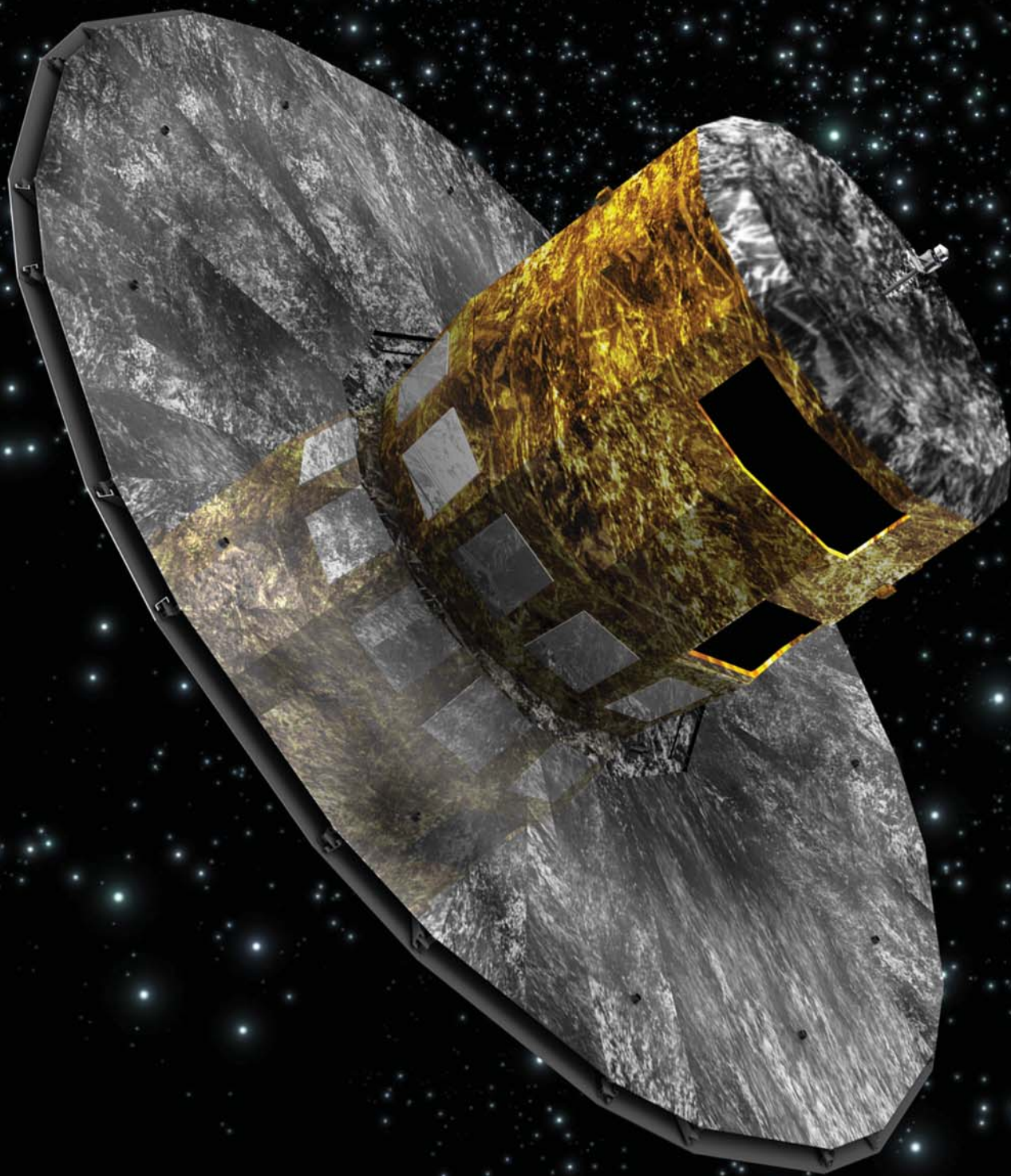


Pinpointing the Milky Way



*Jane Douglas, Jos de Bruijne, Karen O'Flaherty
& Timo Prusti*

Research and Scientific Support Department,
Directorate of Scientific Programmes, ESTEC,
Noordwijk, The Netherlands

William O'Mullane & Uwe Lammers

Gaia Science Operations, Directorate of
Scientific Programmes, ESAC, Villafranca, Spain

François Mignard

Observatoire de la Côte d'Azur, Nice, France

Ronald Drimmel

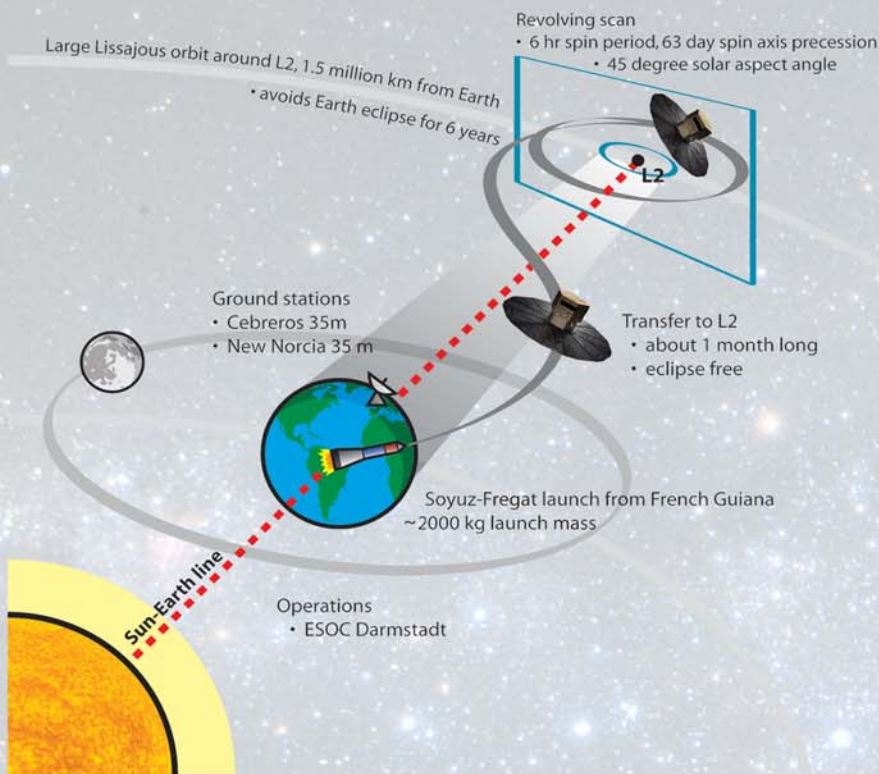
INAF Osservatorio Astronomico di Torino, Turin,
Italy

In the course of Gaia's 5-year astronomical survey, the equivalent of around 20 000 DVDs of raw information on our Galaxy will be harvested and transmitted to Earth. Sophisticated processing is needed to distill this flood of complex data into the final Gaia Catalogue of about 1000 million celestial objects. A group of more than 300 European scientists and software developers is rising to the challenge: the Data Processing and Analysis Consortium is already preparing for Gaia's launch in 2011.

Introduction

Gaia is a pioneering astronomy mission set to revolutionise our view of the Galaxy, the Milky Way, with a precise and detailed survey of the 1000 million brightest celestial objects. With high-accuracy 'astrometry', Gaia will pinpoint the position of each star and track its movement across the sky. Measuring its light spectrum will reveal how fast the star is approaching or receding (its 'radial velocity'). Gaia will also gather 'photometric' data, measuring the brightness of a star in a few dozen colours. This array of data will reveal a

The Formidable Challenge of Processing Gaia's Data



Gaia's launch and operations. (ESA/Astrium)

moving 3-D Milky Way map of unprecedented scope and precision, as well as providing profiles of the physical properties of each star, including temperature and elemental composition.

By surveying all celestial bodies down to the very faint magnitude of +20, Gaia will cover a representative fraction of the Milky Way's population, providing scientists with the material to tackle unanswered questions about our home galaxy, potentially revealing its history, current state and future.

This catch-all survey will naturally include many objects besides stars. Gaia will spot several thousand brown dwarfs (failed stars, too small to ignite) and extrasolar planets, and map out our immediate neighbourhood in great detail, detecting hundreds of thousands of minor Solar System bodies.

Beyond the Milky Way, Gaia will observe bright objects like supernovae and quasars, as well as many distant galaxies.

As a complete sky survey without pre-programmed targets, the discovery

potential of Gaia is profound. Who knows what it might find?

Through sophisticated processing and analysis, the raw data will be translated into the mission's final product: the Gaia Catalogue, an extensive galactic census, rich in scientific content. The unprecedented accuracy and unbiased nature of this full-sky survey will prove valuable, even revolutionary, to a huge range of scientific disciplines besides galaxy studies. Gaia's wealth of data will inform and invigorate scientific areas as diverse as the life cycles of stars, the distribution of 'dark matter' and Einstein's theory of general relativity.

Selected as an ESA Cornerstone mission in 2000, Gaia is now moving from its design phase into development; launch is set for late 2011.

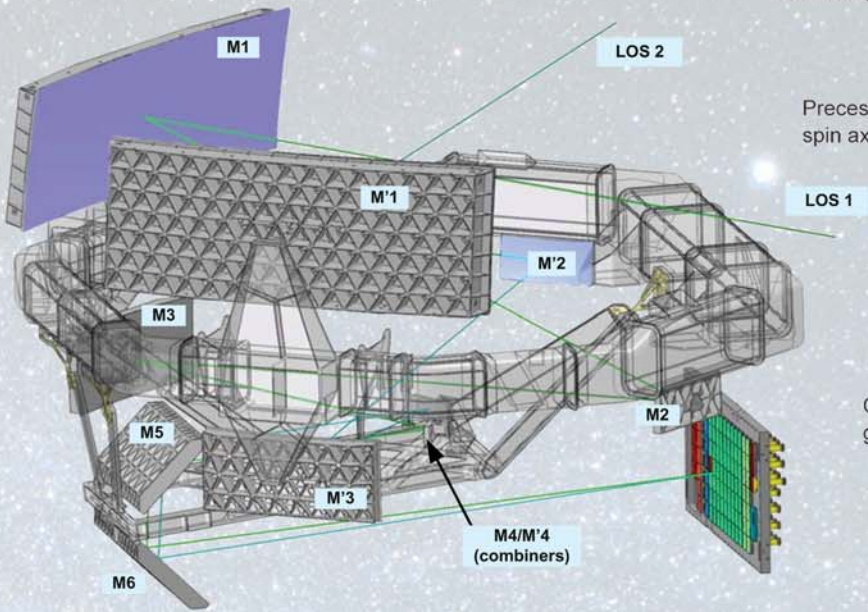
Gaia continues a European tradition in pioneering astrometry, building on the expertise generated by the very first space-based astrometry mission, Hipparcos. Flying between 1989 and 1993, Hipparcos mapped 99% of stars down to magnitude +11 with unprece-

dent accuracy. Its astrometric achievement is still unchallenged. Gaia will outdo its predecessor by orders of magnitude in terms of accuracy, sensitivity and quantity. While Hipparcos followed a programme of pre-selected objects to observe, Gaia's survey is complete and unbiased.

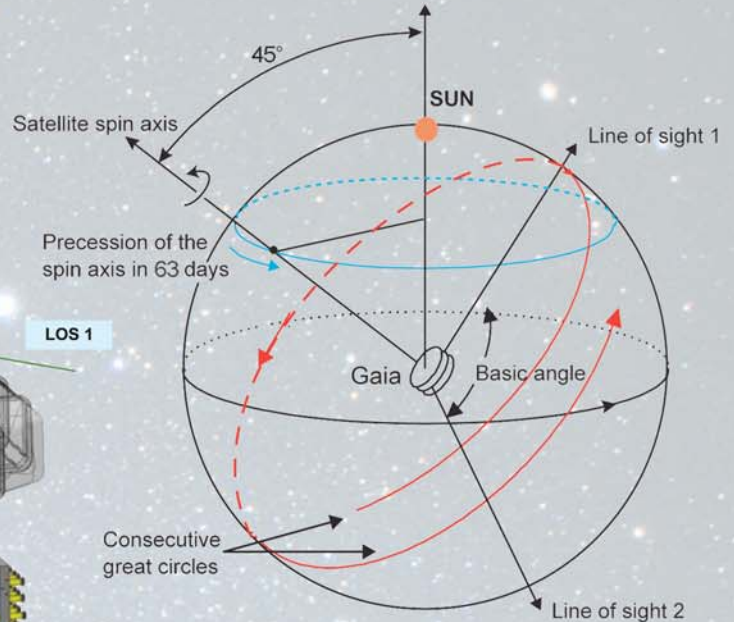
Advanced astrometric instruments and processing will enable Gaia to discern the positions of stars and their minute movements with incredible accuracy: around 25 microarcseconds (about 0.00000001°) for stars of magnitude +15 – roughly the ability to discern a single human hair from a distance of 500 km. A magnitude +15 star is already extremely faint, around 4000 times dimmer than the faintest stars visible to the human eye. Gaia's census will be complete for all celestial bodies down to magnitude +20 (about 400 000 times dimmer than the faintest naked-eye stars), which equates to around 1000 million objects.

Following launch, Gaia will travel 1.5 million kilometres from Earth, away from the Sun, to circle the second





The complex arrangement of Gaia's mirrors. LOS is the line-of-sight, the viewing directions of the telescopes. (EADS Astrium)



The 'scanning law' that allows Gaia to view the entire sky in a series of overlapping great circles

Gaia deploys its sunshield. (ESA/C. Carreau)



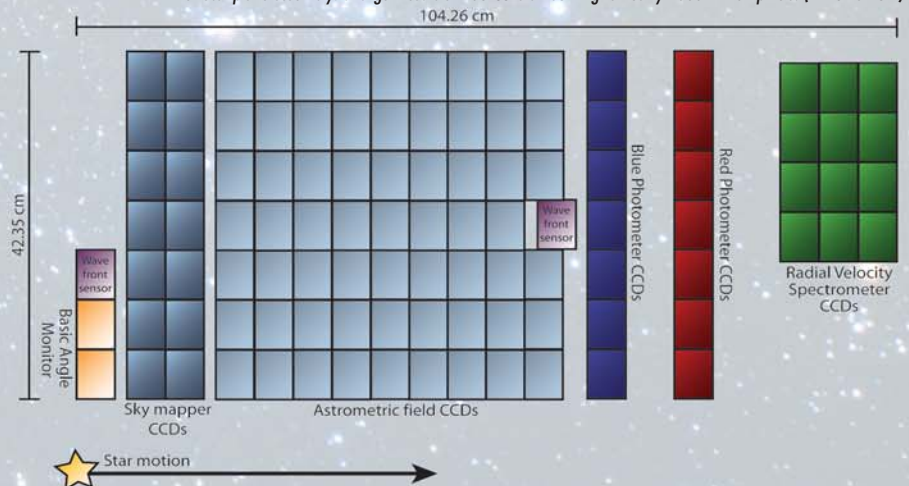
Lagrangian point of the Sun-Earth system. Known as L2, this is a point in space where the combined gravitational pulls of the Earth and Sun balance out, allowing a spacecraft to hold position. Shaded from the light of the Sun, Moon and Earth by a large shield, the satellite will revolve slowly, with one full revolution every 6 h, thereby scanning great circles across the sky.

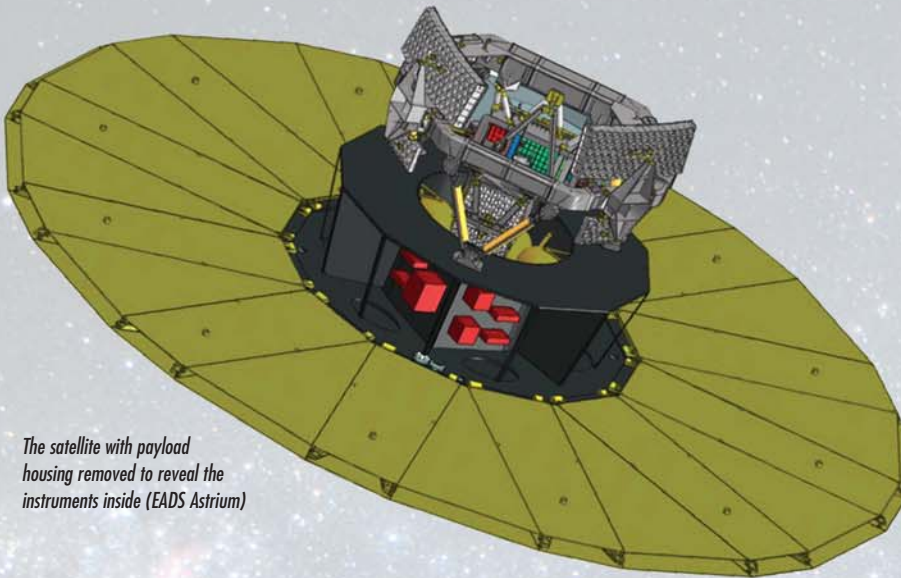
Inside the satellite, Gaia's instruments are mounted on a hexagonal optical bench. Two telescopes sharing a focal plane look out through apertures in the

payload housing. The two viewing directions are separated by a fixed basic angle. Light from a celestial object enters through an aperture, and strikes the large primary mirror opposite (either M1 or M'1).

The light is bounced by a series of mirrors along a total focal length of 35 m, with the two paths meeting at the M4/M'4 beam combiner before finally reaching the shared focal plane. Having two different lines of sight allows Gaia to measure the relative separations of the thousands of stars simultaneously

The focal plane assembly: a large mosaic of 106 CCDs amounting to nearly 1000 million pixels. (EADS Astrium)





The satellite with payload housing removed to reveal the instruments inside (EADS Astrium)

present in the combined fields of view. These wide-angle measurements build up a rigid network of relative star positions, contributing to Gaia's exceptional accuracy. The 'folded-up' focal length of 35 m is required to attain Gaia's high resolution.

At the focal plane is a large mosaic of sophisticated, custom-built charge coupled devices (CCDs), light detectors of essentially the same kind as found in a digital camera. Containing 106 CCDs, the focal plane assembly comprises a total of nearly 1000 million pixels (a 'gigapixel'), compared to the few million of a typical digital camera.

As Gaia slowly rotates, the image of a celestial object crosses the focal plane. The path it takes is recorded as astrometric data. Red and blue photometry prisms between the beam combiner and the focal plane enable Gaia to measure the brightness of an object in a number of 'colours', or bands of wavelength.

A radial velocity spectrometer grating, also located between the beam combiner and focal plane, will perform spectrometry: measuring an object's spectrum of light, in this case specifically to observe how spectral lines are shifted, from which the object's radial velocity can be inferred.

This arrangement means that the focal plane used for astrometry also serves the photometric and spectroscopic instruments, with certain CCDs assigned specifically to photometry and spectro-

scopy. The instruments steadily scan the sky as Gaia spins and gradually 'precesses' (the spin axis itself moves in a circle every 63 days). In this way, the whole sky is eventually covered, with each part being observed around 70 times in the course of Gaia's lifetime.

The Gaia Data Challenge

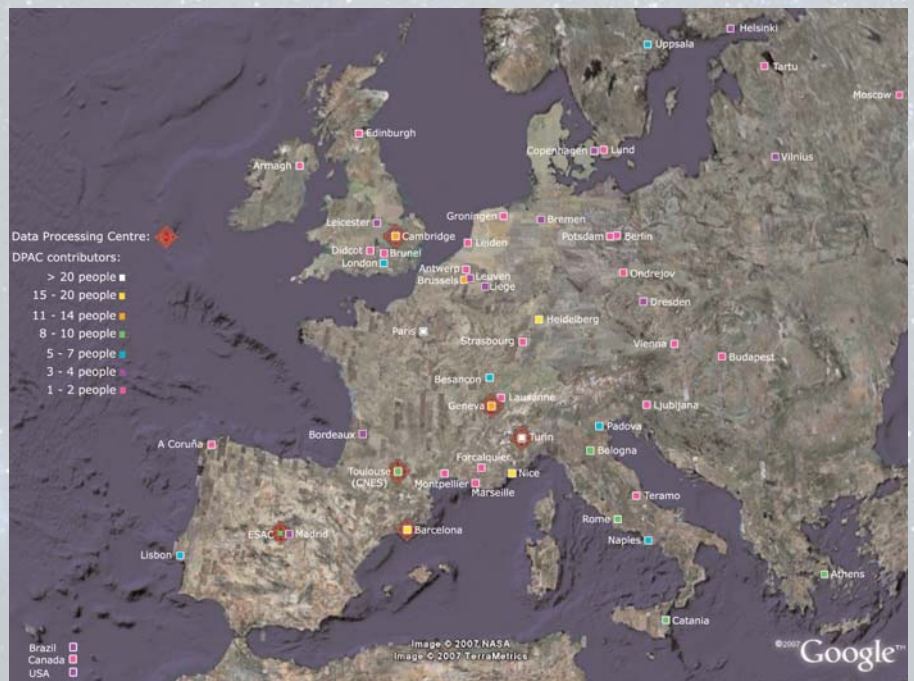
By its very nature, Gaia will acquire an enormous quantity of complex, extremely precise data, representing the

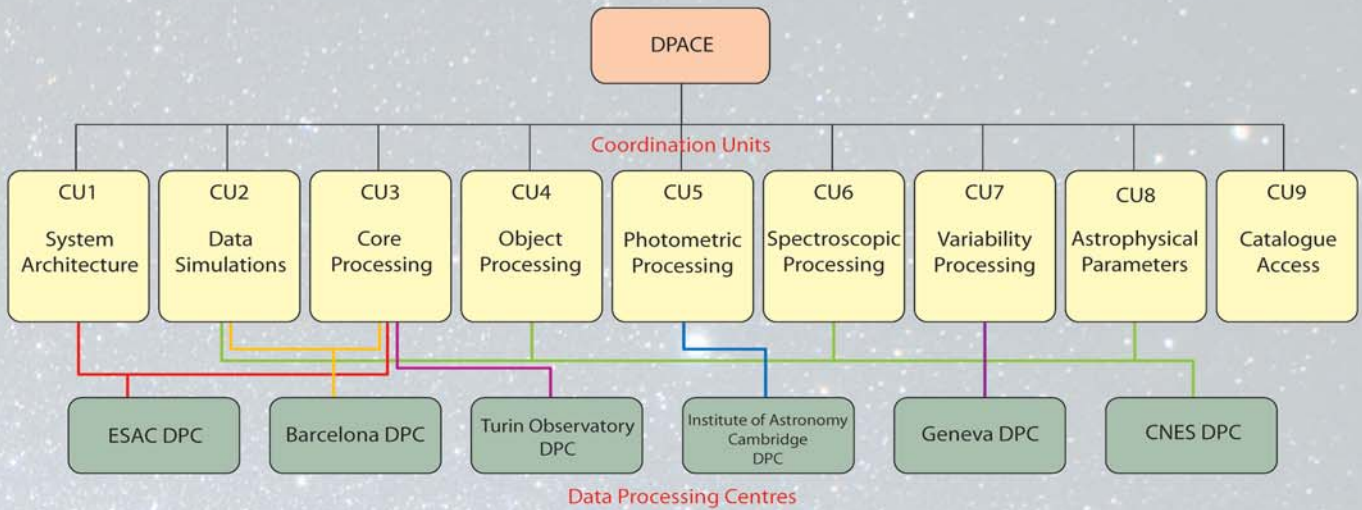
observations of 1000 million diverse objects dozens of times each by a 'double vision' instrument that is spinning and precessing. These data will be transmitted daily to a ground station, either Cebreros in Spain, or New Norcia in Australia. Both have dishes 35 m in diameter to catch the faint radio whisper from space. Downlink speeds and station visibility constraints will result in a daily raw telemetry volume of roughly 50 Gbytes. By the end of Gaia's operational life, around 100 terabytes (10^{14} bytes) will have been channelled to Earth: the equivalent of more than 20 000 DVDs, and some 1000 times the raw volume from Hipparcos.

In its original format, Gaia's data is unintelligible. Not only because it is squeezed into packets by numerical coding but also because of the way Gaia scans the sky, picking up fragments of information at each transit of the 1000 million objects.

It is the combination of data complexity and sheer volume that presents such a demanding task: extensive, sophisticated treatment is required to yield meaningful results, by piecing together the fragments and translating them into

DPAC membership is distributed across more than 15 countries





The DPAC organigram shows the relation of the Coordination Units, the Data Processing Centres and the DPAC executive committee (DPACE)

real positions, motions and other physical information.

In terms of dedicated computing power alone, Gaia's requirements are considerable. The centre of an image moving across the focal plane needs to be determined to within 1% of a pixel size. With several thousand million images registering over the course of the 5-year lifetime, if each one took a mere millisecond to process, it would still require about 30 years to deal with the entire set one image at a time. With the aim of releasing a final Gaia Catalogue only a few years after the satellite has finished collecting data, the need for distributed or parallel processing is clear.

The data treatment, as well as the instruments themselves, ensures that Gaia performs as accurately as intended. The orientation of the spacecraft must be known to microarcsecond accuracy about all three principal axes. Not only must Gaia be extremely physically stable, but the data processing must also work to this highly precise scale.

Microarcsecond-astrometry raises a further unique challenge: at this level, ordinarily negligible 'relativistic' effects come into play. According to general relativity, the path of light is bent by the gravitational attraction of a massive body such as the Sun. This bending appears to shift a star from its true position. At Gaia's level of accuracy, this is constantly significant, distorting



the apparent positions of celestial bodies across the entire sky. Highly precise relativistic corrections must therefore be calculated and introduced into the data processing.

It is clear that the Gaia data challenge is an enormous task in terms of expertise, effort and dedicated computing power. In late 2006, ESA's Announcement of Opportunity for Gaia's data processing was released, calling for a proposal to build and operate a single processing pipeline leading to the intermediate and final mission products. The announcement expected the system to be developed as a collaboration between ESA's Gaia Science Operations Centre and a broad scientific community supported by national funding agencies.

In response, a large pan-European team of expert scientists and software developers submitted their proposal for a comprehensive system capable of handling the full size and complexity of the Gaia data. In May 2007, ESA's Science Programme Committee approved the proposal put forward by the Data Processing and Analysis

Consortium (DPAC). At this point, DPAC became officially responsible for Gaia data processing and analysis.

DPAC

DPAC is a collaboration drawing its membership from all over Europe, including a diverse community of more than 300 scientists and software engineers, spread throughout more than 15 countries, and six large Data Processing Centres. The consortium brings together skills and expertise from across the continent; its international nature and cooperative spirit reflects that of ESA itself.

The consortium is divided into specialist units known as Coordination Units (CUs). These are the building blocks of DPAC, with each unit assigned a unique set of processing tasks. They are supported by the Data Processing Centres (DPCs), the centres with the actual computer hardware. While the CUs are structured for developing the software, each is closely associated with at least one DPC, where their 'algorithms' are actually used.

Besides the technical challenge faced by DPAC, the sociological challenge of coordinating the efforts of 300 people undertaking an unprecedented feat in data processing for astronomy is not to be underestimated. Effective communication, organisation and maintaining motivation for the large consortium as essential.

The Gaia Data Flow

Gaia's processing is broken down into components to facilitate distributed development, identifying the major parts of the system that may operate relatively independently. This approach is also driven by the fact that the system is to be developed in many countries by a number of teams, each with different levels and fields of expertise.

Over the 5-year mission, Gaia will yield a total uncompressed data volume of around 100 Tbytes. The satellite will have contact with the ground station once a day for an average 11 h, during which an uncompressed volume of roughly 50 Gbytes will be downlinked at a rate of a few Mbits per second.

Within an hour, the Mission Control System at the Mission Operations Centre in Germany will receive Gaia's housekeeping telemetry that reveals the health of the craft and its payload. Meanwhile, on the Science Operations Centre (SOC) side in Spain, the science telemetry is received for preliminary processing in the Initial Data Treatment. This is the first step in unravelling the complex scientific data; this preliminary processing decodes and decompresses the telemetry, initially matches observations to known objects, and works out a 'rough' (sub-arcsecond) satellite attitude.

The SOC also carries out 'First Look' processing – this is Gaia's regular health check. While the Mission Control System does basic system monitoring, the First Look indicates any anomalies in the science output that can be corrected on board. The Flight Control Team resolves any anomalies through commands to the satellite.

DPAC focuses on processing the data into science products rather than the presentation of the final catalogue. Production of the actual catalogue will be covered by a future Announcement of Opportunity. The ninth Coordination Unit, 'Catalogue Access' is a placeholder, to be activated at a later date.

Scientific Processing

After the complex raw data from Gaia's



instruments have been untangled, the intermediate data requires scientific processing as the next step towards the final data product. Basic meaningful scientific data will be extracted via 'iterative processing', with the output of

one cycle forming the input of the next, gradually moving closer to the result. Unlike the near-realtime processing outlined above, each iteration will take around 6 months or more. This will continue after the satellite has finished



An artist's impression of our Galaxy (R. Hurt/JPL-Caltech/NASA)

its job, with each CU sending its output to the overarching Main Database.

The Main Database is the hub of all the data in the system. It is planned that this database will be 'versioned' every 6 months or so, with the final Gaia

Catalogue being derived from by about 2020.

Since the science processing is iterative, each new version of the database will be derived from the preceding version. On detecting something of immediate interest to the broader scientific community, DPCs may issue occasional 'science alerts' in the course of the processing. These will be sent to the SOC for immediate dissemination.

One of the most intensive tasks is the 'astrometric core solution', which is key to unlocking the data potential of Gaia. This produces the calibration data and attitude solution needed for all the other treatments, as well as the accurate positions and motions of about 100 million well-behaved primary objects. After this is performed and the products stored in the central database, the more specialised tasks can begin: photometric and spectroscopic processing, and the treatment of all the difficult objects, such as multiple stars.

Further specialised analyses deal with the types of variable stars and retrieving stellar astrophysical information such as temperature and chemical composition.

The Astrometric Core Solution

The vital astrometric core solution is obtained via the Astrometric Global Iterative Solution (AGIS). This is the foundation of DPAC's scheme for reconstructing results from raw data, a procedure that iteratively adjusts unknown parameters to converge gradually on an optimal solution. Variants of the Global Iterative Solution will be used in the photometric and spectroscopic processing.

AGIS will be applied to around 10% of the objects – that means 100 million objects, and hundreds of millions of unknowns. AGIS begins by calculating astrometric parameters (positions and motions) for the predictable, well-behaved primary objects by firstly assuming that attitude and calibration parameters are known. Assuming then that the astrometric and calibration parameters are known, the attitude is estimated. Then, with the assumption

Key terms

Astrometry: the precise measurement of the positions and motions of celestial bodies.

Photometry: the measurement of the brightness of a celestial body over wide bands of wavelength.

Spectrometry: the measurement of the spectrum of light emitted by a celestial body.

Radial velocity: the speed at which a celestial body approaches or recedes from the observer.

Magnitude: used here to mean 'apparent magnitude', this is the brightness of an object as seen by the observer. The scale is negative and logarithmic: the higher the number, the dimmer the object appears. The Sun has an apparent magnitude of -26.7 ; the next brightest star in the sky, Sirius, is -1.4 . The dimmest stars our eyes can see are around $+6$.

Dark matter: hypothetical matter of uncertain composition that potentially accounts for the majority of mass in the observable Universe. Its presence is inferred from gravitational effects on visible matter; neither emitting nor reflecting light, it cannot be observed directly.


Quasar: an extremely bright, active core of a very distant young galaxy.

Supernova: an exceptionally bright explosion at the death of a massive star.

that astrometric and attitude parameters are known, calibration parameters can be estimated.

The whole sequence will be repeated several times in the course of the processing, perhaps once every 6 months during the initial accumulation of observations. Iteration is necessary as many times as it takes to converge upon a solution that fits the data.

Conclusion

DPAC has a pivotal role in fulfilling Gaia's potential. Excellent progress has already been made and further challenging work still lies ahead. In the course of the processing, intermediate data releases with valuable scientific content are expected. When the Gaia Catalogue is finally published in around 2020, DPAC's work will be complete, and Gaia's processed data will be made freely available for investigation by the world's entire scientific community. 

Further information on Gaia and its mission can be found at <http://www.esa.int/science/gaia>