

GAIA

Composition, Formation and Evolution of the Galaxy

Results of the Concept and Technology Study

compiled by the
GAIA Science Advisory Group

Draft Version 1.6

March 31, 2000

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 assessments, 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

I	Scientific Case	21
1	Science Goals	23
1.1	Overview	23
1.2	Structure and Dynamics of the Galaxy	24
1.2.1	Galactic Structure	24
1.2.2	The Bulge	25
1.2.3	Large-Scale Structure of the Disk	27
1.2.4	Small-Scale Structure of the Disk	31
1.2.5	The Thick Disk	36
1.2.6	The Stellar Halo	38
1.2.7	Globular Clusters	42
1.2.8	Dark Matter and the Mass of the Milky Way	46
1.2.9	The Stellar Radiation Field and the Interstellar Medium	49
1.3	The Star Formation History of the Milky Way	52
1.3.1	Method	52
1.3.2	Disk, Bulge and Halo	54
1.3.3	Recent Birth-Rate of Stars in the Solar Neighbourhood	56
1.3.4	Local Luminosity Function	56
1.3.5	Initial Mass Function in Clusters and Associations	57
1.4	Stellar Astrophysics	59
1.4.1	Stellar Structure and Evolution	59
1.4.2	Luminosities Through Distances	61
1.4.3	Variability	61
1.4.4	Physics of Stellar Interiors	62
1.4.5	Stellar Ages, Galactic Evolution, and Age of the Universe	64
1.4.6	Isolated Brown Dwarfs	66
1.4.7	White Dwarfs	67
1.4.8	Other Specific and Rare Stellar Types	69
1.4.9	Cosmic Distance Scale	72
1.4.10	Masses from Microlensing	75
1.5	Binaries and Multiple Stars	78
1.5.1	Census of Binaries	78
1.5.2	Proper Motion Bias from Undetected Binaries	81
1.5.3	Masses from Visual Binaries	81
1.5.4	Fundamental Data from Spectroscopic and Eclipsing Binaries	82

1.5.5	Interacting Binaries, Accretion Disks, and Black Holes	83
1.5.6	Low-Mass Stars	85
1.6	Brown Dwarfs and Planetary Systems	86
1.6.1	Brown Dwarfs in Binaries	87
1.6.2	Planetary Systems	87
1.6.3	Planet Detection and Orbit Determination	88
1.6.4	Formation and Stability of Multi-Planet Systems	91
1.6.5	Transits from Photometric Data	92
1.7	Solar System	92
1.7.1	Minor Planets	92
1.7.2	Trojans in the Inner Solar System	94
1.7.3	Trans-Neptunian Objects: the Kuiper Belt	95
1.7.4	Stellar Encounters and Perturbations of the Oort Cloud	98
1.8	Galaxies, Quasars, and the Reference Frame	99
1.8.1	The Magellanic Clouds	99
1.8.2	Internal Dynamics of the Dwarf Satellites of the Milky Way	101
1.8.3	Rotational Parallax Distances to the LMC, M33, and M31	102
1.8.4	Warps in M31 and M33	103
1.8.5	Orbits in the Local Group: Gravitational Instability in the Early Universe .	103
1.8.6	Galaxies	104
1.8.7	Supernovae	106
1.8.8	Photocentric Motions from Relativistic Jets of Active Galactic Nuclei . . .	107
1.8.9	Quasars	107
1.8.10	Galactocentric Acceleration	110
1.8.11	The Radio/Optical Reference Frame	110
1.9	Fundamental Physics: General Relativity	112
1.9.1	Light Bending in the Solar System	112
1.9.2	Perihelion Precession of Minor Planets	115
1.9.3	Secular Change of the Gravitational Constant	116
1.9.4	Cosmological Shear and Rotation, Mach's Principle	117
1.9.5	Gravitational Waves	117
1.10	Scientific Topics Beyond GAIA	120

II Technical Design 123

2 Overall Design Considerations and Concept Development 125

2.1	Astrometry	125
2.1.1	From Milliarcseconds to Microarcseconds	125
2.1.2	Theoretically Achievable Accuracy	127
2.1.3	Matching Optics and Detector	129
2.1.4	Astrometric Requirements	130

Executive Summary

The GAIA mission has been under study by ESA over the past two 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:

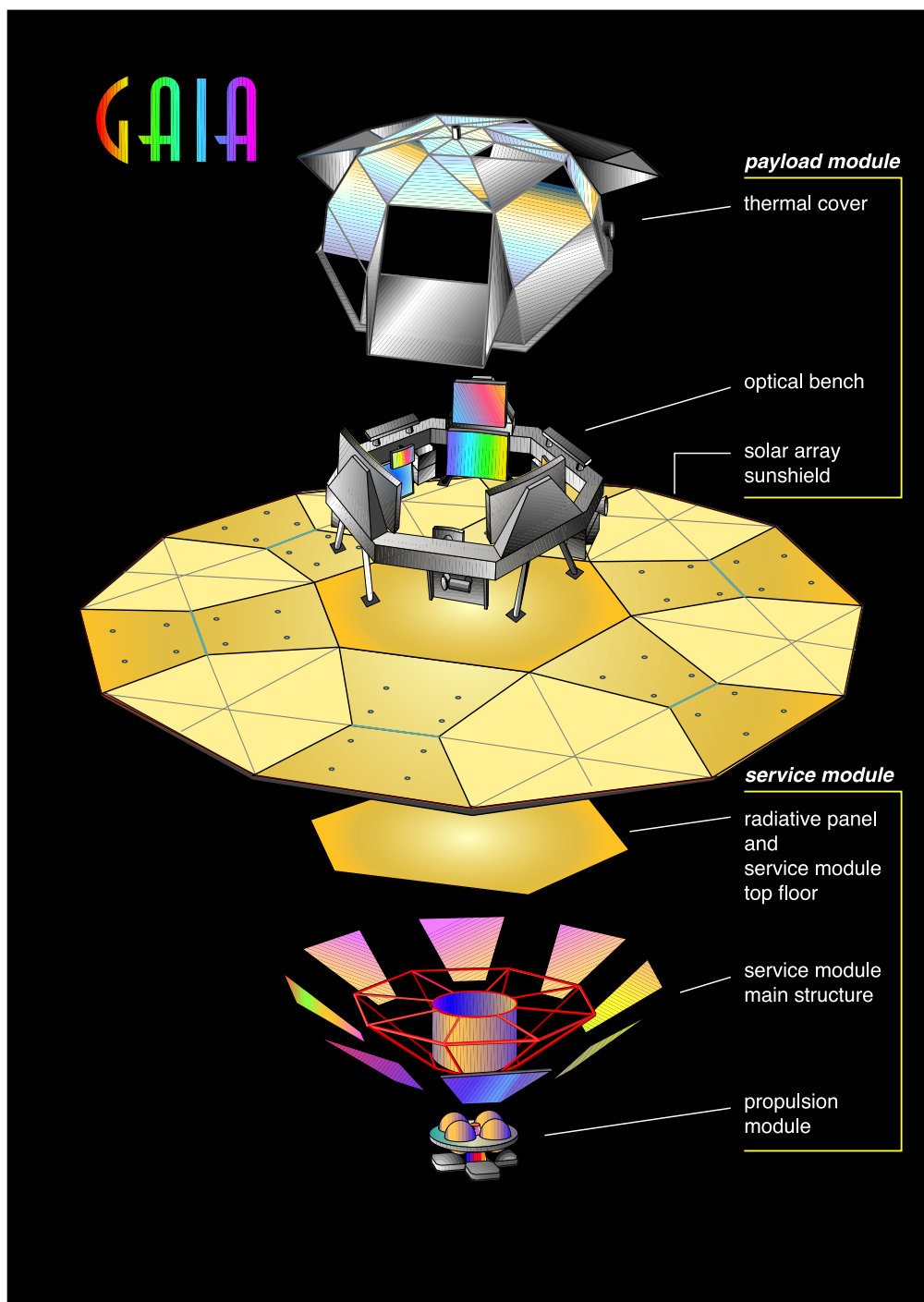
Part I: Science Case Section 1 presents the science case for GAIA. The central theme is to derive the formation and evolution of the Galaxy through a massive stellar population census covering all types of object.

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: $4\,\mu\text{as}$ at 10 mag; $11\,\mu\text{as}$ at 15 mag; $160\,\mu\text{as}$ 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 $0.5\,\text{km s}^{-1}$; 85 million better than $1\,\text{km s}^{-1}$; 210 million better than $3\,\text{km s}^{-1}$; 300 million better than $5\,\text{km s}^{-1}$; 440 million better than $10\,\text{km s}^{-1}$
- **radial velocity accuracies:** $1\text{--}10\,\text{km s}^{-1}$ to $V = 16\text{--}17$ 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 $\sim 5 \times 10^{-7}$; β to $3 \times 10^{-4}\text{--}3 \times 10^{-5}$; solar J_2 to $10^{-7}\text{--}10^{-8}$; $\dot{G}/G$ to $10^{-12}\text{--}10^{-13}\,\text{yr}^{-1}$; constraints on gravitational wave energy for $10^{-12} < f < 4 \times 10^{-9}\,\text{Hz}$; constraints on Ω_M and Ω_Λ from quasar microlensing
- **specific objects:** $10^6\text{--}10^7$ resolved galaxies; 10^5 extragalactic supernovae; 500 000 quasars; $10^5\text{--}10^6$ (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: $1.7 \times 0.7 \text{ m}^2$
 - 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: $\sim 200 \text{ K}$
 - focal plane detectors: CCDs operating in TDI mode
 - pixel size along-scan: $9 \mu\text{m}$
- spectrometric instrument:
 - spectrometer for radial velocity measurements + medium-band photometry
 - entrance pupil: $0.75 \times 0.70 \text{ m}^2$
 - field of view: 2.6 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 = $120 \text{ arcsec s}^{-1}$, 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 ($9\ \mu\text{m}$) 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 ~ 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 & Perryman 1995; 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 μ as at 15 mag, and a target list of more than 50 million objects. The substantial science case for a μ as 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 μ as 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 μ as 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.

¹The 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 $100 \mu\text{as yr}^{-1}$, 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 $250 \mu\text{as yr}^{-1}$. 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 $\mu\text{as yr}^{-1}$). 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} $\mu\text{as/yr}$	(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 tri-axial 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; Freudreich 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, $0.008 \leq Z \leq 0.05$).

Mass Model	Axial ratios	Major Axis (kpc)	Bar Position Angle ($^{\circ}$)	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 $[\alpha\text{-elements}/\text{Fe}] = 0.0$ 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 $[\text{Fe}/\text{H}]$ is about -0.25 , with a large spread, ranging from $[\text{Fe}/\text{H}] = -1$ to 0.45 . For an old population, an uncertainty as large as $\Delta[\text{Fe}/\text{H}] = 0.1$ 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).

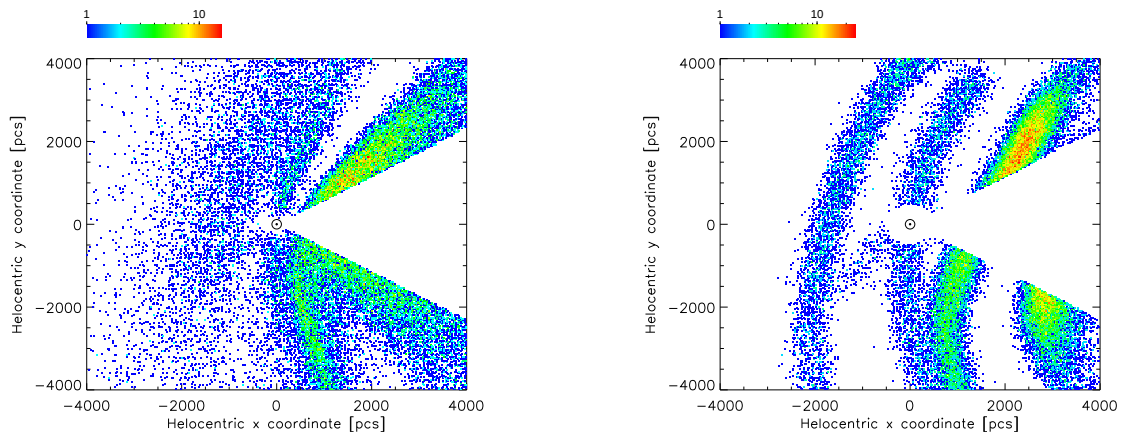


Figure 1.1: Distribution in the Galactic plane of a hypothetical spectroscopic and photometric statistically complete survey of 50 000 OB stars (coordinates are heliocentric, with the Galactic Centre at 8000,0) with photometric distances (left) and with GAIA parallaxes (right), as generated by a synthetic 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 ($\sim 10 \text{ km s}^{-1}$), 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 ~ 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 breakthrough. 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

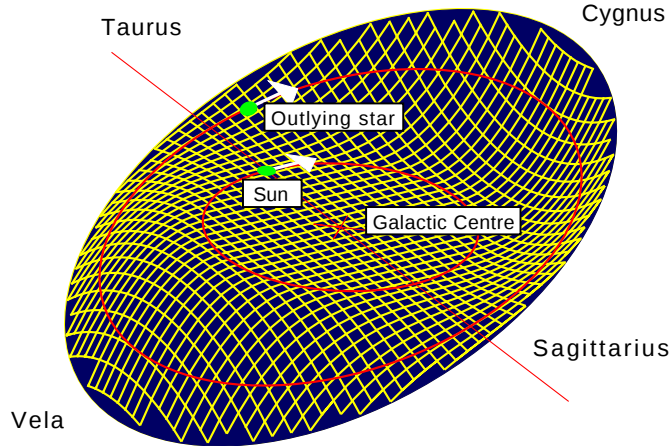


Figure 1.2: The disk of the Galaxy based on HI observations (from Smart et al. 1998). The vertical axis is exaggerated by a factor of 10. The arrows show the motion of the Sun and an outlying star on their orbits in the Galaxy. The outlying star has an upward motion as seen from the Sun. Directions in 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 ($\ell = 90^\circ$) the HI disk curls steadily up to over 3 kpc above the plane at $R \sim 16$ kpc. In the south ($\ell = 270^\circ$) 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 ~ 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 $\sim 90 \mu\text{as yr}^{-1}$ in latitude and $\sim 600 \mu\text{as yr}^{-1}$ 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 $100 \mu\text{as}$, so that resolution of the warp within 10 per cent implies distance accuracies of $10 \mu\text{as}$ 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 $N(\text{H I})$, 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 μ as (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 (Lyngå 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.

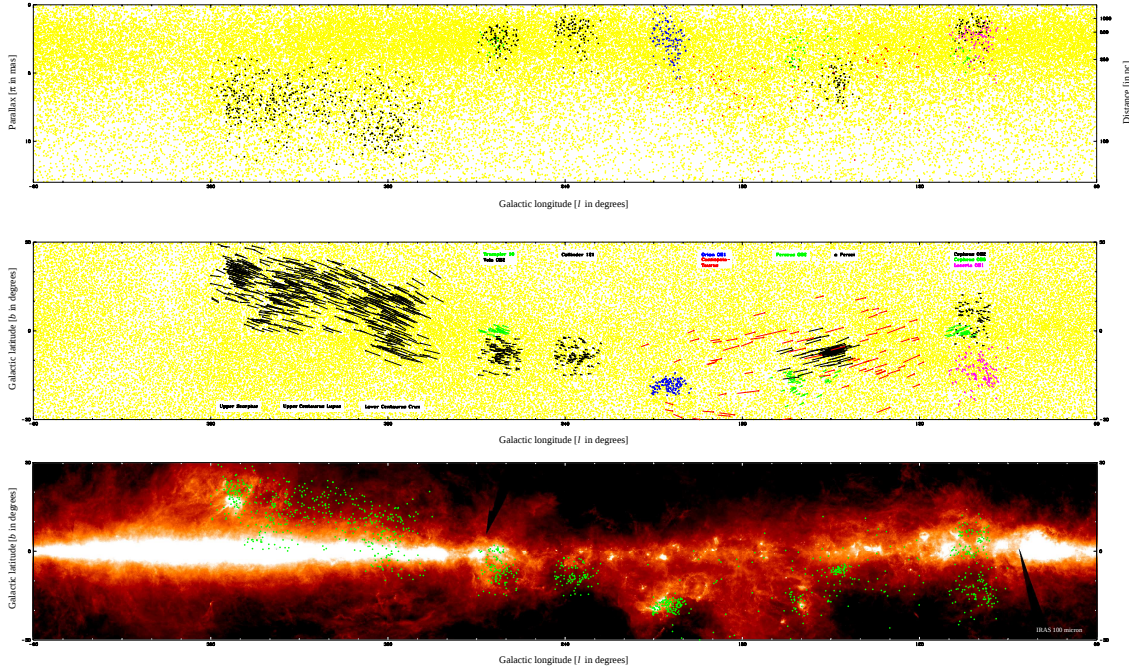


Figure 1.3: Kinematic selection of nearby associations using Hipparcos positions, parallaxes, and proper motions: (a) parallaxes of the OB association members, superimposed on all stars observed by Hipparcos in the range $-30^\circ < b < 30^\circ$; (b) positions and proper motions of the members; (c) 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 granular 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 ($4\text{--}8\text{ km s}^{-1}$) 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 $r < 300\text{ pc}$, 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 (Linblad 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 $\sim 100\text{ pc}$, while the more distant part should be at a maximum distance smaller than $800\text{--}1000\text{ pc}$. 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 & Chernin 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.

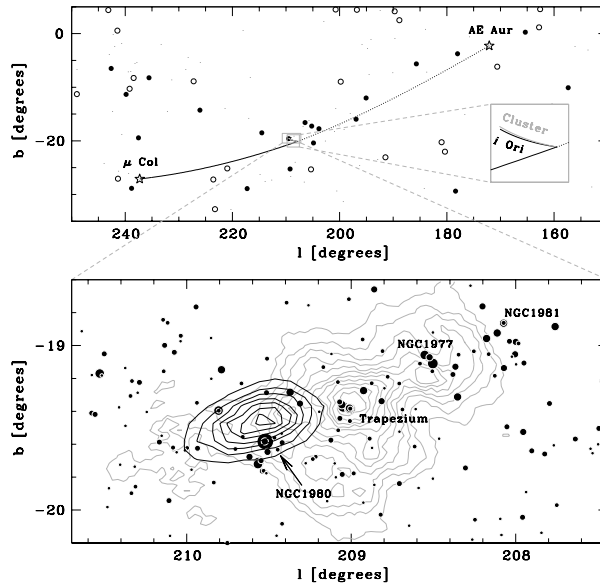


Figure 1.4: Top: orbits, calculated back in time, of the runaways AE Aur (dotted line) and μ Col (solid line) and the binary ι Ori projected on the sky. The starred symbols depict the present position of the two runaways. The stars met ~ 2.2 Myr ago. Using the conservation of linear momentum principle 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 μ m 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 occurrence 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 $(M_{\text{galaxy}}/M_{\text{satellite}}) \cdot t_{\text{cross}}$, where t_{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 occurring 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 $[\text{Fe}/\text{H}] < -3.5$ (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.

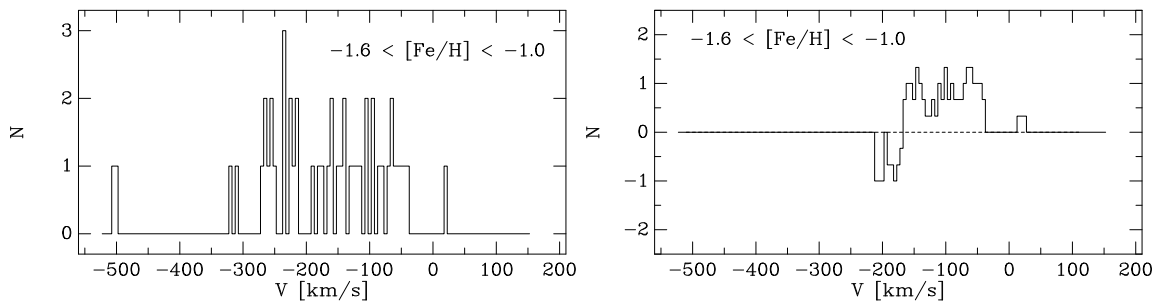


Figure 1.5: A sample of local subdwarfs with *Hipparcos* parallaxes and proper motions from Figure 3 of Fuchs et al. (1998), with metallicities and radial velocities from Carney et al. (1994). Left: all stars in the metallicity range indicated. Right: the ‘excess population’ over a symmetric distribution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

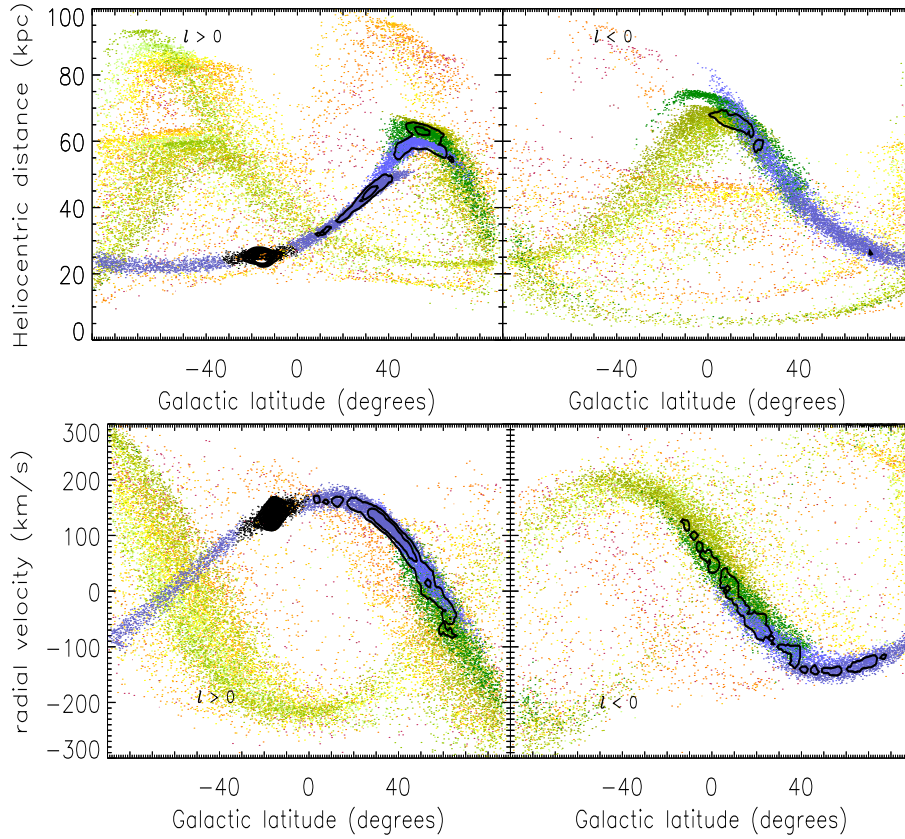


Figure 1.6: An example of the phase-space distribution of stars originating in a disrupted satellite for the case of Sagittarius (Helmi & White 2000). Top: the predicted distribution in heliocentric distance versus latitude. Bottom: the predicted distribution in radial velocity versus latitude. Different 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 $4\text{--}26 \mu\text{as}$ and $3\text{--}16 \mu\text{as yr}^{-1}$ respectively, the uncertainty in two of the velocity components is in the range $1\text{--}5 \text{ km s}^{-1}$, whereas for the radial velocity (third component) it is $3\text{--}10 \text{ km s}^{-1}$. 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).

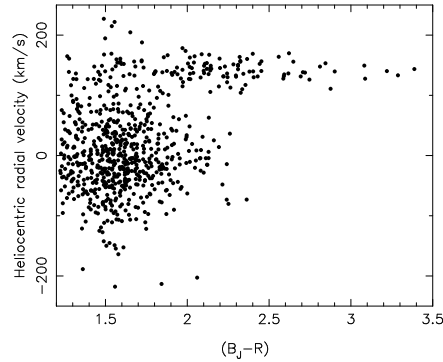


Figure 1.7: The discovery colour-velocity phase-space slice for Sgr, illustrating the need for, and efficiency of, the combination of kinematics and photometry in isolating Sgr and foreground bulge members (Ibata et al. 1995).

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 successfully 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 $\sim 6 \times 10^8 M_\odot$ 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 $2100 \mu\text{as yr}^{-1}$ (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 ($I = 14 - 15 \text{ mag}$) is $40 \mu\text{as}$. While the measuring error ($10\text{--}20 \mu\text{as}$) 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 $20 \mu\text{as}$, 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 $37 \pm 7 \text{ kpc}^{-3}$ (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 an 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).

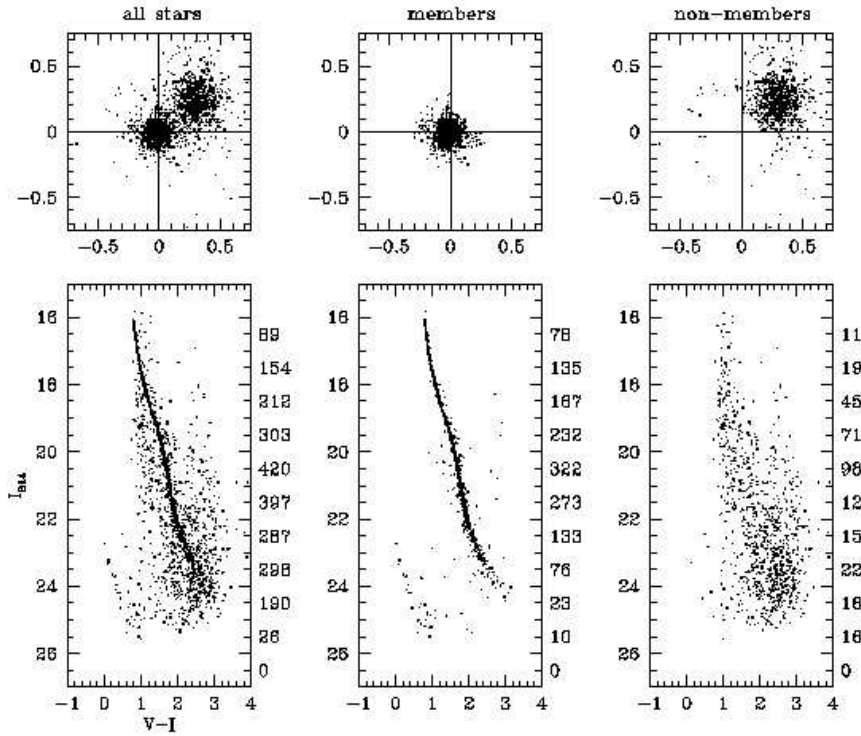


Figure 1.8: Kinematic member selection illustrated by Hubble Space Telescope data for NGC 6397, from King *et al.* (1998).

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 $\mu\text{as s}^{-1}$) 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 $10 \mu\text{as yr}^{-1}$ 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 $V = 15 \text{ mag}$ (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 $50 \mu\text{as}$ 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).

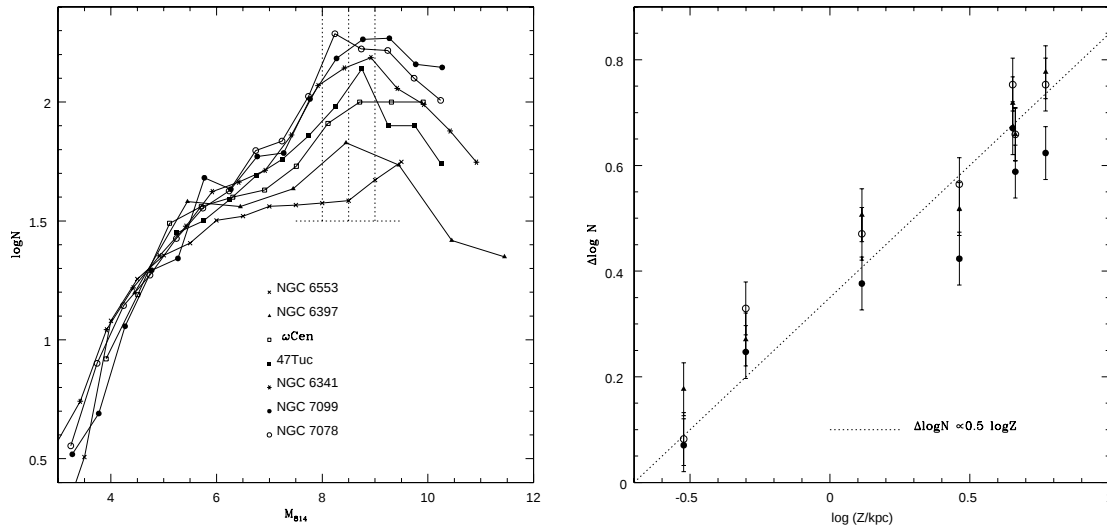


Figure 1.9: (a) *HST* luminosity functions for seven Galactic globular clusters; (b) relative flattening of the faint end of the luminosity functions in (a), measured at $M_{814} = 8.0, 8.5, 9.0$ (filled circles, open circles, triangles) plotted against distance above the Galactic plane.

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 ($-0.3 < [\text{Fe}/\text{H}] < -2.5$ 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.

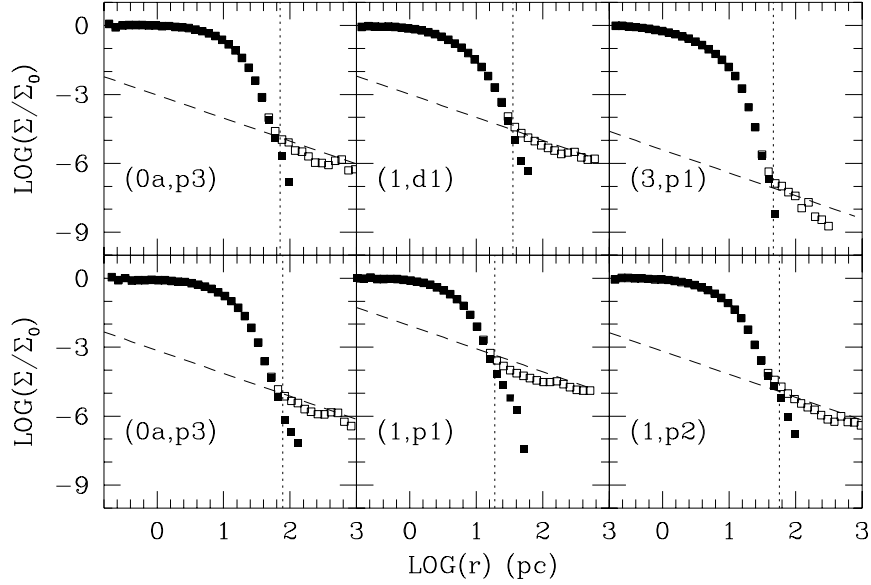


Figure 1.10: Simulations of a variety of realistic globular clusters evolving in the Galactic tidal field. In each case the solid points indicate stars bound to the cluster, open squares stars which have been unbound by the Galactic tide, and the vertical dotted line what would be classified as the ‘tidal 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 $\sim 20 \text{ km s}^{-1}$; for Sgr the radial velocity dispersion is 11 km s^{-1} , while the inner Galactic disk dispersion is $\sim 50 \text{ km s}^{-1}$. A single point accuracy of $< 5 \text{ km s}^{-1}$ is then sufficient to distinguish cluster and field stars. This corresponds to $40 \mu\text{as yr}^{-1}$ for Sgr, $20 \mu\text{as yr}^{-1}$ in the LMC and SMC, and $100 \mu\text{as yr}^{-1}$ 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 $v_c \sim 180 \text{ km s}^{-1}$ (Zaritsky 1999). Out to a distance of $\sim 300 \text{ kpc}$ —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 $M_\odot \text{ pc}^{-3}$, 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 $M_\odot \text{ pc}^{-2}$, 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 $\sim 50 M_\odot \text{ pc}^{-2}$ (Kuijken & Gilmore 1989b; Kuijken & Gilmore 1989a; Bahcall et al. 1992). The best recent determination of ρ_o , using Hipparcos data, puts it at $\sim 0.1 M_\odot \text{ pc}^{-3}$ (Cr     et al. 1998; Bienaym   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 $v_{c,\text{disk}} \sim 150 (\Sigma_{\text{local}}/60 M_\odot \text{ pc}^{-2})^{1/2} \text{ km s}^{-1}$. The local circular velocity is $\sim 220 \text{ km s}^{-1}$, 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, Cr     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.

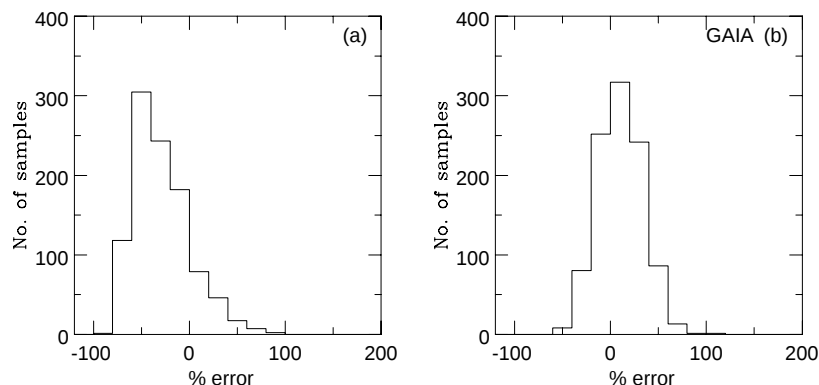


Figure 1.11: Histograms showing how many out of 1000 data sets yielded a given percentage error in the mass of the Milky Way. The left panel shows the present-day situation with 30 data points with radial velocities only, the right panel shows the situation after GAIA with 30 data points with radial 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 $10 \mu\text{s yr}^{-1}$ in proper motion accuracy at $V = 15$ mag and $100\text{--}200 \mu\text{s yr}^{-1}$ 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 $\sim 240 \text{ km s}^{-1}$. 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 $\sim 1.7 \text{ arcmin}^2$. There are ~ 50 stars brighter than $V = 20$ mag visible on the colour-magnitude diagram. Leo I subtends perhaps $\sim 10 \text{ arcmin}^2$ 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 $\sim 14 \text{ km s}^{-1}$ 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 $\sim 1 \text{ km s}^{-1}$. 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 ~ 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. HI 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 HI column density $N_H = \int n_H ds$ where n_H is the atomic hydrogen density. When an HI 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 HI 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 H₂ has no electric dipole moment, radiative transitions of H₂ 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 H₂ 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 ($\sim 10^{-3} \text{ cm}^{-3}$), hot gas ($\sim 10^6 \text{ K}$), 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 $\sim 9 \text{ km s}^{-1}$) 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 $Ly\alpha$ 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 $H\alpha$ 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 ($\propto$ distance²) 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 $H\alpha$ recombination flux. If the HI is opaque, the $H\alpha$ 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.

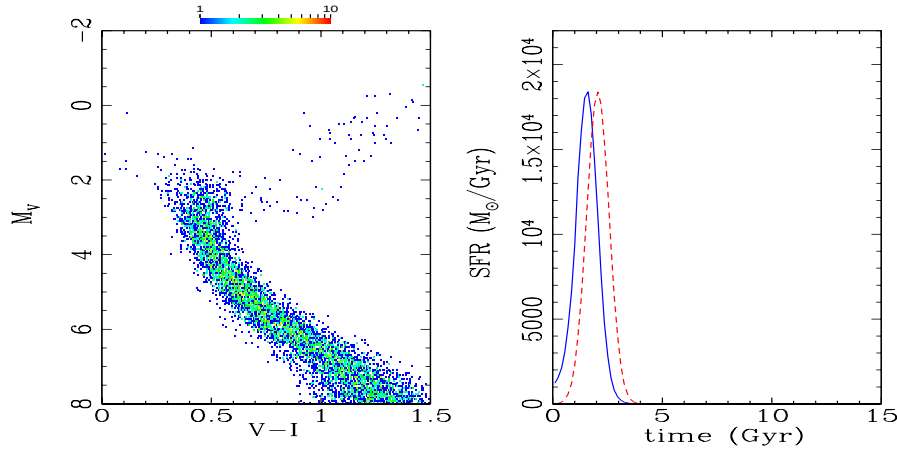


Figure 1.12: Left: synthetic H - R diagram for an ‘arm’ component, using isochrones of Solar metallicity, and including a 10 per cent observational error. Right: inferred star formation rate versus time for the ‘arm’ component, assuming a metallicity slightly off, of $[\text{Fe}/\text{H}] = 0.18$, solid 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.

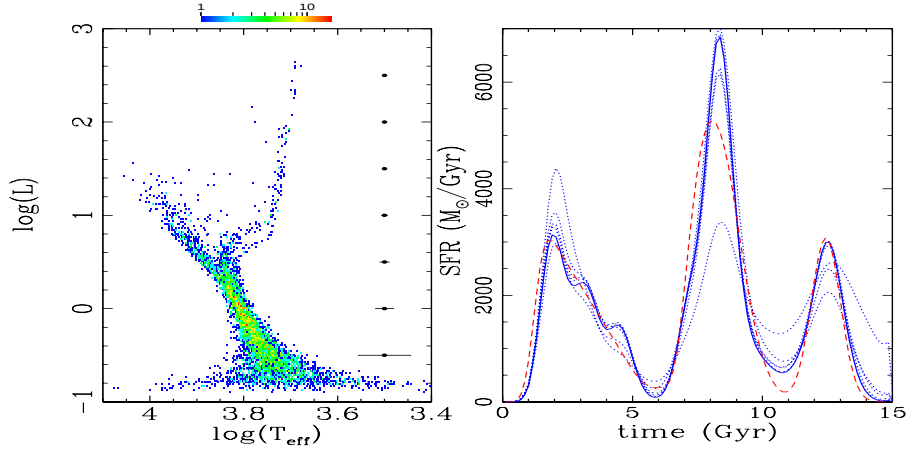


Figure 1.13: Left: synthetic H - R diagram appropriate for good quality GAIA data. The length of the bars to the right corresponds to 2σ in the $\log(L)$ error to each side of the dots. Right: the derived star formation history following inversion of the data in the left panel. The true input star formation 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.

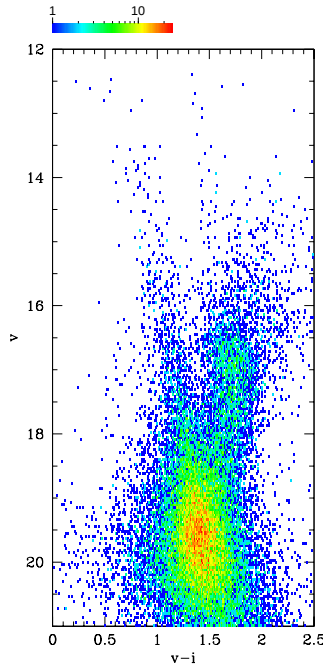


Figure 1.14: An observed colour-magnitude diagram for a field in Baade’s Window, with similar completeness and photometric accuracy to the GAIA data. The bulge turnoff appears at $V \sim 19.5$ mag. The stars in the blue ‘plume’ from $(V, V-I) = (12, 0.5)$ to $(19, 1.5)$ are a mixture of foreground young 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 $25 \text{ km s}^{-1} \text{ kpc}^{-1}$ (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 ~ 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.

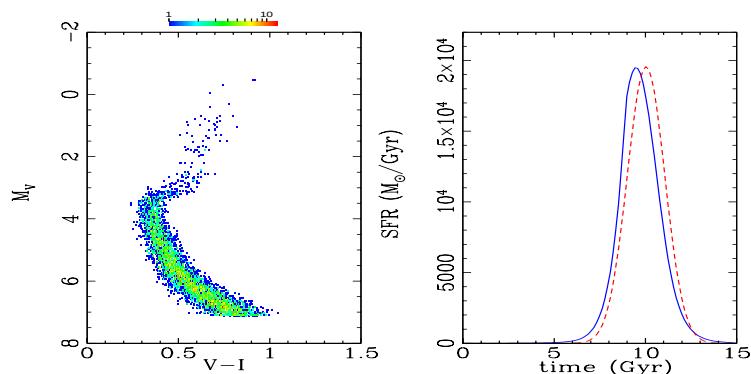


Figure 1.15: Left: synthetic H-R diagram for an ‘old bulge’ component, using isochrones of $[Fe/H] = -1.5$, and including a 10 per cent observational error. Right: the inferred star formation history, assuming a correct metallicity of $[Fe/H] = -1.5$, solid curve. The dotted curve shows the model history. 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 Neighbourhood

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 ~ 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 ~ 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

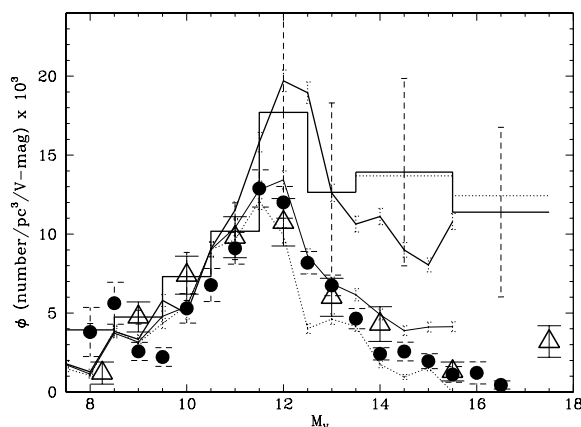


Figure 1.16: The present status of knowledge of the Solar neighbourhood stellar luminosity function. The various curves are as follows: solid histogram: derived from stars within 5.2 pc, contained in the Gliese catalogue (Kroupa 1995); dotted histogram, as for the solid histogram, but illustrating the 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.

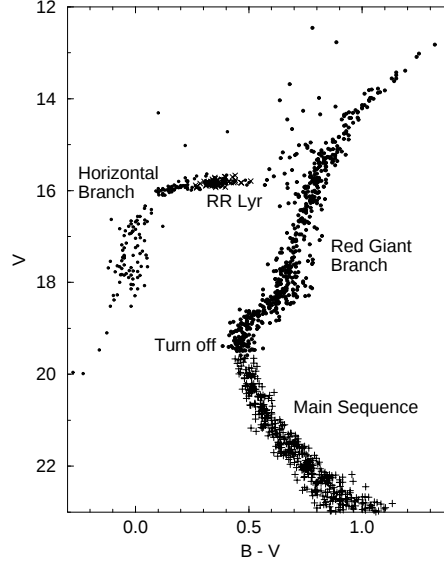


Figure 1.17: Colour-magnitude diagram of the globular cluster M15 (Durrell & Harris 1993).

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_{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 oscillation 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_{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_{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). $N(\text{HIP})$ is the approximate number of Hipparcos stars with the given relative accuracy, while $N(\text{GAIA})$ 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	2700
			0	12	2700
			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	2700
			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 photometry 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 physics 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_{lim} [mag]	V_{lim} [mag]	d_{lim} [pc]	Limiting factor
-5	O V	-0.3	12.2	12.2	27 000	d
	B0-G0 Ib	to				
	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 viscous 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 asteroseismological 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 $0.7\text{--}0.8 M_{\odot}$ 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

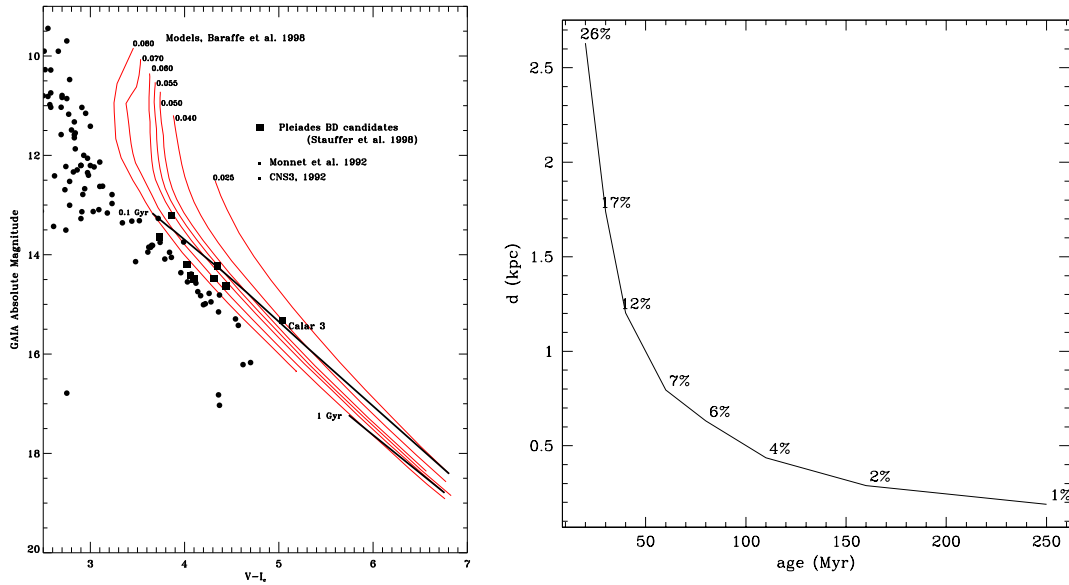


Figure 1.18: (a, left) The bottom of the H-R diagram. Evolutionary sequences of brown dwarfs from Baraffe et al. (1998), labeled with their masses ($M_{\odot}$), are plotted together with data for M dwarfs (dots), brown dwarf candidates in the Pleiades (squares) from Stauffer et al. (1998a) and Calar 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 $\sim 0.1 \text{ mas yr}^{-1}$ and parallaxes accurate to $\sim 0.1 \text{ mas}$. Labels give the percentage accuracy $\Delta\pi/\pi$ for $I \sim 20 \text{ mag}$ 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 $0.003 M_{\odot} \text{ pc}^{-3}$, 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, $< 80 \text{ pc}$, 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 successfully 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, $\dot{G}$. Using the white dwarf luminosity function García-Berro et al. (1995) derived an upper bound of $\dot{G}/G \leq -(1 \pm 1) \times 10^{-11} \text{ yr}^{-1}$, 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 ($\lesssim 4 \cdot 10^6$ 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 ($0.52\text{--}0.56 M_\odot$) 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 ~ 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 ~ 20 with $\sigma_\pi/\pi < 10$ % per star ~ 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 $[\text{Fe}/\text{H}] = -1.6$	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	A+
Cluster kinematics	0.67 ± 0.10	B+	small number statistics 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 dependence of Cepheids properties. In addition to the zero-point calibration, another controversy in the Cepheid luminosity calibration is the issue of metallicity dependence (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 foreseeable 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

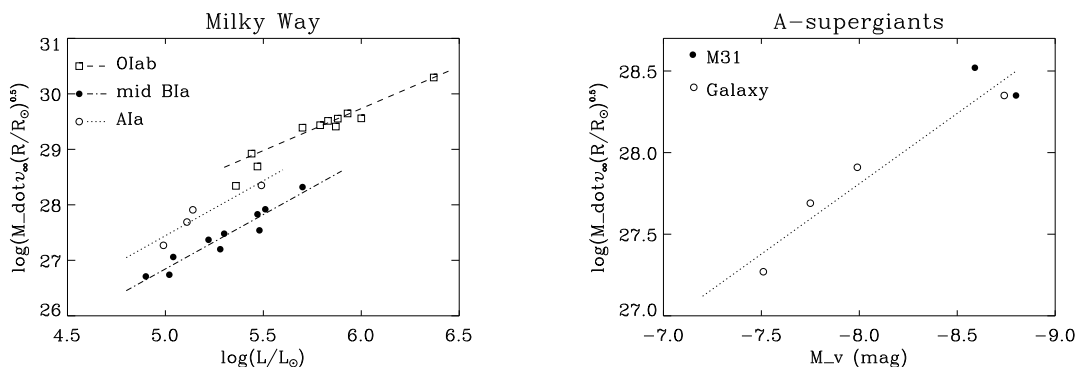


Figure 1.19: (a) The observed wind-momentum luminosity relationship of the brightest galactic supergiants of spectral type O, mid B and early A. Data from Puls et al. (1996) and Kudritzki et al. (1999). Wind momenta are given in cgs-units. (b) Wind momenta of two A-supergiants in M31 and galactic A-supergiants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

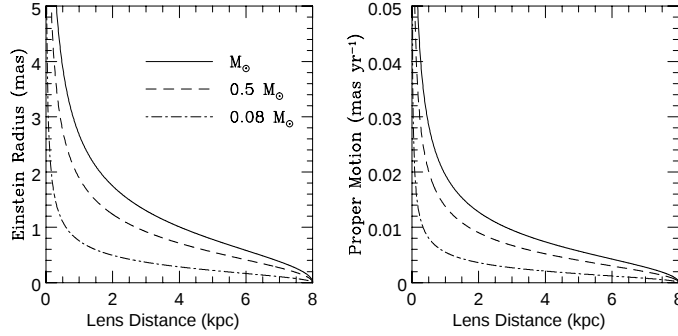


Figure 1.20: The left panel shows the Einstein ring radius and the right panel the relative proper motion between the lens and the source as a function of the lens distance. The source star is in the bulge and the different curves correspond to different lenses with masses $1M_{\odot}$, $0.5M_{\odot}$ 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 τ_{as} is related to the photometric optical depth by $\tau_{\text{as}} \sim \tau_{\text{ph}} \theta_{\text{E}}^2 / \theta_{\text{min}}^2$ (Miralda-Escude 1996; Paczyński 1998), where θ_{E} is the angular Einstein radius, and an event is detected if the maximum angular deflection is greater than θ_{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 τ_{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 $D_s \sim 8$ 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 θ_{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 parallax 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 parallax 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.

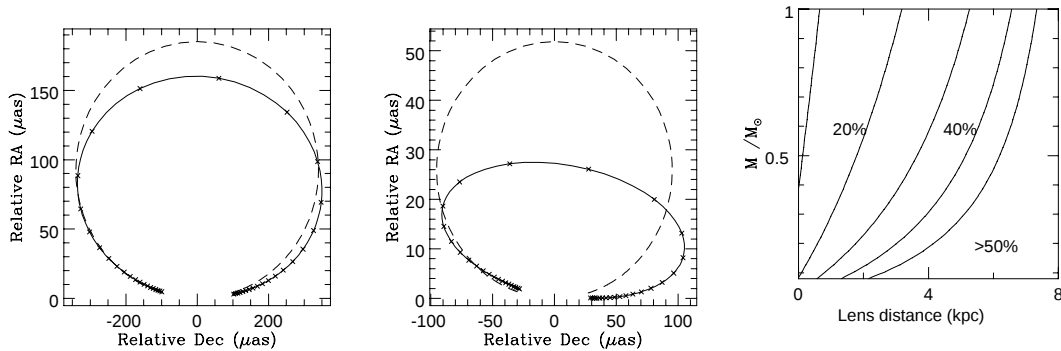


Figure 1.21: (a) Two simulated events for microlensing towards the bulge. The left panel shows a lens of mass $1M_{\odot}$ at 4 kpc, while the right panel a brown dwarf lens of mass of $0.08M_{\odot}$. The astrometric deflection is shown for a barycentric observer (dashed lines) and a terrestrial observer (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 $100 \mu\text{as}$, 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 $> 0.2 \text{ arcsec yr}^{-1}$ — 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 ~ 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 $10 \mu\text{as}$ 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 (Paczynski 1998) which showed that, using the 12 204 Hipparcos Catalogue stars with distances $< 100 \text{ pc}$ and proper motions $> 100 \text{ mas yr}^{-1}$, about 10^{-7} of the whole sky would be subject to displacements $> 100 \mu\text{as}$ 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

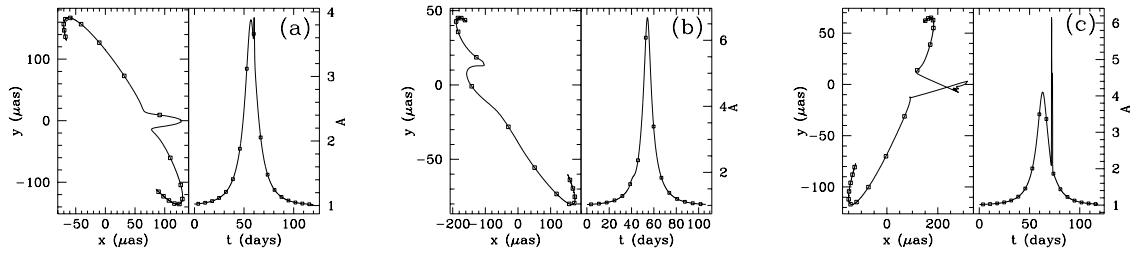


Figure 1.22: Planet lensing examples from Safizadeh et al. (1999). All examples assume $q=10^{-3}$ (where q is the planet to lens mass ratio), with a primary lens Einstein radius of $550 \mu\text{as}$, corresponding to a Saturn mass planet. Squares are plotted one per week: (a) $x_p = 1.3$; (b) $x_p = 0.7$; (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 $\sim 20 \text{ mas}$ 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.

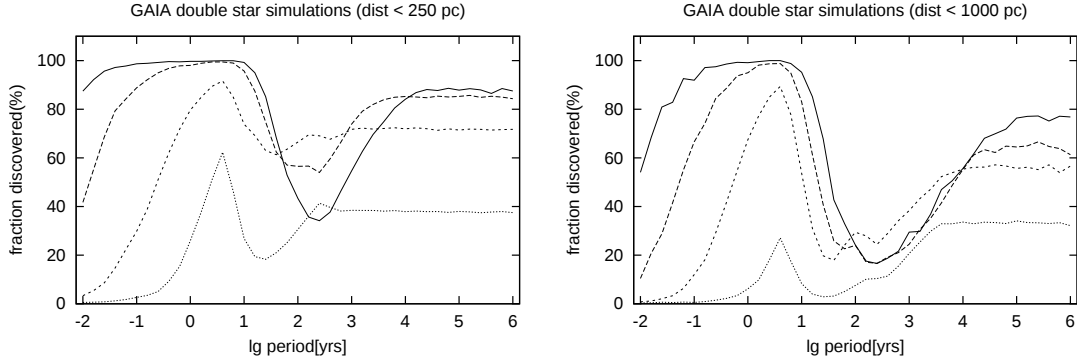


Figure 1.23: Total fraction of the double stars that are observed as non-single by GAIA, as a function of the binary period. The four curves are for the successive magnitude ranges 10–12.5, 12.5–15, 15–17.5 and 17.5–20 mag. (a) stars nearer than 250 pc. (b) stars nearer than 1000 pc.

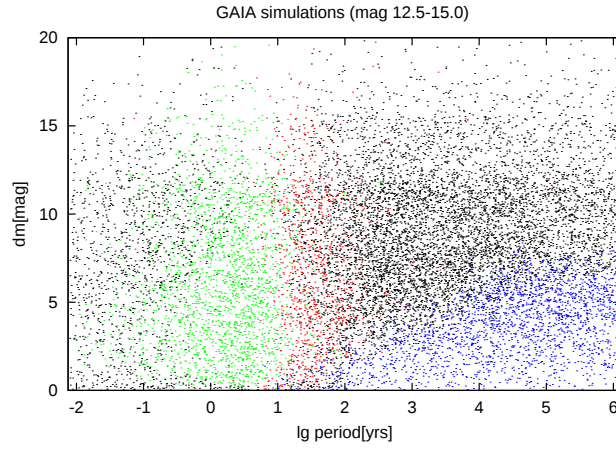


Figure 1.24: Simulated bright binaries (black) in a $\log P$ versus Δm diagram. The other colours denote systems detected as non-single by GAIA: blue for resolved system, red for ‘quadratic’, and green for ‘stochastic’ proper motion deviations.

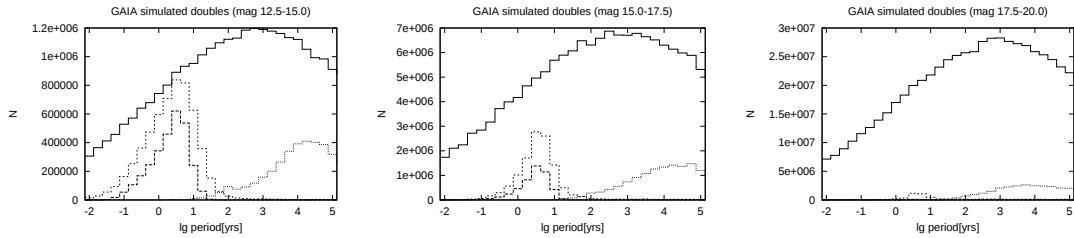


Figure 1.25: Total numbers of binaries per 0.25 dex in log period (solid curves) and the numbers resolved (dotted) or detected (lighter dashed) as astrometric binaries. The lower dashed curve gives the number of successful astrometric orbit determinations. Apparent magnitude intervals: (a) 12.5–15.0;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

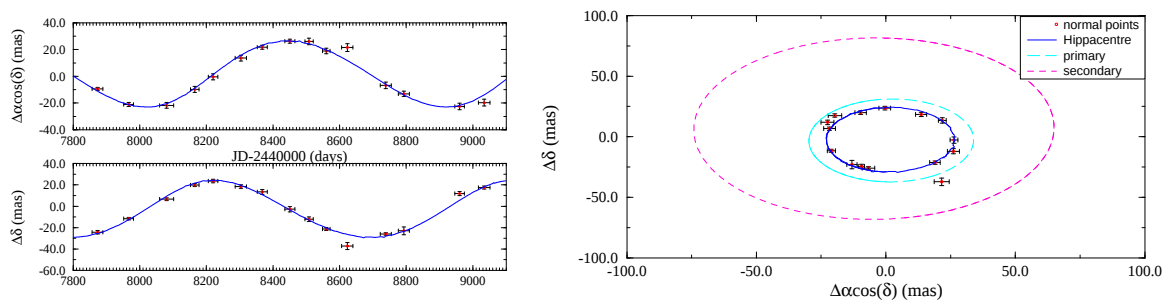


Figure 1.26: *HIP 39903: an example of a new astrometric binary from Hipparcos data. The left panels show the fits in right ascension and declination, and the right panel shows the reconstructed orbit in the $\Delta\alpha, \Delta\delta$ plane. Combining the astrometric data with the spectroscopic data 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, $e = 0.112$, $i = 21^{\circ}.8$. 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

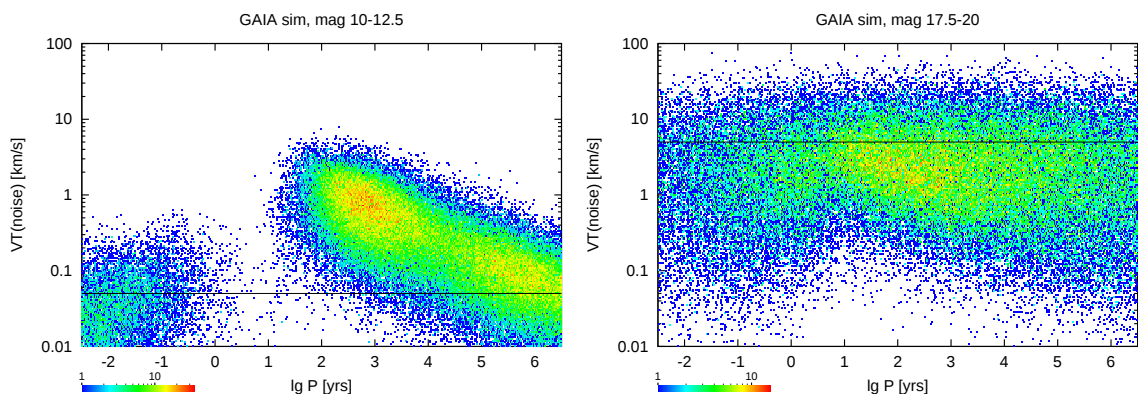


Figure 1.27: Simulated proper motion errors (converted to velocity via the observed parallax) for the photocentres of undiscovered binaries with different periods: (a) in the range 10–12.5 mag; (b) in the range 17.5–20 mag. The horizontal lines indicate a median mean error in the absence of orbital motions.

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

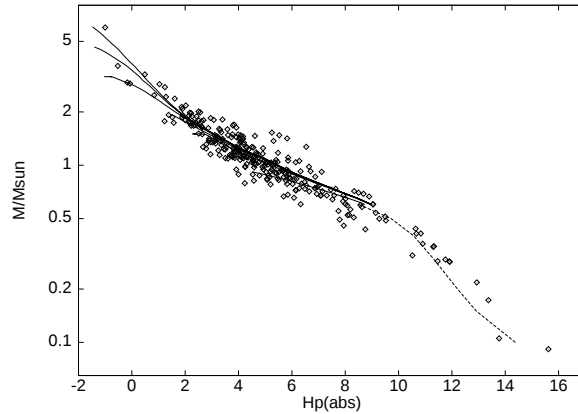


Figure 1.28: The mass-luminosity relation for 276 main sequence stars with mass uncertainties below 30 per cent, as derived from the Hipparcos data by Söderhjelm (1999). Theoretical isochrones are shown for $\log(\text{age}) = 7.5, 8.0, 8.5, 9.0, 9.4, \text{ and } 10.2$.

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 $10 \mu\text{s}$ 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, $(\text{period}) \propto m/r^3$, 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), $i \leq 17^{\circ}$, period = 100 d (Mercury’s orbit) $i \leq 0.34^{\circ}$, period = 1 yr, $i \leq 0.11^{\circ}$. 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 $10^{-3}(1/d(\text{AU}))(M_*/0.1)$ 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 ~ 4000 F–G–K dwarfs and sub-dwarfs to 200 pc, brighter than $V = 13$ mag, and with metallicity $[\text{Fe}/\text{H}] < -1.0$. By looking at stars in the age range $\sim 1\text{--}8$ 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.)

No detailed simulations have been made, but there may in the end be hundreds of small mass-functions available to clarify the brown-dwarf mass-distribution without the complications from their rapid evolution in luminosity.

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 ν 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 $\sim 10M_{\oplus}$), 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’, $\alpha = (M_p/M_s)(a_p/d)$, where M_p and M_s are the planet and stellar masses, a_p is the planet orbital radius, and d is the distance; α is in arcsec if a_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 ν And system, implying a mass of $10.1^{+4.7}_{-4.6} M_J$ 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 $\sim$ 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.

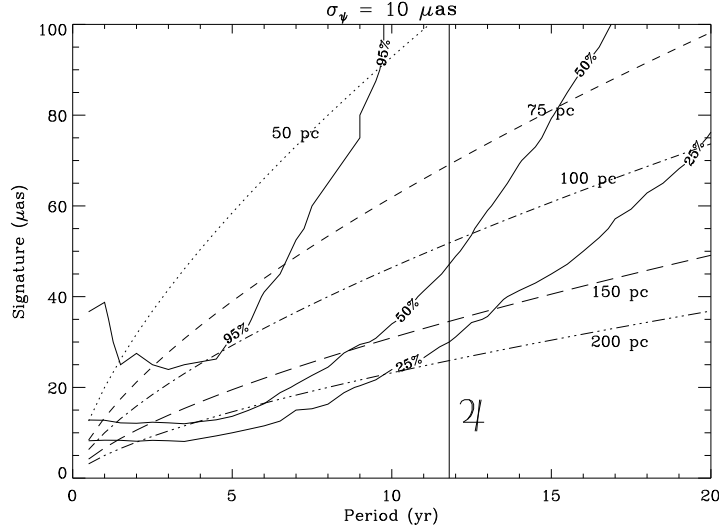


Figure 1.29: Iso-probability contours (solid lines) for 25, 50, and 95 per cent of detection probability, compared with Kepler's third laws (dotted/dashed lines) for systems with Jupiter-Sun masses at $D = 50, 75, 100, 150,$ and 200 pc.

In the same way that the presence of convection currents on the star's surface may limit radial velocity determinations to no better than $\pm 1 \text{ m s}^{-1}$, 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 ($\sim 5 \mu\text{as}$ at 10 pc for a star of radius $1 R_{\odot}$).

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: $0^\circ \leq i \leq 90^\circ$, $0 \leq e < 1$, $0 \leq \Omega$ (position angle of the ascending node) $\leq 2\pi$, $0 \leq \omega$ (argument of periastron) $\leq 2\pi$, $0 \leq \tau$ (periastron epoch) $\leq P$. 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.

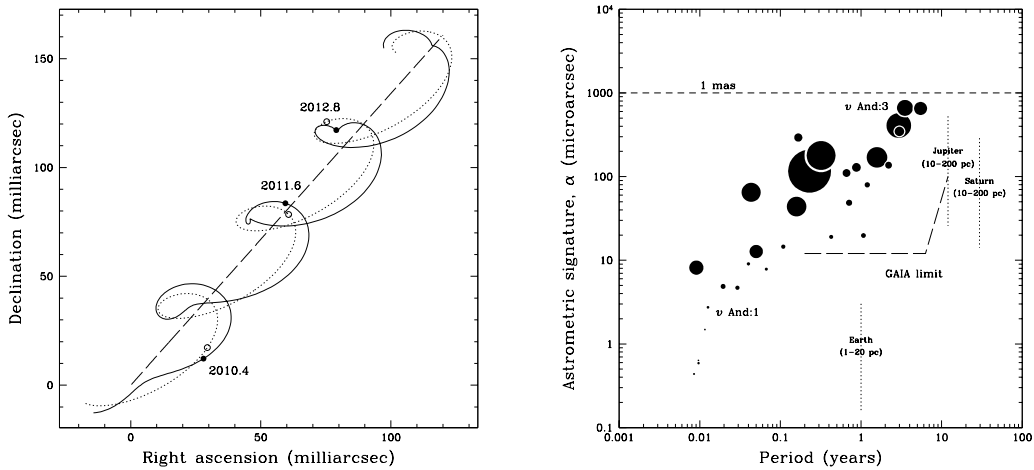


Figure 1.30: Left: the modelled path on the sky of a star at a distance of 50 pc, with a proper motion of 50 mas yr⁻¹, and orbited by a planet of $M_p = 15 M_J$, $e = 0.2$, and $a = 0.6$ AU. The straight dashed line shows the system's barycentric motion viewed from the Solar System barycentre. The dotted 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 $\times 30$ 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 (~ 300 stars surveyed) for giant planets (defined as having masses in the range $0.5\text{--}5 M_J$) 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 ~ 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_*	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 ν 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 $0.6 M_J$ object in a 4.6-day orbit, the two more distant planets have $M_p \sin i$ of 2.0 and $4.1 M_J$, 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 $10.1^{+4.7}_{-4.6} M_J$ using Hipparcos astrometry, compared to $M_p \sin i = 4.1 M_J$, 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 $4.9 \pm 2.3 M_J$ 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; Krymowski & 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 be 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 $M - p \sim 8 M_J$, twice the minimum value for an edge-on orbit. Assuming a Jupiter-like albedo $p = 0.55$ yields $R_p \sim 1.6 - 1.8 R_J$, slightly larger than the structural and evolutionary predictions of Guillot et al. (1997) of $\sim 1.4 - 1.1 R_J$ 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 $V \gtrsim 16 \text{ mag}$, 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 held 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.

As for direct orbit determinations of known asteroids, 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, and present instrument performances. Another set of simulated ground-based observations covering the period 1996–2015 (the latter year being assumed to be the nominal end of the GAIA mission) 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. In other words, after the collection of the GAIA data, all the results of more than one century of ground-based asteroid astrometry would be largely superseded. This impressive result is a consequence of the wonderful astrometric accuracy of the GAIA observations, combined with the fact that each asteroid should be observed several hundreds of times on the average over the mission lifetime.

New Objects 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 all 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 objects.

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

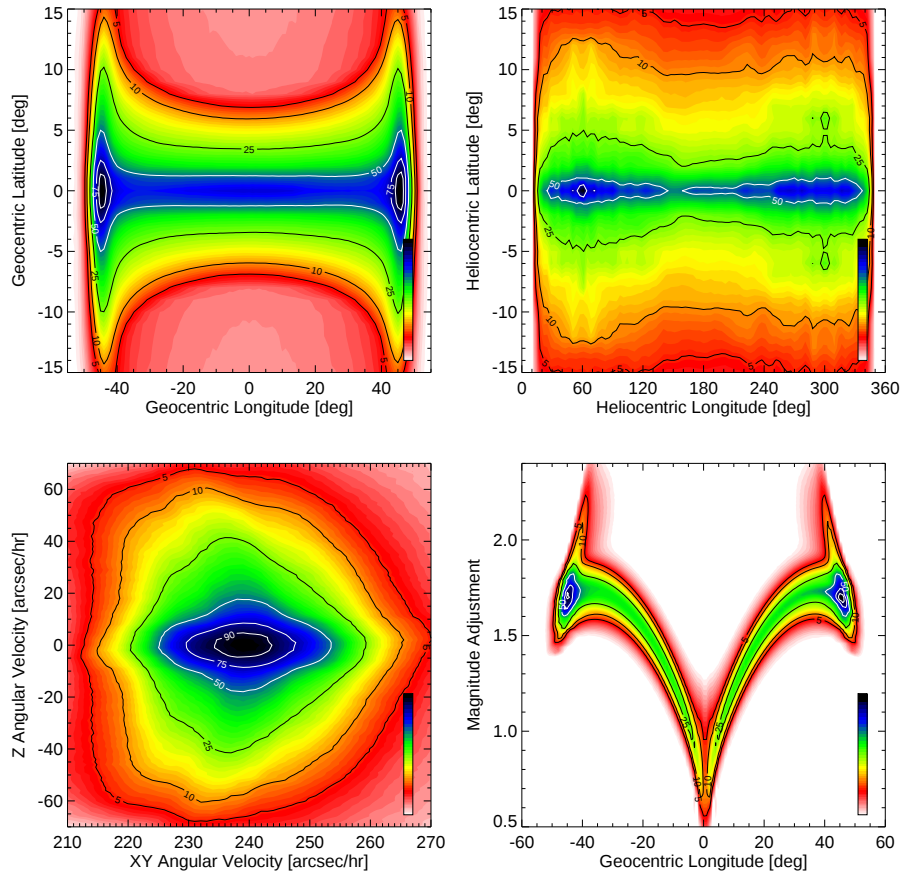


Figure 1.31: Predictions of the expectation of co-orbiting asteroids: (a) in the plane of geocentric ecliptic latitude and longitude; (b) heliocentric ecliptic latitude and longitude; (c) the distribution as a function of proper motions in the plane of, and perpendicular to, the ecliptic; (d) the distribution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

close to the Sun. These are the asteroids on horseshoe orbits at superior conjunction. Even though they are furthest 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 $\sim \pm 45^\circ$. 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 ($\sim 10\,000$ 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

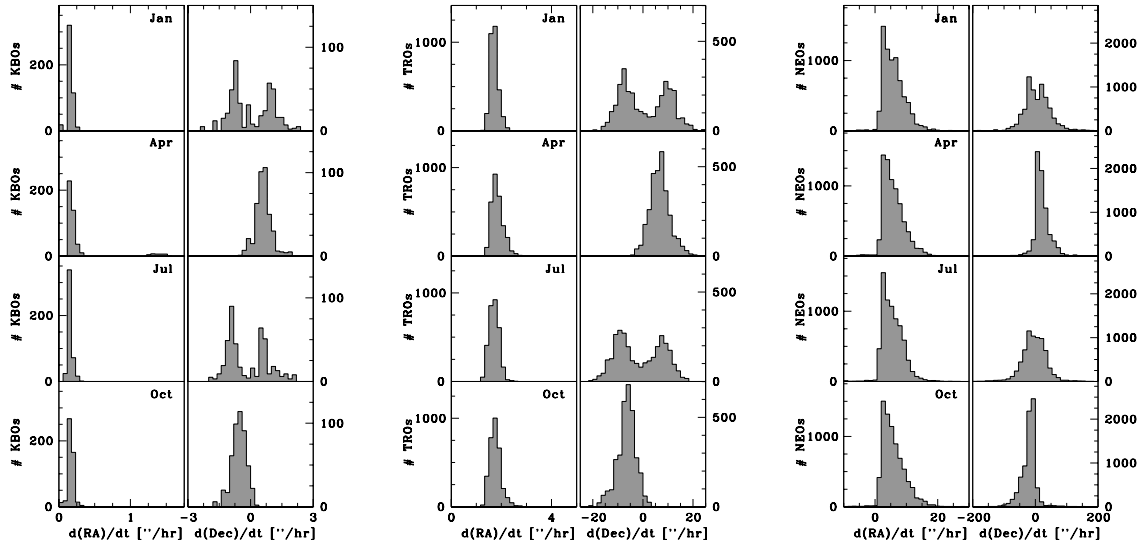


Figure 1.32: Mean angular motions of (a) Kuiper Belt objects, (b) Trojans, and (c) Near-Earth asteroids.

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 $\sim 32^\circ$). 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 TL₆₆, 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 $\sim 0.1M_E$ inside the 30–50 AU region, and perhaps a few $\times 0.1M_E$ 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 $\sim 90^\circ$ elongation (where GAIA will be looking) is small: the known KBOs have $da/dt = 0.02 - 1.0$ 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),

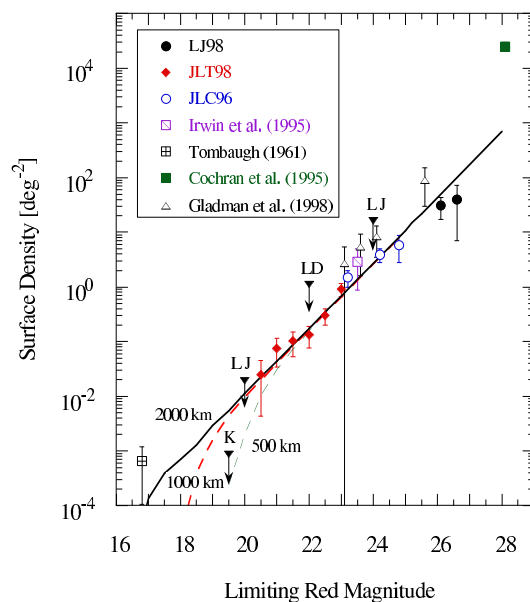


Figure 1.33: Surface density of Kuiper Belt objects as a function of R magnitude, based on the objects discovered before 1999.

implying that GAIA should discover ~ 300 KBOs with $V \leq 20$ (~ 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 ~ 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 $\sim 1M_E$ 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 TL₆₆).

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 neighbourhood (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.

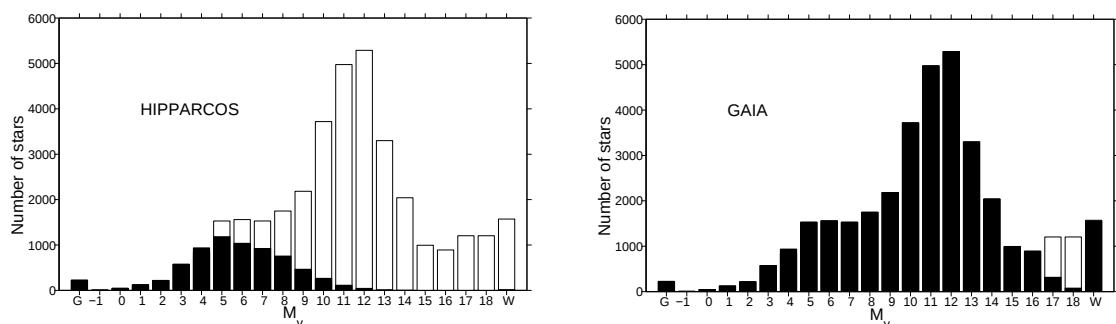


Figure 1.34: Comparison of the completeness from Hipparcos (left) with the expected completeness from GAIA (right). The plots show the number of star systems (individual stars or binary/multiple systems) within 50 pc of the Sun, as a function of absolute magnitude M_V , predicted by a systemic luminosity 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 geologically 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 μ as)

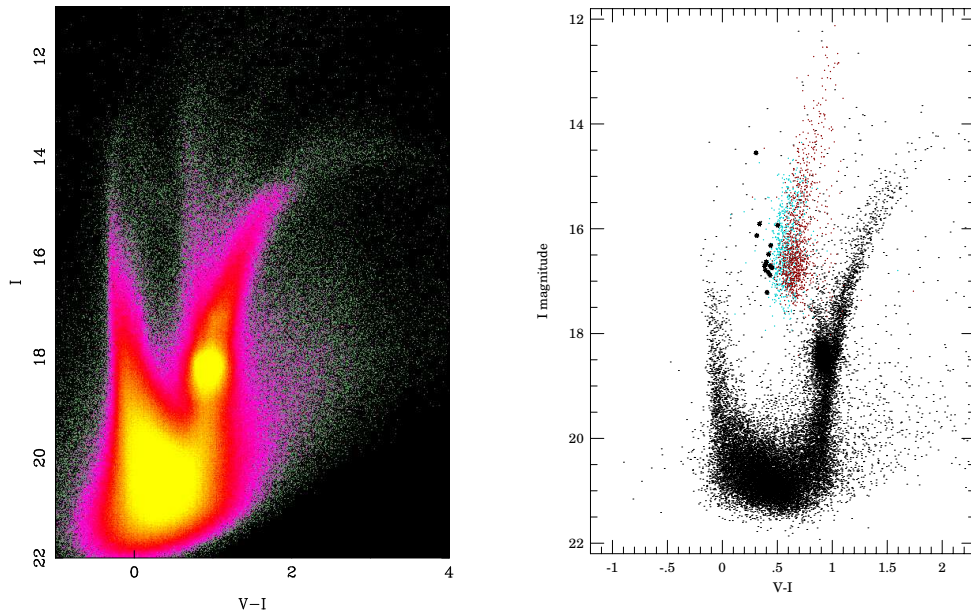


Figure 1.35: Left: A colour-magnitude diagram for the centre of the LMC, from Zaritsky et al. (1997). This is for a field of about $4 \times 4 \text{ deg}^2$, centred 3° north of the bar. It contains 2.4 million stars brighter than $I = 20 \text{ mag}$ (data provided by the Magellanic Cloud Photometric Survey, courtesy of Dennis Zaritsky). Right: Colour-magnitude diagram for an area of $14 \times 57 \text{ arcmin}^2$ in the SMC bar. There are 45 500 stars with $I < 20 \text{ mag}$. 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 $I = 12 - 16 \text{ mag}$, will have transverse velocities determined to $\sim 1 - 2 \text{ km s}^{-1}$ ($\approx 20 \mu\text{as yr}^{-1}$). 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

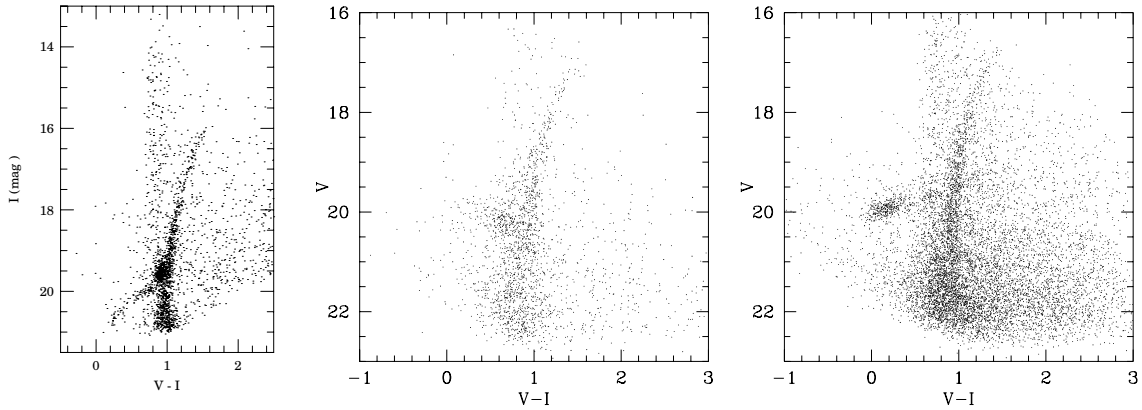


Figure 1.36: Colour-magnitude diagrams for (left) the central regions of the Carina dwarf. This field covers 14×14 arcmin², and has 2210 stars brighter than $I = 20$ mag; (middle) the central 8 arcmin radius of the Draco dSph galaxy; (right) the Ursa Minor dSph galaxy. GAIA will observe a few thousand 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 $10 - 30 \text{ km s}^{-1}$ (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 $3 - 5 \text{ km s}^{-1}$, 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 $\sim 10 \text{ km s}^{-1}$ (Hargreaves et al. 1994), $30 \mu\text{as yr}^{-1}$, 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_{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 ($^\circ$)	b ($^\circ$)	$E(B-V)$ (mag)	$(m-M)_0$ (mag)	Distance (kpc)	V_{lim} (mag)	N(stars) ($V < 20$)	V_r (helio)	$\mu_{v_t-v_r}$ ($\mu\text{as/yr}$)
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 $1 \mu\text{as yr}^{-1}$ 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

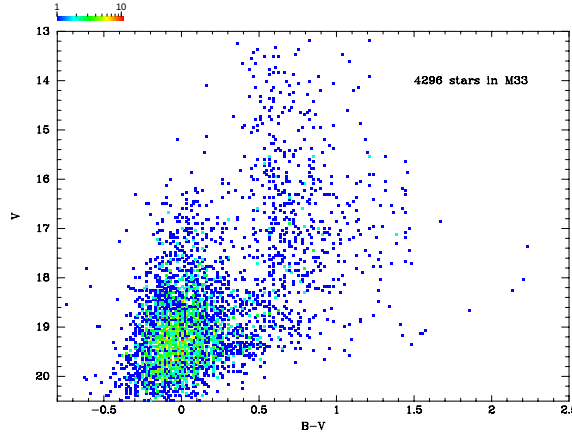


Figure 1.37: Colour-magnitude diagram for stars in M33. This shows 4300 stars brighter than $V = 20$ mag (courtesy of Phil Massey).

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 ($\sim 265 \text{ km s}^{-1}$) corresponds to a transverse proper motion of $\sim 100 \mu\text{as yr}^{-1}$. In the warped outer parts, inclination effects reduce this to $\sim 20 \mu\text{as yr}^{-1}$. 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 $10\text{--}20 \mu\text{as yr}^{-1}$, 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 $(\ell, b) = (268^\circ, 27^\circ)$. 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 ($m = 1, 2, \dots$). 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.

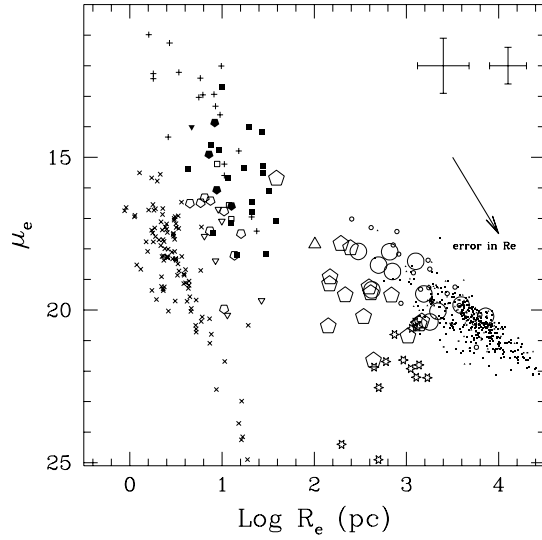


Figure 1.38: Surface brightness μ_e (mag arcsec^{-2}) in the visual band ($F606W$) at effective radius r_e for a sample covering all types of galaxies (from Carollo 1999).

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 $|b| > 15^\circ$.

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 $100 \mu\text{as}$) 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 $300 \mu\text{as yr}^{-1}$, 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 $\sim 25 \text{ deg}^{-2}$ 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 $300 - 900 \text{ nm}$, 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 $\sim 1 \text{ deg}^{-2}$, 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

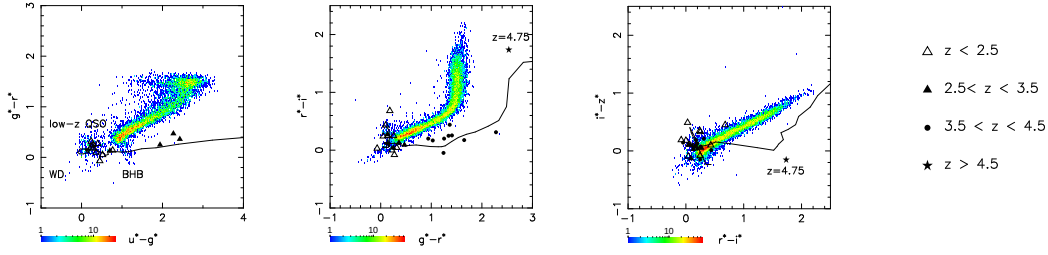


Figure 1.39: An example of the use of broad-band filter photometry to supplement proper-motion searches for quasars. The combination of zero proper motion and colour data will substantially improve the detection efficiency of quasars, and hence their utility for clustering studies. This figure uses Sloan 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 Ω_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 ($\sigma^* = 225 \text{ km s}^{-1}$, corresponding to a value of the efficiency parameter for macro-lensing $F = 0.025$) and of 70 per cent of spiral ($\sigma^* = 144 \text{ km s}^{-1}$, $F = 0.007$; 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 $2\theta_E = 8\pi D_{ds}\sigma^2/D_{os}c^2$. 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 $\Omega_M = 1$, $\Omega_\Lambda = 0$. This represents 96 per cent of the total number of expected lenses. (If the field of view is further reduced to 2 or 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 $\Omega_M = 0$, $\Omega_\Lambda = 1$, 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 $l = 90^\circ$, $b = 0^\circ$. 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 $l = 90^\circ$, $b = 0^\circ$. 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 $6 \text{ mm s}^{-1} \text{ yr}^{-1}$. A change in velocity by 6 mm s^{-1} causes a change in aberration of the order of $4 \mu\text{as}$. The apparent proper motion of a celestial object caused by this effect always points towards the direction of the Galactic centre. Its size is $4 \sin \zeta \mu\text{as yr}^{-1}$, 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 $4 \mu\text{as yr}^{-1}$ corresponds to a transverse velocity of $30\,000 \text{ km s}^{-1}$ at $z = 0.3$ for $H_0 = 75 \text{ km s}^{-1} \text{ Mpc}^{-1}$. 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 $\sim 20 \text{ Mpc}$. 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 $500 \mu\text{as}$. 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 $0.4 \mu\text{as yr}^{-1}$ 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 $14 \mu\text{as yr}^{-1}$, independent of the distance, while if it is equal to the mean velocity dispersion of galaxies, it would be significantly smaller than $0.1 \mu\text{as yr}^{-1}$ 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 $\sim 60 \mu\text{as}$ 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 $|b| < 20^\circ$, 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 μ_{l*} , μ_b 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 $\sigma_0 = 10 \mu\text{as yr}^{-1}$ was assumed, while for Table 1.16 the assumption was a pessimistic $\sigma_0 = 100 \mu\text{as yr}^{-1}$. 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 $1 \mu\text{as yr}^{-1}$ corresponding to $4.606 \times 10^{-11} \text{ m s}^{-2}$). The galactocentric acceleration of the Sun is expected to be $2 \times 10^{-10} \text{ m s}^{-2} \sim 4 \mu\text{as yr}^{-1}$, 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 $\mu\text{as yr}^{-1}$, 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. $\lesssim 0.2 \mu\text{as yr}^{-1}$ 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 $1 \mu\text{as}$ (Sazhin et al. 1998). The typical effect on the mean proper motion over the GAIA lifetime will be $\lesssim 1 \mu\text{as yr}^{-1}$; (3) most of the optical emission from a quasar comes from a region of $\lesssim 1 \text{ pc}$, corresponding to $\lesssim 200 \mu\text{as}$ at 1 Gpc. Assuming that the photocentre moves randomly within this region, a mean proper motion of $\lesssim 50 \mu\text{as yr}^{-1}$ 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 $100 \mu\text{as yr}^{-1}$, 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

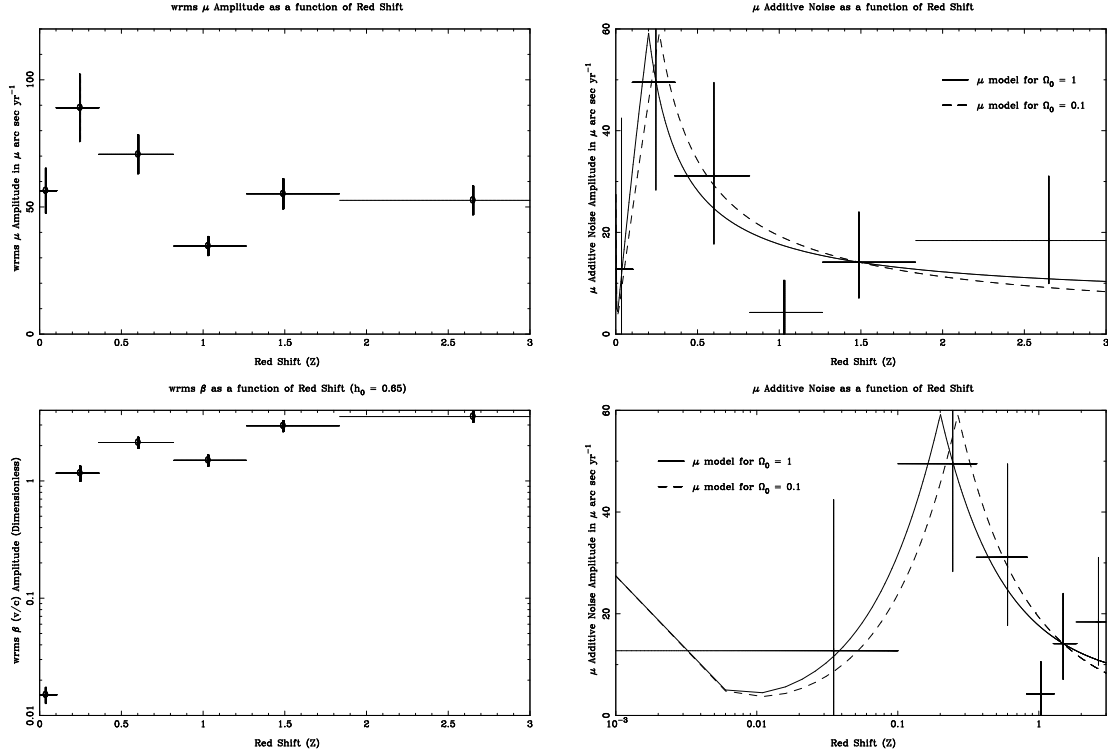


Figure 1.40: Proper motion residuals, σ_μ , computed according to different hypotheses about the source motion, and for different cosmological models (from Eubanks, private communication). See text for further details.

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 $\sim 10 \mu\text{as yr}^{-1}$. 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 $10\text{--}100 \mu\text{as yr}^{-1}$, 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 excess 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 gravitational 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_{QSO} = expected number of recognised quasars with $z < 2.2$ and $|b| > 20^\circ$; $\sigma_{\mu, \text{tot}}$ = mean standard errors in the proper motion per object and coordinate, including an assumed contribution of $\sigma_0 = 10 \mu\text{as yr}^{-1}$ 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_{QSO}	$\sigma_{\mu, \text{tot}}$ ($\mu\text{as yr}^{-1}$)	$\sigma(\omega_1)$	$\sigma(\omega_2)$	$\sigma(\omega_3)$	$\sigma(a_1/c)$	$\sigma(a_2/c)$	$\sigma(a_3/c)$
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 $\sigma_0 = 100 \mu\text{as yr}^{-1}$ from source instability.

B (mag)	P	N_{QSO}	$\sigma_{\mu, \text{tot}}$ ($\mu\text{as yr}^{-1}$)	$\sigma(\omega_1)$	$\sigma(\omega_2)$	$\sigma(\omega_3)$	$\sigma(a_1/c)$	$\sigma(a_2/c)$	$\sigma(a_3/c)$
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:

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

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:

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

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	$\chi = 45^\circ$ μas	$\chi = 90^\circ$ μas	$\chi_{\max}$ μas	J_2	Grazing μas	$\chi = 1^\circ$ 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

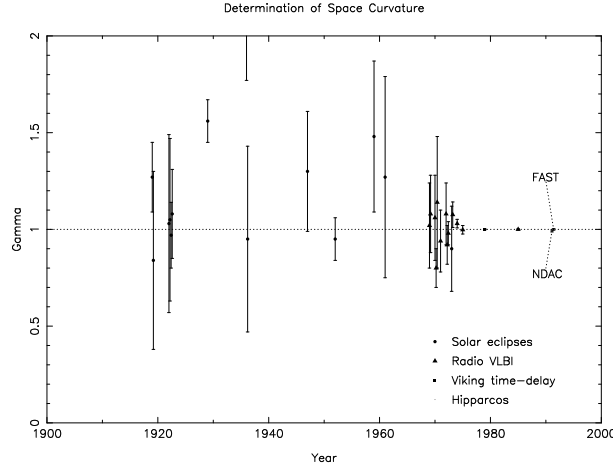


Figure 1.41: Determinations of the post-Newtonian parameter γ , representing the deviation of the gravitational light bending from Newtonian theory. According to General Relativity, $\gamma = 1$. Non-Hipparcos data are taken from Soffel (1989), and the Solar eclipse observations include the most recent 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):

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

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 $G = \text{constant}$ 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 $\dot{G}/G \leq -1.6 \times 10^{-12} \text{ yr}^{-1}$;

distance measurements to the Viking landers (Hellings 1987; Reasenberg 1983) with upper bounds in the range $\dot{G}/G \leq -(3 \pm 0.6) \times 10^{-11} \text{ yr}^{-1}$, and limited by the accuracy of the mass and distribution 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 $\dot{G}/G \leq -(1.10 \pm 1.07) \times 10^{-11} \text{ yr}^{-1}$, and limited by the proper motion adopted for the pulsar; and from white dwarf studies (García-Berro et al. 1995) giving $\dot{G}/G \leq -(1 \pm 1) \times 10^{-11} \text{ yr}^{-1}$ 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 $10^{-12} - 10^{-13} \text{ yr}^{-1}$, 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 $|\dot{g}_p/g_p| < 1 \times 10^{-15} \text{ yr}^{-1}$ and $|\dot{\alpha}/\alpha| < 5 \times 10^{-16} \text{ yr}^{-1}$ 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 $50 \mu\text{as yr}^{-1}$, 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 $600 \pm 900 \text{ km s}^{-1}$; (ii) our peculiar acceleration is $< 5 \times 10^{-10} \text{ m s}^{-2}$, constraining the mass of our Galaxy to $< 2.4 \times 10^{11} M_\odot$; (iii) the local velocity shear is $< 3 \times 10^{-10} \text{ yr}^{-1}$; (iv) the local vorticity is $< 9 \times 10^{-11} \text{ yr}^{-1}$; (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 $\sim 30 \mu\text{as yr}^{-1}$ (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:

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

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 $\vec{\delta}$ into vector spherical harmonics, as follows:

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

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

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

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 $\Omega_{\text{gw}}(f)$, the ratio of the energy density in gravitational waves to the energy density needed to close the Universe:

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

Here $\rho_{\text{gw}}(f)$ 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 $\Omega_{\text{gw}}(f)$ by:

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

where $H_0 \equiv h_{100} 100 \text{ km s}^{-1} \text{ Mpc}^{-1}$. Defining Ω_{gw} to be the total integrated gravitational wave energy in all bands, $\Omega_{\text{gw}} \equiv \int \Omega_{\text{gw}}(f) d \ln f$, and $\Omega_{\text{gw}}^{GAIA}$ to be that part of Ω_{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:

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

Big Bang nucleosynthesis gives a strong constraint on the total gravitational-wave energy in all bands: $\Omega_{\text{gw}} < 10^{-5}$ (Allen 1997). This assumes the waves are generated in the early Universe, before Big Bang nucleosynthesis. The strongest limits on $\Omega_{\text{gw}}(f)$ come from millisecond pulsar timing: $h_{100}^2 \Omega_{\text{gw}}(f \sim 4 \times 10^{-9} \text{ Hz}) < 10^{-8}$, or (roughly) $h_c < 5 \times 10^{-14}$ at $f \sim 4 \times 10^{-9} \text{ Hz}$ (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 $\sim 10^8 \text{ s}$.) Limits from millisecond pulsar timing are much weaker for frequencies lower than $4 \times 10^{-9} \text{ Hz} \sim 1/(\text{total observation time})$, 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: $h_{100}^2 \Omega_{\text{gw}}(f) < 0.04$ for $10^{-11} < f < 4.4 \times 10^{-9} \text{ Hz}$ and $h_{100}^2 \Omega_{\text{gw}}(f) < 0.5$ for $10^{-12} < f < 10^{-11} \text{ Hz}$ (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 $10^{-11} < f < 4.4 \times 10^{-9} \text{ Hz}$ (Kopeikin 1997). VLBI astrometry currently gives the constraint $h_{100}^2 \Omega_{\text{gw}}(f) < 10^{-1}$ in (roughly) the range $10^{-17} < f < 10^{-9} \text{ Hz}$ (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 $0.1 \mu\text{as yr}^{-1}$, and pulsar timing measurements strongly suggest that they will be less than $\sim 0.002 \mu\text{as yr}^{-1}$. 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 Ω_{gw} in the band $10^{-12} < f < 4 \times 10^{-9} \text{ Hz}$. Roughly, the GAIA limit on $\langle \dot{\delta}^2 \rangle^{1/2}$ 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 Ω_{gw} scales like $1/N_s$. Assuming $N_s \sim 10^8$ with a ‘typical’ accuracy of $100 \mu\text{as yr}^{-1}$, it follows that GAIA could set an upper limit of roughly $\Omega_{\text{gw}} < 10^{-6} - 10^{-7}$ 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 Ω_{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 $10^{-12} < f < 10^{-10} \text{ Hz}$ band, the best upper limit on Ω_{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 $\Omega_{\text{gw}}(f)$, 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 Ω_{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 Ω_{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 $\Omega_{\text{gw}}(f)$ is predicted to be roughly 10^{-13} for frequencies $10^{-16} < f < 10^{10} \text{ Hz}$. 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 $\Omega_{\text{gw}} \sim 10^{-14}$ (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 $200 \times 200 \times 200$ pc³ 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 $\sim 2.5 \times 10^{-11}$ pc, which corresponds to ~ 0.05 μ as 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_⊙ (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 μ as. The same number is found for a giant molecular cloud of 10^6 M_⊙ 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_⊙ (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_⊙) in the Galactic disk. The existence of wide binaries excludes any objects of mass larger than 2 M_⊙ 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 $\sim 2 \times 10^{-12} \text{ pc}^{-3}$ 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 $\sim 2 \times 10^{-14} \text{ pc}^{-3}$, 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 $1 \mu\text{as}$, could determine the transverse motions of external galaxies and quasars routinely, and determine their kinematic properties independently of a dynamical model of the Universe.

