A Determination of the Thick Disk Chemical Abundance Distribution: Implications for Galaxy Evolution Astron J* **109** 1095–1111

Gilmore G, Wyse R F G & Kuijken K 1989 *Kinematics, Chemistry, and Structure of the Galaxy Ann Rev Astron Astrophys* **27** 555–627

Girardi L, Bressan A, Chiosi C, Bertelli G & Nasi E 1996 *Evolutionary Sequences of Stellar Models with New Radiative Opacities. VI. Z = 0.0001 Astron Astrophys Suppl* **117** 113–125
Gliese W & Jahreiß H 1991 *Catalogue of Nearby Stars* Astron. Rechen-Institut, Heidelberg
Gnedin O & Ostriker J 1997 *Destruction of the Galactic Globular Cluster System Astrophys J* **474** 223–255
Goldman T, Pérez-Mercader J, Cooper F & Martín-Nieto M 1992 *Phys Lett B* **281** 219
Gómez A E, Luri X, Mennessier M et al. 1997 *The Luminosity Calibration of the HR Diagram Revised by Hipparcos* in M A C Perryman & P L Bernacca, eds., *Hipparcos Venice 97* ESA SP-402 pp. 207–212 ESA, Noordwijk
Gott J R & Thuan T 1978 *Angular Momentum in the Local Group Astrophys J* **223** 426–436
Gould A, Bahcall J N & Flynn C 1997 *M Dwarfs from Hubble Space Telescope Star Counts. III. The Groth Strip Astrophys J* **482** 913–918
Gould A, Bahcall J N, Maoz D & Yanny B 1995 *Star Counts from the HST Snapshot Survey. 2. Wide Binaries Astrophys J* **441** 200–208
Gratton R G, Fusi-Pecci F, Carretta E, Clementini G, Corsi C E & Lattanzi M 1997 *Ages of Globular Clusters from Hipparcos Parallaxes of Local Subdwarfs Astrophys J* **491** 749–771
Gray R O & Corbally C J 1994 *The Calibration of MK Spectral Classes using Spectral Synthesis. 1: The Effective Temperature Calibration of Dwarf Stars Astron J* **107** 742–746
Grenon M, Jordi C, Figueras F & Torra J 1999a *Stellar Classification from GAIA Intermediate Band System* Technical Report MGUB-PWG-003 Obs. de Geneve
Grenon M, Jordi C, Torra J & Figueras F 1999b *GAIA Broad Band Photometry* Technical Report MGUB-PWG-004 Obs. de Geneve
Grillmair C, Freeman K C, Irwin M & Quinn P 1995 *Globular Clusters with Tidal Tails: Deep Two-Color Star Counts Astron J* **109** 2553–2585
Groenewegen M A T, Schrijver H, de Jong T & Slijkhuis S 1997 *Some Hipparcos Results for Late-Type Stars* in M A C Perryman & P L Bernacca, eds., *Hipparcos Venice 97* ESA SP-402 pp. 323–326 ESA, Noordwijk
Guenther D B & Demarque P 1996 *Seismology of η Bootis Astrophys J* **456** 798–810
Guenther D B, Kim Y C & Demarque P 1996 *Seismology of the Standard Solar Model: Tests of Diffusion and the OPAL and MHD Equations of State Astrophys J* **463** 382–390
Guenther D B, Krauss L M & Demarque P 1998 *Testing the Constancy of the Gravitational Constant Using Helioseismolog Astrophys J* **498** 871–876
Guillot T, Saumon D, Burrows A, Hubbard W B, Lunine J I, Marley M S & Freedman R S 1997 *On the Nature of the Newly Discovered Extrasolar Planets* in C B Cosmovici, S Bowyer & D Werthimer, eds., *Astronomical and Biochemical Origins and the Search for Life in the Universe, IAU Coll. 161* pp. 343–350 Editrice Compositori, Italy
Guillout P, Haywood M, Motch C & Robin A C 1996 *The Stellar Content of Soft X-ray Surveys. I. An Age Dependent Numerical Model Astron Astrophys* **316** 89–101
Guillout P, Sterzik M F, Schmitt J H M M, Motch C & Neuhaeuser R 1998 *Discovery of a Late-Type Stellar Population Associated with the Gould Belt Astron Astrophys* **337** 113–124
Gulati R K, Gupta R, Gothoskar P & Khobragade S 1994 *Stellar Spectral Classification Using Automated Schemes Astrophys J* **426** 340–344
Gunn J E 1974 *A Worm's-Eye View of the Mass Density in the Universe Comm Ap Sp Sci* **6** 7–14
Gunn J E & Stryker L L 1983 *Stellar Spectrophotometric Atlas, 3130 to 10800 A Astrophys J Suppl* **52** 121–153
Gwinn C R, Eubanks T M, Pyne T, Birkinshaw M & Matsakis D N 1997 *Quasar Proper Motions and Low-Frequency Gravitational Wave Astrophys J* **485** 87–91
Habing H J 1996 *Circumstellar Envelopes and Asymptotic Giant Branch Stars Astron Astrophys Review* **7** 97–207
Hakkila J, Myers J M, Stidham B J & Hartmann D H 1997 *A Computerized Model of Large-Scale Visual Interstellar Extinction Astron J* **114** 2043–2053
Hardy T, Murowinsky R & J. D M 1998 *IEEE Trans Nuclear Sci* **45** 154
Hargreaves J, Gilmore G, Irwin M & Carter D 1994 *A Dynamical Study of the Ursa-Minor Dwarf Spheroidal Galaxy Mon Not R Astr Soc* **271** 693–705
Harris W E 1996 *A Catalog of Parameters for Globular Clusters in the Milky Way Astron J* **112** 1487–1488
Hartwick F D A & Schade D 1990 *The Space Distribution of Quasars Ann Rev Astron Astrophys* **28** 437–489
Haywood M, Robin A & Crézé M 1997 *The Evolution of the Milky Way Disc. II. Constraints from Star Counts at the Galactic Poles Astron Astrophys* **320** 440–459
Hechler M 1997a *FIRST/PLANCK and GAIA Mission Analysis: Launch Windows with Eclipse Avoidance*

Manoeuvres Technical Report MAS 402 ESOC
— 1997b *GAIA/FIRST Mission Analysis: Ariane and the Orbits around L2* Technical Report MAS 393 ESOC
Hellings R W 1987 in V Melnikov, ed., *Problems in Gravitation* p. 46 Moscow State University Press, Moscow
Helmi A & White S D M 1999 *Building up the Stellar Halo of the Galaxy Mon Not R Astr Soc* **307** 495–517
— 2000 *Mon Not R Astr Soc* **in press**
Helmi A, White S D M, de Zeeuw P T & Zhao H S 1999a *Debris Streams in the Solar Neighbourhood as Relicts from the Formation of the Milky Way Nature* **402** 53–55
Helmi A, Zhao H S & de Zeeuw P T 1999b *Detecting Halo Streams with GAIA* in B K Gibson, T S Axelrod & M E Putman, eds., *The Third Stromlo Symposium: The Galactic Halo: Bright Stars and Dark Matter* ASP Conf. Ser. 165 p. 125 Astronomical Society of the Pacific, San Francisco
Henry T J & McCarthy D W 1993 *The Mass-Luminosity Relation for Stars of Mass 1.0 to 0.08 Solar Mass Astron J* **106** 773–789
Hernandez X, Gilmore G & Valls-Gabaud D 1999a *Mon Not R Astr Soc* **in press**
Hernandez X, Valls-Gabaud D & Gilmore G 1999b *Deriving Star Formation Histories: Inverting Hertzsprung-Russell Diagrams Through a Variational Calculus Maximum Likelihood Method Mon Not R Astr Soc* **304** 705–719
Hernquist L & Mihos J C 1995 *Excitation of Activity in Galaxies by Minor Mergers Astrophys J* **448** 41–63
Hernstein J R, Moran J M, Greenhill L J et al. 1999 *A Geometrical Distance to the Galaxy NGC 4258 from Orbital Motions in a Nuclear Gas Disk Nature* **400** 539–541
Hertz J, Krogh A & Palmer R G 1991 *Introduction to the Theory of Neural Computation* Addison-Wesley, Redwood City
Høg E 1993 *Astrometry and Photometry of 400 Million Stars Brighter than 18 mag* in I I Mueller & B Kolaczek, eds., *Developments in Astrometry and their Impact on Astrophysics and Geodynamics* IAU Symp. 156 pp. 37–45 Kluwer, The Netherlands
— 1997a *Full Sky Survey and Sloan Photometry with the SMs* Technical Report SAG-CUO-022 Copenhagen University Observatory
— 1997b *Limiting Magnitudes for Sky Mapper Observations* Technical Report SAG-CUO-013 Copenhagen University Observatory
— 1997c *Limiting Magnitudes for Sky Mapper Observations II* Technical Report SAG-CUO-019 Copenhagen University Observatory
— 1997d *On-Board Detection* Technical Report SAG-CUO-008 Copenhagen University Observatory
— 1997e *On-Board Detection III: Focal Plane Assembly* Technical Report SAG-CUO-010 Copenhagen University Observatory
— 1997f *Photometry by ARVI* Technical Report SAG-CUO-021 Copenhagen University Observatory
— 1998a *APT and ASTRO Sky Mappers* Technical Report SAG-CUO-036 Copenhagen University Observatory
— 1998b *ARVI Photometry* Technical Report SAG-CUO-023 Copenhagen University Observatory
— 1998c *Estimation of Background. APT Photometry* Technical Report SAG-CUO-026 Copenhagen University Observatory
— 1998d *GAIA at High Star Densities* Technical Report SAG-CUO-031 Copenhagen University Observatory
— 1998e *The Sampling in GAIA Photometry* Technical Report SAG-CUO-051 Copenhagen University Observatory
— 1998f *Spectrophotometry or Filter Photometry with GAIA?* Technical Report SAG-CUO-045 Copenhagen University Observatory
— 1998g *Star Detection in the Astrometric Sky Mapper* Technical Report SAG-CUO-046 Copenhagen University Observatory
— 1999 *Self-Consistency in the Photometric Design* Technical Report SAG-CUO-068 Copenhagen University Observatory
Høg E & Arenou F 2000 *GAIA Meeting at the Paris Observatory* Technical Report IWG-OPM-003 Copenhagen University Observatory
Høg E & Fabricius C 1998 *Simulation of Final GAIA Photometry* Technical Report SAG-CUO-043 Copenhagen University Observatory
Høg E, Fabricius C, Flynn C, Knude J & Makarov V V 1999a *A Table of Kinematic Tracers* Technical Report SAG-CUO-070 Copenhagen University Observatory
Høg E, Fabricius C, Knude J & Makarov V V 1998a *GAIA Surveys of Nebulae and Sky Background* Technical Report SAG-CUO-032 Copenhagen University Observatory

— 1998b *GAIA Surveys of Surface Brightness* Technical Report SAG-CUO-039 Copenhagen University Observatory
— 1999b *Design of Narrow-Band Photometry* Technical Report SAG-CUO-055 Copenhagen University Observatory
— 1999c *Sky Survey and Photometry by the GAIA Satellite Baltic Astronomy* **8** 25–56
Høg E, Fabricius C & Makarov V V 1997 *Astrometry from Space: New Design of the GAIA Mission Experimental Astronomy* **7** 101–115
— 1999d *The Velocity Field and 3-D Structure of the Universe Baltic Astronomy* **8** 233–237
Høg E, Fabricius C & V. M V 1999e *Design of Medium-Band Photometry: II* Technical Report SAG-CUO-067 Copenhagen University Observatory
Høg E & Knude J 1998 *Accuracy of Photometry with GAIA* Technical Report SAG-CUO-048 Copenhagen University Observatory
— 1999 *Photometry from Space and from Ground* Technical Report SAG-CUO-062 Copenhagen University Observatory
Høg E & Lindegren L 1994 *ROEMER Satellite Project: the First High-Accuracy Survey of Faint Stars* in L V Morrison & G Gilmore, eds., *Galactic and Solar System Optical Astrometry* pp. 246–252 Cambridge University Press
Høg E & Makarov V V 1998 *Simulation of BBP Observations for Single and Double Stars* Technical Report SAG-CUO-042 Copenhagen University Observatory
Høg E, Novikov I D & Polnarev A G 1995 *MACHO Photometry and Astrometry Astron Astrophys* **294** 287–294
Høg E & Petersen J O 1997 *Hipparcos Parallaxes and the Nature of δ Scuti Stars Astron Astrophys* **323** 827–830
Høg E & Sørensen H J 1997 *Focal Plane Assembly for GAIA* Technical Report SAG-CUO-018 Copenhagen University Observatory
Holland A, Holmes-Siedle A, Johlander B & Adams L 1991 *Techniques of Minimizing Space Proton Damage in Scientific CCDs IEEE Trans Nuc Sci* **38(6)** 1663–1670
Holman M, Touma J & Tremaine S 1997 *Chaotic Variations in the Eccentricity of the Planet Orbiting 16 Cygni B Nature* **386** 254–256
Holtzman J A, Light R M, Baum W A, Worthey G, Faber S M, Hunter D A, O'Neil E. J. J, Kreidl T J, Groth E J & Westphal J A 1993 *Wide Field Camera Observations of Baade's Window Astron J* **106** 1826–1838
Hoogerwerf R & Aguilar L A 1999 *Identification of Moving Groups and Member Selection using Hipparcos Data Mon Not R Astr Soc* **306** 394–406
Hoogerwerf R, den Hollander R, de Bruijne J H J & de Zeeuw P T 2000 *Astron Astrophys* **in press**
Hopkins I H et al. 1994 *Proton-Induced Charge Transfer Degradation in CCDs for Near-Room Temperature Applications IEEE Nuc and Space Radiation Effects Conference*
Hopkinson G R 1996 *Proton Effects in CCDs IEEE Trans Nuc Sci* **43(2)** 614
— 1999 *GAIA CCD Radiation Study* Technical report Sira Electro-Optics
Huang S & Carlberg R 1997 *Sinking Satellites and Tilting Disk Galaxies Astrophys J* **480** 503–523
Hummell C A, Mozurkewich D, Elias N. M. I, Quirrenbach A, Buscher D F, Armstrong J T, Johnston K J, Simon R S & Hutter D J 1994 *Four Years of Astrometric Measurements with the Mark III Optical Interferometer Astron J* **108** 326–336
Ibata R & Gilmore G 1995 *The Outer Regions of the Galactic Bulge - II. Analysis Mon Not R Astr Soc* **275** 605–627
Ibata R, Gilmore G & Irwin M 1994 *A Dwarf Satellite Galaxy in Sagittarius Nature* **370** 194
— 1995 *Sagittarius: the Nearest Dwarf Galaxy Mon Not R Astr Soc* **277** 781–800
Ibata R A, Wyse R F G, Gilmore G, Irwin M & Suntzeff N 1997 *The Kinematics, Orbit, and Survival of the Sagittarius Dwarf Spheroidal Galaxy Astron J* **113** 634–655
Iess L, Giampieri G, Anderson J D & Bertotti B 1999 *Doppler Measurement of the Solar Gravitational Deflection Class Quantum Grav* **16** 1487–1502
Irwin M J 1985 *Automatic Analysis of Crowded Fields Mon Not R Astr Soc* **214** 575–604
Isern J, García-Berro E, Hernanz M, Mochkovitch R & Torres S 1998 *The Halo White Dwarf Population Astrophys J* **503** 239–246
Isern J, García-Berro E, Hernanz M & Salaris M 2000 in *The Galactic Halo: From Clusters to Field Stars* p. in press
Isern J, Hernanz M & García-Berro E 1992 *Axion Cooling of White Dwarfs Astrophys J* **392** L23–L25
Isern J, Hernanz M, García-Berro E & Mochkovitch R 1999 *The Local Star Formation Rate from White Dwarfs* in S E Solheim & E G Meistas, eds., *11th European Workshop on White Dwarfs* ASP Conf.

Ser. 169 p. 408 Astronomical Society of the Pacific, San Francisco
Jaschek C & Gómez A E 1998 *The Absolute Magnitude of the Early Type MK Standards from Hipparcos Parallaxes Astron Astrophys* **330** 619–625
Jeffries R D 1995 *The Kinematics of Lithium-Rich, Active Late-Type Stars: Evidence for a Low-Mass Local Association Mon Not R Astr Soc* **273** 559–572
Jewitt D, Luu J & Trujillo C 1998 *Large Kuiper Belt Objects: The Mauna Kea 8K CCD Survey Astron J* **115** 2125–2135
Johnston K 1998 *A Prescription for Building the Milky Way's Halo from Disrupted Satellites Astrophys J* **495** 297–308
Johnston K V, Sigurdsson S & Hernquist L 1999a *Measuring Mass-Loss Rates from Galactic Satellites Mon Not R Astr Soc* **302** 771–789
Johnston K V, Zhao H S, Spergel D N & Hernquist L 1999b *Tidal Streams as Probes of the Galactic Potential Astrophys J* **512** L109–L112
Jones B F 1976 *Gravitational Deflection of Light: Solar Eclipse of 30 June 1973 Astron J* **81** 455–463
Kahn F & Woltjer L 1959 *Intergalactic Matter and the Galaxy Astrophys J* **130** 705–717
Kaiser N & Jaffe A 1997 *Bending of Light by Gravity Waves Astrophys J* **484** 545–554
Kaluzny J, Kubiak M, Szymanski M, Udalski A, Krzeminski W & Mateo M 1995 *OGLE Catalogue of Variable Stars in the Sculptor Dwarf Spheroidal Galaxy Astron Astrophys Suppl* **112** 407–428
Kapteyn J C 1905 *Brit Assoc Adv Science* p. 257
Kent S M, Dame T M & Fazio G 1991 *Galactic Structure from the Spacelab Infrared Telescope. II - Luminosity Models of the Milky Way Astrophys J* **378** 131–138
Kent S M, Stoughton C, Newberg H, Loveday J, Petravick D, Gurbani V, Berman E, Sergey G & Lupton R 1994 *Sloan Digital Sky Survey* in D R Crabtree, R J Hanisch & J Barnes, eds., *Third Annual Conference on Astronomical Data Analysis Software and Systems* ASP, San Francisco
Kenyon S J & Luu J X 1998 *Accretion in the Early Kuiper Belt. I. Coagulation and Velocity Evolution Astron J* **115** 2136–2160
— 1999 *Accretion in the Early Kuiper Belt. II. Fragmentation Astron J* **118** 1101–1119
King I R, Anderson J, Cool A M & Piotto G 1998 *The Luminosity Function of the Globular Cluster NGC 6397 near the Limit of Hydrogen Burning Astrophys J* **492** L37–L41
Kirkpatrick J D, McGraw J T, Hess T R, Liebert J & McCarthy D W 1994 *The Luminosity Function at the End of the Main Sequence: Results of a Deep, Large-Area, CCD Survey for Cool Dwarfs Astrophys J Suppl* **94** 749–788
Klioner S A & Kopeikin S M 1992 *Microarcsecond Astrometry in Space - Relativistic Effects and Reduction of Observations Astron J* **104** 897–914
Klypin A, Kravtsov A V, Valenzuela O & Prada F 1999 *Where Are the Missing Galactic Satellites? Astrophys J* **522** 82–92
Knox R, Hawkins M R S & Hambly N C 1999 *A Survey for Cool White Dwarfs and the Age of the Galactic Disc Mon Not R Astr Soc* **306** 736–752
Knude J & Høg E 1998 *Interstellar Reddening from the Hipparcos and TYCHO Catalogues. I. Distances to Nearby Molecular Clouds and Star Forming Regions Astron Astrophys* **338** 897–904
— 1999 *Interstellar Reddening from the Hipparcos and TYCHO Catalogues. II. Nearby Dust Features at the NGP Associated with Approaching HI Gas Astron Astrophys* **341** 451–457
Kochanek C 1995 *Evidence for Dark Matter in MG 1654+134 Astrophys J* **445** 559–577
Kochanek C S 1996 *The Mass of the Milky Way Astrophys J* **457** 228–243
Kochanek C S, Kolatt T S & Bartelmann M 1996 *Proper Motions of VLBI Lenses, Inertial Frames, and the Evolution of Peculiar Velocities Astrophys J* **473** 610–619
Kopeikin S M 1997 *Binary Pulsars as Detectors of Ultralow-Frequency Gravitational Waves Phys Rev D* **56** 4455–4469
Kopeikin S M & Schäfer G 1999 *Lorentz Covariant Theory of Light Propagation in Gravitational Fields of Arbitrarily-Moving Bodies Phys Rev D* **60** 124002
Kopeikin S M, Schäfer G, Gwinn C R & Eubanks T M 1999 *Astrometric and Timing Effects of Gravitational Waves from Localized Sources Phys Rev D* **59** 84023
Kovalevsky J, Lindegren L, Perryman M A C et al. 1997 *The Hipparcos Catalogue as a Realisation of the Extragalactic Reference System Astron Astrophys* **323** 620–633
Krauss L M & White M 1992 *Gravitational Lensing, Finite Galaxy Cores, and the Cosmological Constant Astrophys J* **394** 385–395
Krełowski J & Greenberg J M 1999 *Diffuse Band Shifts: a Possible Explanation Astron Astrophys* **346** 199–204
Krisciunas K, Margon B & Szkody P 1998 *The Recognition of Unusual Objects in the Sloan Digital Sky*

Survey Color System Publ Astr Soc Pac **110** 1342–1355
Kroupa P 1995 *Unification of the Nearby and Photometric Stellar Luminosity Functions Astrophys J* **453** 358–368
— 1998 *The Stellar Mass Function* in R Rebolo, E L Martin & M R Zapatero Osorio, eds., *Brown Dwarfs and Extrasolar Planets* ASP Conf. Ser. 134 p. 483 Astronomical Society of the Pacific, San Francisco
Kroupa P, Petr M G & McCaughrean M J 1999 *Binary Stars in Young Clusters: Models Versus Observations of the Trapezium Cluster New Astronomy* **4** 495–520
Kroupa P, Tout C & Gilmore G 1993 *The Distribution of Low-Mass Stars in the Galactic Disc Mon Not R Astr Soc* **262** 545–587
Krymolowski Y & Mazeh T 1999 *Studies of Multiple Stellar Systems. II. Second-Order Averaged Hamiltonian to Follow Long-Term Orbital Modulations of Hierarchical Triple Systems Mon Not R Astr Soc* **304** 720–732
Kudritzki R P 1998 *Quantitative Spectroscopy of the Brightest Blue Supergiant Stars in Galaxies* in A Aparicio, A Herrero & F Sánchez, eds., *VIII Canary Islands Winter School of Astrophysics on Stellar Astrophysics for the Local Group* p. 149 Cambridge University Press
Kudritzki R P, Lennon D J, Haser S M, Puls J, Pauldrach A, Venn K & Voels S A 1996 in P Benvenuti et al., eds., *Science with the Hubble Space Telescope II* pp. 285–296
Kudritzki R P, Pauldrach A W A, Puls J & Abbott D C 1989 *Radiation-Driven Winds of Hot Stars. VI - Analytical Solutions for Wind Models Including the Finite Cone Angle Effect Astron Astrophys* **219** 205–218
Kudritzki R P, Puls J, Lennon D J, Venn K A, Reetz J, Najarro F, McCarthy J K & Herrero A 1999 *The Wind Momentum-Luminosity Relationship of Galactic A- and B-Supergiants Astron Astrophys* **350** 970–984
Kuijken K 1996 *Observational Evidence for a Bar in the Milky Way* in R Buta, D A Crocker & B G Elmegreen, eds., *Barred Galaxies* IAU Coll. 157 p. 504 Astronomical Society of the Pacific, San Francisco
Kuijken K & Gilmore G 1989a *The Mass Distribution in the Galactic Disc - Part Two - Determination of the Surface Mass Density of the Galactic Disc Near the Sun Mon Not R Astr Soc* **239** 605–649
— 1989b *The Mass Distribution in the Galactic Disc. I - A Technique to Determine the Integral Surface Mass Density of the Disc Near the Sun. II - Determination of the Surface Mass Density of the Galactic Disc Near the Sun. III - The Local Volume Mass Density Mon Not R Astr Soc* **239** 571–603
Kupka F 1999 *Computing Solar and Stellar Overshooting with Turbulent Convection Models. First Tests of a Fully Non-Local Model* in A Gimenez, E F Guinan & B Montesinos, eds., *Theory and Tests of Convection in Stellar Structure* ASP Conf. Ser. 173 p. 157 Astronomical Society of the Pacific, San Francisco
Kurtz M J 1982 Ph.D. thesis Dartmouth College
Kurucz R L 1992 *Model Atmospheres for Population Synthesis* in B Barbuy & A Renzini, eds., *Stellar Populations of Galaxies* IAU Symp. 149 p. 225 Kluwer, Dordrecht
Lahav O, Naim A, Sodré L & Storrie-Lombardi M C 1996 *Mon Not R Astr Soc* **283** 207
Lang K R 1971 *Interstellar Scintillation of Pulsar Radiation Astrophys J* **164** 249–264
— 1978 *Astrophysical Formulae* Springer-Verlag
Larson R B 1990 *Galaxy Building Publ Astr Soc Pac* **102** 709–722
LaSala J 1994 in C J Corbally, R O Gray & R F Garrison, eds., *The MK Process at 50 Years: A Powerful Tool for Astrophysical Insight* ASP Conf. Ser. 60 p. 312 Astronomical Society of the Pacific, San Francisco
Latham D W, Mazeh T, Stefanik R P, Mayor M & Burki G 1989 *The Unseen Companion of HD114762: A Probable Brown Dwarf Nature* **339** 38–40
Lattanzi M G, Gai M, Cecconi M, Cesare S & Mana G 1997 in *Third International Conference on Space Optics, Toulouse*
Lattanzi M G, Spagna A, Sozzetti A & Casertano S 2000 *Space-Borne Global Astrometric Surveys: The Hunt for Extra-Solar Planets Mon Not R Astr Soc* **in press**
Laughlin G & Adams F G 1999 *Stability and Chaos in the υ Andromedae Planetary System Astrophys J* **526** 881–889
Lebreton Y 2000 *Stellar Structure and Evolution: Deductions from Hipparcos Ann Rev Astron Astrophys* **38** in press
Lebreton Y, Gómez A E, Mermilliod J C & Perryman M A C 1997 *The Age and Helium Content of the Hyades Revisited* in M A C Perryman & P L Bernacca, eds., *Hipparcos Venice 97* ESA SP-402 pp. 231–236 ESA, Noordwijk
Lebreton Y, Michel E, Goupil M J, Baglin A & Fernandes J 1995 *Accurate Parallaxes and Stellar Age*

Determinations in E Høg & P K Seidelmann, eds., *Astronomical and Astrophysical Objectives of Sub-Milliarcsecond Optical Astrometry* IAU Symp. 166 p. 135 Kluwer, Dordrecht

Lebreton Y, Perrin M N, Cayrel R, Baglin A & Fernandes J 1999 *The Hipparcos HR Diagram of Nearby Stars in the Metallicity Range: $-1.0 < [Fe/H] < 0.3$. A New Constraint on the Theory of Stellar Interiors and Model Atmospheres Astron Astrophys* **350** 587–597

Leinert C, Bowyer S, Haikala L K et al. 1998 *The 1997 Reference of Diffuse Night Sky Brightness Astron Astrophys Suppl* **127** 1–99

Lenz D, Newberg H, Rosner R, Richards G & Stoughton C 1998 *Photometric Separation of Stellar Properties Using SDSS Filters Astrophys J Suppl* **119** 121–140

Lindblad P O, Palous J, Lodén K & Lindegren L 1997 *The Kinematics and Nature of Gould's Belt: a 30 Myr Old Star Forming Region* in M A C Perryman & P L Bernacca, eds., *Hipparcos Venice 97* ESA SP-402 pp. 507–512 ESA, Noordwijk

Lindegren L 1978 *Photoelectric Astrometry - a Comparison of Methods for Precise Image Location* in F V Prochazka & R H Tucker, eds., *Modern Astrometry* IAU Coll. 48 p. 197 University Observatory, Vienna

— 1997a *GAIA: Choice of Basic Angles for Three Viewing Directions* Technical Report SAG-LL-010 Lund Observatory

— 1997b *Why More than Two Viewing Directions for GAIA?* Technical Report SAG-LL-009 Lund Observatory

— 1998a *Calculation of Chromatic Displacement* Technical Report SAG-LL-024 Lund Observatory

— 1998b *Charge Trapping Effects in CCDs for GAIA Astrometry* Technical Report SAG-LL-022 Lund Observatory

— 2000 *Detection of Faint Galaxies with GAIA* Technical Report SAG-LL-029 Lund Observatory

Lindegren L, Bastian U, Gilmore G, Halbwachs J L, Høg E, Knude J, Kovalevsky J, Labeyrie A, van Leeuwen F, Pel J W, Schrijver H, Stabell R & Thejll P 1993a *Roemer: Proposal for the Third Medium Size ESA Mission (M3)* Technical report Lund Observatory

Lindegren L & Perryman M A C 1994 *The GAIA Mission: Internal Report Submitted to the ESA Horizon 2000+ Survey Committee* Technical report Lund/ESA

— 1995 *A Small Interferometer in Space for Global Astrometry: the GAIA Concept* in E Høg & P K Seidelmann, eds., *Astronomical and Astrophysical Objectives of Sub-Milliarcsecond Optical Astrometry* IAU Symp. 166 p. 337 Kluwer, Dordrecht

— 1996 *GAIA: Global astrometric interferometer for astrophysics. Astron Astrophys Suppl* **116** 579–595

Lindegren L, Perryman M A C, Bastian U, Dainty J C, Høg E, van Leeuwen F, Kovalevsky J, Labeyrie A, Loiseau S, Mignard F, Noordam J E, Le Poole R S, Thejll P & Vakili F 1993b *GAIA — Global Astrometric Interferometer for Astrophysics Proposal for a Cornerstone Mission concept submitted to ESA in October 1993* Technical report Lund

Lindegren L, Perryman M A C & Loiseau S 1995 *Global Astrometric Interferometer for Astrophysics (GAIA)* in R D Reasenberg, ed., *Spaceborne Interferometry II* SPIE 2477 pp. 91–103

Lissauer J J 1999 *Three Planets for Upsilon Andromedae Nature* **398** 659–660

Luhman K L & Rieke G H 1998 *The Low-Mass Initial Mass Function in Young Clusters: L1495E Astrophys J* **497** 354–369

Luri X, Gómez A E, Torra J, Figueras F & Mennessier M O 1998 *The LMC Distance Modulus from Hipparcos RR Lyrae and Classical Cepheid Data Astron Astrophys* **335** L81–L84

Luyten W J 1976 *Catalogue of High Proper Motion Stars* University of Minnesota Press

Lydon T J & Sofia S 1996 *A Measurement of the Shape of the Solar Disk: The Solar Quadrupole Moment, the Solar Octopole Moment, and the Advance of Perihelion of the Planet Mercury Phys Rev Lett* **76** 177–179

Lynden-Bell D 1982 *The Genesis of the Local Group* in H Bruck, G Coyne & M S Longair, eds., *Astrophysical Cosmology* pp. 85–108 Vatican

LyngAstron. Astrophys. G 1987 *Catalogue of Open Cluster Data, 5th edition* Centre de Données Stellaires, Strasbourg

Ma C, Arias E F, Eubanks T M et al. 1998 *The International Celestial Reference Frame as Realized by Very Long Baseline Interferometry Astron J* **116** 516–546

Majewski S 1993 *Galactic Structure Surveys and the Evolution of the Milky Way Ann Rev Astron Astrophys* **31** 575–638

Majewski S R 1999 *The Role of Accretion in the Formation of the Halo: Observational View* in B K Gibson, T S Axelrod & M E Putman, eds., *The Third Stromlo Symposium: The Galactic Halo* ASP Conf. Ser. 165 p. 76 Astronomical Society of the Pacific, San Francisco

Majewski S R, Munn J A & Hawley S 1996 *Absolute Proper Motions to B=22.5: Large-Scale Streaming Motions and the Structure and Origin of the Galactic Halo Astrophys J* **459** L73–L77

Makarov V V 1998 *Absolute Measurements of Trigonometric Parallaxes with Astrometric Satellites Astron Astrophys* **340** 309–314
Malhotra R 1995 *The Origin of Pluto's Orbit: Implications for the Solar System Beyond Neptune Astron J* **110** 420–429
Mao S, Mo H & White S D M 1998 *The Evolution of Galactic Discs Mon Not R Astr Soc* **297** L71–L75
Mao S & Paczyński B 1991 *Gravitational Microlensing by Double Stars and Planetary Systems Astrophys J* **374** L37–L40
Marciano W J 1984 *Time Variation of the Fundamental 'Constants' and Kaluza-Klein Theories Phys Rev Lett* **52** 489–491
Marcy G W & Butler R P 1998 *Detection of Extrasolar Giant Planets Ann Rev Astron Astrophys* **36** 57–97
Marcy G W, Butler R P & Fischer D A 1999 *Three Jupiter-Mass Companions Orbiting υ Andromedae Bull Amer Astron Soc* **194, 14.02**
Marcy G W, Cochran W D & Mayor M 2000 *Extrasolar Planets Around Main Sequence Stars* in V Mannings, A P Boss & S S Russell, eds., *Protostars and Planets IV* University of Arizona Press, Tucson
Martín E L 1997 *Lithium in Young Clusters and Pre-Main Sequence Stars* in G Micela, R Pallavicini & S Sciortino, eds., *Cool Stars in Clusters and Associations: Magnetic Activity and Age Indicators* volume 68 of *Mem. Soc. Astron. Ital.* p. 905
Mateo M 1998 *Dwarf Galaxies of the Local Group Ann Rev Astron Astrophys* **36** 435–506
Mathewson D S, Cleary M N & Murray J D 1974 *The Magellanic Stream Astrophys J* **190** 291–296
Matteucci F 1998 *The Chemical Evolution of the Galactic Bulge* in M Spite & F Crifo, eds., *Galaxy Evolution: Connecting the Distant Universe with the Local Fossil Record* p. 78 Kluwer, Dordrecht
May A & Binney J J 1986 *Solar-Neighbourhood Observations and the Structure of the Galaxy Mon Not R Astr Soc* **221** 857–877
Mayor M & Queloz D 1995 *A Jupiter-Mass Companion to a Solar-Type Star Nature* **378** 355–359
Mazeh T, Latham D W & Stefanik R P 1996 *Spectroscopic Orbits for Three Binaries with Low-Mass Companions and the Distribution of Secondary Masses Near the Substellar Limit Astrophys J* **466** 415–426
Mazeh T, Zucker S, Torre A D & van Leeuwen F 1999 *Analysis of the Hipparcos Measurements of υ Andromedae – A Mass Estimate of its Outermost Known Planetary Companion Astrophys J* **522** L149–L151
McCarthy J K, Kudritzki R P, Lennon D J, Venn K A & Puls J 1997 *Mass-Loss Rates and Stellar Wind Momenta of A-Supergiants in M31: First Results from the Keck HIRES Spectrograph Astrophys J* **482** 757–764
McWilliam A & Rich R M 1994 *The First Detailed Abundance Analysis of Galactic Bulge K Giants in Baade's Window Astrophys J Suppl* **91** 749–791
Méndez R & van Altena W 1996 *Galactic Structure Toward the Open Clusters NGC 188 and NGC 3680 Astron J* **112** 655–667
Mera D, Chabrier G & Baraffe I 1996 *Determination of the Low-Mass Star Mass Function in the Galactic Disk Astrophys J* **459** L87–L90
Metcalfe N, Shanks T, Fong R & Roche N 1995 *Galaxy Number Counts - III. Deep CCD Observations to B=27.5 Mon Not R Astr Soc* **273** 257–276
Micela G, Sciortino S & Favata F 1993 *Stellar Birthrate in the Galaxy - Constraints from X-ray Flux-Limited Surveys Astrophys J* **412** 618–624
Michel E, Belmonte J A, Alvarez M, Jiang S Y, Chevreton M, Auvergne M, Goupil M J, Baglin A, Mangeney A, Roca Cortes T, Liu Y Y, Fu J N & Dolez N 1992 *Multi-Periodicity of the Delta Scuti star GX Pegasi - Second Photometry Campaign of the STEPHI Network Astron Astrophys* **255** 139–148
Mighell K J 1997 *WFPC2 Observations of the Carina Dwarf Spheroidal Galaxy Astron J* **114** 1458–1470
Mighell K J & Butcher H R 1992 *Star Formation Histories of Galaxies from Their Stellar Luminosity Functions Astron Astrophys* **255** 26–34
Mignard F 1998 *Ephemeris Requirements for GAIA* Technical Report SAG-FM-004 OCA-CERGA
Miller A S 1993 *A Review of Neural Network Applications in Astronomy Vistas in Astronomy* **36** 141–161
Miralda-Escude J 1996 *Microlensing Events from Measurements of the Deflection Astrophys J* **470** L113–L116
Moisson X 1997 *New Solution for the Planetary Motion* in J Vondrák & N Capitaine, eds., *Journées des Systèmes de Référence* p. 1
Molaro P, Bonifacio P, Castelli F & Pasquini L 1997 *New Beryllium Observations in Low-Metallicity Stars Astron Astrophys* **319** 593–606
Moore B, Governato F, Quinn T, Stadel J & Lake G 1998 *Resolving the Structure of Cold Dark Matter Halos Astrophys J* **499** L5–L8

Morel P & Baglin A 1999 *Microscopic Diffusion and Subdwarfs Astron Astrophys* **345** 156–162
Morgan W W, Keenan P C & Kellman E 1943 *An Atlas of Stellar Spectra with an Outline of Spectral Classification* University of Chicago Press
Morrison H L 1993 *The Local Density of Halo Giants Astron J* **106** 578–590
Mould J R, Han M, Stetson P B, Gibson B, Graham J A, Huchra J, Madore B & Rawson D 1997 *The Age of the Large Magellanic Cloud Cluster NGC 1651 Astrophys J* **483** L41–L44
Mould J R, Huchra J P, Freedman W L et al. 2000 *The HST Key Project on the Extragalactic Distance Scale. XXVIII. Combining the Constraints on the Hubble Constant Astrophys J* **in press**
Munari U 1999a *GAIA Spectroscopy: Proposing the 8500–8750 Å Region and Evaluating the Performance Baltic Astronomy* **8** 73–96
— 1999b *A Modular and Consistent Photometric System for GAIA Baltic Astronomy* **8** 123–138
Munari U & Castelli F 1999 *High-Resolution Spectroscopy Over 8500–8700 Åfor GAIA. II. Synthetic Model Spectra Astron Astrophys Suppl* **141** 141–148
Murdoch K & Hearnshaw J B 1993 *The Orbit of the Spectroscopic Binary HR 3220 The Observatory* **113** 126–127
Murray C A 1983 *Vectorial Astrometry* Adam Hilger
Newberg H J, Richards G T, Richmond M & Fan X 1999 *Catalog of Four-Color Photometry of Stars, Galaxies, and QSOS Using SDSS Filters Astrophys J Suppl* **123** 377–435
Ng Y K, Bertelli G, Bressan A, Chiosi C & Lub J 1995 *The Galactic structure towards the Galactic Centre. I. A study of the Palomar-Groningen Field 3. Astron Astrophys* **295** 655–677
Ng Y K, Bertelli G, Chiosi C & Bressan A 1996 *The Galactic Structure Towards the Galactic Centre. III. A Study of Baade's Window: Discovery of the Bar Population? Astron Astrophys* **310** 771–796
— 1997 *Probing the Galaxy. I. The Galactic Structure Towards the Galactic Pole Astron Astrophys* **324** 65–79
Nissen P & Schuster W 1997 *Chemical Composition of Halo and Disk Stars with Overlapping Metallicities Astron Astrophys* **326** 751–762
Nordlund A & Stein R F 1999 *Convection Simulations* in A Gimenez, E F Guinan & M Montesinos, eds., *Theory and Tests of Convection in Stellar Structure* ASP Conf. Ser. 173 p. 91 Astronomical Society of the Pacific, San Francisco
Nordtvedt K 1995 *The Relativistic Orbit Observables in Lunar Laser Ranging Icarus* **114** 51–62
Norris J 1994 *Population Studies. 12: The Duality of the Galactic Halo Astrophys J* **431** 645–657
North P, Jaschek C & Egret D 1997 *Delta Scuti Stars in the HR Diagram* in M A C Perryman & P L Bernacca, eds., *Hipparcos Venice 97* ESA SP-402 pp. 367–370 ESA, Noordwijk
Odenkirchen M & Brosche P 1999 *The Proper-Motion Signal of Unresolved Binaries in the Hipparcos Catalogue Astr Nach* **320** 397–400
Okamura S, Yasuda N, Shimasaku K, Yagi M & Weinberg D H 1999 *Retrieving Bulge and Disk Parameters and Asymptotic Magnitudes from the Growth Curves of Galaxies Publ Astr Soc Pac* **111** 31–44
Olive K A, Schramm D N, Steigman G & Walker T P 1990 *Phys Lett B* **236** 454
O'Mullane W & Lindegren L 1999 *An Object-Oriented Framework for GAIA Data Processing Baltic Astronomy* **8** 57–72
Oort J H 1928 *BAN* **4** 269
— 1932 *BAN* **6** 249
Oppenheimer B R & Kulkarni S R 2000 in V Mannings, A P Boss & S S Russell, eds., *Protostars and Planets IV* University of Arizona Press, Tucson
Ortolani S, Renzini A, Gilmozzi R, Marconi G, Barbuy B, Bica E & Rich R M 1995 *Near Coeval Formation of the Galactic Bulge and Halo Inferred from Globular Cluster Ages Nature* **377** 701
Oudmaijer R D, Groenewegen M A T & Schrijver H 1998 *The Lutz-Kelker Bias in Trigonometric Parallaxes Mon Not R Astr Soc* **294** L41–L46
Paczyński B 1997 *The Future of Massive Variability Searches* in R Ferlet, J P Maillard & B Raban, eds., *Proc. 12th IAP Astrophysics Coll. 1996: Variable Stars and Astrophysical Returns from Microlensing Surveys* p. 357 Editions Frontières
— 1998 *Gravitational Microlensing with the Space Interferometry Mission Astrophys J* **494** L23–L26
Paczyński B, Stanek K Z, Udalski A, Szymanski M, Kałużny J & Kubiak M 1994 *The Distribution of Galactic Disk Stars in Baade's Window Astron J* **107** 2060–2066
Pauldrach A W A, Puls J & Kudritzki R P 1986 *Radiation-Driven Winds of Hot Luminous Stars - Improvements of the Theory and First Results Astron Astrophys* **164** 86–100
Paunzen E 1999 *A Comparison of Different Spectral Classification Systems for Early-Type Stars Using Hipparcos Parallaxes Astron Astrophys* **341** 784–788
Peebles P J E 1971 *Physical Cosmology* Princeton University Press

— 1996 *Dynamics of the Relative Motions of the Galaxies in and Near the Local Group* in O Lahav, E Terlevich & R J Terlevich, eds., *Gravitational Dynamics* p. 219 Cambridge University Press, Cambridge

Peek H L, Verbugt D W, Beenhakkers M J, Huinink W F & Kleimann A C 1996 *An FT-CCD Imager with True* 2.4×2.4 *μm Pixels in Double Membrane Poly-Si Technology IEEE IEDM*

Perryman M A C 1994 *A Galactic Population Census in Support of Astrometric Measurements* in E Høg & P K Seidelmann, eds., *Astronomical and Astrophysical Objectives of Sub-Milliarcsec Astrometry* IAU Symp. 166 pp. 211–215 Kluwer

— 2000 *Near-Earth Objects Observable with GAIA* Technical Report SAG-MP-002 ESTEC

Perryman M A C, Brown A G A, Lebreton Y et al. 1998 *The Hyades: Distance, Structure, Dynamics and Age Astron Astrophys* **331** 81–120

Perryman M A C, Lindegren L, Arenou F et al. 1996 *Hipparcos Distances and Mass Limits for the Planetary Candidates 47 UMa, 70 Vir and 51 Peg Astron Astrophys* **310** L21–L24

Petersen J O & Christensen-Dalsgaard J 1996 *Pulsation Models of δ Scuti Variables Astron Astrophys* **312** 463–474

Pickles A J 1998 *A Stellar Spectral Flux Library: 1150-25000 A Publ Astr Soc Pac* **110** 863–878

Pinsonneault M H, Stauffer J, Soderblom D R et al. 1998 *The Problem of Hipparcos Distances to Open Clusters. I. Constraints from Multicolor Main-Sequence Fitting Astrophys J* **504** 170–191

Piotto G, Cool A & King I 1997 *A Comparison of Deep HST Luminosity Functions of Four Globular Clusters Astron J* **113** 1345–1352

Platais I, van Leeuwen F & Mermilliod J C 1995 *High Accuracy Astrometry of Star Clusters and Associations* in M A C Perryman & F van Leeuwen, eds., *Future Possibilities for Astrometry in Space* ESA SP-379 pp. 71–76 ESA, Noordwijk

Pochoda P & Schwarzschild M 1964 *Variation of the Gravitational Constant and the Evolution of the Sun Astrophys J* **139** 587–593

Pont F, Mayor M, Turon C & Vandenberg D A 1998 *Hipparcos Subdwarfs and Globular Cluster Ages: the Distance and Age of M92 Astron Astrophys* **329** 87–100

Popowski P & Gould A 1999 *The RR Lyrae Distance Scale* in A Heck & F Caputo, eds., *Post-Hipparcos Cosmic Candles* Astrophys. Sp. Sci. 237 p. 53 Kluwer, Dordrecht

Pöppel W G L 1997 *The Gould Belt System and the Local Interstellar Medium Fundamental of Cosmic Physics* **18** 1–271

Popper D M 1997 *Orbits of Detached Pre-Main Sequence Eclisping Binaries of Types Late F to K. II. UV Leonis, UV Piscium and BH Virginis Astron J* **114** 1195–1205

Puls J, Kudritzki R P, Herrero A, Pauldrach A, Haser S M, Lennon D J, Gabler R, Voels S A, Vilchez J M, Wachter S & Feldmeier A 1996 *O-Star Mass-Loss and Wind Momentum Rates in the Galaxy and the Magellanic Clouds Observations and Theoretical Predictions. Astron Astrophys* **305** 171–208

Putman M E, Gibson B K & Staveley-Smith L 1998 *Tidal Disruption of the Magellanic Clouds by the Milky Way Nature* **394** 752–754

Pyne T, Gwinn C R, Birkinshaw M, Eubanks T M & Matsakis D N 1996 *Gravitational Radiation and Very Long Baseline Interferometry Astrophys J* **465** 566–577

Quevedo H & Parkes L 1989 *General Relativity and Gravitation* **21** 1047–1072

Quist C F 2000 *Brown Dwarf Detection Possibilities from GAIA Astron Astrophys* **in press**

Ratnatunga K, Griffiths R E & Ostrander E 1999 *The Top 10 List of Gravitational Lens Candidates from the Hubble Space Telescope Medium Deep Survey Astron J* **118** 2010–2023

Ratnatunga K U, Bahcall J N & Casertano S 1989 *Kinematic Modeling of the Galaxy. I - The Yale Bright Star Catalogue Astrophys J* **339** 106–125

Ratnatunga K U & Freeman K C 1985 *Kinematics of K Giants in the Outer Galactic Halo Astron J* **291** 260–269

Reasenberg R D 1983 *The Constancy of G and Other Gravitational Experiments Phil Trans Soc London A* **310** 227–238

Reasenberg R D, Babcock R W, Murison M A, Noecker M C, Phillips J D, Schumaker B L & Ulvestad J S 1994a *POINTS: an Astrometric Spacecraft with Multifarious Applications* in J B Breckinridge, ed., *Amplitude and Intensity Spatial Interferometry II* SPIE 2200 pp. 2–17

Reasenberg R D, Babcock R W, Phillips J D, Johnston K J & Simon R S 1994b *Newcomb, a Scientific Interferometry Mission at Low Cost* in J B Breckinridge, ed., *Amplitude and Intensity Spatial Interferometry II* SPIE 2200 pp. 18–26

Rebolo R, Martin E L, Basri G, Marcy G W & Zapatero-Osorio M R 1996 *Brown Dwarfs in the Pleiades Cluster Confirmed by the Lithium Test Astrophys J* **469** L53–L56

Reid I N 1997 *Younger and Brighter - New Distances to Globular Clusters Based on Hipparcos Parallax Measurements of Local Subdwarfs Astron J* **114** 161–179

— 1999 *The HR Diagram and the Galactic Distance Scale After Hipparcos Ann Rev Astron Astrophys* **37** 191–237
Reid I N, Kirkpatrick J D, Liebert J, Burrows A, Gizis J E, Burgasser A, Dahn C C, Monet D, Cutri R, Beichman C A & Skrutskie M 1999 *L Dwarfs and the Substellar Mass Function Astrophys J* **521** 613–629
Renzini A 1993 *Formation and Evolution of Stars in Galactic Bulges* in H DeJonghe & H Habing, eds., *Galactic Bulges* IAU Symp. 153 p. 151 Kluwer, Dordrecht
Reshetnikov V & Combes F 1998 *Statistics of Optical Warps in Spiral Disks Astron Astrophys* **337** 9–16
Reuland M, Hoogerwerf R, de Bruijne J H J & de Zeeuw P. T. 2000 *Astron Astrophys* **in press**
Rich R M 1988 *Spectroscopy and Abundances of 88 K giants in Baade's Window Astron J* **95** 828–865
— 1992 in L Blitz, ed., *In the Centre, Bulge and Disc of the Milky Way* p. 47 Kluwer
Richer H B, Fahlman G G, Ibata R A et al. 1997 *White Dwarfs in Globular Clusters: Hubble Space Telescope Observations of M4 Astrophys J* **484** 741–760
Richer H B, Fahlman G G, Rosvick J M & Ibata R 1998 *The White Dwarf Cooling Age of M67 Astrophys J* **504** L91–L94
Rickett B J 1970 *Interstellar Scintillation and Pulsar Intensity Variations Mon Not R Astr Soc* **150** 67–91
Ridge D, Becker D, Merkey P & Sterling T 1997 *Beowulf: Harnessing the Power of Parallelism in a Pile of PCs* in *Proc. IEEE Aerospace*
Rivera E & Lissauer J J 2000 *Stability Analysis of the Planetary System Orbiting υ Andromedae Astrophys J* **530** 454–463
Rix H & Zaritsky D 1995 *Nonaxisymmetric Structures in the Stellar Disks of Galaxies Astrophys J* **447** 82–102
Robichon N & Arenou F 2000 *HD 209458 Planetary Transits from Hipparcos Photometry Astron Astrophys* **355** 295–298
Robichon N, Arenou F, Mermilliod J C & Turon C 1999 *Open Clusters with Hipparcos. I. Mean Astrometric Parameters Astron Astrophys* **345** 471–484
Robin A & Crézé M 1986 *Stellar Populations in the Milky Way - A Synthetic Model Astron Astrophys* **157** 71–90
Robin A & Oblak E 1987 *A Kinematic Model of the Galaxy* in J Palous, ed., *X European IAU Assembly, Prague, Vol. 4* Czechoslovak Academy of Sciences
Rodriguez-Canabal J 1984 *SOHO Mission Analysis: Halo Orbit* Technical Report MAO 206 ESOC
Rogers F J 1994 *Equation of State of Stellar Plasmas* in G Chabrier & E Schatzman, eds., *The Equation of State in Astrophysics* IAU Coll. 147 p. 16 Cambridge University Press, Cambridge
Röser S, Bastian U, de Boer K S et al. 1997 *DIVA - Towards Microarcsecond Global Astrometry* in M A C Perryman & P L Bernacca, eds., *Hipparcos Venice 97* ESA SP-402 pp. 777–782 ESA, Noordwijk
Rosvick J M & VandenBergh D A 1998 *BV Photometry for the 2.5 Gyr Open Cluster NGC 6819: More Evidence for Convective Core Overshooting on the Main Sequence Astron J* **115** 1516–1523
Ruciński S 1994 *MV = MV(log P, log Te) Calibrations for W Ursae Majoris Systems Publ Astr Soc Pac* **106** 462–471
Rudnick G H & Rix H W 1998 *Lopsidedness in Early-Type Disk Galaxies Astron J* **116** 1163–1168
Ryan S G & Norris J E 1991 *Subdwarf Studies. II - Abundances and Kinematics from Medium Resolution Spectra. III - The Halo Metallicity Distribution Astron J* **101** 1835–1878
Ryan S G, Norris J E & Beers T C 1996 *Extremely Metal-Poor Stars. II. Elemental Abundances and the Early Chemical Enrichment of the Galaxy Astrophys J* **471** 254–278
Safa F 2000 *GAIA: Gravity and Inertia Effects in Orbit* Technical report MMS, Toulouse
Safizadeh N, Dalal N & Griest K 1999 *Astrometric Microlensing as a Method of Discovering and Characterizing Extra-Solar Planets Astrophys J* **522** 512–517
Saint-Pé O 1998 *CCD Technology Survey for GAIA* Technical Report GAIA/MMS/NT/046.98 Matra Marconi Space
Salaris M, Degl'Innocenti S & Weiss A 1997 *The Age of the Oldest Globular Clusters Astrophys J* **479** 665–672
Sanchez-Salcedo F J 1999 *Massive Dark Clusters and the Heating Problem of Galactic Discs Mon Not R Astr Soc* **303** 755–772
Sansò F, Betti B & Migliaccio F 1989 *Experiments Towards a Global Sphere Solution* in *The Hipparcos Mission: Pre-Launch Status* ESA SP-1111, Vol III pp. 437–455
Santiago B, Elson R & Gilmore G 1996 *HST Photometry of 47 Tuc and Analysis of the Stellar Luminosity Function in Milky Way Clusters Mon Not R Astr Soc* **281** 1363–1374
Santolaya-Rey A E, Puls J & Herrero A 1997 *Atmospheric NLTE-models for the Spectroscopic Analysis of Luminous Blue Stars with Winds Astron Astrophys* **323** 488–512

Sasselov D D, Beaulieu J P, Renault C et al. 1997 *Metallicity Effects on the Cepheid Extragalactic Distance Scale from EROS Photometry in the Large Magellanic Cloud and the Small Magellanic Cloud Astron Astrophys* **324** 471–482

Sazhin M V, Zharov V E, Volynkin A V & Kalinina T A 1998 *Microarcsecond Instability of the Celestial Reference Frame Mon Not R Astr Soc* **300** 287–291

Scalo J 1998 *The IMF Revisited: A Case for Variations* in G Gilmore & D Howell, eds., *The Stellar Initial Mass Function* ASP Conf. Ser. 142 p. 201 Astronomical Society of the Pacific, San Francisco

Schlesinger F 1917 *On the Secular Changes in the Proper-Motion other Elements of Certain Stars Astron J* **30** 137–138

Schmidtobreick L, Vallenari A & Bertelli G 1998 *The Structure of the Galaxy from Stellar Photometry Astronomische Gesellschaft Meeting* **14** 63

Schneider D P, Schmidt M & Gunn J E 1994 *Spectroscopic CCD Surveys for Quasars at Large Redshift. 3: The Palomar Transit Grism Survey Catalog Astron J* **107** 1245–1269

Scholz R D, Meusinger H & Irwin M 1997 *A UBV/Variability/Proper Motion QSO Survey from Schmidt Plates. I. Method and Success Rate Astron Astrophys* **325** 457–472

Schramm K J et al. 1993 *Recent Activity in the Optical and Radio Lightcurves of the Blazar 3C345: Indications for a 'Lighthouse Effect' due to Jet Rotation Astron Astrophys* **278** 391–405

Schwarzschild K 1907 *Über die Eigenbewegungen der Fixsternen Nachr Kgl Ges Wiss Gottingen Math Phys Klasse* **5** 614

Schwarzschild M 1958 *Structure and Evolution of the Stars* Princeton Univ. Press

Sciortino S, Favata F & Micela G 1995 *The Stellar Coronal Component of the Galaxy. II. An Analysis of the Stellar Content of the Einstein Extended Medium Sensitivity Survey Astron Astrophys* **296** 370–379

Searle L & Zinn R 1978 *Compositions of Halo Clusters and the Formation of the Galactic Halo Astrophys J* **225** 357–379

Seltzer S M 1980 *SHIELDOSE: a Computer Code for Space Shielding Radiation Dose Calculations* Technical Report TN-1116 NBS

Sembach K R & Savage B D 1992 *Observations of Highly Ionized Gas in the Galactic Halo Astrophys J Suppl* **83** 147–201

Sembach K R, Savage B D, Lu L & Murphy E M 1999 *Highly Ionized High-Velocity Clouds: Intergalactic Gas in the Local Group or Distant Gas in the Galactic Halo? Astrophys J* **515** 108–127

Shao M 1993 *Orbiting Stellar Interferometer* in R D Reasenberg, ed., *Spaceborne Interferometry* SPIE 1947 pp. 89–90

— 1995 *Prospects for Ground-Based Interferometric Astrometry Astrophys Space Sci* **223** 119–123

Shao M & Colavita M M 1992 *Potential of Long-Baseline Infrared Interferometry for Narrow-Angle Astrometry Astron Astrophys* **262** 353–358

Shapiro I I 1989 in X Ashby, X Bartlett & X Wyss, eds., *General Relativity and Gravitation* pp. 313–330 Cambridge University Press

Shapiro S L & Teukolsky S A 1983 *Black Holes, White Dwarfs and Neutron Stars* Wiley, New York

Smart R L, Drimmel R, Lattanzi M G & Binney J J 1998 *Unexpected Stellar Velocity Distribution in the Warped Galactic Disk Nature* **392** 471–473

Smecker-Hane T A, Stetson P B, Hesser J E & Lehnert M D 1994 *The Stellar Populations of the Carina Dwarf Spheroidal Galaxy. 1: A New Color-Magnitude Diagram for the Giant and Horizontal Branches Astron J* **108** 507–513

Soderblom D R, Duncan D K & Johnson D R H 1991 *The Chromospheric Emission-Age Relation for Stars of the Lower Main Sequence and its Implications for the Star Formation Rate Astrophys J* **375** 722–739

Söderhjelm S 1999 *Visual Binary Orbits and Masses Post Hipparcos Astron Astrophys* **341** 121–140

Soffel M 1989 in *Relativity in Astrometry* Spinger-Verlag, Heidelberg

Sofia S, Heaps W & Twigg L W 1994 *The Solar Diameter and Oblateness Measured by the Solar Disk Sextant on the 1992 September 30 Balloon Flight Astrophys J* **427** 1048–1052

Sommer-Larsen J, Beers T C, Flynn C, Wilhelm R & Christensen P R 1997 *A Dynamical and Kinematical Model of the Galactic Stellar Halo and Possible Implications for Galaxy Formation Scenarios Astrophys J* **481** 775–781

Spaenhauer A, Jones B F & Whitford A E 1992 *Proper Motions of Bulge Stars Astron J* **103** 297–302

Spitzer L 1956 *On a Possible Interstellar Galactic Corona Astrophys J* **124** 20–34

Spitzer L & Schwarzschild M 1953 *The Possible Influence of Interstellar Clouds on Stellar Velocities. II Astrophys J* **118** 106–112

Stauffer J R, Schild R, Barrado y Navascues D, Backman D E, Angelova A M, Kirkpatrick J D, Hambly N & Vanzi L 1998a *Results of a Deep Imaging Survey of One Square Degree of the Pleiades for Low-Luminosity Cluster Members Astrophys J* **504** 805–820

Stauffer J R, Schulz G & Kirkpatrick J D 1998b *Keck Spectra of Pleiades Brown Dwarf Candidates and a Precise Determination of the Lithium Depletion Edge in the Pleiades Astrophys J* **499** L199–L203
Sterken C 1992 *On the Future of Existing Photometric Systems Vistas in Astronomy* **35** 139–155
Stern S A & Colwell J E 1997 *Collisional Erosion in the Primordial Edgeworth-Kuiper Belt and the Generation of the 30-50 AU Kuiper Gap Astrophys J* **490** 879–882
Storrie-Lombardi M C & Lahav O 1994 in M A Arbib, ed., *Handbook of Brain Theory and Neural Networks* MIT Press, Boston
Stothers R B & Chin C 1995 *Tests of Two Convection Theories for Red Giant and Red Supergiant Envelopes Astrophys J* **440** 297–302
Straižys V 1992 in *Multicolour Stellar Photometry* p. 66 Pachart Publ. House, Tucson
Straizys V 1999 *Stromvil Photometry: Peculiar Stars and Anomalous Reddening Baltic Astronomy* **8** 109–121
Straižys V & Høg E 1995 *An Optimum 8-colour Photometric System for a Survey Satellite* in M A C Perryman & F van Leeuwen, eds., *Future Possibilities for Astrometry in Space* ESA SP-379 pp. 191–196 ESA, Noordwijk
Szalay A S & Brunner R J 1998 *Astronomical Archives of the Future: a Virtual Observatory* in *Future Generation Computer Systems* Elsevier
Tabachnik S & Evans N W 1999 *Cartography for Martian Trojans Astrophys J* **517** L63–L66
Tada H Y & Carter J R 1982 *Solar Cell Radiation Handbook* Technical Report 82-69 JPL
Tagliaferri G, Cutispoto G, Pallavicini R, Randich S & Pasquini L 1994 *Photometric and Spectroscopic Studies of Cool Stars Discovered in EXOSAT X-ray Images II. Lithium Abundances Astron Astrophys* **285** 272–284
Tanvir N R 1997 in M Livio, M Donahue & N Panagia, eds., *The Extragalactic Distance Scale* p. 91 Cambridge University Press, Cambridge
— 1999 *Distance Determination with Cepheid Variables* in A Heck & F Caputo, eds., *Post-Hipparcos Cosmic Candles* Astrophys. Sp. Sci. 237 pp. 17–25 Kluwer, Dordrecht
Taylor J H & Cordes J M 1993 *Pulsar Distances and the Galactic Distribution of Free Electrons Astrophys J* **411** 674–684
Terlevich E 1987 *Evolution of N-Body Open Clusters Mon Not R Astr Soc* **224** 193–225
Terndrup D M 1988 *The Structure and Stellar Population of the Galactic Nuclear Bulge Astron J* **96** 884–908
Thorsett S E & Dewey R J 1996 *Pulsar Timing Limits on Very Low Frequency Stochastic Gravitational Radiation Phys Rev D* **53** 3468–3471
Tinsley B M 1980 *Evolution of the Stars and Gas in Galaxies Fund Cosmic Phys* **5** 287–388
Tolstoy E 1995 Ph.D. thesis Groningen University, The Netherlands
Tolstoy E & Saha A 1996 *The Interpretation of Color-Magnitude Diagrams through Numerical Simulation and Bayesian Inference Astrophys J* **462** 672–683
Tombaugh C W 1961 in G P Kuiper & B M Middlehurst, eds., *Planets and Satellites* pp. 12–30 University of Chicago Press, Chicago
Torra J, Chen B, Figueras F, Jordi C & Luri X 1999 *Predicting GAIA Observations from a Star Count Model Baltic Astronomy* **8** 171–179
Torra J, Gómez A E, Figueras F et al. 1997 *Young Stars: Irregularities of the Velocity Field* in M A C Perryman & P L Bernacca, eds., *Hipparcos Venice 97* ESA SP-402 pp. 513–518 ESA, Noordwijk
Torres S, García-Berro E & Isern J 1999 *Neural Network Identification of Halo White Dwarfs Astrophys J* **508** L71–L74
Tremaine S 1993 in S Holt & F Verter, eds., *Back to the Galaxy* p. 599 AIP, New York
Tsujimoto T, Miyamoto M & Yoshii Y 1998 *The Absolute Magnitude of RR Lyrae Stars Derived from the Hipparcos Catalogue Astrophys J* **492** L79–L82
Tucholke H J & Brosche P 1995 *Globular CLusters and Dwarf Spheroidals: from Internal Dynamics to the Mass of the Galaxy* in M A C Perryman & F van Leeuwen, eds., *Future Possibilities for Astrometry in Space* ESA SP-379 pp. 77–82 ESA, Noordwijk
Tulloch S 1996 in *ESO Detector Workshop* ESO
Turon C 1999 *Hipparcos: A New Basis for Calibrating Distance Indicators* in A Heck & F Caputo, eds., *Post-Hipparcos Cosmic Candles* Astrophys. Sp. Sci. 237 p. 1 Kluwer, Dordrecht
Turon C & Perryman M A C 1999 *Global Space Astrometry: Impact on Cosmic Distance Scale* in D Egret & A Heck, eds., *Harmonizing Cosmic Distance Scales in a Post-Hipparcos Era* ASP Conf. Ser. 167 pp. 1–12 Astronomical Society of the Pacific, San Francisco
Unavane M, Gilmore G, Epchtein N, Simon G, Tiphene D & de Batz B 1998 *The Inner Galaxy Resolved at IJK Using DENIS Data Mon Not R Astr Soc* **295** 119–144

Unavane M, Wyse R F G & Gilmore G 1996 *The Merging History of the Milky Way Mon Not R Astr Soc* **278** 727–736

Vaccari M & Høg E 1999a *Simulated GAIA Observations of Galaxies* Technical Report SAG-CUO-069 Copenhagen University Observatory

— 1999b *Statistical Model of Galaxies* Technical Report SAG-CUO-061 Copenhagen University Observatory

van Altena W F, Lee J T, Hanson R B & Lutz T E 1988 *Parallax Calibration of the Population II Main Sequence II The Effect of Changes in the Corrections to Absolute Parallax* in A G Davis Philip, ed., *Calibration of Stellar Ages* p. 175 Davis Press, Schenectady, New York

van Eck S, Jorissen A, Udry S, Mayor M & Pernier B 1998 *The Hipparcos Hertzsprung-Russell Diagram of S Stars: Probing Nucleosynthesis and Dredge-Up Astron Astrophys* **329** 971–985

van Leeuwen F 1999 *Hipparcos Distance Calibrations for 9 Open Clusters Astron Astrophys* **341** L71–L74

van Leeuwen F, Feast M W, Whitelock P A & Yudin B 1997 *First Results from Hipparcos Trigonometrical Parallaxes of Mira-Type Variables Mon Not R Astr Soc* **287** 955–960

Vannier M 1998a *Correction of GAIA Chromatic Displacements using Linear Combinations of Sloan Colour Indices* Technical Report SAG-MV-003 ESTEC

— 1998b *Scintillation Effects for Microarcsec Astrometry* Technical Report SAG-MV-001 ESTEC

— 1999 *Noise of the GAIA Astrometric Instrument* Technical Report SAG-MV-004 ESTEC

Vasilevskis S, Klemola A & Preston G 1958 *Relative Proper Motions of Stars in the Region of the Open Cluster NGC 6633 Astron J* **63** 387–395

Vauterin P & Dejonghe H 1998 *On the Kinematic Signature of a Central Galactic Bar in Observed Star Samples Astrophys J* **500** 233–240

Veillet C, Ashby N, Damour T et al. 1993 *TROLL: Test of Relativity in Orbit by Laser Light. Proposal submitted to ESA in response to a call for mission ideas for M3* Technical report Observatoire de la Côte d'Azur

Velazquez H & White S D M 1999 *Sinking Satellites and the Heating of Galaxy Discs Mon Not R Astr Soc* **304** 254–270

Vette J I 1989 *Trapped Radiation Models in Development of Improved Models of the Earth's Radiation Environment* Technical Report 1, Chapter 4 NASA

Viateau B & Rapaport M 1998 *The Mass of (1) Ceres from its Gravitational Perturbations on the Orbits of 9 Asteroids Astron Astrophys* **334** 729–735

Vieira E F & Pons J D 1995 *Automated Classification of IUE Low-Dispersion Spectra. I. Normal Stars Astron Astrophys Suppl* **111** 393–398

von Hippel T, Gilmore G, Tanvir N, Robinson D & Jones D 1996 *The Metallicity Dependence of the Stellar Luminosity and Initial Mass Function: HST Observations of Open and Globular Clusters Astron J* **112** 192–200

von Hippel T, Storrie-Lombardi L, Storrie-Lombardi M C & Irwin M 1994 *Automated Classification of Stellar Spectra - Part One - Initial Results with Artificial Neural Networks Mon Not R Astr Soc* **269** 97–104

Wagner S J, Camenzind M, Dreissigacker O, Borgeest U, Britzen S, Brinkmann W, Hopp U, Schramm K J & von Linde J 1995 *Simultaneous Optical and Gamma-Ray Flaring in PKS 0420–014. Implications for Emission Processes and Rotating Jet Models Astron Astrophys* **298** 688–698

Wagner S J, Witzel A, Krichbaum T P, Wegner R, Quirrenbach A, Anton K, Erkens U, Khanna R & Zensus A 1993 *Intraday Variability in the BL Lac Object 0954+658 Astron Astrophys* **271** 344–347

Walker A R 1998 *The Distances of the Magellanic Clouds* in A Heck & F Caputo, eds., *Post-Hipparcos Cosmic Candles* Astrophys. Sp. Sci. 237 p. 125 Kluwer, Dordrecht

Walker I, Mihos J C & Hernquist L 1996 *Quantifying the Fragility of Galactic Disks in Minor Mergers Astrophys J* **460** 121–135

Walker T P, Steigman G, Schramm D, Olive K A & Kang H S 1991 *Primordial Nucleosynthesis Redux Astrophys J* **376** 51–69

Walter F M, Brown A, Mathieu R D, Myers P C & Vrba F V 1988 *X-ray Sources in Regions of Star Formation. III - Naked T Tauri Stars Associated with the Taurus-Auriga Complex Astron J* **96** 297–325

Ward W R 1997 *Survival of Planetary Systems Astrophys J* **482** L211–L214

Wasserman I & Weinberg M D 1991 *Wide Binaries in the Woolley Catalog Astrophys J* **382** 149–167

Weaver W B & Torres-Dodgen A V 1995 *Neural Network Classification of the Near-Infrared Spectra of A-Type Stars Astrophys J* **446** 300–317

— 1997 *Accurate Two-dimensional Classification of Stellar Spectra with Artificial Neural Networks Astrophys J* **487** 847–857

Weiland J L, Arendt R G, Berriman G B, Dwek E, Freudenreich H T et al. 1994 *COBE Diffuse Infrared Background Experiment Observations of the Galactic Bulge Astrophys J* **425** L81–L84

Weinberg M 1999 *Effects of the Magellanic Clouds on the Milky Way Disk and VICE* in B K Gibson, T S Axelrod & M E Putman, eds., *The Third Stromlo Symposium: The Galactic Halo: Bright Stars and Dark Matter* ASP Conf. Ser. 165 p. 100 Astronomical Society of the Pacific, San Francisco

— 2000 *Astron J* **in press**

Wetherill G W 1994 *Possible Consequences of Absence of 'Jupiters' in Planetary Systems Astrophys Space Sci* **212** 23–32

Whitelock P, Feast M & Catchpole R 1991 *IRAS Sources and the Nature of the Galactic Bulge Mon Not R Astr Soc* **248** 276–312

Whitelock P A, van Leeuwen F & Feast M W 1997 *The Luminosities and Diameters of Mira Variables from Hipparcos* in M A C Perryman & P L Bernacca, eds., *Hipparcos Venice 97* ESA SP-402 pp. 213–218 ESA, Noordwijk

Whitney C A 1983 *Principal Components Analysis of Spectral Data. I - Methodology for Spectral Classification Astron Astrophys* **51** 443–461

Wielen R 1995 *Statistical Aspects of Stellar Astrometry and their Implication for High-Precision Measurements* in M A C Perryman & F van Leeuwen, eds., *Future Possibilities of Astrometry in Space* volume ESA SP-379 pp. 65–68 ESA, Noordwijk

— 1997 *Principles of Statistical Astrometry Astron Astrophys* **325** 367–382

Wielen R, Dettbarn C, Jahreiss H, Lenhardt H & Schwan H 1999 *Indications on the Binary Nature of Individual Stars Derived from a Comparison of their Hipparcos Proper Motions with Ground-Based Data. I. Basic Principles Astron Astrophys* **346** 675–685

Wielen R, Jahreiß H & Kruger R 1983 *The Determination of the Luminosity Function of Nearby Stars* in A G David Philip & A R Upgren, eds., *The Nearby Stars and Stellar Luminosity Function* IAU Coll. 76 Davis Press, Schenectady, New York

Wilkinson M I & Evans N W 1999 *The Mass Of The Milky Way Halo* in D R Merritt, M Valluri & J A Sellwood, eds., *Galaxy Dynamics,* ASP Conf. Ser. 182 p. 371 Astronomical Society of the Pacific, San Francisco

Will C M 1981 *Theory and Experiment in Gravitational Physics* Cambridge University Press

— 1987 in S W Hawking & W Israel, eds., *300 Years of Gravitation* p. 80 Cambridge University Press, Cambridge

Williams J G, Newhall X X & Dickey J O 1996 *Relativity Parameters Determined from Lunar Laser Ranging Phys Rev D* **53** 6730–6739

Windhorst R A, Franklin B E & Neuschaefer L W 1994 *Removing Cosmic-Ray Hits from Multi-Orbit HST Wide Field Camera Images Publ Astr Soc Pac* **106** 798–806

Winget D, Kepler S O, Kanaan A, Montgomery M H & Giovannini O 1997 *An Empirical Test of the Theory of Crystallization in Stellar Interiors Astrophys J* **487** L191–L194

Winget D E, Hansen C J, Liebert J, van Horn H M, Fontaine G, Nather R E, Kepler S O & Lamb D Q 1987 *An Independent Method for Determining the Age of the Universe Astrophys J* **315** L77–L81

Woolf N J & Angel J R P 1998 *Astronomical Searches for Earth-Like Planets and Signs of Life Ann Rev Astron Astrophys* **36** 507–537

Wyse R F G 1997 *From Proper Motions to Galaxy Formation and Evolution* in R M Humphreys, ed., *Proper Motions and Galactic Astronomy* ASP Conf. Ser. 127 pp. 183–190 Astronomical Society of the Pacific, San Francisco

Wyse R F G & Gilmore G 1995 *Chemistry and Kinematics in the Solar Neighborhood: Implications for Stellar Populations and for Galaxy Evolution Astron J* **110** 2771–2787

Wyse R F G, Gilmore G & Franx M 1997 *Galactic Bulges Ann Rev Astron Astrophys* **35** 637–675

Ya'ari A & Tuchman Y 1996 *Long-Term Nonlinear Thermal Effects in the Pulsation of Mira Variables Astrophys J* **456** 350–355

— 1999 *On the Pulsation Mode of Mira Variables: Nonlinear Calculations Compared with Radii Observations Astrophys J* **514** L35–L37

Yoshizawa M, Sato K, Nishikawa J, Fukushima T & Miyamoto M 1997 *An Optical/Infrared Astrometric Satellite Project LIGHT* in M A C Perryman & P L Bernacca, eds., *Hipparcos Venice 97* ESA SP-402 pp. 795–798 ESA, Noordwijk

Zappalà V & Cellino A 2000 *On a Possible NEO Experiment with GAIA* Technical Report SWG-VZ-001 Torino Astronomical Observatory

Zaritsky D 1999 *The Mass and Extent of the Galactic Halo* in B K Gibson, T S Axelrod & M E Putman, eds., *The Third Stromlo Symposium: The Galactic Halo: Bright Stars and Dark Matter* ASP Conf. Ser. 165 p. 34 Astronomical Society of the Pacific, San Francisco

Zaritsky D, Harris J & Thompson I 1997 *A Digital Photometric Survey of the Magellanic Clouds: First Results from One Million Stars Astron J* **114** 1002–1013
Zensus A 1997 *Parsec-Scale Jets in Extragalactic Radio Sources Ann Rev Astron Astrophys* **35** 607–636
Zhao H 1998 *The Survival of the Sagittarius Dwarf Galaxy and the Flatness of the Rotation Curve of the Galaxy Astrophys J* **500** L149–L152
Zhao H S 1996 *A Steady-State Dynamical Model for the COBE-Detected Galactic Bar Mon Not R Astr Soc* **283** 149–166
Zhao H S & de Zeeuw P T 1998 *The Microlensing Rate and Mass Function Versus Dynamics of the Galactic Bar Mon Not R Astr Soc* **297** 449–461
Zhao H S, Rich R M & Biello J 1996 *Proper-Motion Anisotropy, Rotation, and the Shape of the Galactic Bulge Astrophys J* **470** 506–512
Zhao H S, Spergel D N & Rich R M 1994 *Signatures of the Bulge Triaxiality from Kinematics in Baade's Window Astron J* **108** 2154–2163

During the Concept and Technology Study for GAIA, approximately 200 scientific and technical notes were prepared by members of the GAIA Scientific Advisory Group, and by members of the various Working Groups. All are included in the Documentation Management System (DMS) maintained in ESTEC for the GAIA project, accessible to participating scientists via the WWW.

Part VII
Index