Science with Gaia: 2021

Essays on the scientific results from Gaia
michaelperryman.co.uk

Michael Perryman

Essays 1–52 (Jan – Dec 2021)

Preface

GAIA IS a mission of the European Space Agency dedicated to astrometry – the measurement of the positions of celestial bodies. The satellite was launched in 2013, and operated until January 2025. Gaia provides the distances and motions (and much other related data) of more than two billion stars in our Galaxy and beyond, with an unprecedented accuracy barely imaginable 25 years ago. It builds on the success of ESA's pioneering Hipparcos mission, which was operated in orbit between 1989–93. The Hipparcos Catalogue of nearly 120 000 stars was published in 1997.

Because of the enormous amount of data processing involved, improved Gaia catalogues are being released progressively, with Data Release 1 in 2016, Data Release 2 in 2018, Early Data Release 3 in 2020, and Data Release 3 in June 2022. Data Releases 4 and 5 are scheduled for the end of 2026, and 2030, respectively.

SINCE THE beginning of 2021, I have been writing a (mostly weekly) 2-page 'essay', picking out some scientific highlights of the mission as they are emerging, or as they caught my attention, and mixing them with occasional asides on related topics, including the history of astrometry, and some more technical, managerial, or developmental aspects of both Hipparcos and Gaia.

Who are they written for? Anyone who might have a general interest in science and astronomy, including amateur astronomers, young scientists starting out on their careers, mid-career scientists looking in on Gaia for the first time to get a feeling of what is possible, and even specialists looking in from different areas of astronomy, or physics more generally.

THE SCIENTIFIC TOPICS I select each week are not necessarily the most important. They do not follow any specific sequence. They are not a complete review of a given topic. Many will be quickly superseded by new results. But together, they are a look at what this long journey of space astrometry is achieving. They offer a snapshot of some of the discoveries that Gaia is making, and written in a form that I hope will be reasonably accessible to those not so deeply involved. I post these weekly essays on my own www site michaelperryman.co.uk.

In each, I have included a footer (DR1, DR2, EDR3, DR3) to indicate which of the (latest) data releases the essay refers to. I have used DR0 to signify technical or historical material not connected with any specific data release. This is intended to communicate how current (or out of date!) any particular essay is likely to be.

Only a few references are included, and these are (generally) 'discreetly' hyperlinked for those who want to read more. Where references appear in the form (Einstein 1908) or www.gaia.com, clicking on the text (even though not generally highlighted) should lead to the relevant (ADS) online article.

HERE, I gather all of my essays from one year in a single compilation, which can also be displayed on-screen in a 'flip book' format.

Michael Perryman

Contents

1. The measurement of angles

ASTROMETRY IS the branch of astronomy dealing with the positions of celestial objects. Its history is a large and multiply-connected field, having its origins in the earliest records of astronomical observations more than two thousand years ago, and extending to the high accuracy observations being made from space today. Over the centuries, improved star positions led to remarkable and revolutionary advances in understanding our place in the Universe.

Perhaps foremost amongst these was the transformative understanding of the motion of the Earth, and the associated acceptance of the heliocentric hypothesis, and our understanding of the scale of the solar system. Measurements led an understanding of the motion of the Sun and Earth through space, and the comprehension and acceptance of Newtonianism. They also proved crucial to the practical task of maritime navigation.

Another challenging task which has underpinned the field for the past 400 years, and which continues to do so to this day, was that of determining distances to the stars, making use of the extended measurement baseline given by the Earth's orbit around the Sun.

When quantified for the first time in the 1830s, stellar distances revealed, at a stroke, the utter vastness of the Universe. What followed was a focus on determining the distances, motions, and physical properties of stars, and on the resulting ability to characterise the structure, dynamics – and eventually origin – of our Galaxy.

After a period in the middle years of the twentieth century in which accuracy improvements were greatly hampered by the perturbing effects of the Earth's atmosphere, ultra-high accuracies of star positions from space platforms have led to a renewed advance in this fundamental science over the past few years.

ASTROMETRY, THEN, is concerned with the accurate measurement of the positions and motions of celestial objects. This includes the positions and motions of the planets and other solar system bodies, stars within our Galaxy and, in principle, galaxies and clusters of galaxies within the Universe.

Since recording and refining the positions of the stars and planets was one of the few investigations of the heavens open to the ancients, astronomy and astrometry were largely synonymous until a little more than a century ago, when other types of astronomical investigation, such as spectroscopy, became possible.

Astrometry therefore has a remarkably long scientific history, while it remains acutely topical today. Over more than two millennia of recorded history, star positions have been measured with progressively increasing accuracy and fundamental advances in our understanding of the Universe have accompanied this progress.

As in all sciences, advances have traced out a perpetual contest between theory and observation. New ideas down the centuries demanded better observations to confirm them, while instrument advances provided new empirical evidence, stimulating new ideas. Throughout this history, advances in astrometry have benefitted from telescope improvements, from the control of measurement errors and from the ability to graduate and further subdivide angular arcs on the celestial sphere.

Recording star positions progressed through naked eye observations, later assisted by optical telescopes, through to the large-scale recording of stellar images on photographic plates of the late nineteenth century, to the high-efficiency detectors of the last twenty years.

SOME 40–50 YEARS AGO, progress ran into almost insurmountable problems imposed by the Earth's atmosphere. Scientific advances associated with improved astrometric measurements faltered, and astrometry consequently receded in global scientific importance in the face of many other rapidly-advancing branches of observational and theoretical astronomy, such as spectroscopy and cosmology. It took a back seat as observations opened up in other electromagnetic frequencies, such as in the radio, infrared, and X-ray.

But in the past two decades, unprecedented positional observations made from two satellites placed above the Earth's atmosphere have ushered in a renewed and unparalleled advance in measurement capability.

ASTROMETRIC MEASUREMENTS essentially involve determining the position of a star as it appears projected on the celestial sphere. The star's distance being, at least to a first approximation, unknown, positions at any time can be simply and uniquely specified by the two angular coordinates of spherical geometry, precisely corresponding to latitude and longitude on Earth.

The origin of the coordinate system is a delicate issue in practice, but the principle is straightforward: just as for geographical latitude, one coordinate can be tied to the extension of the Earth's equatorial plane and, as for geographical longitude and the choice of the prime meridian as its origin, the other is referenced to some arbitrary but well-defined direction in space.

The basis of astrometric measurements, then, is the accurate measurement of tiny angles that divide up the sky. Dividing a circle, whether on paper or on an imaginary sweep of the celestial sky, is a task well-posed in principle. Practical techniques for doing so aside, it is only necessary to agree on the unit of subdivision. Although scientific users today work and calculate angles in radians, the commonly accepted choice of 360 degrees in a circle was made for us long ago.

Ascribed to the Sumerians of ancient Babylonia, more than 2000 BCE, it was perhaps guided by the number of days in a year. One degree was subdivided into sixty minutes of arc, and each minute of arc was divided still further into sixty seconds of arc. The choice of sixty rests on the number itself being highly composite: it has many divisors, which facilitated calculations with fractions performed by hand.

TO VISUALISE these angles, the Sun and the Moon both cover the same *angle* on the sky, about half a degree. The much smaller one second of arc corresponds to a linear distance of one meter viewed from a distance of about two hundred kilometers. This very small angle turns out to be a particularly convenient angular measure and benchmark in astronomy, and it has been used to construct the very basic measure of astronomical distances, the parsec.

In very round numbers, one second of arc is also the angle to which astronomers can measure, with relative ease today, the position of a star at any one moment from telescopes sited below the Earth's atmosphere. It is the shimmering atmosphere which has most recently pushed these measurements to space, and a little background is useful to explain more carefully why.

The lowest portion of our atmosphere is known as the troposphere (from the Greek 'tropos' for 'turning' or 'mixing'). Extending to a height of about ten kilometers, it contains three quarters of the atmosphere's mass, and within it turbulent mixing of the air, due to convective heating rising from the Earth's surface, plays an important part in our atmosphere's structure and behaviour.

Turbulence affects light rays passing through the atmosphere, and causes the familiar twinkling of star light. Already Eratosthenes (276–194 BCE) had commented on their 'tremulous motion'. To minimise this, astronomers build their telescopes at high mountain sites where the thinner atmosphere and smaller turbulence gives more stable images. At good sites, the dancing motion might drop below a second of arc, but not by much more.

The human eye imposes its own limit to measuring angles of about one minute of arc. Mainly determined by the small diameter of the pupil through which light enters the eye, this limit is many times worse than that imposed by the atmosphere.

UNTIL THE INVENTION of the telescope, observations by eye could only place far less stringent limits on the accuracy of star positions. The introduction of the telescope, credited to Dutch opticians in the opening years of the 17th century, but more famously improved upon by Galileo in 1609, brought with it two distinct improvements. First, it could detect much fainter objects, revealing vastly more than were visible by eye. The larger diameter of the telescope aperture also gave improved positional accuracies.

Making telescope mirrors larger improves the accuracy, but only up to the point that the atmospheric turbulent motion sets in at around one second of arc. Above that, even the very large 10-m class telescopes on the ground generally fail to break through the accuracy limit on angular positions set by the atmosphere.

For this reason, there is an important distinction between the angular resolution of an optical telescope, and the formal accuracy in positional location on the one hand, and the relative positional accuracies that can be achieved over large angles across the sky on the other.

SO MUCH FOR one second of arc. It is a tiny angle, corresponding to the size of a Euro coin viewed from a distance of 5 km. It remains problematic enough to measure, and one which proved a great challenge for the astronomical instrument makers of earlier centuries.

Space astrometry took a giant leap with Hipparcos in 1989, reaching accuracies of one thousandth of one second of arc. This angle corresponds to the size of an astronaut on the Moon viewed from Earth, a golf ball in New York viewed from Europe, the diameter of human hair seen from 10 km, or the (angular) growth of human hair *in one second* when viewed from a distance of 1 m.

The Gaia satellite, launched in 2013, is advancing this by a further factor one hundred, targeting accuracies of a few microseconds of arc, corresponding to one Bohr radius viewed from a distance of 1 m. Such accuracies, naturally, pose extreme engineering challenges for optical quality, detector performance, and gravitational and thermal instrumental flexure.

2. Why measure star positions?

MEASURING AN accurate position *per se* is rarely the ultimate objective of astrometry. Rather, the positions of stars in the sky vary minutely with time for a number of reasons. The crucial point is that repeatedly measuring a star's position over a period of months and years can track certain tiny motions which prove central to understanding their nature. It turns out that measuring the positions of stars offers deep insight into their properties, with cascading implications on diverse topics such as the structure and origin of our Galaxy, and the origin and age of the Universe.

Those with even a little familiarity with the night sky will know that the stars do not appear in the same position from night-to-night, nor even between the start of the night and the end, but *appear* to move slowly across the night sky. To a first approximation, the stars occupy fixed positions relative to each other, and it is simply the spinning of the Earth on its axis, once every 24 hours, combined with its motion around the Sun, once per year, which together give the stars their apparent collective movement. The apparent rotation of the heavens is just a consequence of the rotating Earth.

EVERY STAR IS moving through space. We know this now, although 300 years ago scientists did not. As a result, over many decades or centuries, small displacements of some of the most swiftly moving stars do begin to be discernible. Manifestation of these motions was first reported by Edmond Halley in 1718. Following Halley's discovery, other stellar motions were soon reported, and the collective study of stellar motions was born. Today, ultra-high accuracy astrometric measurements allow these star motions through space to be detected over relatively short periods of months or years. But to the human eye, relative star positions remain fixed without change over hundreds if not thousands of years.

Measuring how their angular positions on the celestial sphere change with time gives is called the star's 'proper motion'. The name is a little cryptic, probably drawn from the French 'propre' for 'own', but was used to make clear that what is being measured is the motion through space of the star itself, and not that due to other effects like the Earth's rotation. It also reminds us that what has been measured is an *angular* shift over time, and not the actual speed of the star through space.

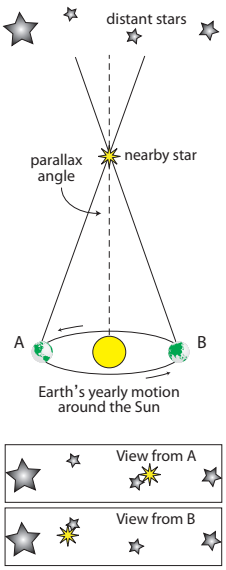
The two are closely related, but the distinction is important. What we see is simply the star's movement projected onto the celestial sphere, which we can only describe in terms of an angular motion. Without knowledge of the star's distance, its true velocity through space cannot be inferred: a star whose position has changed by a certain amount over a few years might be a relatively nearby star moving slowly through space, or a star at a greater distance moving more rapidly.

In practice, stars with large proper motions do tend to be nearby, and searching for high proper motion objects, by comparing photographs of the sky a few years apart, has proven a bountiful way of sifting out potentially nearby stars. But most stars are far enough away that their angular proper motions are very small.

A star's distance is therefore needed to convert its angular motion across the celestial sphere into a true space velocity. And knowledge of a star's distance is needed to convert its observed properties, and in particular its apparent brightness (which varies considerably between stars), into true physical quantities, notably its intrinsic luminosity. It is these basic physical properties of each star which are essential ingredients in putting together a picture of its composition and its internal structure, its age and its past and future evolution.

HERE WE ENCOUNTER a major problem, for the distances to even the nearest stars are truly vast. They are enormously and extravagantly large, and there are no analogies that really allow us to comprehend them. John Herschel (1792–1871), son of the illustrious William, attempted to describe the unimaginable distance scales: *"To drop a pea at the end of every mile of a voyage on a limitless ocean to the nearest fixed star, would require a fleet of ten thousand ships, each of six hundred tons burthen."* In another attempted analogy, the world's population of seven billion people, spaced out one every

five thousand kilometers, would just about stretch to the nearest star. This preposterous extent of space, but more poignantly its stark emptiness, can perhaps be conveyed by a scale model in which the Sun is shrunk to a marble 1 cm in size: the Earth would be a grain of salt one meter from it, and Pluto would sit far beyond at forty meters. In this grand orrery, the *nearest* stars, Proxima and α Centauri, would be 200 km away.



The key to measuring stellar distances is based on the classical surveying technique of triangulation. It is based on the fact, known since Copernicus, that the Earth moves around the Sun, taking one year to complete its orbit.

This yearly motion provides slightly different views of space as we move around the Sun. The nearest stars then *appear* to move back and forth with respect to the more distant ones over this annual cycle.

The problem is that the back-and-forth motion is minuscule. Picturing a grain of salt orbiting a marble at a distance of a meter, and using this perspective change to measure a point of light 200 km away, describes the challenge. Astronomers struggled for centuries to measure the first stellar distances, not so surprising in view of the colossal (and unknown) problem facing them.

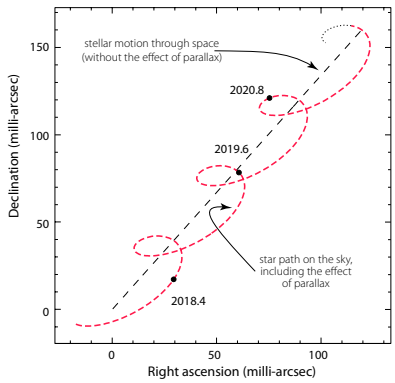
In terms of the Earth's orbital motion around the Sun, then, each star has its own 'parallax angle', corresponding to the ratio of the Earth–Sun distance to that of the star. If measured during this yearly motion, nearby stars appear to oscillate slightly more, back and forth, compared to the more distance stars. The underlying principle of measuring stellar distances, then, is actually rather straightforward; it's just the small size of this parallax motion that makes the task so challenging.

We now know that stars nearest the Sun, like Proxima and α Centauri, have parallaxes of around one second of arc, while more distant stars are even smaller. Down the centuries, attempts to measure this parallax effect failed repeatedly because the relevant angles were so small, almost as if the stars were points of light at infinite distance. The effect was first measured only in the 1830s.

It's one of Nature's coincidences that atmospheric blurring is about the same size as the parallax of nearby stars. Parallaxes of more distant stars required accessing angles of one tenth or one hundredth of a second of arc, and this became feasible only during that latter part of the 20th century. Measuring even more distant stars requires accuracies of around one thousandth of a second of arc, or even better. Before the advent of space measurements, such goals remained well beyond reach.

THE PARALLAX-BASED distance measurement technique is so basic that the fundamental unit of distance measurement in astronomy beyond the solar system is based upon it. Conveying the essentials of 'parallax' and 'second' of arc it is referred to as the *parsec*. The name was proposed by H. H. Turner in 1913, and subsequently adopted by the International Astronomical Union in 1922. One parsec is simply the distance at which a star has a parallax angle of one second of arc as the Earth moves in its orbit around the Sun.

The light-year is another convenient description of astronomical distances, and the two are often used side-by-side (although only the parallax can be measured directly). The light-year is the distance covered by light, which travels at nearly 300 000 km per second, in one year. One parsec is a little more than three light years, and one light-year is some ten *million million* kilometers. The nearest stars are at a distance of a little more than 1 pc, or around four light-years. In terms of parallax angles then, astronomers needed to master measurement accuracies of around one second of arc to measure the distances to the nearest stars. And the more distant the star, the smaller the parallax.



The nearest star cluster to our Sun, the Hyades, lies at a distance of around 40 pc, or a little more than a hundred light years. The spiral arms of our Galaxy closest to us are at around 500 parsec, and the centre of our Galaxy is nearly 10 000 parsec distant, or a colossal thirty thousand light years. Our nearest neighbouring galaxies, the Magellanic Clouds, are some fifty thousand parsecs. Beyond that, great galaxy clusters stretch out to distances of tens of millions of parsecs or more—their light taking tens of millions of years to reach us.

WE CAN SUMMARISE our present knowledge, and hence the context in which astrometry has advanced through history, as follows. Viewed under a celestial magnifying glass, each star in the sky has a tiny component of angular motion due to its velocity through space, plus a minuscule apparent oscillatory motion due the Earth's annual motion around the Sun.

As accuracies improve, various other details become discernible, including the effects of binary companions, planets in orbit around them, and general relativistic light-bending. Quantifying these are the goals of contemporary astrometry.

3. A history of astrometry

Developments in Ancient Greece

THE FLOURISHING of western astronomy over the past few hundred years has its origins in much earlier bursts of scientific activity. Prehistoric sites revealing celestial alignments, such as Newgrange in Ireland and Stonehenge in England, date from around 3000 BCE.

The first recorded developments emerged in Mesopotamia around 1000 BCE where, in the land between the Tigris and Euphrates rivers now occupied by southern Iraq, Assyro–Babylonian astronomers observed the night skies, building on common lore already conscious of the changing daylight over the year.

They observed, measured, and recognised, for the first time, that certain celestial phenomena were periodic: amongst them the regular appearance of Venus, and the eighteen year cycle of lunar eclipses. Their careful records formed the basis for later developments, not only in ancient Greek and Hellenistic astronomy, but also in classical Indian and medieval Islamic astronomy.

EARLY GREEK philosophers, the Pythagoreans amongst them, played a key part in astronomy's earliest awakening. They believed that the underlying regularities, or laws of nature, were discoverable by reason. As part of this philosophical school, astronomers of ancient Greece tried to understand the Universe based on principles of 'cosmos', or order. The revolutionary idea that the Earth might be spherical began to replace the pre-Socratic view that its surface was flat.

Plato (427–347 BCE) and his contemporaries knew that the heavens rotated night after night with constant speed, the 'fixed' stars preserving their relative positions as the heavens turned. But moving in a complex and unfathomable way were the seven wanderers—the Greek *planetes*—the Sun, the Moon, and the planets visible to the naked eye: Mercury, Venus, Mars, Jupiter and Saturn.

Seen from Earth, their positions trace out complex and convoluted paths, sometimes with even backward 'retrograde' loops. As described by Goodman & Russell in The Rise of Scientific Europe 1500–1800 (1991) *'Their erratic behaviour had baffled and infuriated generations of Greek thinkers, up to Plato himself. It seemed impossible to reconcile their celestial meanderings with either the supposed divinity of heavenly bodies or with any simple concept of circular motion.'*

Scientific thinking was dominated by the idea that the Earth lay fixed at the centre of the Universe. This fundamental tenet in mankind's early views completely obstructed the correct interpretation of planetary motions. We now know that the apparently complex paths of the planets follows from the rotation of the Earth, combined with the orbits of the Earth and other planets around the Sun. When interpreted correctly in a heliocentric system, and with elliptical orbits, the motions are simple. But in a system in which the Earth is fixed they are not.

Heraclides had hinted at a Sun-centred system in the fourth century BCE, but his view failed to find support in a culture generally attached to the idea of an Earth fixed in space, which would continue to hold sway, erroneously, for a further two millennia.

Aristarchus of Samos (circa 310–230 BCE) made one of the first attempts to determine the distances and sizes of the Sun and Moon. He deduced the ratio of their distances using trigonometry, by measuring the angle between them when the Moon is exactly half lit. He also argued in favour of the heliocentric, Sun-centred system, a view supported by Seleucus of Seleucia around the second century BCE. But these ideas found little favour at the time, and they remained lost amongst the geocentric, Earth-centred system still being championed by most of his contemporaries.

To explain the complex apparent motions of the planets and the varying speed of the Moon, geocentric proponents could not appeal to planetary orbits which were simply circular. They had to introduce complex epicyclic motions—patterns traced out by circles turning around the circumference of larger circles. Contrived though they were, they broadly explained the irregular speeds of the planets across the sky throughout the year, occasionally even tracing backward loops with respect to the background stars.

At around the same time, Eratosthenes (276 BCE–194 BCE) invented a system of latitude and longitude, and used the varying elevation of the Sun to estimate the size of the Earth, deriving a value which would be used for centuries afterwards.

Eratosthenes argued that on the summer solstice at local noon in Swenet (Aswan) the sun appeared at the zenith, while in Alexandria, assumed to lie due south, the angle of elevation of the sun was 1/50th of a great circle south of the zenith at the same time. He concluded that the distance from Alexandria to Swenet must therefore be 1/50 of the total circumference of the Earth.

Hipparchus (c. 160–126 BCE) is credited with a number of advances in astronomy, although most of what is known about his work is handed down from Ptolemy's second century thirteen-volume *Megale Syntaxis*, or 'Great Compilation'. This became better known as the *Almagest*, 'The Greatest', as assigned by 9th century Arabic translators. Ptolemy pioneered the classification of star brightness still in use today, dividing them into six groups, the brightest designated as first magnitude (the first to be seen at dusk), and the faintest as sixth.

He followed the ancient Babylonians in dividing a circle into 360 degrees, each of 60 minutes of arc, and he compiled the first systematic star catalogue, recording star positions with an accuracy of about one degree. He was the first to describe the precessional motion of the fixed stars, that is the steady wobbling of their positions over decades due to the steady change in the position of the Earth's spin axis in space, just like a spinning top.

BUT HIPPARCHUS incorrectly continued to uphold the geocentric system. His argument was that a precisely circular orbit of the Earth around the Sun failed to explain the planetary motions. We know now that the planetary orbits are elliptical, so that his argument was compelling, but fallacious. Nevertheless his views, and his authority, effectively ensured that the heliocentric hypothesis would lay discarded for many centuries.

More than two hundred years later, in the second century CE, Ptolemy would put forward his own variant of this geocentric view, and would also invoke the epicyclic motions to predict, successfully even if based on flawed models, the future positions of the planets.

Greek scientific activity came to a fairly abrupt end. According to Goodman & Russell (1991): *'…the most likely reasons seem to be the paucity of scientists and their isolation…Education in Greek schools concentrated on music, poetry and gymnastics, not on science…For Europe to have developed the sciences further from these Greek foundations, knowledge of Greek, close contact with the Greek scientific texts, and sustained interest in what they might teach were all necessary. But in the centuries after the fall of the Roman Empire in the west, none of these conditions was satisfied.'*

The 'Dark Ages' in Europe: 200–1500 CE

THE SUBSEQUENT decline of the Roman empire, in population, economic and political order, precipitated by barbarian attacks, decimating epidemics, and inability to provide for the succession of government, ushered in the 'Dark Ages'.

To the East, meanwhile, China's first major economic burst under the Han dynasty (206 BCE–220 CE) nurtured a philosophical period roughly coincident with the innovative centuries of Greek philosophical and scientific thought. In India the Gupta empire, from around 320 CE, also stimulated navigation and advances in numeracy, embracing the concept of 'zero' as well as the use of Arabic numbers. Astronomy was recognised as a separate discipline, and around 500 CE Aryabhata held that the Earth was a sphere rotating on its axis.

Much later in China, under the Sung emperors in the 11th and 12th centuries, scientific advances flourishes. Star catalogues, as well as records of sunspots and comets, have been handed down to us from this time. But China's separation from the west, buffered by the nomadic tribes of central Asia, meant that their records of planetary motions, novae, and supernovae had little immediate influence on Europe's scientific re-awakening.

THE BURGEONING Islamic culture was to dominate the world's economic development from the 7th century for the next 300 years. Geographically closer than China and India, it had more of a direct influence on the west, and played an important part in reviving scientific enquiry in Europe. Supported by the patronage of the Caliphs, Islamic scholars transmitted, translated, and criticised the ancient Greek texts. And knowledge of astronomy was inspired by practical needs: to establish each mosque's direction to Mecca, the timing of daily prayers, and the precise beginning and end of Ramadan.

Amongst their achievements, Al Battani, around 900 CE at his observatory on the Euphrates, refined Ptolemy's description of the orbits of the Sun and Moon. Ibn Yunus (c. 950–1009) described planetary alignments and lunar eclipses accurate enough for a great figure in late nineteenth century astrometry, Simon Newcomb, to use them for his own theories of lunar motion.

Ulugh Beg, grandson of the Mongol conqueror Tamerlane, constructed a sextant of 36 m radius in Samarkand in 1428, a circular arc between marble walls. His catalogue of 994 stars, with positions accurate to about one degree, was the greatest star catalogue between those of Hipparchus and Tycho Brahe.

European revival: 1500–1700 CE

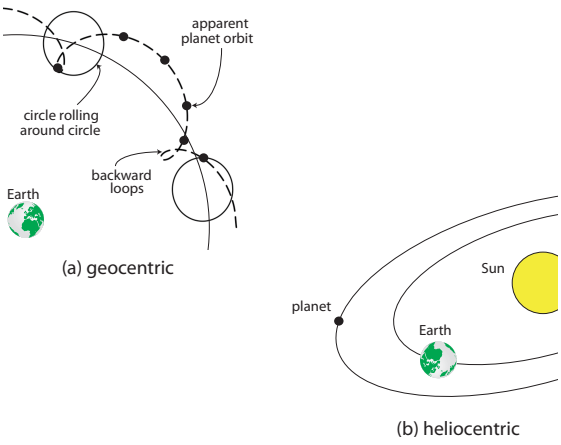
EUROPE'S SLOW EMERGENCE from the 'Dark Ages' began to gather pace under the Carolingian dynasty, from around 730 CE onwards. Trade and towns in western Europe started to revive, and economic life progressively shifted from the Mediterranean to the North Sea and Atlantic coast.

From this renewed prosperity, and improved political stability, the foundations of the modern age slowly emerged. Indeed, research over the past fifty years has quite dispelled the idea that Europe between 500–1500 CE was intellectually and technologically stagnant.

A NEW CURIOSITY about the heavens surfaced and thrived. The basic imponderable of astronomy and cosmology until the Middle Ages, that of the inexplicable motion of the five planets known at the time, was picked up again after a pause of a full millennium.

Nicholas Copernicus (1473–1543) openly drew on this rich medieval tradition, and finally laid the secure foundations for a credible heliocentric world model, in which the Earth moves in orbit around the Sun rather than vice versa. Little new observational evidence motivated his thinking: he lived before the invention of the telescope, and his best observational accuracy was only about ten minutes of arc. Rather, rediscovery and reinterpretation of the ancient texts played a major part in the origins of Renaissance culture in general, and astronomy in particular. According to Sir Thomas Heath (1913): *'Copernicus himself admitted that the [heliocentric] theory was attributed to Aristarchus.'*

Copernicus proposed that the Earth, far from being fixed in space, was actually subject to three kinds of motion. The first was an annual orbit around the Sun. The second was a daily rotation accounting for day and night, but about an axis tilted with respect to its orbit plane which would account for the changing seasons. Third was a more complex and very long period wobble of the Earth's axis as it spins, known as precession.



His *De Revolutionibus Orbium Coelestium*, 'Concerning the Revolutions of the Heavenly Bodies' of 1543, marked the beginning of Europe's scientific awakening.

In a (hypothetical) Earth-centred (geocentric) system, the apparent motions of the planets, as viewed from the Earth, could only be explained as a superposition of complex 'epicyclic' curves. In our present understanding the planets, including the Earth, orbit the Sun in elliptical paths, and the solar system can be explained as a set of planetary masses orbiting the Sun according to Kepler's laws.

THE ACCEPTANCE of a Sun-centred solar system accounted for the most extreme contributions to the backward looping motions of the outer planets.

But Copernicus still needed highly contrived epicycles to explain their detailed motions, albeit of a smaller magnitude than those invoked by Ptolemy in his Earth-based system. Other subtleties were needed to match the known orbits of the planets, for Copernicus was erroneously trying to fit a series of circular motions to their yet-to-be discovered elliptical paths.

Only with the later work of Johannes Kepler, Galileo Galilei and Isaac Newton, and the realisation and understanding that the planetary orbits were elliptical rather than circular, could the need for epicyclic motions be discarded altogether.

By demonstrating that the motions of celestial objects could be explained without putting the Earth at rest in the centre of the Universe, the work of Copernicus stimulated further scientific investigations and became a landmark in the history of modern science.

IN 1610, Galileo Galilei published his *Sidereus Nuncius*, which described the surprising observations made with his newly-invented telescope—mountains on the Moon, moons around Jupiter, and patchy nebulae for the first time resolved into innumerable faint stars.

His support for Copernican heliocentrism set in train a lengthy and well-documented conflict with the Catholic Church, leading to suggestions of heresy, and his eventual trial and house arrest in 1633.

Giordano Bruno (1548–1600) was a proponent of heliocentrism and the infinity of the universe, who had earlier burned at the stake albeit for other more extreme theological heresies. Such were the harsh penalties for questioning the authority of the Holy Scriptures, which decreed that the Earth was the centre of the Universe, and that all heavenly bodies revolved around it.

From the altar of St Peter's Basilica in Rome in March 2000, Pope John Paul II issued an apology for the errors of the Church over the last two millennia, including the trial of Galileo: *'The error of the theologians of the time, when they maintained the centrality of the Earth, was to think that our understanding of the physical world's structure was, in some way, imposed by the literal sense of Sacred Scripture.'*

JOHANNES KEPLER (1571–1630) was as an important figure in the 17th century astronomical revolution, best known for his eponymous laws of planetary motion. He defended heliocentrism from both a theoretical and theological perspective. His observational work with Tycho Brahe encouraged his own protracted attempts to calculate the orbit of Mars around the Sun.

Eventually, in 1605, he found that while a circular orbit did not match the observations, an elliptical one did. It was a simple answer which he had previously assumed too straightforward for earlier astronomers to have overlooked. He concluded that all planets move in ellipses, with the Sun at one focus. This deduction, his first law of planetary motion, provided a foundation for Newton's theory of gravitation.

Aside from his mathematical skills, Kepler lived at a time when there was no clear separation between the science of astronomy and the pseudoscience of astrology, and he also had a reputation as a skilful astrologer.

ISAAC NEWTON (1642–1727) occupies a lofty pedestal in the history of science, and his *Philosophiae Naturalis Principia Mathematica* of 1687 is arguably its most influential book. He bestowed on mathematics and physics a rich collection of new ideas. Together, and in a stroke, his laws of motion, gravitational attraction, and the inverse square law of gravity gave an explanation of the motions of all celestial bodies.

But this package of new ideas, Newtonianism, was not the only scientific movement competing for support, and it was not accepted immediately. Rivals included Hutchinsonianism in England, centred around the Trinitarian theology of John Hutchinson, and Cartesianism in France, based on the influential philosophical doctrine of René Descartes.

The heliocentric hypothesis eventually prevailed, and Newtonian gravity along with it. With their joint acceptance came an inevitable consequence, a conclusion that would mark a fundamental turning point in science. For if the Earth indeed moves in orbit around the Sun, then the 'fixed' stars cannot remain truly fixed in space.

Unless they were at infinite distance, they would have to possess a parallax motion—an oscillation of their *apparent* position arising from the Earth's annual motion around the Sun. To be sure, neither Aristarchus nor Copernicus had observed the effect, and this fact alone implied that the distances to the stars must dwarf even the colossal distance scale of the solar system.

THE CONCLUSION that the parallax effect had to exist therefore seemed inescapable. A renewed push to detect it began, armed with the certain knowledge that the effect being sought would be tiny. Great improvements in measurement accuracy would be needed before the effect could be measured.

Newtonianism and parallax: 1600–1850 CE

STARTING SOME three or four centuries ago, the search for parallax, the further comprehension and definitive acceptance of Newtonianism, and understanding the precise nature of the Earth's motion through space were interwoven, and together motivated the progressive improvement of angular measurements. A related but more urgent practical problem came to a head at the same time: the navigational problems associated with the determination of longitude.

For most of history, explorers in general and mariners in particular had struggled to determine their precise longitude, their point east or west of some reference point on the Earth. Latitude has the Earth's equator as a natural reference plane, and it can be determined by observing the altitude of the Sun or stars using specialised protractor-like instruments like the quadrant or sextant, or the astrolabe, a sort of analogue calculator capable of working out different kinds of problems in spherical astronomy. There is, however, no such unique reference position for longitude, and no practical means for its direct estimation.

For a ship lost at sea on the slowly-spinning Earth, estimating longitude was frequently a matter of life or death. But it was tied directly to the knowledge of time. Without time, there was no hope of determining longitude: any uncertainty in the local time corresponds to an uncertainty in a star's transit across the local meridian, and an equivalent uncertainty in the observer's longitude.

THE PROBLEM WAS urgent, and the economic consequence of ships, cargos and lives lost at sea was substantial. In France, Louis XIV promoted the construction of the Paris Observatory, established in 1667 under director Giovanni Domenico Cassini, with the express purpose of extending France's maritime power and expanding her international trade.

In England, King Charles II was similarly moved to found the Royal Greenwich Observatory in 1675, with the purpose of compiling detailed star maps for navigational purposes. He instructed the first Astronomer Royal, John Flamsteed, *'to apply himself with the most exact care and diligence to the rectifying of the tables of the motions of the heavens, and the places of the fixed stars, so as to find out the so much-desired longitude of places for perfecting the art of navigation.'*

In 1725 Flamsteed's *Historia Coelestis Britannica* was published posthumously, containing his catalogue of 2935 stars. It was the first significant contribution of the Greenwich Observatory, and a landmark in the history of astrometry—positions, accurate to around ten or twenty seconds of arc, were the first measured with telescopic sights, and a major improvement over earlier work.

STAR CHARTS ALONE, however, could not provide a solution to the problem of navigation. Without a clock that could keep accurate time over months of an ocean voyage, there was no practical way of establishing what time it was at the reference point. With Galileo's discovery of the four brightest moons of Jupiter in 1610, named by him as the Medicean stars after his patron but subsequently named the Galilean moons in his honour, it became possible in theory to deduce the time on board ship by observing when the satellites appeared from behind the planet—the events occurred frequently and, more importantly, predictably.

The world's first national almanac, the *Connaissance des Temps*, giving these eclipse timings, was published from 1679. Tables could then be consulted to see when these events were due to occur as measured at the prime meridian. Galileo himself pursued this approach to navigation during his lifetime, and even petitioned King Philip III of Spain who had also offered a financial reward for a breakthrough in determining longitude. Yet such measurements could only be made at night, were much at the mercy of the weather, and quite impossible from a rolling boat in high seas, and it failed to provide a practical solution. Before the middle of the eighteenth century, most sailors continued to use a variant of dead reckoning to try to keep track of their position. Galileo died in 1642, before his method became widely used by cartographers on land.

The search for a solution was spurred on by the Longitude Act of 1714, during the reign of Queen Anne. The British Parliament offered a prize of £20 000, a fortune of some £6 million in present worth, for a method that could determine longitude within thirty nautical miles. A solution was eventually found through the use of accurate celestial charts and lunar tables, in combination with the measurement of precise time.

WITH THE SUCCESS of the marine chronometer in the 1760s, pioneered by English clockmaker John Harrison, time could at last be carefully measured and accurately transported throughout a long voyage. Accurate clocks eventually became commonplace. The problem of navigation at sea was considered as solved, and the Board of Longitude was dissolved in 1828 (the story is told in the popular account by Dava Sobel). Not until 1884, however, was the International Meridian Conference meeting in Washington DC to adopt the meridian passing through Greenwich as the universal, if quite arbitrary and long contested, zero point of longitude. France abstained, maintaining her preferred use of the Paris meridian until 1911 for timekeeping purposes, and until 1914 for navigation.

From 1767, the Nautical Almanac has been published annually. From 1958, the US Naval Observatory and the HM Nautical Almanac Office have jointly published a unified volume, for use by the navies of both countries. It still tabulates the positions of the Sun, Moon, planets, and a number of stars selected for ease of identification and widely spaced across the sky. To find the position of a ship or aircraft by celestial navigation follows the method unchanged for more than two centuries: the navigator uses a sextant to measure the height of a chosen star above the horizon, notes the time from a chronometer, and deduces location by comparing the star's position with that given in the almanac for that time. Thousands of lives and considerable fortunes had been lost before star charts in combination with transportable time could be used for reliable navigation.

Well into the 1800s, star positions provided the most accurate means of determining geographical coordinates, and with them the distance between cities or the position of national borders. An interesting parallel occurs today: the huge civilian, commercial, and military reliance on global satellite navigation, notably GPS, depends crucially upon the inclusion of the delicate effects of Einstein's special and general relativity: omit them from consideration, and positions would be several kilometers in error after only a few hours. In this area alone, astronomy and relativity have proven indispensable to this important social and commercial venture.

AS COPERNICANISM spread throughout Europe, and the heliocentric cosmos gained acceptance, the race to measure parallax gathered pace. Even before 1600, astronomers were in agreement that the crucial evidence needed to detect the Earth's motion around the Sun was the measurement of trigonometric parallax. The early British Astronomers Royal, amongst others, appreciated the importance of measuring stellar distances, and had devoted much energy and ingenuity to the task. For example, according to Allan Chapman (1990) *'Though the application of the telescope sight to angular measurement from the 1660s constituted a major technical breakthrough, the optics involved were simple, conservative in type, and secondary to the engraved divisions. This becomes apparent from the letters, notebooks, and Gresham College lectures of John Flamsteed, the first Astronomer Royal, for while his early decades at Greenwich were beset with instrumental problems, they were almost exclusively of a mechanical nature. The equality of scale degrees, or the regularity of a micrometer screw, claimed more attention than the resolving power of telescope lenses, and nowhere in his extensive writings is more than passing attention paid to optical resolution'.*

Indeed it was improved angular measurement, not enhanced visual acuity, that held the key to a range of astronomical problems from the sixteenth to the early nineteenth centuries. But it was still to take a further two hundred and fifty years, and failure upon failure, until the first star distances were measured.

Advances in positional accuracy: 1500–1700 CE

FROM ANCIENT TIMES through to the start of the twentieth century, the measuring of celestial positions had always been central to astronomical research. The quality of the instruments determined the accuracy of the measurements.

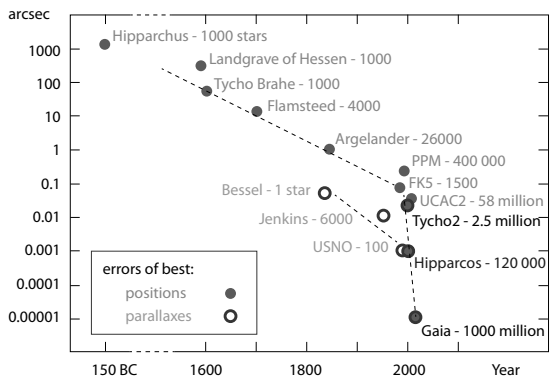
The art of dividing a physical circular scale into degrees and minutes of arc was but a practical problem, essentially one of accurately marking off successively smaller angles. But it was one of such technical complexity that it now presented the principal barrier to advancing research.

During the Middle Ages, European and Islamic astronomers adopted a brute force approach to the problem. They constructed observing circles with very large radii (and therefore very large physical dimensions) such that they could more easily inscribe and further dissect more precise angles on their annular limbs.

Tycho Brahe (1546–1601), whose observations provided the basis of Kepler's laws of planetary motion, employed such an instrument. His Great Quadrant had a radius of fourteen cubits, around seven meters, and probably reached an accuracy of around six minutes of arc, one fifth the Moon's diameter.

At his lavish observatory of Uraniborg on the Danish island of Hven, developed under the patronage of Frederick II, King of Denmark and Norway, he used his families of sextants, armillary spheres, and quadrants. By the last decade of the sixteenth century, he was reaching an unsurpassed accuracy of around twenty seconds of arc.

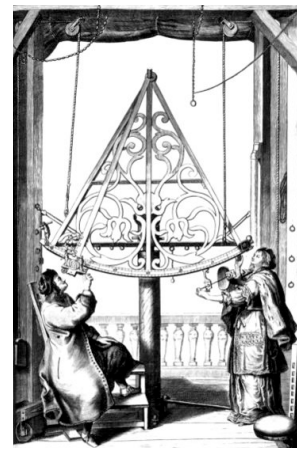
Despite his observational skills and his extravagant funding, Tycho attempted, but also failed, to detect parallax motion. But the accuracies that he achieved allowed him to deduce that the stars must lie several thousand times more distant than the Earth from the Sun. These distances were so immense that he was convinced Copernicus must be in error, and that the Earth was indeed fixed at the centre of a modified 'Tychonic' system.



Astrometric accuracy versus time

IN REALITY, with even the nearest stars having a parallax angle of only one second of arc, Tycho's accuracy was still twenty times too poor, and even his careful measurements could not but have failed to detect its effects. Nevertheless by the end of the sixteenth century, his catalogue of a thousand stars, and a similar effort by Landgrave (Baron) Wilhelm the Wise of Hesse (1532–1592), set the standard for future surveys.

The sextant and quadrant were protractor-like instruments designed to measure angles between pairs of stars, of up to sixty and ninety degrees respectively. Catalogues were built up from many pairs of separations. Portable versions were later fixed in the meridian plane—the imaginary circle perpendicular to the celestial equator and horizon.



The sextant of Johannes Hevelius

Observations with wall-mounted 'mural' instruments began with Tycho's large meridian quadrant. Fixed to the local horizon, stars appear to drift past the local meridian as the Earth spins: this gave one part of the star's coordinates (the equivalent of geographical longitude, or right ascension) from the timing of its transit, and the other (the geographical latitude, or declination) from the graduated instrument itself.

These were later replaced by the meridian circles, consisting of a horizontal axis in the east–west direction resting on fixed supports, about which a telescope mounted at right angles could revolve freely.

Until the late eighteenth century, the art of graduating circular scales into ever finer subdivisions was pursued in earnest, but carried out largely in secrecy to thwart the competition. Wider exposition of practical methods accelerated when the Board of Longitude, which had been formed in 1714 to solve the problem of finding longitude at sea, persuaded John Bird to publish his methods in 1767. In the following decades, Jesse Ramsden, John Smeaton, and Edward Troughton continued the advance of angular measurements.

Prestigious Fellowships of the Royal Society were awarded for their instrument advances, underlining the importance with which the measurement of stellar positions was held, and testament to their innovation. In his chronicle of the rise and fall of economies throughout history, for example, Peter Jay (2000) includes Ramsden's dividing machine for accurate graduation of circles for navigational and surveying instruments as one of the inventions which contributed to the productivity gain that signaled the Industrial Revolution.

During the later parts of the 17th and early 18th century, other instruments were added to the arsenal of techniques for measuring star positions. These included the transit telescope, which added a regulator clock to time the passage of stars across the Earth's meridian. Its more specialised form, the zenith sector, was used by Robert Hooke (1635–1703), one of the most important scientists of his age, in his own attempts to measure the parallax of the bright star Gamma Draconis.

Gamma Draconis is a giant star in the constellation of Draco, and it has been a notable object throughout recorded history.

According to Allen (1899): *'Its rising was visible about 3500 BCE through the central passages of the temples of Hathor at Denderah and of Mut at Thebes. And Lockyer [Sir Joseph Norman Lockyer, 1836–1920] says that thirteen centuries later it became the orientation point of the great Karnak temples of Rameses and Khons at Thebes, the passage in the former, through which the star was observed, being 1500 feet in length; and that at least seven different temples were oriented toward it. When precession had put an end to this use of these temples, others are thought to have been built with the same purpose in view; so that there are now found three different sets of structures close together, and so oriented that the dates of all, hitherto not certainly known, may be determinable by this knowledge of the purpose for which they were designed. Such being the case, Lockyer concludes that Hipparchus was not the discoverer of the precession of the equinoxes, as is generally supposed, but merely the publisher of that discovery made by the Egyptians.'*

The interest of the star Gamma Draconis to the seventeenth and eighteenth century parallax hunters was simply that it lay almost exactly in the zenith of Greenwich, minimising refraction by the atmosphere, and conveniently studied by a fixed telescope pointing straight up—Hooke had cut a hole in the roof of his apartment to observe it. In 1674 he claimed the detection of a parallax for Gamma Draconis of roughly thirty seconds of arc, and with it proof of the Copernican system. Later work showed that his results were in error.

Proper motion and stellar aberration: 1700–1800 CE

A REMARKABLE AND crucial breakthrough came in 1718. Edmond Halley, who had been comparing contemporary observations with those that the Greek Hipparchus and others had made, announced that the bright stars Aldebaran, Arcturus, and Sirius were displaced from their expected positions by large fractions of a degree. He deduced that each star had its own distinct velocity across the line of sight, or proper motion. It was the first convincing experimental suggestion that stars were moving through space.

Halley's scientific achievements were many and varied. He predicted the return in 1758 of a periodic comet which now bears his name, identified solar heating as a cause of atmospheric turbulence, and suggested a measurement of the distance between the Earth and the Sun by timing the transit of Venus. Less successful was his suggestion, to explain anomalous compass readings, that the Earth was a hollow shell some eight hundred kilometers thick. This example also shows the limits in scientific understanding that existed a mere three hundred years ago.

By 1725, instrumental advances had reached accuracies of a few seconds of arc. The Reverend James Bradley, England's third Astronomer Royal, was immersed in his own efforts to measure parallax, and was also focusing his attention on Gamma Draconis. His attempts were unsuccessful, for the star is too distant for the effect to show up at the accuracy then available. But they pushed his own estimates of the nearest stellar distances out to nearly half a million times that of the Earth from the Sun.

More importantly, Bradley's experiments yielded an unexpected surprise: the detection of a small systematic shift in his star positions, of a form very different from that expected from the effects of parallax, and which he eventually correctly attributed as resulting from the addition of the velocity of light to the Earth's velocity as it moves in orbit around the Sun. The usual analogy is that when rain is falling straight down, and you're walking briskly ahead, you tilt an umbrella forward slightly to intercept the apparent direction of the rainfall. It's a consequence of adding two velocities.

Dinghy sailors know the effect well: the flag atop the mast doesn't indicate the wind direction, but that of the wind and boat speed combined. Bradley had pondered the meaning of his perplexing star measurements for three years before enlightenment struck, his insight precipitated by observing such a moving vane on a sail boat on the River Thames.

Bradley's observations of this effect, known as stellar aberration, or the aberration of starlight, was announced in 1729, and arguably rates as one of the most significant discoveries in the history of astronomy. It provided the first direct proof that the Earth was moving through space. His results therefore supported the Copernican theory, that the Sun, rather than the Earth, was the centre of the solar system.

But it confirmed, at the same time, Danish astronomer Ole Rømer's discovery of the finite velocity of light fifty years earlier. Rømer had been observing the eclipses of Jupiter's moons as part of the ongoing challenge to establish a practical method to determine longitude. His own conclusion that the velocity of light was finite, rather than propagating at infinite speed, wasn't fully accepted until Bradley's measurement of aberration provided crucial supporting evidence.

By failing to detect the parallax of Gamma Draconis, even at the unprecedented level of about one second of arc, Bradley's observations went further in confirming Newton's hypothesis of the enormity of stellar distances, and confirmed that the measurement of parallax would continue to pose a technical challenge of inordinate delicacy. In parallel with the direct search for parallax were less direct estimates of stellar distances, for example those made by Newton and others by appeal to the inverse square law, an approach resting on the simplistic (but incorrect) hypothesis that all stars had luminosities comparable to that of the Sun.

This method was extended to circumvent the difficulties posed by the extremely bright Sun by the use of Jupiter as a (reflecting) intermediate calibrator, as first used for Sirius by James Gregory in 1668, and for Vega by John Michell in 1767.

Nevil Maskelyne, England's fifth Astronomer Royal, spent seven months on the remote island of Saint Helena in 1761, a crucial staging and rendezvous point for sailing ships in the South Atlantic. He had been despatched by the Royal Society to observe the transit of Venus, and thereby to improve knowledge of the Earth's distance from the Sun and the scale of the solar system. He used a zenith sector and plumb-line in an unsuccessful attempt to measure the parallax of Sirius during the same expedition.

During the eighteenth century, after Halley's first detection of stellar motions, the movements of many more stars were being announced. In 1783 William Herschel found that he could partly explain these collective motions by assuming that, in addition to the Earth's motion around the Sun, the Sun itself was moving through space. With his sister Caroline, Herschel made numerous important advances: he discovered Uranus in 1781, two moons of Uranus and two of Saturn between 1787–89, and discovered infrared radiation.

Herschel observed and catalogued binary stars, detecting the first orbital motions and, in the process, the first proof that Newton's laws of gravitation applied outside the solar system. He was a prolific telescope maker, and also sought to detect a parallax shift from measurements repeated over the course of a year.

Yet in this, even armed with his largest telescope, a primary mirror more than a meter in diameter and a colossal twelve meter focal length, he too failed. As he wrote in 1782: *'To find the distance of the fixed stars has been a problem which many eminent astronomers have attempted to solve; but about which, after all, we remain in a great measure still in the dark.'*

Meanwhile, another important step in expanding ever larger star surveys was the work of Jérôme Lalande (1732 –1807) in France. His *Histoire Céleste Française* of 1801, gave the places of 50 000 stars with an accuracy of around three seconds of arc.

The symbolic if arbitrary figure of one second of arc was now within sight, and attempts to measure parallax intensified. But since the distances to even the nearest stars were still unknown, nobody could predict what angular accuracy would be needed for the effect to be detected. The topic was the focus of many learned papers published in the opening decades of the 1800s.

The failures of Tycho, Hooke, Flamsteed , Bradley, Maskelyne, Herschel and many others, were followed by a renewed flurry of measurements and false claims: amongst them by Giuseppe Piazzi in Palermo, Giuseppe Calandrelli in Rome, François Arago in Paris (later Prime Minister of France), Baron Bernhard von Lindenau in Gotha, Johan Schröter in Lilienthal, and John Brinkley in Dublin.

In the words of Alan Hirshfel (2001): *'Each claimed victory in what astronomers increasingly perceived as a parallax race. But instead of glory, the recent parallax competitors gained only the suspicion, if not the contempt, of their colleagues.'*

The first parallaxes: 1800–1850

WHAT WAS URGENTLY needed were criteria for selecting stars likely to be close to the Sun, to avoid time wasted in trying to measure distant stars. In 1837, German-born Wilhelm Struve, working at Dorpat in Russia (now Tartu in Estonia), gave three suggestions: the star should be bright; it should be moving with a large angular rate across the sky (although this *could* be a rapidly moving star at a large distance, it was more likely to be 'nearby'); and if the star was one of a binary pair, the two components should be well separated as judged by the time taken to orbit each other.

Struve drew up a list of stars satisfying these criteria. Our present-day knowledge confirms that astronomers were, at last, able to select some of the very nearest stars on which to focus their painstaking measurements.

After many unsuccessful attempts, the very first stellar parallaxes were measured and reported during a burst of activity in the 1830s, two hundred years after Isaac Newton had removed any final doubt that the Earth was in motion around the Sun. After this protracted marathon to detect the first parallax, three scientists breasted the winning tape almost together.

Wilhelm Struve had selected the bright, high proper motion star Vega for study. At his disposal in Dorpat was a twenty-four centimeter aperture refractor, manufactured by the German physicist and craftsman Joseph Fraunhofer, and the largest instrument of its kind in the world. Equipped with a 'filar micrometer', long used for measuring separations of double stars, two tiny parallel wires or threads, often of fine but immensely strong spider silk, could be moved by the observer using a screw

mechanism. The changing separations between the target star and nearby comparison stars could be tracked.

Struve's results from seventeen observations starting in 1835 were announced two years later, giving a parallax of one-eighth of a second of arc, close to the present value. But since there had been a long history of fallacious claims to the measurement of parallax, others remained sceptical, and Struve continued his measurements until, in 1840, he gave the results from nearly a hundred observations.

Wilhelm Struve's 24-cm refractor, Tartu

Friedrich Bessel is generally credited as being the first to publish a reliable parallax, spurred on in his measurements by correspondence with Struve and the latter's preliminary result for Vega. From observations made between 1837–38, Bessel tracked the detailed path of the fast-moving binary star known as 61 Cygni, using the heliometer at Königsberg (now Kaliningrad), also manufactured by Fraunhofer.

The heliometer had originally been designed to measure the Sun's angular diameter, and hence the name. Its sixteen centimeter diameter refractor lens had been sliced in half, each segment mounted side-by-side, so forming a pair of images which could be adjusted laterally by turning a thumbscrew. Bessel used it to follow the slowly changing angles between his chosen target and a comparison star close by on the sky. Careful monitoring over the course of a year would show a varying separation if the accuracies were sufficient to discern the parallax wobble of the nearby binary.

In Alan Hirshfeld's 'Parallax' (2001), a readable account of this protracted race he describes Bessel's precision instrument as *'almost painfully beautiful: a copper-shaded, mahogany-veneer tube; burnished knobs, gears, and wheels; and a wooden equatorial mount that descended to Earth through a complex of gracefully splayed struts and stout beams.'* To guarantee stability *'the central part of the [telescope] tower's base was filled with five feet of masonry. Atop this were slabs of sandstone and a layer of timbers. Bolted to the timbers were a series of iron-reinforced beams that rose to the upper reaches of the tower and supported the platform on which the heliometer rested.'*

It was an excellent piece of engineering, and with it pointed to the heavens the first star parallax was measured: in 1838, Bessel announced that 61 Cygni had a parallax of 0.314 seconds of arc, placing it at a distance of three parsecs, or ten light-years. What convinced others that a star distance had been measured for the first time was the match between theory and the expected pattern of separations as the Earth moved in its annual orbit around the Sun.

Hot on Bessel's heels was the work of Thomas Henderson, first Astronomer Royal for Scotland, who published a parallax for the nearby star Alpha Centauri in 1839, derived from observations made even earlier in 1832–33 at the Cape of Good Hope. Although the star is particularly close to the Sun, and its parallax angle therefore amongst the very largest of all stars in the sky, it is only observable from southern latitudes.

With the exception of occasional southern expeditions, such as Halley's and Maskelyne's to Saint Helena, and Abbé Nicolas Louis de Lacaille's catalogue of more than 10 000 stars observed from the Cape of Good Hope in the 1750s, the southern skies had received but scant attention. The situation was addressed by England's Board of Longitude which set up a dedicated observatory at the Cape under its first director, the Reverend Fearon Fallows, whom Henderson replaced in 1832.

Henderson returned to England barely a year later, dissatisfied with working conditions at the Cape. But included amongst his observations, made with an ordinary mural circle and yet to be analysed, were a series of careful measurements of Alpha Centauri. The star was bright, with a large proper motion, and also one component of a binary with a large separation. It thereby handsomely fulfilled all three of Wilhelm Struve's criteria of likely proximity.

The announcements of Bessel and Struve, and the star's probable proximity, prompted him to re-examine his own observations from which he duly determined its parallax. Still today, the binary pair of Alpha Centauri, and their fainter companion Proxima Centauri, remain the nearest known stars to our Sun. Pin-pointed from the Hipparcos space measurements, Alpha Centauri has a parallax of 0.742 seconds of arc, which corresponds to a distance of 1.35 parsecs, or 4.396 light-years—just over forty million million kilometers.

WHAT HAD AT LAST come together was the understanding that distances could be measured using the Earth's motion around the Sun. Those most promising to measure on account of their likely proximity could be pin-pointed. Improvements in telescope size, quality, and accuracy, inspired and drove the relentless pursuit.

These first parallax measurements provided the very first rigorous determination of the distances to the stars. The confirmation that they lay at very great, yet not infinite, distances represented a turning point in the understanding of the Universe. The moment when distances to the stars, and the enormous scale of space, were suddenly and unambiguously revealed must rank as one of the most pivotal in the entire history of science.

John Herschel, President of the Royal Astronomical Society at the time, congratulated members of the society that they had [quoted by Hoskin 1997] *'lived to see the day when the sounding line in the universe of stars had at last touched bottom.'* In awarding the society's gold medal to Friedrich Bessel in 1841, he described it as *'the greatest and most glorious triumph which practical astronomy has ever witnessed.'*

The refractor used by Struve to measure the parallax of Vega still resides in the museum of the Old Observatory in Tartu, Estonia. Bessel's heliometer, along with the observatory and city of Königsberg, was destroyed in the war-time ravages of 1944–45.

Developments 1850–1980

OVER THE PERIOD of three hundred years leading up to the detection of the first parallax in 1838, the measurement of star positions had actually followed two somewhat separate branches. The first of these concentrated on the measurement of parallax, exemplified by the pioneering works of Bradley and Bessel.

In parallel were the much larger sky surveys, like those of Flamsteed in the early 1700s at Greenwich, and Lalande in the early 1800s in Paris. For these, the very highest accuracy of individual measurements was sacrificed, and parallaxes were not part of the design. The goal was rather the charting of large numbers of star positions and motions, the motivation being a better understanding of their distribution and their motions through space.

Over the last hundred and fifty years or so, these two branches of star measurements have really split more convincingly into three: small numbers of stars measured with the highest relative accuracy to fix more parallax distances; others spread over the sky and measured with a very good absolute accuracy to give an overall stellar reference frame; and large surveys aimed at elucidating the structure and properties of our Galaxy from the distribution and motion of the stars.

Parallax measurements: 1850–1990 CE

THE FIRST OF THESE measurement branches focused on a concerted effort to determine more, and more accurate, parallax distances. In the years following the first success of Bessel, initial excitement at the prospect of staking out the space distribution of many more stars was overtaken by the bleak realisation that the majority of bright stars lay at colossal distances that still could not be discerned. Observers had to continue to select target stars fastidiously with the best possible prospects of being nearby, while attention still had to be lavished on a relatively small number of candidates.

The measurements remained delicate and time consuming. The highest instrument qualities, meticulous checks for any possible errors, and multiple observations throughout the year were all mandatory.

Visual observations using heliometers continued to dominate until the dawn of the twentieth century. A copy of Joseph Fraunhofer's Königsberg heliometer was installed in Bonn in 1848, and a still larger instrument delivered to Wilhelm Struve's group at the imperial Russian observatory in Pulkovo. Others were procured by observatories at Oxford, Stuttgart, Leipzig, Göttingen, and Bamberg in Europe, with the largest such instrument ever made, eight and a half inches in aperture and ten feet long, installed at the Kuffner observatory in Vienna in 1896. David Gill began a heliometer programme in the southern hemisphere at the Cape of Good Hope, and the first in America was started by W. Lewis Elkin at Yale in 1885.

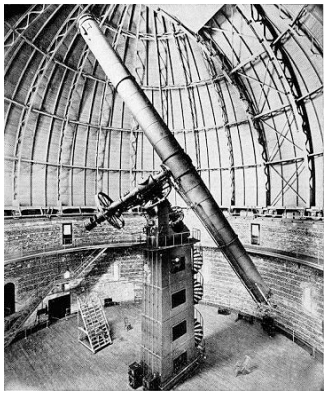
SLOWLY THE NUMBER of star distances grew. But progress remained painfully sluggish, and lengthy discussions of the errors reinforced the continuing very great difficulty of the task. Indeed to some it appeared that the era of star parallax measurements was already effectively over; astronomers again, in the words of Hirshfeld, *'defeated by the sheer immensity of the realm they were attempting to chart'.*

What came to the rescue was the new medium of photography. The earliest commercially viable photographic process, daguerrotype, was used by Harvard astronomers J.A. Whipple and William Cranch Bond to capture the first photographic image of the bright star Vega in July 1850. More efficient photographic processes appeared, and early celestial astrophotography by amateur Warren De la Rue in England was followed by the first photographic parallaxes by Charles Pritchard at Oxford in 1886.

Jacobus Kapteyn in Groningen published a list of just 58 parallaxes in 1901. Meridian circles at Leiden and Heidelberg, and photographic plates from Pulkovo and Cambridge, upped the total to 365 by 1910. Yet Kapteyn remained far from satisfied: *'Up to the present and for obvious reasons, parallax observers have devoted their labours exclusively to the bright and swiftly moving stars. In our opinion the time has come for a change of tactics. We need the average parallax of the faint stars and of those with moderate and small proper motion as sorely as the rest."* His urgent plea was to *'extend the investigations into the arrangement of the stars in space.'*

A NEW ERA in photographic parallax determinations was duly opened up by Frank Schlesinger (1871–1943). Astronomy was developing on many fronts, and knowledge of stellar distances became of pressing importance. Schlesinger was born in New York, and his

PhD at Columbia University had made use of an unusual benefaction: in 1890, the university had received from the pioneering amateur astrophotographer Lewis Morris Rutherfurd more than a thousand photographic plates of the Sun, Moon, planets and stars taken between 1858 and 1877. Acquired with a thirteen-inch refractor, a particular type of telescope which uses lenses rather than mirrors to focus the starlight, Schlesinger's experience with the plates convinced him that with a high quality telescope of considerable focal length, parallaxes could be determined more economically, more conveniently, and more accurately than by any other method.



The Yerkes 40-inch refractor in 1897

So it was that at the Yerkes observatory in Wisconsin in 1903, Schlesinger started a parallax programme using their recently completed forty-inch refractor. This giraffe of a telescope, which remains the largest refractor in the world, was designed around a very long focal length to provide the highest magnification of a small carefully-chosen region of the sky, the easier to discern the tiny parallax wobble. Measurements under his direction started at the observatories of Allegheny in Pennsylvania, where he served as director from 1905 to 1920, and were continued by his successors at the observatories of Yerkes in Chicago, Van Vleck in Connecticut, and McCormick in Virginia. His classic papers appeared in print in 1910 and 1911, detailing the results for just twenty eight stars.

WITHIN THE NEXT decade, such was his influence, and such was the importance of the task, that eight observatories had made parallax determinations a prominent part of their astronomy programmes.

His first task as director of Yale university observatory, a position he held between 1920 and 1941, was to plan a new telescope to further the onslaught. A new twenty six-inch photographic refractor of thirty six feet focal length was designed. This time it was destined for the southern hemisphere, to Johannesburg, where it would carry out for the southern skies a programme similar to that at Allegheny for the north.

Schlesinger went to Johannesburg in 1924 to supervise the observatory construction, and its subsequent dedication by the Prince of Wales. At the time of his death twenty years later, a remarkable fifty thousand plates had been exposed, and shipped back to Yale university in New Haven for measurement. From this mountain of glass, a further sixteen hundred precious star distances were distilled. Moved to Australia in 1952 due to deteriorating sky conditions, the telescope was destroyed by a fierce forest fire in January 2003—a tragic ending for an instrument which had pinned down the distances of so many of the brightest stars.

In 1924 Schlesinger published his *General Catalogue of Stellar Parallaxes*, advancing the total known to just short of two thousand, and extending the frail stellar distance network out to a few tens of light-years. His life's work brought him the gold medal of the Royal Astronomical Society in 1927 and the Bruce medal, another of the highest honours in the field of astronomy, in 1929.

For almost a century thereafter, parallax determinations were led by American astronomers. It was said of his methods that they were *'basic and complete, and that no major improvements are possible.'* Grand praise, and no great surprise therefore that almost all other parallax programmes of the same era would follow his approach.

Outside the United States, Sir Frank Watson Dyson, England's Astronomer Royal from 1910 to 1933, published twelve years of parallax observations from Greenwich in 1925. His successor as Astronomer Royal, Sir Harold Spencer Jones (1890–1960), published a number of parallaxes of southern hemisphere stars from observations acquired at the Royal Observatory established at the Cape of Good Hope.

At a time when many astronomers were moving to newer – and more glamorous – fields of astrophysics, a few still dedicated their careers to astrometric measurements of the very highest calibre.

IN ASTRONOMY it often happens that some individual will take the initiative, and rise to the challenge, of making a compilation of all the different work going on around the world in a particular field. With various observatories contributing more distances, often of different quality, and sometimes duplicating attempts at measuring the same star with different instruments, a critical compilation of parallaxes was needed. Louise Freeland Jenkins at Yale stepped in to fill a much-needed gap.

Jankins brought out a new edition of Schlesinger's *General Catalogue of Trigonometric Stellar Parallaxes* in 1952, with distances for just under six thousand stars based on photographic determinations of the Schlesinger era. A supplement in 1963 raised the total to nearly six and a half thousand.

A further update appeared in 1995. The *Yale Trigonometric Parallax Catalogue* of just over eight thousand stars was pieced together by Yale astronomer William van Altena. It was the catalogue that the world's astronomers consulted near the end of the second millennium if they wanted to know the distance to a star. It was also to be the final collection of ground-based parallaxes before those from the European Space Agency's Hipparcos satellite.

At the time of the push to space, the total number of known star distances was certainly respectable, and had been extremely hard won. But even amongst the nearby stars it was a paltry sampling, let alone amongst the hundred billion stars in our Galaxy as a whole. Crucial and niggling were the plethora of discrepancies and errors arising from the shimmering atmosphere. Accuracies were supposedly around one hundredth of a second of arc, but in reality were often much poorer. This made it difficult to rely on published values, and dangerous to draw wide-reaching scientific conclusions. A new approach to measuring distances was sorely needed.

The stellar reference frame: 1850–1990 CE

A SECOND MEASUREMENT branch was devouring enormous efforts, in parallel with the work on parallax, to set up the best possible stellar reference frame—to measure and list the positions of a number of agreed reference stars ranged across the entire sky.

To determine a chosen star's distance, repeated positions measured with respect to some other star nearby on the sky would hopefully reveal its parallax motion over the course of a year, but the position of the reference star itself was quite irrelevant. A celestial reference frame demanded, in contrast, a network of precise positions of stars over the entire sky—a set of agreed reference beacons, with positions and motions well nailed down. Hipparchus and Ulugh Beg, Tycho, Flamsteed and Bradley had typified the earliest efforts to establish a stellar reference system across the celestial sphere.

BY THE SECOND half of the nineteenth century, a multitude of studies clamoured for a much improved grid of astral trig points. It was needed as a reference frame for the much fainter star surveys starting up to probe the Galaxy's structure, and for studies of the motions of the planets and the rotation of the Earth. These needs turned to meridian circle instruments to give the best positions for a relatively small numbers of stars.

The problem was one that Hipparcos would be well set-up to solve properly later on: that of linking together observations made at different geographic locations and at different times. The reference frame demanded positions of the stars, linked through to the planets, the Moon, and the Sun. A perfidious complication was the fact that the measurement platform, the Earth itself, was slowly 'wobbling' due to effects of precession, nutation, and short-term and unpredictable polar motion.

In Germany, a sequence of whole-sky star catalogues, named the FK series after the German Fundamental Katalog, began with the work of Arthur von Auwers (1838–1915) in the late 1870s and early 1880s. Their work was to dominate the field for over a century, although a parallel American effort started with Simon Newcomb's accurate charting of just over a thousand stars in 1899, and continued with Benjamin Boss's influential General Catalogue of 1937.

Auwers had started out on his own career at Königsberg, using Bessel's original heliometer. He made his own measurements of a small number of parallaxes, and established the orbital motion of the binary companion of Sirius based on many thousands of meridian circle observations taken over six years. There were, however, no nearby suitable comparison stars for Sirius, and his experiences in constructing a reference system based upon earlier observations led to the catalogue construction work which would dominate the rest of his life. Auwers began by returning to the very accurate observations made by James Bradley over the years 1750–62, comparing them with more modern observations to determine star motions. This piece of work alone would occupy him from 1866 for a further ten years.

BY SUCCESSIVE steps, Auwers established a system of just thirty six benchmark stars, with longitudes across the sky fully consistent with each other. Their origin was set by Bradley's observations of the Sun a century before. Into this he folded other observations, of Bradley's own zenith sector measurements acquired from Greenwich, and others by Nevil Maskelyne around the 1760s and by Stephen Groombridge around 1810.

The result was a reference system across the sky of just over three thousand stars. All were reobserved from Greenwich around 1865, to give the most accurate motions of stars to date, pinned down from the grand lever arm of a century and a half of meticulous observation.

These motions would form the basis of many pioneering researches into star movements carried out over many decades, including Simon Newcomb's revision of the Earth's wobbling motion, and Jacobus Kapteyn's investigations into the rotating Galaxy. The resulting catalogue was published by the Saint Petersburg Academy of Sciences in 1888.

In a later collaboration with David Gill in 1889 to refine the distance to the Sun, Auwers provided observational skills much needed by Gill, which the latter acknowledged in his obituary: *'Such cooperation proves, if proof is necessary, that science knows no nationality, and that common pursuit of truth for truth's sake affords one touch of nature which makes the whole world kin.'*

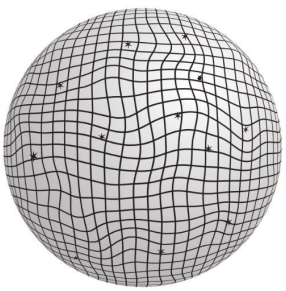
These 'fundamental' catalogues, it should be stressed, charted only a rather small number of reference stars, dictated by the brightness of stars which could be observed by the meridian circles of the day. They gave only one star every six degrees or so on the sky, or just a handful across the whole of Europe if thought of as a mapping of Earth. Successive catalogues added more observations, and slowly yielded a better grid, although rejecting inferior observations also whittled down the number of quality reference stars.

More could be interpolated from meridian circle or photographic observations, but inherent distortions would ultimately rest on the quality of the primary grid.

The final catalogues in the series were prepared at the Astronomisches Rechen-Institut in Heidelberg: the FK4 led by August Kopff and Walter Fricke was published in 1963, and the FK5 after a quarter of a century devoted to its upgrade, led by Walter Fricke and published in 1988. The work required to create these catalogues extended over many years of careful observation and critical analysis. The FK5 catalogue was the culmination of a compilation of about 260 individual catalogues, observed mostly with meridian circles and some astrolabes. Like the FK4 it contained just 1535 stars.

But the scientific importance of these catalogues was nevertheless substantial: they alone provided the reference grid into which the positions of very much fainter star images, captured on photographic plates in their hundreds of thousands during the early 1990s, and in their tens of millions in the later years of the 20th century, could be interpolated.

WHILE INDIVIDUAL star positions in these reference catalogues reached accuracies of several hundredths of a second of arc, and notwithstanding the massive effort and observations invested, evidence still suggested that there were significant hidden errors depending on their sky position. Years before, Kapteyn said in 1922: *'I know of no more depressing thing in the whole domain of astronomy, than to pass from the consideration of the accidental errors of our star places to that of their systematic errors. Whereas many of our meridian instruments are so perfect that by a single observation they determine the coordinates of an equatorial star with a probable error not exceeding two or three tenths of a second of arc, the best result to be obtained from a thousand observations at all of our best observatories together may have a real error of half a second of arc and more.'*



Star positions in a warped reference system

Like ancient maps of Earth, the star charts were topologically correct, but stretched and squeezed over the sky in ways that could be guessed but not fully fathomed, hidden errors which proved impossible to track down and remove. They were only fully apparent once the Hipparcos space results were published.

Meridian circles remained the instrument of choice for the highest accuracy surveys until the late twentieth century. Most were phased out after the Hipparcos catalogue was published in 1997, but the automatic 18 cm aperture Carlsberg meridian telescope is one of a few exceptions. It was moved to La Palma in 1984, refurbished with a CCD detector in 1998, and continues to operate remotely, turning out more than a hundred thousand star observations each night, their positions locked into the Hipparcos grid.

The FK5 catalogue of 1535 stars was the final word on the celestial reference frame before the launch of Hipparcos. It was the state-of-the-art in star charting until the space-based positions appeared. But it was far too sparse, and inaccurate, to satisfy modern needs. Like the parallax catalogues, it was impossible to rely on published positions.

This was not a good situation for a field so basic. As for the star distances, a new approach to the measurement of a celestial reference frame was required.

Large-scale surveys: 1850–1990 CE

THE THIRD MEASUREMENT branch of relevance to astrometry over the last century is represented by the large-scale photographic surveys. This branch traces its roots to the early 1600s, when Galileo used the newly-invented telescope to observe the Milky Way, and found that it could be resolved into innumerable faint stars. By the mid-eighteenth century astronomer Thomas Wright (1711–1786) had described the Milky Way as a flattened disk of stars in which the Sun is itself confined, *'an optical effect due to our immersion in what locally approximates to a flat layer of stars.'*

Philosopher Immanuel Kant (1724–1804) developed these ideas, and also postulated the existence of other 'island universes' distributed throughout space at enormous distances. William Herschel counted the number of stars in different sky regions to deduce the relative dimensions of our Galaxy. These gave valuable insights, but with conclusions founded on the crucial but incorrect assumption that all stars had the same absolute brightness. It was nevertheless becoming clear that large stellar surveys could have much to say about our Galaxy's basic properties such as its size and its shape.

Many large and enormously influential surveys have been made over the last 150 years. Important amongst the earliest were the huge three-part 'Durchmusterung', named for the German for survey, a word capturing the grandeur of the enterprise. The first two parts were the last of the great star maps to be made visually, pre-dating the use of photography—assistants recorded the positions and magnitudes of stars as the Earth spun and the sky drifted across the fixed telescope field surveying successive latitude zones.

The series started with the northern sky surveyed from Bonn by Friedrich Argelander and Eduard Schönfeld. Published between 1852 and 1859, this gave the positions of more than 324 000 stars of the northern the sky.

The extension southwards was surveyed from Córdoba in Argentina by John Thome starting in 1892.

The new medium of photography had burst onto the astronomical scene in the late 1800s. Hand-in-hand with the meridian circles giving the highest accuracy reference grid for the brightest stars, photography was to dominate surveys of the skies for the next century. The switch to photography also represented a change in methodology: until then, position measurements had been made by eye, then transcribed to make a star chart. With photography, a chart of the sky was captured directly, and the positions of the stars deduced from them.

AMONGST THE EARLIEST of these was the southward extension of the Bonn and Córdoba Durchmusterung, covering the southernmost skies from the Cape of Good Hope. The results of the work, led by Sir David Gill (1843–1914) and influential Dutch astronomer Jacobus Kapteyn, were published around the turn of the century. Positions were around one second of arc, limited by the twin barriers of atmospheric turbulence and photographic plate quality. The vast Durchmusterungen triptych was only eventually transcribed to computer form in a 15-year effort in the 1980s.

These first truly large-scale surveys provided the foundations on which many later investigations would build their own views of the changing positions of the stars. Thereafter new and deeper surveys from many different observatories around the world contributed to the growing edifice.

PHOTOGRAPHY ALLOWED the positions of stars to be measured wholesale. With the large telescopes and long exposures of the later 1900s, deep sky images several degrees in extent could yield thousands or millions of star images per plate. The technique was straightforward in principle: exposed at the focus of a telescope tracking the apparent motion of the celestial sky, the plates provided images of stars in huge numbers.

Positions on the plates could then be measured and recorded, duly transformed to provide immense catalogues of star positions. In practice, good images require excellent high-altitude observing sites, excellent telescope optics, and accurate and smooth drive mechanisms to track the rotating sky. But they could never eliminate the straightjacket imposed by the atmosphere.

Photographic plates store well for decades, and astronomical libraries and archives across the world preserve a record of how the skies appeared over the past century. As the technology reached its peak in the 1970s and 1980s elaborate and fast automatic measuring machines scanned new and ancient archive plates wholesale. Together they have captured and stored the results in the form of huge digital catalogues of the night sky which will be preserved indefinitely.

Yet fundamental distortions due to the telescope optics have always confounded the ultimate accuracies, while the Earth's atmosphere, and the ever-so-slightly dancing images seen through it, still imposes its ever impenetrable barrier.

Although the highest positional accuracies were therefore sacrificed in favour of quantity of stars, massive sky surveys using photographic plates nevertheless changed the course of astronomy. There were various reasons for this impact. First off, simply counting stars to different brightness limits in different directions of the Galaxy has provided many clues as to its structure and dimensions. The technique is especially powerful when interpreted alongside other knowledge, such as the type or temperature of the stars from spectroscopy.

Measuring the same region of sky over many years or decades is particularly effective at revealing the motions of many stars. Photographic surveys, carefully calibrated and repeated decades later, have turned the early detections of star motions by Halley and others into a large-scale discovery factory on an industrial scale.

Repeating exposures of the stars over intervals of months or years has another important spin-off: it has led to the discovery of huge numbers of variable stars, their variability over time encoding clues as to their masses, luminosities, and evolutionary states. Star colours measured from different filters and photographic emulsions also provide a wealth of indicators such as their temperature and gravity.

STAR POSITIONS in large numbers allowed astronomers to embark on a new, more quantitative discussion of our Galaxy's structure. In 1904, studying the Cape Photographic Durchmusterung, which he had worked on in collaboration with David Gill, Kapteyn found that the motions of stars were not random, but could be divided into two streams, moving in nearly opposite directions in different parts of the sky—the first hint of the rotation of our Galaxy.

In a summary of his life's work published in 1922, Kapteyn described the Galaxy as a lens-shaped island universe in which the density of stars decreased away from its centre. His Galaxy was some 40 000 light-years in size, not so far from present ideas. But, as if clutching at long-held belief that the Earth must occupy some privileged place in the Universe, Kaptyen held that the Sun was close to its centre, at around 2000 light-years.

The size of the Galaxy, and the distance scale within it, became issues of great debate. It was not easy to infer the structure of the Galaxy from star counts alone, and there were many complications. Great clouds of dense interstellar gas occupy various pockets within our Galaxy's disk, and these block out the more distant light from stars beyond. It's not so different to looking at the night sky covered by thin cloud.

With no simple means to identify the gas, seeing only a few stars along a particular sight line might suggest that the Galaxy was only thinly populated by stars in that direction, while the very opposite might be true. Another tricky problem was caused by the growing realisation that stars were of many different types, with hugely varying luminosities and very different types of motion through space. So evident in retrospect, trying to figure out the properties of our Galaxy from an erroneous census was doomed to fail.

So it was that even into the 1920s, the detailed structure of our Galaxy, and the relationship between it and those that we now know lie far beyond, remained a puzzle. The uncertainties precipitated an exchange which has gone down as astronomy's Great Debate, which took place on 26 April 1920 in the Smithsonian Museum of Natural History in Washington DC.

Harlow Shapley, of the Mount Wilson Observatory, argued that our Sun lay far from the centre of a single Great Galaxy, in which spiral nebulae such as Andromeda were simply part of our own. Heber Curtis, of the Allegheny Observatory, disagreed. He held that the Sun was near the centre of a relatively small Galaxy, with the entire Universe composed of many other galaxies somewhat like our own. It was a debate deeply rooted in the uncertainty of the scale of the Universe which had still not been resolved.

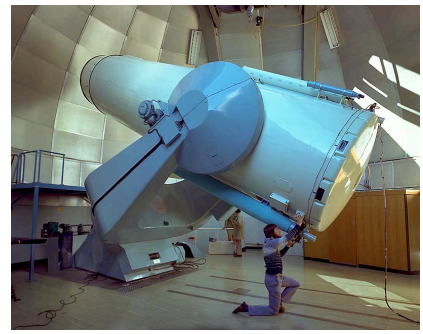
EDWIN HUBBLE'S identification of pulsating Cepheid variables in the Andromeda nebula in the mid-1920s confirmed that it was a distant galaxy much like our own, but far beyond. Like brilliant lighthouses pulsing across the depths of space, these standard candles illuminated our understanding of the scale on which the Universe is constructed. Shapley was proven more correct about the size of our Galaxy and the Sun's location in it. But Curtis's view that the Universe was composed of many more galaxies, and that 'spiral nebulae' were galaxies just like our own, was corroborated.

With almost a century's hindsight, the debate is important, in the words of Frank Shu (1982): *'not only as a historical document, but also as a glimpse into the reasoning processes of eminent scientists engaged in a great controversy for which the evidence on both sides is fragmentary and partly faulty.'*

SCHMIDT TELESCOPES appeared on the scene in the second half of the twentieth century, and brought their own revolution. Named after their optical designer Bernhard Schmidt, a cleverly-designed 'corrector' lens positioned in front of the primary reflecting mirror resulted in strongly reduced image aberrations over unprecedentedly large fields of view of several degrees on a side (specifically, the design allows very fast focal ratios, while controlling coma and astigmatism).

This made it possible to observe a substantially larger region of the sky, several times the diameter of the full Moon, in a single exposure. As a result, Schmidt telescopes contributed a flood of high-quality observations that brought positional astronomy back to the fore.

Monumental surveys were carried out from Palomar Mountain in California from 1949, in a grand programme funded by a grant from the National Geographic Society to the California Institute of Technology. The southern skies were surveyed from the European Southern Observatory's La Silla observatory in Chile from 1973, and from the UK's observatory in Australia about the same time.



AAO/David Malin

UK 1.2-m Schmidt telescope, AAO, Australia

THE SURVEYS produced thousands of meticulously exposed plates which were themselves reproduced photographically, and circulated in limited editions to the world's astronomical institutes for detailed scrutiny. Collectively, they comprise hundreds of billions of star images, an archival view of the celestial sky as it will never be seen again. The resulting vast catalogues, of more than a billion stars across the sky, are used for countless astronomical projects, including pointing their way around the sky by the great space observatories, the Hubble Space Telescope amongst them.

Photographic plate surveys made far in the past—a century or more ago—remain of value to present day astronomy, for a repeat survey today will easily identify the most rapid movers with the largest motions. Catalogues of stellar motions continue to be constructed from various combinations of these photographic plates, using the same technique which allowed Edmond Halley to identify the first stellar motions three hundred years ago.

The Carte du Ciel: 1850–1950 CE

IN THIS CONTEXT, one remarkable project deserves specific mention: the imposingly named *Carte du Ciel*, the Map of the Heavens. It is noteworthy not so much for its profound scientific achievements, but rather for its hugely ambitious scale. This vast and unprecedented international star-mapping project was initiated by ex-naval officer and Paris Observatory director Rear Admiral Amédée Mouchez, in collaboration with Sir David Gill, Her Majesty's Astronomer at the Cape of Good Hope at the time.

Mouchez had started his career with hydrographic studies of the ocean depths, tides and currents along the coasts of Korea, China and South America and later, during the Franco–Prussian War, led a heroic defence of Le Havre. Taking the helm at the Paris Observatory, correspondence between Mouchez and Gill led to the *'assembling of a great international conference'*, the Astrographic Congress of more than fifty astronomers held in Paris, on 16 April 1887. Participants included Auwers from Germany, Kapteyn from The Netherlands, Struve from Russia, and William Christie, the Astronomer Royal from England.

The new medium of astronomical photography offered a remarkable possibility to carry out a celestial survey totally unprecedented in the history of astronomy, and astronomers seized the opportunity. The objectives of this first ever international astronomical collaboration on a massive scale were hugely ambitious but would prove to be overwhelming. The idea was to build up and deploy a system of identical telescopes straddling the full range of latitudes on Earth, survey the sky, and build up a monumental star catalogue as a result.

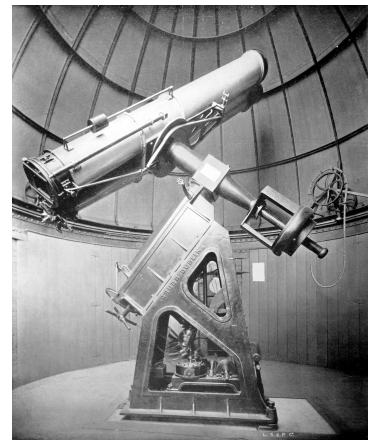
According to H. H. Turner's highly-readable description of the project from 1912: *'The discussions were, to say the least of it, animated. There are no universal rules for conducting public business, and astronomers from one country were not familiar with rules in use elsewhere. It interested Englishmen, for instance, who are accustomed to have resolutions moved by anyone rather than the chairman, to learn that this was by no means a universal rule. On the contrary the chairman of the first conference considered it part of his duties to move all the resolutions. After listening to a discussion, he took it to be his function to summarise the sense of the meeting in a resolution which he put from the chair and in favour of which he held up his own hand. Unfortunately for his success his was sometimes the only hand held up, and the discussion was necessarily resumed.'* Turner considered that the conference was: *'. . . a remarkable meeting, the first of its kind in the history of astronomy; and it has shown the way for subsequent gatherings. . . On all of these occasions the French have acted as hosts and have discharged these duties with a cordiality and hospitality that has never failed to impress their colleagues from the most distant parts of the world.'*

THE AMBITIOUS enterprise had two separate yet connected parts. The first, the Astrographic Catalogue, would photograph the entire sky to 11 magnitude, thereby picking out stars a hundred times fainter than the feeblest seen by the unaided eye. It would provide a plentiful reference catalogue much denser than anything observed by transit instruments.

Twenty observatories around the world participated, each choosing a strip of sky convenient in latitude. Each would procure the necessary astrograph (a telescope designed specifically for the purpose of astrophotography), suitably equipped and staffed. Then collectively they would expose, for six minutes each, more than twenty thousand glass plates of the night sky. Turner estimated the total weight of these plates at three tons.

A key agreement, and one essential to the survey uniformity, was to use similar telescopes. Around half of the observatories eventually procured astrographs from the Henry brothers in France, with the others coming from the firm of Howard Grubb in Dublin. The different observatories were assigned different latitude strips to photograph: Greenwich, the Vatican, Catania, Helsing, Potsdam and Hyderabad would cover the northern sky. Uccle, Oxford, Paris, Bordeaux, Toulouse, Algiers, San Fernando and Tacuba would span the equatorial regions. Córdoba, Perth, Cape of Good Hope, Sydney, and Melbourne would survey the southern skies.

The first plate was taken in August 1891 at the Vatican Observatory. The exposures there, taken by the hands of a single observer, took more than twenty seven years to complete. The very last plate was finally exposed in December 1950 at the Uccle Observatory in Bruxelles.



Graham Dolan

The Greenwich astrograph, c1900

The plates were in due course photographed, measured, and the results published in their entirety, providing star positions with an accuracy of about half a second of arc. In practice, the measurements were a highly protracted affair, with the tasks around the world assigned to willing—and in some cases unwilling—assistants.

ADRIAAN BLAAUW recalls that Pieter van Rhijn (1886–1960), Kapteyn's successor as director of the Astronomical Institute in Groningen and who Blaauw himself knew well, had told him that Kapteyn had numerical computations of star coordinates carried out by prisoners in Groningen. According to Blaauw: *'A number of these tables still exist and are now part of the Kapteyn legacy collection kept in the Groningen University Library where they can be consulted. They are a marvel of neatness and accuracy. The people who made them must have taken great pride in delivering them and one can imagine that it must have given them great satisfaction to contribute in this way to Kapteyn's scientific work.'*

Doubts were raised about the role of prisoners at the Kapteyn Legacy Symposium in 2000, there being no written documentation, but Blaauw, who I got to know well in the 1980s during his role in the preparation of the Hipparcos observing programme, vouched for the story's pedigree.


Torino Observatory

Nuns measuring the Vatican plates, c1900

Measurements of the star images were made by eye, and recorded by hand. In several observatories (Paris, Melbourne, Perth, Cape, Toulouse and others) twenty or thirty women (the original 'computers') assisted with the herculean task. For the Vatican plate collection, archival photographs from Torino Observatory show nuns from the Congregation of the Child Mary at work measuring the plates.

Turner commented that *'each observatory has thus to measure about half a million star images… These measures took a staff of four or five people at Oxford some ten years or so to complete: and the printing of them another four years.'* In total, nearly five million stars were recorded. Publication of the various parts proceeded from 1902 to 1964, and resulted in a massive two hundred and fifty four printed volumes.

FOR THE SECOND part of the conference goals of 1887, a further set of plates, with longer exposures but minimal overlap, would photograph all stars to 14 magnitude, corresponding to stars a thousand times fainter than those that can be seen with the naked eye. Most of these plates used three exposures of twenty minutes each, displaced to form a small triangle with sides of ten seconds of arc, making it easier to distinguish stars from plate flaws, and to differentiate stars from the more rapidly-moving asteroids.

The grand idea was that exposed plates would be reproduced and distributed as a set of charts, the *Carte du Ciel*. However, reproduction of the charts, originally to be undertaken using engraved copper plates, proved to be prohibitively expensive, and many zones were either not completed or not properly published.

Despite, or perhaps because of, its vast scale, the project was only ever partially successful, even though many committed individuals had devoted decades of their careers to its success. The *Carte du Ciel* component was never completed, and the Astrographic Catalogue lay largely ignored for nearly a century. Its star positions were difficult to work with because they were not available in computerised form, and neither were they listed in convenient coordinates.

SOME HISTORIANS of science have classified this vast project as the story of how the best European observatories of the nineteenth century lost their leadership in astronomy by committing vast resources to a somewhat misguided undertaking.

Long portrayed as an object lesson in over-ambition, languishing lost and forgotten for a century, the Astrographic Catalogue made a remarkable reappearance on the world's astronomical stage at the turn of this century. The Hipparcos catalogue positions could be used, in combination with each star's proper motion, to provide a reference frame back at the time when the Astrographic Catalogue plates were taken. So calibrated, they gave the places of all catalogue stars which they occupied in the sky some one hundred years before.

Combining those with the satellite positions nearly a century later gave extremely accurate motions for two and a half million stars: the Hipparcos satellite-based Tycho 2 Catalogue, led by my long-time colleague, the leading Danish astrometrist Erik Høg.

Like the ancient catalogue of Hipparchus dusted off and used to reveal star motions by Halley, the Astrographic Catalogue is a remarkable example of an all-but-abandoned project, for whom so many had toiled for so long, waiting patiently to prove its inestimable value generations afterwards.

Other Photographic sky surveys: 1900–1980 CE

NUMEROUS OTHER large-scale photographic astrometric sky surveys were carried out in the twentieth century. The following chronology of some of the major developments in twentieth century astrometric surveys is intended only to set the context.

AGK2: between 1928 and 1931, the sky north of declination $-5°$ was photographed on 1940 glass plates each covering over $5° \times 5°$ with two dedicated astrographs located in Bonn and Hamburg, Germany. Two exposures, one of 3 minutes and one of 10 minutes, were made on each plate, and reached about 12 mag. During the 1930s–1950s the measuring and reduction of the brighter stars were carried out, by hand, resulting in the AGK2 Catalogue.

AGK3R and AGK3: after a proposal that the AGK2 Catalogue should be observed again at Hamburg to provide proper motions, an extensive international programme of meridian observations at ten observatories was organised, under IAU Commission 8, to provide a reference star catalogue, AGK3R, which was then used for the reduction of the photographic work carried out at Hamburg between 1956–63. This resulted in the AGK3 Catalogue, containing proper motions for all stars, which was subsequently used as the stellar reference frame in the northern hemisphere.

SAO: by the mid-1960s a high density catalogue of star positions was needed for satellite tracking. This was compiled by the Smithsonian Astrophysical Observatory for more than 250 000 stars. In each declination zone, preference was given to source catalogues with proper motions, namely the Yale Photographic Catalogues in the north, and the Cape Catalogues in the south. The resulting SAO Catalogue was limited by the generally poor quality of the first epoch material in both hemispheres (the AGK3 not yet being available in the north). Not surprisingly, in view of the inhomogeneous source material used in the construction of the SAO, the differences with the later Hipparcos results show various large distortion patterns.

SRS: the success of the AGK3R programme led to plans for a similar campaign in the southern hemisphere, formulated by the International Astronomical Union in 1961. The resulting Southern Reference Star (SRS) Catalogue was constructed from observations made with 13 transit circles, with observations extending from 1961 for about two decades. The International Reference Stars (IRS Catalogue) comprises the combination of the resulting reference stars observed from both hemispheres, i.e. the AGK3R in the north, and the SRS in the south.

CPC2: to complement the AGK3 in the northern hemisphere, the Second Cape Photographic Catalogue, CPC2, was constructed from 5820 southern hemisphere plates taken with a new astrograph at the Cape Observatory during 1962–1972 (mean epoch 1968), and scanned with the GALAXY plate measuring machine at the Royal Greenwich Observatory, Herstmonceaux. This resulted in a catalogue of 276 131 stars in the range 6.5 – 10.5 mag.

MANY OF THESE (and other) grand twentieth century photographic surveys have been revitalised by the results of the Hipparcos satellite mission. The new reference system from space can be propagated backwards in time using the measured proper motions, to give an improved reference system for the years that the plates were taken. The improved reference system then gave much better positions for the large numbers of other stars on the plates. This, in turn, has led to vastly improved star motions tracked between the times of the earliest photographic plates a century ago, and the measurements from space made in the last decade of the second millennium.

Solar system measurements: 1800–1990 CE

A FINAL MIX OF curious phenomena showed up in the measurement of the accurate positions of the stars and the planets over the last couple of centuries, bringing us back, in full circle, to the earliest of the Greek studies of the fixed stars and the wandering planets.

Objects in our daily lives are generally not massive enough, or the effects not measurable accurately enough, for Newton's Law to be examined for real flaws or imperfections. But the motions of the planets provide a miraculous laboratory for observing the most delicate touches of gravity. Alongside innumerable other successes of Newtonian gravity was its part in the discovery of the planet Neptune.

In the middle of the nineteenth century French mathematician Urbain Le Verrier (1811–1877), working under François Arago at the Paris Observatory, had been making a careful study of the orbit of Uranus. There were small but systematic discrepancies between its observed orbit, and that predicted by Newtonian theory—its measured position was consistently off from where theory forecast it should be. Something was wrong.

Newtonian gravity had proven itself repeatedly and was not the suspect. Le Verrier was forced to conclude that an undiscovered planet existed out in the far reaches of the solar system, giving erratic tugs at Uranus during its journey around the Sun. He could predict a position for an unknown object which, he believed, must be responsible for disturbing its orbit. Neptune, as it would be called, was duly discovered by Johann Galle and Heinrich d'Arrest, within one degree of his predicted location, on 23 September 1846.

It was a triumph for Newtonian gravity, and a sensational result for Le Verrier, who became director of the Paris Observatory in 1854, following in the footsteps of Cassini and Lalande. A source of debate ever since has been the extent to which John Couch Adams, who had made similar calculations even earlier, should also be credited with Neptune's discovery.

The earliest and most worrying sign that all was not completely well with Newtonian theory was the detailed motion of our innermost planet. Mercury circles the Sun in a tight, bakingly-hot elliptical orbit of just ninety days. Its point of closest approach advances around the Sun by a small amount each year, about one minute of arc, due to various effects, including the gravitational pull of the other planets.

Le Verrier noticed that the slowly changing shift could not be fully explained by Newton's laws. There was a tiny mismatch of a little less than half a second of arc per year, an almost undetectable amount, except for the fact that it rolls up and accumulates with time, to nearly forty three seconds of arc each century. In 1843, inspired by his success with Neptune, Le Verrier published his interpretation of the mismatch as being due to a hypothetical inner planet, which he named Vulcan.

This precipitated a search for the new planet, and a wave of false detections that would follow unabated over the next sixty years. One Edmond Lescarbault was even awarded France's prestigious *Légion d'honneur* for his claimed sighting of the non-existent body.

In 1915, while the searches were in full swing, Albert Einstein published his general theory of relativity. This describes gravity as a basic property of the geometry of space and time, a distortion in their very fabric due to the presence of mass. It superseded Newton's law of universal gravitation as 'the' theory of gravity. Mathematicians admire its elegance, and physicists like it because it gives hints as to why this force exists. Mostly the predictions of Newton and Einstein agree. But in certain situations they differ, slightly but significantly, and tests to confirm or repudiate it were eagerly sought.

The orbit of Mercury was an obvious target. It was Einstein himself who showed that his theory explained exactly the discrepancy, important evidence that he had identified the correct form of the equations describing gravity. The effect, referred to as perihelion precession, has also been seen for Venus and Earth. In a very close binary pulsar system, discovered in 1974, the effect is a hundred thousand times larger. In all cases, theory and observation are in precise accord.

Le Verrier died in 1877 still convinced that he had detected a second planet. Yet while most of the interest in Vulcan evaporated, claims and counter-claims of asteroid transits, and searches for Vulcanoid asteroids orbiting close to the Sun, continue to the present.

Another test proved to be still more compelling. According to the prescriptions of general relativity, starlight should be deflected by a very tiny but entirely predictable amount as it passes from a distant star close to the limb of the massive Sun on its way to an observer on Earth. The size of the deflection was predicted to be very small, just over one second of arc at the limb of the Sun where the effect would be largest. Barely at the limit of the dancing motion of the atmospheric ripples, it would demand careful measure, and an excellent knowledge of the undeflected star image positions to compare with.

It would be impossible to measure shifts of faint stars close to the limb of the brightest object in the entire sky except, perhaps, if they could exploit the exceptional conditions of a total solar eclipse. This was American solar astronomer George Ellery Hale's proposal to Einstein when asked to suggest an appropriate test. A German–USA expedition planned for an eclipse passing over Crimea in 1914 was foiled by the outbreak of war.

The first observations of this light bending were eventually made during the total eclipse of 29 May 1919. Astronomer Royal Sir Frank Watson Dyson had identified this as an auspicious celestial alignment because the Sun and Moon would pass in front of the bright Hyades cluster, more bright stars making it easier to detect changes in their position. The undeflected star positions that would later be observable close to the Sun's limb during the eclipse had been observed six months previously by night.

Arthur Eddington and Edwin Cottingham from Cambridge journeyed to the West African island of Príncipe in the Gulf of Guinea, while Andrew Crommelin and Charles Davidson from the Royal Greenwich Observatory set up their base near the Brazilian town of Sobral—the two observing stations chosen to improve prospects of observing the eclipse in case of poor weather.

During the eclipse, as the sky was plunged into darkness, a few bright stars popped into view and remained visible for two or three minutes. This time, their positions would be minutely deflected by the presence of the Sun's huge gravitating mass along the light path from the distant stars behind the Sun to observers on Earth.

The agreement between the small extra shifts observed on the one hand, and Einstein's theory on the other, was very much at the limit of star measurement accuracies of the time.

Confirmation of the predicted bending was duly claimed, and widely greeted as spectacular news. It made the front page of major newspapers, making the theory of general relativity world famous, and Einstein himself even more so. When asked what he would have said had his theory not been proven by the observation, Einstein notoriously replied *'I would have had to pity our dear Lord. The theory is correct all the same.'*

Debate about the quality of these early observations has continued, in the sense of how convincingly they confirmed the predictions of general relativity but, nonetheless, the theory itself is now unquestioned.

Better measurements for other solar eclipses, including one in June 1973 by Texan astronomers from a desert site near Chinguetti in Mauritania, sightings of quasars at radio frequencies, gravitational lenses observed in astronomy in the 1980s, gravitational redshift as perfectly accounted for by GPS navigation satellites, the first direct detection of gravitational waves generated by the merger of two black holes in 2015, and many other more subtle manifestations, have confirmed general relativity as our best description of gravity to date.

The 1980s, and solid-state detectors

In the last 20–30 years, photographic plates have all but disappeared from astronomy, going the way of sextants and quadrants and most meridian circles before them. In their place the CCD, the ultra-sensitive solid-state silicon detectors, of the type used in digital cameras (and comparable infrared-sensitive detectors), has taken over the challenge, and has brought with it another revolution in surveying the skies.

The full-sky surveys of the US Naval Observatory, notably USNO B and UCAC2, and the Sloan Digital

Sky Survey supported by the Alfred P. Sloan Foundation (a philanthropic structure set up by the one-time President of General Motors), have led this new wave, leading to deeper exposures, and more stars, than ever before. Other comparable surveys have also been carried out in the near infrared, notably the 2MASS infrared sky survey led by the University of Massachusetts.

Other very-large scale CCD or infrared sky surveys have recently come on line, notably VST (the ESO VLT Survey Telescope), VISTA (the ESO Visible and Infrared Survey Telescope for Astronomy), and Pan-STARRS (the Panoramic Survey Telescope and Rapid Response System), while yet grander projects (notably LSST, the Large Synoptic Survey Telescope) will soon be operational. They are located at premier high-altitude sites such as in the Atacama desert or perched in the mountain top observatories of Hawaii.

The emphasis has evolved somewhat, to surveying the sky as quickly as possible in as many colour filters as technically feasible. They fall almost exclusively into the category of large-scale surveys (rather than parallax or reference-frame surveys).

STATE-OF-THE-ART astrometric accuracy is not their primary objective, and all have based their overall reference frame on the positional network provided by the Hipparcos Catalogue derived from the first astrometric survey from space. New challenges come as these unprecedented surveys scan the night skies, over and over, with a speed and sensitivity inconceivable only a couple of decades before.

Nearby Stars

THE DEFINITION of the nearby stellar population figures in many areas of astronomical research, ranging from studies of star formation to the statistical occurrence of extra-solar planets. It remains, however, a difficult task to establish a complete census of stars within the immediate solar neighbourhood, even out to distances of only 10–20 parsec.

One of the first attempts to compile a census of stars in the solar neighbourhood, largely based on trigonometric parallaxes, was Woolley's *'Catalogue of Stars within Twenty-Five Parsecs of the Sun'*, while a growing compilation has been maintained by the Astronomisches Rechen-Institut in Heidelberg over the last 50 years. The 1957 *'Katalog der Sterne näher als 20 Parsek für 1950.0'* contained 915 single stars and systems within 20 parsec. The 1969 *'Catalogue of Nearby Stars'*, or CNS2, had a slightly enlarged distance limit of 22.5 parsec.

CNS3 extended the census to some 1700 stars nearer than 25 parsec, while the as-yet-unpublished CNS4 incorporates data from the Hipparcos catalogue, and provides a major development in the comprehensive inventory of the solar neighbourhood up to a distance of 25 parsec from the Sun.

Other compilations include Northern Arizona University 'NStars Database', dating from 1998, which maintains a compilation of all stellar systems within 25 parsec, while Georgia State University's 'Research Consortium on Nearby Stars' (RECONS) aims to discover and characterise 'missing' stars within 10 parsec, using astrometry, photometry, and spectroscopy.

While the earliest ground-based parallax surveys were very successful in identifying nearby very bright stars, problems still persist for stars of very low intrinsic luminosity, where a complete parallax survey even out to only 10 pc remains impossible. The advent of accurate all-sky multi-colour surveys has facilitated the direct search for nearby, low-luminosity stars.

As Wilhelm Struve had originally suggested almost two centuries ago, surveys searching for high-proper motion stars have long been used to detect nearby candidate stars which were then added to parallax programmes, including the Hipparcos Input Catalogue in the early 1980s. Although these high-proper motion surveys imply a strong bias towards high-velocity objects, frequently part of the extended spherical 'halo' component of our Galaxy's stellar population, the latest deep digital sky surveys continue to discover faint high proper motion stars, and specific attempts to determine their parallaxes are being made with the objective of completing the census of stars nearest to the Sun.

Narrow-Field Astrometry

ANOTHER SPECIALISED and productive field of astrometry over the past century or more has been the study of binary and multiple stars. Many stars are born as members of a binary system (or less commonly as a triple or quadruple system), and the relative motions of their individual components, or their photocentre, has led to an enormous body of data on binary and multiple star orbits. Traditionally, long-focus telescopes with a large photographic plate scale were used. Reasonably high *relative* positional accuracy could be achieved because the atmosphere does not impose the same type of deleterious random image motion on very small angular scales (say, within 5–10 seconds of arc), as it does on larger angular scales. Accordingly, while not providing information on parallaxes, or on the celestial reference frame, this approach has provided a wealth of data on higher-order positional effects that modify relative positions on small angular scales.

Within the last 10 years or so, this technique is being further applied to narrow-field astrometry using optical or infrared interferometers on Earth. Relative accuracies of order one thousandths of an second of arc

or better have been achieved, while efforts are ongoing to drive these narrow-field astrometric measurements to perhaps some 10 millionths of a second of arc (as targeted by VLTI–GRAVITY). Such accuracies would greatly assist in characterising the properties of the extra-solar planets now being discovered.

The Move to Space

TWO THOUSAND YEARS of charting the stars has led us on a remarkable voyage of discovery. The Earth, as we now know, is not at all at the centre of the Universe, but a spinning body of unremarkable mass which orbits the Sun. Billions of other stars, as well as planets, interstellar gas and dust, radiation, and invisible material are bound together to form our Galaxy—a magnificent disk spiral system, prevented from collapsing by its own rotation. Our Sun lies way out in one of the spiral arms, thirty thousand light-years from the centre. Around us the stars, at truly immense distances, move along their own eternal paths. Beyond our own island universe, the Milky Way, a seeming infinity of other galaxies recede from us at astonishing speeds, pointing their fingers backwards in time to the dawn of creation.

Many of these advances in our understanding have accrued from a steady refinement in measuring star positions. Over the past century, improvements advanced along a very high accuracy branch for a very few stars, culminating in the compilations of parallax distances for around eight thousand stars. A medium accuracy branch for a thousand or so stars gave our very best, but still troublingly inadequate, celestial reference system.

The lower accuracy branch developed progressively from Tycho's catalogue of 1000 stars with an accuracy of fifty seconds of arc in around 1600, Flamsteed's survey of 3000 stars to twenty seconds of arc around 1700, Lalande's 50 000 stars at three seconds of arc around 1780, and Argelander's survey of more than 300 000 stars at one second of arc around the 1850s. Billion star surveys were compiled from the world's arsenal of Schmidt telescopes in the late 1900s, but despite their colossal strength in numbers, positions were only marginally better than the surveys of more than a century before.

AT THE DAWN of the third millennium, the quality of star positions lagged far behind the progress achieved in other areas of astronomy. Accurate distances were still only known for a few hundred nearby stars, a severe barrier to understanding the physical processes within them. Accuracies from the large photographic surveys were strongly limited by the atmosphere. Proper motions were known for millions of stars, but with systematic errors over the sky which confounded their interpretation. Distances needed to transform them to space motions was all but lacking.

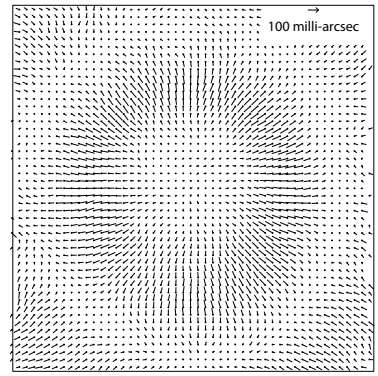
By the second half of the 20th century the steady advance in the accuracy of stellar positions was running into a number of insurmountable barriers. The biggest problem was the bending and twinkling effects of the atmosphere, condemning star images to their eternal and unpredictable wobbling dance. New thin-mirror telescope technologies have had great success in correcting effects over small angles, but all attempts to nail down large angles across the sky failed miserably.

In addition, there were the tiny variations in telescope alignment as the mountain-top observatories went through their endless day and night cycles of warming and cooling. The variable flexing of telescopes under their own weight as the huge supporting structures were steered to observe different parts of the sky added other unpredictable distortions.

YET ANOTHER complication was that any telescope on Earth can observe only part of the sky at any one time: a telescope in the northern hemisphere only ever sees the northern skies. Even so, it still requires a year to elapse for the entire region to be observable by night. It follows that a reference grid of star positions spanning the entire sky could only be constructed from a vast spider web of thousands of geometrical triangulations from separate telescopes observing accessible portions of the sky at different times.

However, between the various observations which had to be carefully patched together, all of the star images had moved by tiny but discernible amounts – due to their proper motions and parallaxes.

Like an ancient cartographic survey of the Earth made with primitive surveying instruments, the result of centuries of effort was a map of the sky, but one which was highly distorted and unpredictably warped. At accuracies below a second of arc, it was simply unreliable. Star positions were plagued by unfathomable errors which could not be unravelled.



Plates distortions (6° × 6°) from Hipparcos

Their space motions were, in consequence, of variable and sometimes questionable quality. More importantly, distances remained largely unknown, the tiny signatures of their minuscule parallaxes buried under a shroud of error-prone measurements imposed by the flickering atmosphere.

A fundamentally new approach to measuring star positions was desperately required.

THE PROPOSAL to make these delicate observations from space was the next master stroke of instrumental creativity. It was first formally laid out in the mid-1960s by 61-year old French astronomer Pierre Lacroute. Until then space science, still very much in its first flush of youth, had been somewhat the preserve of magnetospheric experts studying the region of the Earth's environment controlled by its magnetic field, discovered by Explorer–1 in 1958. X-ray astronomers, meanwhile, were eagerly following up their discovery of the first cosmic X-ray source in 1962.

It seems even more remarkable in hindsight that such a specialised goal in space science should have followed, within just a decade, of the first ever artificial satellite, the Soviet Union's Sputnik 1 in 1957.

Lacroute had realised that a space telescope would allow the measurement of arcs and triangulations to be made above the flickering effects of the atmosphere. Also, beyond the buckling forces of Earth's gravity, the telescope would not be sagging unpredictably as it made its cosmic census. Far from the Earth, the satellite would have an uninterrupted view of the entire sky, and it could also be shielded to simulate perpetual night time.

The most ingenious part of Lacroute's idea, however, was to observe in two very widely separated directions at the same time. Combining these two different sight lines into a single telescope focus, by means of a special split mirror looking out in two directions simultaneously, would give a network of wide-angle measurements spanning the whole celestial sphere in its entirety.

The idea of making differential angular measurements was not new in itself, and indeed Friedrich Bessel's first parallax measurements had made use of a somewhat similar approach a century and a half before. The novelty, empowered by the elimination of the atmosphere, was making these differential angular measurements across very wide sweeps of the night sky. From the network of space measurements, strict trigonometric distances could be disentangled. The goal, in short, was to construct a vastly improved census of stellar parallaxes, so that their distances could be measured and their physical properties derived. The satellite concept was duly named Hipparcos, a somewhat contrived, and thereafter rarely used, contraction of 'high-precision parallax collecting satellite', but also paying tribute to the ancient Greek pioneer of celestial mapping.

A LONG PROCESS of lobbying, and detailed design and feasibility study, eventually led to the Hipparcos project's adoption by the European Space Agency in 1980, and the satellite's launch in 1989.

Particularly influential in picking up Lacroute's concept, refining its technical precepts, consolidating its mathematical foundation, and detailing its scientific objectives were the four scientific consortium leaders who dedicated much of their own careers to its successful pursuit – Erik Høg (Copenhagen), Jean Kovalevsky (Grasse), Lennart Lindegren (Lund) and Catherine Turon (Paris–Meudon). A substantial technical and scientific effort underpinned the extensive international collaboration coordinated by ESA and directed by the Hipparcos Science Team, in total comprising some 200 European scientists, 30 European industrial teams, some hundreds of engineers and managers from across the ESA member states, and an overall budget of some €400 million (at year 2000 economic conditions).

Publication of the Hipparcos catalogue in 1997 presented the positions, space motions, and distances of more than 100 000 stars, all measured with equal attention, all accurate to around one thousandth of a second of arc, comprising comprehensive astrometric, photometric, and double star data. Subsequently-published products included the Tycho 2 catalogue of 2.5 million star, and an improvement in the astrometric quality primarily of the brightest stars.

Pre-launch testing of Hipparcos

THE HIPPARCOS satellite mission – two decades of focused work by hundreds of European scientists and engineers – provided not only the most accurate positional survey to date by far. Very significantly, it joined together in a single survey the most delicate work on individual stellar distances, the highest accuracy of the best reference frames, and the formidable large-scale surveys of history's great star charts.

Its substantial leap in accuracy was the largest single advance in astrometry in the entire history of the field, an improvement over its predecessors by a factor of fifty, and with resulting contributions to stellar astrophysics, the distance scale, and Galactic structure and dynamics. Freeman Dyson, in his 1998 book *Infinite in All Directions*, said of it: *'Hipparcos is the first time since Sputnik in 1957 that a major new development in space science has come from outside the United States.'*

Meanwhile, also based in Earth orbit, the NASA/ESA Hubble Space Telescope, launched in 1990, has also provided narrow-field positional accuracies of better than one thousandth of a second of arc on a limited number of stars. Like Hipparcos, this instrumental advance has also further validated the approach of performing high-accuracy astrometric measurements from space.

Gaia in context

HISTORY DID NOT come to an end with the successful completion of the Hipparcos mission! Already by 1997, as the Hipparcos catalogue was being lodged in scientific libraries around the world, astronomers were advancing ideas for yet more ambitious experiments to map the stars from space.

These included both 'pointed' and 'sky-scanning' instrumental approaches, amongst them the German DIVA satellite, NASA's Space Interferometry Mission (SIM), various initiatives from the US Naval Observatory (FAME, AMEX, OBSS, and MAPS), from Russia (OSIRIS and LIDA), and Japan (JASMINE and Nano-JASMINE).

Most of these have since fallen by the wayside due to technological, cost, or political considerations, itself underlining the substantial technical complexity and cost of undertaking astrometric observations from space.

THE FIELD'S next major instrumental advance, Gaia, follows the same principles as Hipparcos, but with both scientific ambition and the experiment itself scaled up to reflect 20 years of progress in astronomy and technology, surpassing the Hipparcos accuracy by a factor 100. It features a much larger lightweight telescope, built from the highly stable ceramic silicon carbide.

Like a massive digital video camera, a carpet of CCD silicon sensors almost a square meter in area records the millions of star images that pass across it as this latest orbiting satellite once more scans the heavens.

The satellite operates far from Earth, 1.5 million km away, at the Sun–Earth Lagrange point. A powerful onboard processor handles a vast cascade of image manipulations before the information stream is despatched to Earth. Its data rate from its distant orbit to the Earth is, at around 5 Mbits per second, more than a hundred times that of its predecessor.

After five years of studies, and after protracted discussion and intense lobby, the European Space Agency's advisory bodies signed up to Gaia in October 2000, twenty years after a very different body of scientists did the same for Hipparcos in 1980. It claimed measurements of ten *millionths* of a second of arc for the brightest stars, a hundred times better than the pioneering results obtained from space by Hipparcos, and for more than a billion stars.

I had been ESA's Project Scientist for Hipparcos for its full 17-year duration (1981–1997), including as overall project manager following launch. Then, failure of the apogee boost motor left it in its unscheduled highly elliptical orbit, leaving countless problems to overcome during its 4-year operational period. But this experience of Hipparcos, from 'cradle-to-grave', had provided me with much experience in all aspects of the formulation and development of Gaia.

GAIA WAS DULY launched from Europe's space port in Kourou, French Guiana, in 2013, almost 25 years since the launch of Hipparcos. After scanning the skies in the opening years of the third millennium, its final harvest will be in scientific hands in the late 2020s.

This next leap in ambition is yielding a scientific harvest which dwarfs that of Hipparcos. Its colossal survey of more than a thousand million stars is providing a defining census of around one per cent of our Galaxy's entire stellar population, pin-pointing them in space right across its vast expanses.

Unimaginable numbers of stellar motions will reveal many more details of the vastly complex motions at play within our Galaxy. It will provide insights ranging from new tests of general relativity to stringent limits on the variation of fundamental physical constants. Even planets circling other stars will appear in their thousands from their tiny wobbling motions, identifying candidate systems for the burgeoning discipline of exoplanetology.

PERHAPS, IN A decade or two from now, some ingenious scientists and engineers will figure out how to build a satellite to measure a thousand times better than Gaia, at the billionth of a second of arc. At that point, distances out across the vast uncharted cosmological expanses of the Universe could be measured directly.

For now, such a possibility remains largely in the realms of science fiction. Indeed, as Danish authority Erik Høg has written after his lifelong contributions to the field: *'The Gaia astrometric survey of a thousand million stars cannot be surpassed in completeness and accuracy within the next forty or fifty years.'*

History is littered with erroneous predictions, so many self-proclaimed seers consistently failing to anticipate the accelerating pace of change. It would take a brave person to wager a significant sum either way… but my tendency would be to side with Erik Høg!

*****

This selective summary is based on my review 'The History of Astrometry', published in The European Physical Journal H (Historical Perspectives on Contemporary Physics), Vol. 37, pp. 745–792 (2012). The basis of this account originally appeared in my popular book describing the Hipparcos project The Making of History's Greatest Star Map, 2010.

My text on the early history draws much on the cited works of David Goodman & Colin Russell (1991), Michael Hoskin (1997), and Allan Chapman (1990). The latter provides a detailed account of angular measurements between 1500–1850.

My coverage of developments over the past century is inevitably incomplete, being intended as an overview of the subject in its broadest outlines rather than a detailed chronicle.

4. Hipparcos: the push to space

BY THE SECOND HALF of the twentieth century the steady advance in the accuracy of stellar positions was running headlong into a number of essentially insurmountable barriers. Progress in telescopes and their instruments seen over the previous two or three centuries was running out of steam. Limited improvement in measuring star positions, in turn, obstructed further progress in fixing star distances and studying their space motions.

For once, telescope size or optical quality were no longer the limiting factors. After two millennia of hard-won improvements, human ingenuity appeared to be finally barred by Nature's innate complexity.

THE BIGGEST PROBLEM was the bending and twinkling effects of the atmosphere, condemning star images to their eternal and unpredictable wobbling dance. New thin-mirror technologies were having some success in correcting effects over small angles, but all attempts to nail down large angles across the sky failed miserably.

In addition, there were the tiny variations in telescope alignment as the mountain-top observatories went through their endless day and night cycles of warming and cooling.

The variable flexing of telescopes under their own weight as the huge supporting structures were steered to observe different parts of the sky added other unpredictable distortions.

Yet another unassailable complication was that any telescope on Earth can observe only part of the sky at any one time: a telescope in the northern hemisphere only ever sees the northern skies. Even so, it still requires a year to elapse for the entire region to be observable by night. A grid of star positions spanning the entire sky could only be constructed from a vast spider web of thousands of geometrical triangulations from separate telescopes observing accessible portions of the sky at different times. However, between the various observations which had to be carefully patched together, all of the star images had moved, all but chaotically, by the tiny amounts which were to be probed.

Like an ancient cartographic survey of the Earth made with primitive surveying instruments, the result of centuries of effort was a map of the sky of sorts, but one which was highly distorted and unpredictably warped. At accuracies below a second of arc, it was simply unreliable. Star positions were plagued by unfathomable errors which could not be unravelled. Their space motions were, in consequence, of variable and sometimes questionable quality. More importantly, distances remained largely unknown, the tiny signatures of their miniscule parallaxes buried under a shroud of error-prone measurements imposed by the flickering atmosphere. A fundamentally new approach to measuring star positions was desperately required.

THE PRECOCIOUS PROPOSAL to make these delicate observations from space was the next master stroke of instrumental creativity.

It was first formally laid out in front of other scientists in the mid-1960s by 61-year old French astronomer Pierre Lacroute, although the idea of astrometric detection of binary stars and planets from space had been aired by Paul Couteau and Jean-Claude Pecker in a restricted bulletin of the Nice Observatory a couple of years before that. Until then space science, still very much in its first flush of youth, had been somewhat the preserve of magnetospheric experts studying the region of the Earth's environment controlled by its magnetic field, discovered by Explorer–1 in 1958. X-ray astronomers, meanwhile, were eagerly following up their discovery of the first cosmic X-ray source in 1962.

It seems even more remarkable in hindsight that such a specialised goal in space science should have followed so closely on the heels, within just a decade, of the first ever artificial satellite to orbit the Earth, the Soviet Union's Sputnik 1 in 1957.

Lacroute's compatriot Pierre Bacchus, assistant professor at the Strasbourg Observatory at the time, and later director of the Lille Observatory, was also involved in these early discussions. But the idea of dedicating an expensive space platform to measure star positions

came out of left field and was, for a number of years, neither enthusiastically received nor widely embraced. Beyond a limited peer group of its active exponents, the proposal most likely appeared to be a misdirection of the limited opportunities of space funding. Seen from outside, it probably didn't seem too difficult to tighten up the measurements of star positions made from the ground; other people's problems are, after all, never quite as taxing as one's own.

Astrometry was also a victim of its own difficulties. Despite the innovative efforts of brilliant instrumentalists, progress had been painfully slow because the problems were so forbidding. As a result, exciting new scientific results flowing from their work had largely dried up, and new creative minds were ill-inclined to enter the field. From the outside, the discipline probably appeared to be one which had run its course.

Wide support for the funding of an expensive space mission would be all the more difficult to engender.

Bordeaux Observatory (Yves Réquième)

Early discussions, Bordeaux, October 1965 (Lacroute seated at front, Bacchus behind)

Director of the Strasbourg Observatory for thirty years until his retirement in 1976, his life's work devoted to the measurement of star positions, Pierre Lacroute had realised that a space telescope would allow the measurement of arcs and triangulations to be made above the flickering effects of the atmosphere. Also, beyond the buckling forces of Earth's gravity, the telescope would not be sagging as it made its cosmic census.

Far from the Earth, the satellite would have an uninterrupted view of the entire sky, and the experiment could also be shielded to simulate perpetual night time. It may seem strange to think of space being so dark, but the back side of a satellite shields the telescope from the Sun in just the same way as the Earth shields us at night.

In fact, the shielding is much better because there is no atmosphere to scatter sunlight back into the telescope. So from a satellite above the Earth, the skies are very dark indeed. There would always be a region in the direction of the bright Sun which could not be observed at any one time. But even this gap would be filled in as the Earth moved on in its annual orbit around the sky.

THE MOST INGENIOUS PART of Lacroute's idea, however, was to observe in two very widely separated directions at the same time. Combining these two different sight lines into a single telescope focus, by means of a special split mirror looking out in two directions simultaneously, would give a network of wide-angle measures spanning the whole celestial sphere in its entirety.

The idea of making differential angular measurements was not new in itself, and indeed Friedrich Bessel's first parallax measurements had made use of a somewhat similar approach a century and a half before. The novelty, empowered by the elimination of the atmosphere, was making these differential angular measurements across very wide sweeps of the night sky. From the network of space measurements, strict trigonometric distances could be disentangled. The goal, in short, was to construct a vastly improved census of stellar parallaxes, so that their distances could be measured and their physical properties derived.

THE SATELLITE CONCEPT was duly named Hipparcos, a somewhat contrived, and thereafter rarely used, contraction of 'high-precision parallax collecting satellite', but also paying tribute to the ancient Greek pioneer of celestial mapping.

Careful mathematical studies and computer simulations showed that the celestial survey would be both superbly accurate and immensely rigid. It would be like replacing a highly-distorted map of the world made from sailing ships in the 1500s, with one established by satellite imagery and GPS positional technology in the year 2000.

LACROUTE'S FIRST PROPOSAL for these space observations was presented at the thirteenth triennial General Assembly of the world's astronomers, the International Astronomical Union, held in Prague in August 1967. There is a time-honoured way in science for gathering ideas to improve on any new experiment, and that is to publish an early concept, and wait for others to circulate criticisms or suggestions for improvements.

The following years accordingly saw Lacroute's ideas becoming increasingly developed within the active French community. But beyond France, while the grand vision of space astrometry attracted interest, some of his central ideas were considered technically unrealistic. It would take almost another decade, and the involvement of other talented instrument designers, to transform the early concepts into something truly feasible.

CNES Publication, February 1993

Pierre Lacroute (1993)

Hipparcos scientific consortia leaders and early proponents: Erik Høg, Jean Kovalevsky, Lennart Lindegren, and Catherine Turon

LACROUT HAD ORIGINALLY hoped to have his ideas supported and carried into space by the French national space agency, CNES. France, as some of the other larger European countries, like Germany, Italy, and UK, had its own vibrant space programme which, before and since, has chalked up a string of impressive scientific, telecommunications, and Earth observation satellites.

But their studies of the early satellite concept suggested that it would be too complex, too expensive, and too risky for France to go it alone. In particular, Lacroute's proposed implementation of the beam-combining mirror which would, with unprecedented acuity and supreme stability, look out in different directions at the same time, was quickly found to be a critical and formidable technical challenge.

Pooling financial resources, management experience, and technical know-how, the European Space Agency (then ESRO) provided the next obvious choice to the growing lobby for space astrometry for a possible funding and development route. French astronomer Jean Kovalevsky, an early supporter of Lacroute's idea, played a key part in converting the project into a wider European venture.

BACK IN 1960, an intergovernmental conference at Meyrin, in Switzerland, had agreed on setting up a pan-European preparatory commission for the coordination of space research. At a 1962 meeting in London, delegates from Belgium, France, Germany, Italy, The Netherlands, and the United Kingdom signed a convention creating the European Launcher Development Organisation, ELDO. Later that year in Paris, delegates from the same countries, along with Denmark, Spain, Sweden, and Switzerland, agreed to the creation of the European Space Research Organisation, ESRO.

ESA itself was formed in 1973, when the European Space Conference meeting in Brussels decided on the merger between these two bodies—ELDO and ESRO—as well as on the start of the Spacelab and Ariane programmes. Today, ESA is an intergovernmental organisation of 22 member states which participate to varying degrees in a range of mandatory and optional space programmes, across several disciplines or directorates.

Opening the project to wider international collaboration was to prove enormously beneficial. Across Europe, those with creative experimental insight added important features to the instrument design, and others assisted with developing the astronomical arguments necessary to gain the support of the wider scientific community. Some would embark on a marathon journey to devote years of their lives on figuring out how the data would be analysed and interpreted once beamed down from space.

BY THE END OF THE 1970S, Hipparcos looked very different from the early ideas laid down by Lacroute. A symposium organised by ESA in October 1974 in Frascati, Italy, tested its wider international interest, and confirmed its potential pan-European appeal. With a number of crucial improvements brought by experienced Danish astrometrist Erik Høg after his introduction to the project in 1975, the satellite had been transformed. It was yet bolder in its objectives, far more efficient, technically simpler, and therefore more feasible.

It had also succeeded in generating a substantial scientific following across Europe, backed by an increasingly vocal international community. After a decade of study, it was at last poised to compete with other proposals for space missions vying for ESA's acceptance.

The project always fell somewhat short on instant glamour. But it garnered wide international acclaim for its profound scientific value. Its lack of allure beyond the immediate specialists would even prove to be a trump card during its execution: it was a scientific vision ignored by the other space agencies and, in consequence, encountered no competition during its development. Europe would duly win the unspoken race to measure the stars from space, a victory clearly facilitated by the absence of other competitors.

THE TRANSITION from initial concept to acceptance within ESA's scientific programme in 1980 followed a precarious path. For any major space facility the process involves considerable preparation and lobbying, and acceptance or rejection follows a long protracted course through the relevant advisory committees.

SCIENCE IS JUST ONE of the broad topical areas of the European Space Agency: others include telecommunications, meteorology, Earth observation, and human space flight. Each has its own advisory committees and decision-making structures. Then, as today, ESA solicited ideas for new missions, studied them intensively, consulted widely, evaluated competing concepts, and duly ruled on the next experiment to be flown in space under the flag of Europe.

The goal was and is still to select, once every few years, ideas which are innovative and scientifically compelling, challenging enough to develop European industrial capabilities while still being technically feasible, and financially acceptable. The competition is always fierce and keenly fought—projects can be studied intensively over a number of years, only to lose out in the face of competing projects as they advance over and around the progressive hurdles of the Agency's advisory bodies.

Within the scientific programme, a hierarchy of external advisory and decision-making structures formulates European space science strategy and policy. At the top, wielding political and financial muscle as well as the ultimate authority of scientific endorsement or veto, the Science Programme Committee bridges national and European interests.

Advising it is the influential Science Advisory Committee, which itself bestrides two further subordinate committees: the Solar System Working Group and the Astronomy Working Group.

The deliberations of these two groups also aims to create some balance in ESA's science programme between space missions focused respectively on the solar system (the Sun, the Sun–Earth system, the inner and outer planets, their moons, asteroids and comets) and the stars and galaxies beyond (as observed by optical, infrared, ultraviolet and X-ray instruments).

All of these committees, including their respective chairs, are populated by scientific advisors external to ESA. Their members are appointed from universities or technical institutes across the various countries to provide objective but in-depth guidance.

The selection process works as follows: the two working groups look at new ideas within their specific area of expertise, and make the first scientific assessment. Their recommendations are passed up one level to the Science Advisory Committee. Smaller in composition, typically more senior in career progression, and expected to be correspondingly more dispassionate in their deliberations, this committee then has to tread the difficult line of weighing and arbitrating across these two somewhat disparate disciplines.

Their mandate is to consider the importance of the project's scientific aims, rather than political or financial issues, although the triplet of constraints are closely welded, and difficult to completely separate.

AFTER DUE DELIBERATION, the Science Advisory Committee provides its scientific advice to the Science Programme Committee, the Agency's most senior external science advisory body.

This final committee stage has the additional delicate task of combining impartial scientific guidance with technical, political, and financial considerations, as well as overall programme balance and national interests. The ingredients, once stirred, can occasionally make for an explosive mix. This senior body is accordingly comprised of all the ESA member state national delegations, where senior scientists and space policy makers sit side by side under the chairmanship of one of the elected member state delegates.

Their formal and decisive meetings are held in ESA's headquarters in Paris three or four times a year, with interventions translated simultaneously into English, French and German, as laid down in its constitution. Through the national delegations, the Science Programme Committee in turn advises ESA's Director of Science as to how it wishes the scientific programme to be executed.

Roger-Maurice Bonnet (2000)

During the 'Hipparcos years', from 1980 to 1997, the post of Director of Science was occupied for the first three by German Ernst Trendelenburg, and thereafter by French solar physicist Roger–Maurice Bonnet.

THIRTY YEARS ON, the main protagonists probably have somewhat different recollections of the most crucial steps that were trodden as Hipparcos inched forward. However, the main decision points in its early path have been documented in the authoritative history of the European Space Agency [*A History of the European Space Agency, Volume 2 (1958–1987)* by J. Krige, A. Russo & L. Sebesta (ESA Publications, 2000)], from which I quote here *in extenso*.

Following earlier recommendations of the advisory structure in the mid-1970s, a number of competing ideas for a new round of space missions had been studied in 1977 and early 1978, aiming for a final decision in 1979. At the first committee stage, the Astronomy and Solar System Working Groups examined the ideas that had been put forward for new missions in their respective areas of expertise.

Hipparcos was one of the two competitors in front of the Astronomy Working Group, the other being a telescope to probe the Universe at ultraviolet wavelength, EXUV. The committee eventually awarded its priority to astrometry. The voting was a reasonably unambiguous eight votes to three, so for Hipparcos, so far so good.

The Solar System Working Group, meanwhile, had to rule on a competition between POLO, a proposed polar orbiting lunar observatory, and Giotto, the mission being proposed to rendezvous with Comet Halley. The struggle was hard fought, but Giotto duly emerged triumphantly as the committee's unequivocal choice.

The recommendation came with its own certain frisson: Giotto would have to be developed and launched with an urgency unprecedented within European space science, with launch just five years hence. Only on this aggressive schedule would there be time for it to race to its rendezvous with Comet Halley which, in its seventy-six year orbit around the Sun, would make its final apparition of the millennium in 1986.

When the Science Advisory Committee assembled for its meeting on 6–7 February 1980, the choice before them was therefore between Hipparcos and Giotto. Astrometry of the stars or rendezvous with a comet? The aspirations of the astronomy community, or the wishes of the solar system community? How could the two choices be compared? Not unexpectedly, discussions were sharply divided. As ESA's authorised history relates:

> The supporters of both missions strongly lobbied to have their pet project approved by ESA's decision-making bodies. Behind Hipparcos were the astronomers and the French delegation to ESA; support for Giotto came from the already established constituency of the ill-fated ESA/NASA cometary mission to Tempel–2, from the German delegation to ESA, and from the influential ESA Director of Science, Ernst Trendelenburg. The Science Advisory Committee also liked Giotto as it considered Hipparcos too costly, and its technical feasibility not completely established.

Two restricted sessions later and the Science Advisory Committee reached a tentative accord: it would select the Giotto comet mission as the Agency's next scientific project, but with a double proviso. Its scientific value would have to be further substantiated over the following three months. And its estimated cost ceiling of €120 million could not be exceeded.

While Hipparcos had convincing advocates in the Science Advisory Committee, including Jean Kovalevsky from France and Gustav Tammann from Switzerland, British physicist Harry Elliott finessed the discussions by arguing forcefully that *"a definite decision to proceed with Hipparcos could not yet be taken because of the absence of complete confidence in the technology."*

Complete confidence in an innovative space mission is a tall order. And the arguments scattered doubts which, once sown, were reflected in the committee's resolution. The committee requested that the three-month interregnum should also be used for further technical studies of the astrometry mission. In addition, it requested the Director of Science to find the ways and means whereby the telescope, so central to the satellite's purpose, could be funded nationally and not by ESA.

It thereby re-opened a thorny debate which had been enacted before, and which would surface again many times in the future. Should ESA simply exist to develop, finance and launch space platforms on which scientists placed their nationally-funded experiments? Or should it fund the experiments also?

Managerially, technically, scientifically, politically, and financially, there are pros and cons for both points of view. The entire Pandora's Box is opened anew at the time of each major selection, those peering inside for the first time usually bemused and perplexed by the gifts lying in wait.

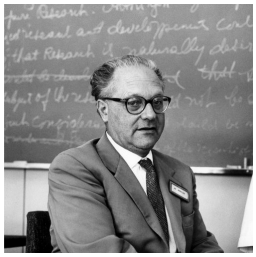
THE COMMITTEE'S RESOLUTION was a potentially crippling blow for Hipparcos. For it concluded that:

> In the event that the Hipparcos payload would need to be funded within the mandatory ESA programme, the committee was divided as to whether Hipparcos should then remain as the Agency's choice, or if EXUV should rather be carried out because this mission was considered by some members to be just as interesting.

ESA's authorised history goes on:

> The Science Advisory Committee decision came as a bombshell in the scientific community. Klaus Pinkau [chairing the meeting] and Ernst Trendelenburg, as well as the Science Programme Committee chairman [Edoardo Amaldi] and the ESA Director General, were flooded by telexes and letters from all over Europe, blaming, on the one hand, the unusual and 'arbitrary' procedure of recommending a project not supported by technical studies and not previously discussed by the Solar System Working Group, and claiming, on the other, the great support that Hipparcos enjoyed within the scientific community.
>
> The chairman of the Hipparcos consortium, [Italian astronomer] Pier Luigi Bernacca, wrote that 170 research proposals for the astrometry mission had been presented by 125 astronomers from twelve countries, recalling that twenty four institutes from eight countries were available to put manpower into hardware and software activities, and five were already working on aspects of hardware and software using their own funds. The cometary lobby was just as active, however, and many telexes arrived expressing satisfaction with the Science Advisory Committee's decision and wholehearted support for Giotto—which was *"a once in a lifetime opportunity."*

SUCH ARE THE CHASMS of impossible decisions across which we string a thin tightrope and insist that our unsuspecting leaders traverse. And thus befell to the Science Programme Committee, at its meeting of 4–5 March 1980, the unenviable task of arbitrating on the selection of ESA's next scientific project. Big money, unique scientific opportunities, and countless careers would ride on the outcome.


Istituto Nazionale di Fisica Nucleare (INFN)

Edoardo Amaldi

The meeting was chaired by Edoardo Amaldi, Italian physicist and scientific statesman, one of the founding fathers of both CERN, the European Organisation for Nuclear Research, and ESRO, the forerunner of ESA.

There was no solution which would please everyone, and the meeting was conducted in an atmosphere which reflected this. Most delegations *"regretted"* the lack of information and the hurried decision that the committee was being asked to confront. The French representatives explicitly challenged the executive for presenting a proposal which was *"not politically advisable since it had not met with a general consensus in the scientific community and could possibly lead to a complete split in the Committee."*

Only Germany and Sweden came down resolutely in favour of the Giotto comet mission. France, Belgium, The Netherlands, Denmark, Italy, Switzerland and Spain supported Hipparcos. The British delegation requested that a decision be deferred until more information was available on each, and the Irish announced an equal interest in both (23rd Meeting of ESA's Science Programme Committee, 4–5 March 1980).

If a vote had been taken by the Science Programme Committee, it seems inescapable that Hipparcos would have been chosen as the Agency's next scientific project. But such a decision, Amaldi recognised, would have left several delegations, and a large fraction of the scientific community, deeply dissatisfied, for there were scientists from most countries who had some interest in each.

INSTEAD, A COMPROMISE was negotiated: Hipparcos was reinstated as the next scientific project, with the provision that ESA should also take technical and financial responsibility for the telescope as well as for the satellite platform. The study of the Halley mission was to be extended for three further months and, if proven technically feasible within a cost envelope of €80 million, would also be included in the programme as a fast-track priority. As ESA's authorised history viewed it

> ...the stars could wait, while the comet could not be stopped in its journey through the solar system, and the two-week launch window of July 1985 could not be missed.

The compromise did not make the Science Advisory Committee happy, at least not its chairman who offered his resignation to the ESA Director of Science, Ernst Trendelenburg. Facing the tight financial situation of the science programme, he did not like the disproportionately large price tag for the astronomy mission, nor the decision to finance the Hipparcos payload out of the ESA science budget.

In the domain of solar system science it was customary for ESA to provide only the satellite platforms, with the experiments funded nationally and provided by scientific laboratories from the various nation states. Dual standards seemed to be operating: Hipparcos was to be the third astronomy programme, following the X-ray observatory EXOSAT and the European Faint Object Camera instrument on the Hubble Space Telescope, for which the experiment costs would be carried by ESA. The inevitable consequence would be a reduction in the funds available for new space projects.

THE FINAL DECISIONS were taken at the Science Programme Committee meeting on 8–9 July 1980 [24th Meeting of ESA's Science Programme Committee]. Reports of the meeting suggest that it was not at all plain sailing either, again ending in an uncomfortable compromise between ESA's two most influential member states at the time.

After lengthy discussions, the German delegation agreed to withdraw its request for alternative funding for the Hipparcos instrument, nonetheless still convinced that the science budget would be blocked for a long period, to the detriment of future launch opportunities. France continued to vote against Giotto, formally on the grounds that the opportunity to provide experiments on board would not be open to scientists from the US.

GIOTTO WAS FINALLY ADOPTED with ten votes for and one against. Trendelenburg stressed that *"the Giotto project was certainly more risky than any other project undertaken by the Agency to date, but believed that ESA had demonstrated that it was technically able to undertake such a project, and hoped delegations would fully support the executive in its endeavours to carry out the mission successfully."*

Hipparcos would follow.

THE ESA AUTHORISED HISTORY sums it up perceptively after chronicling several similarly difficult choices over nearly two decades:

> Choosing a big scientific project is also a matter of confrontation among scientists involved in the decision-making process: members of advisory committees or national delegations, government advisors, and policy makers. At each stage of the process, the traditional ties of cooperation, fellowship and solidarity that characterise the scientific community are strained by the emergence of national interests, disciplinary competitions, personal ambitions, career expectations, and personal relationships. When only one or two big projects can be started every three or four years the stakes are high and scientific objectivity is often a luxury. When making a choice entails some kind of painful discrimination, personal prestige, diplomatic talent, and personal or professional links can play a decisive role.

THE CONSEQUENCES OF THESE DIFFICULT but momentous decisions have echoed on down the years. With almost three decades of hindsight we can see that, ultimately, both projects were highly successful.

Giotto went on to make its spectacular encounter with Halley's comet on 14 March 1986, a challenging project completed in record time. Its remarkable rendezvous was broadcast live around the world, providing humanity's first close up view of a comet nucleus. The historic event put ESA in the spotlight for its stunning technical achievement. Largely unscathed by its close passage, Giotto sped off triumphantly to a new encounter with comet Grigg–Skjellerup in July 1992.

The financial concerns expressed at the time were, however, surely vindicated. Following seven satellites launched by ESRO, the forerunner to ESA, in the period 1968–1972, and three successfully launched by ESA between 1975–1978, only three were launched in the 1980s: the X-ray observatory EXOSAT in 1983, Giotto in 1985, and Hipparcos in 1989.

Far from being a consequence solely of the advisory committee decisions of the first six months of 1980, and attributable in part to the growing complexity of space missions, this was nevertheless a far cry from the target of one launch per year which had been suggested by the Science Advisory Committee as necessary for a viable European space science programme.

THE 1980S, in practice, marked the dawn of a voracious scientific appetite. It was matched with a precocious engineering capability in Europe, driven by many exciting ideas from its member state scientists clamouring for ever more challenging and expensive missions to unravel the secrets of the Universe. A generation of brilliant space engineers were ready to meet the challenges.

And there was a corresponding industrial development, growing in technical stature and the capability to manage the programmes. At the same time, this exciting period was soon to be met by a downturn in ESA's real scientific programme financial budget, and a rise in the administrative structures put in place to execute it.

But Hipparcos, at least, had its feet in the starting blocks. No ESA science mission before, and only one since, had ever been approved, only to be cancelled before launch. With its ticket to space all but guaranteed, celebrations were the order of the day.

DANISH ASTROMETRIST Erik Høg, co-opted onto the Astronomy Working Group between 1976–78, and a member of the Hipparcos scientific advisory team from 1981–1997, has written almost three decades later of his conviction that, if approval had failed back in 1980, then space astrometry would never had happened. His arguments were that (Contributions to the History of Astrometry Number 6: *Miraculous Approval of Hipparcos in 1980* by Erik Høg, 28 May 2008):

> For decades up to 1980 the astrometry community was becoming ever weaker, the older generation retired, and very few young scientists entered the field. I myself would have lost faith that the astrophysicists would ever let such a mission through, and others would also have left the field of space astrometry. If someone would have tried a Hipparcos revival one or two decades later, the available astrometric competence would have been weaker, and where should the faith in astrometry have come from then? When Hipparcos became a European project in 1975, and the hopes were high for its realisation, the competence from many European countries gathered, and eventually was able to carry the mission. This could not have been repeated after a rejection.

Would NASA have picked up such a mission in subsequent years? No, believes Høg for two reasons: first, they had less breadth of competence to draw on than in Europe in this specific field. And, quoting an American colleague, *"You can convince a US Congressman that it is important to find life on other planets, but not that it is important to measure a hundred thousand stars."*

I MET PIERRE LACROUTE for the first time in 1981, just after I had taken up position as ESA's Project Scientist, and just as the detailed satellite design phase had started. Then aged 75, he was still actively writing technical reports, developing a way to improve the measurements by making use of the dynamical stability of the slowly-spinning satellite.

He made several visits from his retirement home in Dijon to ESA's technical centre in The Netherlands to discuss his ideas. He spoke a laboured English when necessary, but preferred French. Always immaculately dressed in a three-piece suit, he was uncommonly distinguished-looking amongst the broader ranks of astronomers who are not renowned for sartorial elegance.

The full acceptance of Hipparcos must have meant much to him, yet in our various meetings he showed little emotion at the impending prospects. In an obituary (Bulletin of the American Astronomical Society, Vol. 25, p1498, 1993) André Heck, who had worked with him in Strasbourg, noted *"I was always impressed by his kindness, although he was moulded in the old-style, somewhat authoritative, managerial approach."*

OVER THE NEXT DECADE, Lacroute watched from his home in Burgundy as Hipparcos slowly took shape, joining us for the launch from Kourou in 1989.

There is a street now named after him in his home town of Dijon. But in the astronomy world, his name will be remembered as the father of space astrometry.

*****

Part of this account appeared in my book describing Hipparcos: 'The Making of History's Greatest Star Map', 2010.

5. An input catalogue, or…

As Gaia scans the sky, it detects and observes everything brighter than a specified threshold, at about 21 mag. This avoids the use of a pre-defined observing programme, and it ensures that all objects bright enough at the time of their observation – whether regular or irregular variables, transients including supernovae or microlensed events, or moving objects in the solar system – are detected and observed.

This powerful system required some clever techniques to implement. And it circumvented one of the very big challenges that its predecessor, Hipparcos, had to tackle: defining the mission's observing programme.

The principles employed for Gaia were driven by this earlier experience. I will give here an overview of the Hipparcos background, in part as a brief historical record, but equally to demonstrate how hard-won experiences translate into ideas for technological advances.

Hipparcos was conceived in the late 1960s, and launched in 1989. But a satellite is built around technologies that must be ready and proven when the detailed design is undertaken, some years before that. Hipparcos, accordingly, was designed and built around the technologies to hand in the early 1980s.

CCDs were arriving, but with readout noise and charge overflow properties totally unsuitable for the mission's goals. Instead, the heart of the instrument's measurement system was a (by today's standards) low-efficiency photocathode-based 'image dissector tube'.

Mounted behind a high-precision 'modulating grid', the detector's 30-arcsec diameter sensitive spot was piloted, electronically, to a given physical location. It tracked this moving spot for a few seconds as the telescope scanned the sky, then jumped to another star within the combined field of view. And so the process repeated indefinitely, interlacing the dwell periods on each star, and switching between them as stars entered the field and, 20 seconds later, exited from it.

All of this had important consequences for the satellite observations. First, in terms of the size of the observing programme, only one star could be observed at a time, and as a consequence of this, the total observing time available had to be carved up between the stars visible in the telescope's combined fields of view at that time. It followed that only a certain subset of all stars could be observed in total.

Second, stars had to be brighter than the instrument's detection threshold (around 11–12 mag), and in view of the available observing time, there could not be too many of these 'faint stars' in any area of the sky.

Finally, the strict demands on the detector piloting, and its sensitive area, meant that each star observed had to have its position known, at the epoch of each observation, to better than about 1 arcsec. To allow an optimum distribution of observing time, magnitudes also had to be known, preferably to better than about 1 mag.

The practical implications were far from straightforward. There would need to be a very careful selection of those stars chosen for measurement by Hipparcos. There would need to be extensive preparations to establish the positions and proper motions of these target stars, and supplemented by new observations if the actual observational knowledge was inadequate. The same was true for the magnitudes of the chosen stars, and all of this was compounded in the case of variable stars, double or multiple stars, selected asteroids, and so on.

To summarise, before the satellite could be set in motion, there was a need for a master catalogue, defining the stars to be observed, and the satellite attitude itself. It would list the star positions at the times of observation, both instructing the satellite's pointing system which part of the sky it was scanning, and informing its detector which stars were next to be observed.

Down to the faint limit of observability of the Hipparcos telescope, of around twelfth magnitude, there are some million or more stars in the sky. It was not difficult to figure out how much time could be given to each as the telescope scanned, and how many could therefore be observed over its lifetime. Hipparcos would have time enough to observe only around one hundred thousand, so a careful selection had to be made.

DECIDING WHICH out of the million possible were to be observed would itself determine which stars would be in the final catalogue, and therefore which would be handed down to future generations with their accurate distances and proper motions.

But which were the most important? Would it be the few known white dwarfs, or the most nearby stars, or those representative of our Galactic disk or its ancient halo? High proper motion stars were important, so too were a long list of binary stars and variable stars. Stars with unusual chemical abundances had to be included, along with the oldest subdwarfs. And so the list went on.

On top of all, there had to be a fairly uniform distribution of stars across the sky to serve as the celestial reference frame. For each star included, ten would have to be excluded. For each scientist pleased with the wisdom of a certain selection, another might be quite dissatisfied with the myopic choice.

An important to bear in mind was that the selection of Hipparcos in the late 1970s and early 1980s was a competitive, protracted and tortuous affair. Coordinating the preparations for its launch throughout the 1980s, I can state that, outside of the traditional astrometric community there were few scientists enthusiastic to see ESA undertake the mission, and indeed some were quite opposed. Amongst the Hipparcos teams at the time, there could be no idea that a follow-on mission would ever be considered, and no idea that the necessary technologies would advance so rapidly.

Hipparcos was seen as the one opportunity to define an observing programme, and reference frame, that would represent the state-of-the-art for a very long time.

CONSTRUCTING THIS starting catalogue – what was called the Hipparcos Input Catalogue – was indeed to prove a mammoth task. It was led by astronomer Catherine Turon of the Paris Observatory in Meudon.

In the late 1970s, Catherine Turon had discovered the intoxicating grandeur of the project through three colleagues: Jean Delhaye, former director of the Paris Observatory, who had conveyed to her his own curiosity about the structure of the Galaxy; Jean-Claude Pecker who, as their paths crossed in the observatory gardens, had asked her to replace him at an early symposium organised by ESRO (the forerunner of ESA) to gauge interest in space astrometry; and Jean Kovalevsky, who had urged her to probe the appeal to French astronomers of large numbers of accurate star parallaxes.

Once committed to the goals, and duly elected to lead the task, she assembled a team of about fifty astronomers ranged across European institutes and observatories to begin the work. Superbly organised, with an encyclopaedic knowledge of the stars, she inspired a large team that would work for more than five years to deliver the starting catalogue.

Always with a smile, always quick to laugh, she had a passion for the task she had undertaken, and a clear view of the final result that she wanted to achieve. People could not wriggle out of commitments they had made, and excuses for anything deviating from perfection, or the pressing schedule, were not well received. '*Mais non*, that was *not* what we agreed!' was heard often, but her reprimands were always issued with a winning smile.

Putting together the satellite's observing list was a balancing act: figuring out scientific priorities of each star, checking the expected performance of a trial catalogue by detailed simulations, assembling the information already known about each object, and setting up new observations using telescopes on ground to fill in missing data needed for the space operations.

THE PRACTICAL problems that had to be tackled in the 1980s are hard to appreciate today. One was that, in the proposals submitted, the same star could appear under many different names. The bright star Procyon, for example, is HD 61421 in the Henry Draper catalogue, GC 10277 in the General Catalogue, FK5 291 in the Fundamental Katalog, LTT 12053 from a high proper motion survey, and so on. Indeed, much of today's catalogue cross-indexes grew out of the Hipparcos work.

One might think that the star's position would resolve this dilemma, but many catalogues at the time did not list accurate positions; indeed, for many stars, accurate positions had often never been measured!

On top of this, all stars have a proper motion, and depending on which reference frame and time standard was used, even the position could differ between catalogues. Reaching the accuracy one second of arc or better by the time of launch, just to point the satellite's detector, was not too difficult in principle. But, in the early 1980s, it was enormously time consuming.

IT HAD BEEN agreed, during an early phase in the project's development, that the wider scientific community would be consulted on their opinions as to which stars should be observed—this was considered a once in a lifetime opportunity for science, and the wisest council was, in consequence, sought.

I steered through a policy paper which had to be debated and endorsed by the ESA advisory committees before we could open this to non-European suggestions. The stars observed would form a legacy for decades, and we wanted to make sure that the most important would be observed. This was no time to be parochial, our argument went, and it would be inappropriate to restrict scientific opinion exclusively to European scientists.

There was the counter view, forcefully expressed, that European nations were paying and that, accordingly, it should only be European scientists sowing the ideas and reaping the rewards.

If this should seem small minded, the logic carried force for those who held authority: national funding bodies would expect to see a return on their investment, in the form of scientific publications citing their own astronomers' work, not somebody else gaining the credits. Both arguments had substance, and had to be debated. The more altruistic camp held sway, and a world-wide call for observing programmes was issued early in 1982.

A CLOSING DATE OF OCTOBER 1982 was announced. The delivery format was carefully specified. Suggested star lists, in their hundreds or thousands, came pouring in. Scientists around the world had taken the opportunity as seriously as we had hoped they might.

We received lists of the stars most likely to provide a maximum scientific insight into their inner workings. Other lists identified for us objects likely to give the most knowledge about the rotation or structure of our Galaxy as a whole. Lists detailing nearby stars, high velocity stars, rare but important stars like the pulsating Cepheids and RR Lyrae stars, others mandatory for defining the stellar reference system, important binary systems, bright stars in the Magellanic Clouds, asteroids, and so forth, all flooded in.

Today, such details could be sent comfortably by e-mail. But in the early 1980s, neither e-mail nor the internet existed. Instead, half the proposals came by post on nine-track magnetic tapes, a bulky storage medium the size of a couple of dinner plates, which could hold a hundred megabytes of data, and which were the state-of-the-art in data storage at the time.

The remainder were sent in on the even bulkier punched cards! These had dominated data entry and computer programming for almost half a century, and although their popularity was waning, they were still in use. Each card, of size 7-3/8 × 3-1/4 inches, encoded up to 80 characters over its 80 columns, each represented by rectangular holes in each of 12 punch locations.

At ESA's technology and research centre in Noordwijk, I found an office for the temporary storage of these tapes and cards, before their onward despatch to the Observatory of Paris in Meudon, which would be the command centre for the next phases. By the time of the proposal deadline, the office was piled high with tapes of different sizes, and punched cards of varying colours. It was an Aladdin's Cave representing humanity's collective knowledge of the stars at that time. I regret not having a photograph to recall the one-time existence of this weighty collection, and of this seemingly primitive way that data was stored such a very short time ago.

TO PASS FROM disparate lists of suggested stars, to a true master list which could be used to operate a satellite, required a huge amount of work. Redundancies had to be eliminated, and obvious omissions rectified. Sky regions too much in demand had to be whittled down. Holes had to be plugged in areas where too few stars had been submitted. Positions had to be checked, and proper motions too. But it was the scientific priorities that would cause the biggest headache.

Adriaan Blaauw, elder statesman in the astronomical world, Director General of the European Southern Observatory between 1970–75 and one-time president of the International Astronomical Union, provided a guiding hand in defining the observing programme.

Following a suggestion by Henk Olthof, the secretary of ESA's Astronomy Working Group at the time, Blaauw was approached, and invited to set up and chair an independent committee to assist. It would be tasked to scrutinise the scientific suggestions, and to assign priorities to the goals laid out. Its brief was to ensure that the starting catalogue observed by the satellite was put together as carefully as possible.

In the early 1980s, Olthof took charge of nurturing all new projects entering, or wishing to enter, the privileged ranks of ESA's science programme, and he moved calmly and confidently through the various communities encouraging and facilitating. He felt that his compatriot's authority and contacts would rise to the challenge.

Mindful of this unique opportunity to get the list of stars to be observed chosen optimally Blaauw, who was nearing 70 at that time, assembled his own team of fifteen prominent figures in the astronomy world, from institutes around Europe, to contribute their impartial advice. Over the next few years, their task would be to debate the state of knowledge—of nearby stars, stars for the reference frame, stars of specific astrophysical interest, and so on—and adjust the observing programme to reflect the results most demanded from the satellite.


Michael Perryman

Selection committee, Paris (April 1987)

Three meetings of the committee over a period of three years were to guide the priorities. Each resulted in a progressive adjustment in the catalogue's contents, all changes to be made in careful dialogue with the leader of the scientific effort, Catherine Turon and her own international team.

OVER SEVERAL YEARS, the team constructing this 'input catalogue' met up for many progress meetings, traveling to one or other of the leading institutes. Astronomers managed to occupy some splendid real estate centuries ago, high points outside major cities, and

many have retained these superb sites down the years. The observatories of Paris and Nice, Rome and Torino, Heidelberg and Edinburgh are just a few that have cornered some of these great locations.

The goal of these meetings was to report on progress, reassess priorities, and debate plans and problems. But they served the additional purpose of bringing team members from different countries into close contact. This fostered a spirit of great collaboration and mutual trust, so essential to the big task building up around us.

ONE PARTICULARLY memorable meeting was of the entire consortium, some fifty people, held in the mountain village of Aussois on the edge of the Vanoise National Park in France, in early spring 1985. Catherine Turon, supported by her executive team and her international steering committee, drew up plans for a one week conference to get a complete picture of the current state of play of the starting catalogue.

Bernard Nicolet, hailing from Switzerland, had brought his impressively dimensioned Alpine horn along with him, and he roused us at sunrise each morning with a haunting reveille performed on the slopes outside. At the close of business each day, we could walk up from the Paul Langevin conference centre at the edge of the village into the still snow-covered alpine pastures from which marmots were starting to emerge.

Later we might be entertained by evening concerts from the more musically accomplished. To talk science for a week in such a location was an inspiration.

THE MEETING coincided with the 79th birthday of Pierre Lacroute, the satellite's originator, who was there. There was a great cake, and even some dancing.

Daniel Egret

Input Catalogue Consortium, Aussois, 1985

On such occasions I could meet with some of the senior figures who had dominated astrometry from the ground over the previous decades, including the influential Heidelberg astronomers Wilhelm Gliese (1915–1993), whose name still eponymises our knowledge of nearby stars, and Walter Fricke (1915–1988), who led the construction of the state-of-the-art catalogue of ground-based positions and proper motions of stars, the FK5.

The FK5 was a small but very select catalogue of just 1535 stars published in 1988, constructed meticulously following an enormous observing effort over decades. It was the authoritative word on the stellar reference frame before Hipparcos started on its own revolution.

Fricke, small in stature, large and jovial in character, had devoted his professional life to ground-based astrometry, but became an enthusiastic convert to its future from space. Despite our forty-year age difference, I found him charming, and encouraging, and he slipped me some valued advice along the way.

Two decades later, we would be looking at thousands of scientific papers making use of the final catalogue. We would be grateful that its unique content had been assembled with such attention and passion. Even before launch, Adriaan Blaauw proclaimed that Hipparcos had served as astronomy's great 'vacuum cleaner', its preparations already providing good positions for many stars hitherto unmeasured, cleaning up the confusing plethora of star names, and consistently identifying and labeling the components of binary star systems.

In a Foreword to the pre-launch mission description in 1989 Blaauw reflected: *All those who have contributed… are to be congratulated for their achievements – achievements that we will remember vividly when the Hipparcos satellite leaves us behind on earth to assume its heavenly high-precision task.*

Thirty years on, I would go further: it was the scientific content of the final catalogue which underpinned a wider appreciation of the importance of space astrometry, and played its part in ensuring the approval of Gaia.

FIVE YEARS AFTER THE WORK on the starting catalogue began, the Hipparcos Input Catalogue was completed. The list of stars to be observed was formally delivered to ESA at a ceremony presided over by its Director General, Reimar Lüst, on 11 April 1988 at Noordwijk.

Roger–Maurice Bonnet as Director of Science participated, and Pierre Lacroute, now 82, was guest of honour. Catherine Turon handed over a magnetic tape to Reimar Lüst in a symbolic gesture of a major milestone: the list of stars that had cost so much effort to prepare.

After the ceremony, Hamid Hassan (ESA's project manager) and I led Reimar Lüst and Roger–Maurice Bonnet, with Pierre Lacroute and Catherine Turon, Erik Høg and Jean Kovalevsky, along the warren of corridors, through the security barriers, and into the integration vault where, coated and masked, we could see Hipparcos being assembled under 'clean room' conditions.

Wires hung everywhere. Motors hummed, and lights flashed as tests and checks progressed. For the last time, we could gaze upon this remakable construction of glass and metal, and reflect on the complex combination of circumstances which had led to its creation.

SUCH WAS THE monumental task that confronted the construction of the 120 000 star Hipparcos Input Catalogue in the 1980s.

It really was not at all difficult to see that a different approach would be needed for Gaia's 20–21 mag, billion star catalogue. Easy to comprehend, but it would prove to be a great technical challenge too.

6. Galactic tracers, by design

IT IS NOT ONLY by good fortune that Gaia is uncovering so much about the structure and kinematics of our Milky Way Galaxy. The design of the scientific instrument, and its operation in orbit, targeted very specific measurement goals, formulated in terms of limiting magnitude, numbers of stars, and their unprecedented astrometric accuracy.

Complementing its astrometric measurements, Gaia was designed, from the start, to include accurate multi-colour multi-epoch photometric measurements that would allow each of its billion or more stars to be characterised, in terms of position in the Hertzsprung–Russell diagram, metallicity, and reddening.

Put simply, measuring the accurate distances and motions of a billion stars in the Galaxy would be a remarkable achievement in itself. But having a meaningful understanding of each star that had been measured was also enormously desirable.

THE PRIMARY OBJECTIVE of the Gaia mission is to observe the physical characteristics, kinematics and distribution of stars over a large fraction of the volume of our Galaxy, with the goal of achieving a major advance in understanding its dynamics and structure, and consequently its formation and history.

Gaia is making this goal possible by providing, for the first time, a catalogue which is sampling a large and well-defined fraction of the stellar distribution, determining the 3-dimensional positions and space velocities of every star observed, from which significant astrophysical conclusions can be drawn for the entire Galaxy.

Hipparcos did this for the solar neighbourhood. Gaia targets this for a large fraction of the Galaxy.

THE MOST CONSPICUOUS component of the Milky Way is its flat disk, which contains some hundred billion stars of all types and ages orbiting the Galactic centre. The Sun is located about 8.5 kpc from the centre. The disk displays spiral structure, and also contains interstellar material, mainly 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 about 3 million solar masses.

The disk and bulge are surrounded by a halo of about 10^9 old and metal-poor stars, as well as around 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 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.

UNDERSTANDING OUR GALAXY requires the accurate measurement of distances and space motions for large and unbiased samples of stars of different mass, age, metallicity, and evolutionary stage. The huge number of stars, impressive accuracy, and faint limiting magnitude of Gaia is able to quantify our understanding of the structure and motions within the bulge, the spiral arms, the disk and the outer halo, and is revolutionising dynamical studies of our Galaxy.

A summary of the main Galaxy components and sub-populations is given in the accompanying table, together with the required astrometric accuracies and limiting magnitudes. Compiled by Gilmore & Høg (1995), this was further described in a technical study note by Erik Høg (SAG–CUO–070, 1999), and appeared both in the Gaia Concept & Technology Study Report (ESA–SCI(2000)4, July 2000), and in Perryman et al. (2001).

The idea is that for each sub-population, and its representative kinematic tracers of known absolute magnitude and estimated distance and reddening, we can estimate the corresponding range of V magnitudes, and required astrometric accuracy, over which the populations must be sampled.

(1) Tracer (d = dwarf; g = giant)	(2) M_V mag	(3) ℓ deg	(4) b deg	(5) d kpc	(6) A_V mag	(7) V_1 mag	(8) V_2 mag	(9) ϵ_T km/s	(10) σ_{μ_1} μas/yr	(11) σ'_{μ_1} –	(12) σ'_{π_1} –
Bulge:											
• gM	–1	0	< 20	8	2–10	15	20	100	10	0.01	0.10
• horizontal branch	+0.5	0	< 20	8	2–10	17	20	100	20	0.01	0.20
• main-sequence 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
• horizontal branch	+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
• horizontal branch	+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
• horizontal branch	+0.5	all	> 20	30	0	13	21	100	35	0.05	1.4
Gravity (K_Z relation):											
• 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

IN THIS TABLE, Column 1 lists the various Galaxy populations, and their target 'tracers', which can be used to characterise both their distribution in space and their kinematics. These tracers include specific stellar types (e.g. Cepheid, RR Lyrae, or Mira variables), and specific spectral types (here B, F, G, K and M) and luminosity classes (d = dwarf; g = giant). Horizontal branch stars may refer to RR Lyrae, red horizontal branch, or clump giants, as appropriate.

For globular clusters, the cluster itself is used as tracer, with the proper motion referring to the mean of many stars. Globular cluster kinematics refers to the internal kinematics of globular clusters of our Galaxy.

Columns 2–8 list the corresponding photometric data: Column 2 is the absolute visual magnitude of a typical tracer star; Columns 3–4 give the typical Galactic coordinates for that tracers; Column 5 is the typical distance (or upper limit) for each tracer; Column 6 gives the typical visual extinction along the line of sight (for low latitudes the extinction in a Galactic window is given); Columns 7–8 then give the resulting typical range of apparent visual magnitudes over which the chosen tracer can be expected to appear in the Gaia survey.

Columns 9–12 list the relevant astrometric parameters: Column 9 gives the expected velocity dispersion for tracer stars of the sub-population, in the proper motion direction; Column 10 gives the expected Gaia proper motion standard error for a single star at the given representative magnitude; Column 11 gives the expected relative proper motion error; Column 12 gives the expected relative error on Gaia astrometric distances at the representative magnitudes.

LOOKING DOWN Columns 7–8, i.e. at the range of apparent visual magnitudes over which the chosen tracer can be expected to appear, we can see that an astrometric survey probing these different populations and tracers must reach at least down to $V = 14-15$ mag, and preferably as faint as $V = 20-21$ mag, to reach those specific population representatives. In short, the faint magnitude limit of Gaia, of around 20–21 mag, is essential for probing the different Galaxy populations.

In terms of target astrometric accuracies, Column 11 demonstrates that the proper motions for stars of all listed sub-populations can indeed be measured by Gaia with a significant, and often a superb, precision.

7. On-board detection

GAIA WAS ACCEPTED by ESA's Science Programme Committee in 2000. This followed a 3-year feasibility study, led by me (as Project Scientist) and Oscar Pace (as Study Manager), supported by a Science Advisory Group and various industrial teams. Findings appeared as a 380-page report, ESA–SCI(2000)4, in July 2000.

We identified 15 preparatory technologies needed to ensure that the satellite could be developed on schedule, and within budget. Five of these related to advanced CCD performances not available at the time, including 3-side buttable, small-pixel, high-performance chips; the large-area highly-integrated focal plane assembly; and high-speed, low-noise detection chains.

Associated developments were required for efficient on-board compression algorithms for the science data; optimisation of the payload data-handling electronics; and a (non-moving) phased-array antenna suitable for the high data-rate transmission from Gaia's L2 orbit.

Other technology advances were required for the large silicon-carbide mirrors and ultra-stable payload structure; the 10-m deployable solar array sunshield; the micro-Newton reaction system for the fine attitude control; and inch-worm actuators for telescope refocusing.

IN NORMAL CCD imaging, individual pixels accumulate photoelectrons during an exposure, after which the entire CCD is 'read out': pixel columns are successively advanced electronically towards a readout register.

In contrast, and central to the use of the Gaia CCDs, is the method of 'time-delayed integration', or TDI. Here, star images move slowly across the detector, and all CCD columns are stepped continuously towards the readout register, at a rate precisely synchronised with the satellite rotation. Exposures build up as the satellite scans the sky. There is no pausing for discrete image read outs.

The whole challenge of onboard source detection is intimately tied to the detailed CCD technical performance, and the entire system function depends on the adopted readout rate of the columns and rows, the readout noise in the electronics detection chain, and a host of other effects such as full-well capacity and charge overflow, and radiation damage and charge-trapping.

WHILE THERE would be some scientific merit in transmitting the entire data stream from the focal plane CCDs to the ground, certain issues conspire to make this approach unrealistic. Full focal plane readout in TDI mode would result in date rates of gigabits per second, compared with a realistic limit on the sustained telemetry rate from the L2 orbit of a few megabits per second. Even in the absence of telemetry rate limits, the rapid read-out rates would result in a high 'read-out noise', and a degraded signal-to-noise ratio per pixel.

Detailed analysis showed that full focal plane readout would, in any case, be of little value: for most of the sky, especially out of the Galactic plane, the fractional area covered by stars is very small, even at 20 mag.

The down-link of small patches of the detector field, centred on each detected object, and extending a little beyond the Airy disk, would instead be a satisfactory solution. Data rates would then drop to manageable levels, and the read-out noise per pixel could be brought down to adequate levels, even for the faintest stars.

Given that the focal plane data on all objects down to about 20–21 mag could be read out and telemetered to ground, the central problem was then how to identify the patches of sky containing these objects. Two options were evaluated: the use of an input catalogue; and the on-board detection of targets.

SEVERAL PROBLEMS made the use of an input catalogue unattractive if not unfeasible. In the first place, there was no existing catalogue which would suit Gaia's needs. And any plausible attempt to create one could never replicate the instrument's broad, red spectral response, and its sub-arcsec angular resolution, with strong implications for biases in the resulting measurements. Another major argument against this approach was the inevitable exclusion of transients, such as variable stars, burst sources, and solar system objects.

Scientifically, the issue was not only an operational one: a clear definition of the selection function used to decide which targets to observe (for example, catalogue completeness as a function of magnitude) would also be a key issue affecting the mission's scientific value.

THE SOLUTION ADOPTED was to detect every prospective target on-board, by means of a dedicated 'sky mapper'. This would also have the merit that the detection is carried out in the same wavelength band, and at the same angular resolution, as the final observations.

Converging on the finally chosen design was the result of more than 100 technical notes from the scientific side, many advanced by Erik Høg (Copenhagen), and many more from the industrial prime contractor, EADS Astrium (now Airbus Group), Toulouse.

In practice, existing catalogues were used to facilitate early attitude determination and source matching.

SUPPORTING THE many design choices were two very detailed accuracy assessment budgets: one maintained on the scientific side, under my responsibility, by Jos de Bruijne, and one on the industrial side by the prime contractor's chief system engineer, Frédéric Safa.

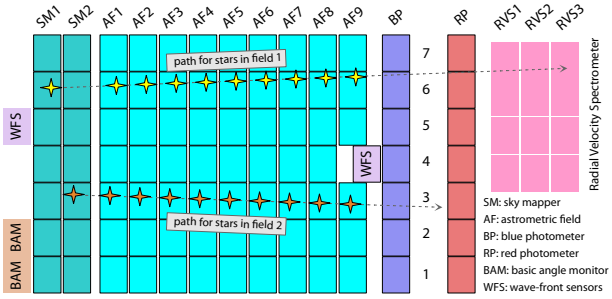
Every possible effect that the scientific and engineering teams could identify that might have an influence on the final astrometric accuracies were assessed, tabulated, and fed into these global accuracy budgets.

IN ITS FINAL configuration, Gaia consists of two telescopes, separated by 106.°5, feeding a common focal plane. The spacecraft rotates around an axis perpendicular to the two viewing directions, once every 6 hours.

The focal plane comprises 106 CCDs in seven rows, each with its own autonomous control unit. It takes a bit more than a minute for star images to drift across the entire 1.°5 focal plane in the along-scan direction. Operating in TDI mode, the scan rate is adjusted to coincide with the clock rate of the CCD readout. Each line of pixels is shifted in a little less than 1 ms.

The first two vertical CCD strips, SM1 and SM2, are the sky mappers, and perform the onboard object detection. Optical baffles ensure that each sees only one telescope field. In the absence of prior star knowledge, all pixels are read out, at the expense of a higher read noise.

Fast onboard analysis then produces a list of point-like objects for observation in the following CCDs (in which both fields are superimposed), rejects isolated cosmic ray events, and determines the scan rate about both axes. This information is provided to the attitude control subsystem, which allows the star positions to be predicted for the remaining focal plane crossing.

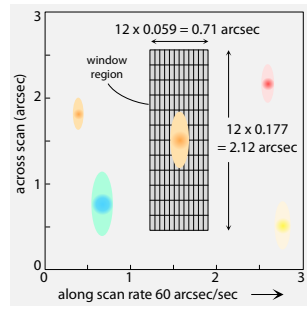


BEYOND THE astrometric field, star images cross the blue and red photometers, where prisms provide low-resolution spectra used to derive star colours. Finally, images reach the radial velocity spectrometer, providing high-resolution spectra for the brighter stars.

Various other tasks rely on measurements carried out in the focal plane. CCD 'gating' restricts the integration time for bright stars, and allows the measurement of objects that would otherwise cause detector saturation. Charge-injection mitigates some of the problems of radiation-induced charge trapping. Two CCDs are dedicated to wave-front sensors which allow telescope focussing. Two further CCDs are dedicated to a specific laser metrology system which monitors any tiny changes in the angle between the two viewing directions.

ONLY SMALL WINDOWS around the predicted positions of each object are read out. These windows are 12 pixels across scan (2.1 arcsec). Along scan, they are either 18 pixels (1.1 arcsec) for $G = 13 - 16$, or 12 pixels (0.7 arcsec) for $G > 16$ mag.

For $G > 13$ mag, windows are binned across scan to reduce the CCD readout noise. Rules determine which windows are used in very dense sky regions, and in conflicts due to close binary stars.



Images typically have a different *transverse* speed for each field, due to the satellite's precessional scan motion. With an integration time of 4.42 seconds per CCD, the transverse motion can reach 4.5 pixels over a single CCD, and can therefore result in significant across-scan smearing. As most samples are binned in this direction, the net effect on the observations is small.

THE ELONGATED CCD pixels, I should emphasise, are a consequence of the telescope's rectangular primary mirrors ($1.45 \times 0.5\,\mathrm{m}^2$), which result in an asymmetric point spread function with aspect ratio 3:1. The image width in the scan direction is about 0.1 arcsec, determining Gaia's ultimate resolution as an imaging system.

The CCDs have 4500 pixels in the along-scan direction, and 1966 pixels in the across-scan direction. Each pixel is rectangular in the same 3:1 ratio, the 10×30 micron pixels corresponding to 59×177 mas^2 on the sky.

Since each source is observed some 70 times with different scan directions during the nominal 5-year mission, the one-dimensional scans can eventually be used to create a two-dimensional reconstructed image.

MANY MORE details and examples of the onboard detection and windowing are given by Fabricius et al. (2016) and Arenou et al. (2018).

8. Why radial velocities?

THE SCIENTIFIC GOALS of ESA's Hipparcos mission, as they were prioritised at the time, were set out in ESA's Phase A study into the mission's feasibility in 1979. Priorities included defining a better stellar reference system, and providing a framework of improved distances, luminosities, and proper motions.

Radial velocities as a crucial observational quantity, and as a complement to the Hipparcos astrometry data, began to be discussed around the same time, especially in France and Switzerland. Indeed, the Phase A study report (pp19–20) made explicit reference to their importance. The compilation of measurements was duly identified as a task for the mission preparation, as described in the proposal for the 'input catalogue' to ESA in 1982.

But funding authorities were unwilling to support dedicated facilities, and acquiring even a limited subset of radial velocities proved to be a challenge.

KNOWING A STAR'S radial velocity is important because classical astrometry cannot measure a star's complete motion in space. And this applies to both Hipparcos and Gaia as well. Specifically, positional measurements can fix the star's position on the sky and its distance, in other words its full three-dimensional coordinate. But they can measure only the *projected* motion of the star across the sky (its proper motion).

The third component of the star's space velocity can only be determined from its radial velocity. Without it, the star's complete space motion is unknown, and kinematic and dynamical insights of individual objects – and entire populations – are accordingly restricted.

Multiple radial velocity measures can yield information on orbits and orbital companions (whether stars or planets), which themselves will distort knowledge of the star's space motion. But even a single radial velocity measure would provide much valuable information.

By 1982, with the solicitation for observing proposals on which the Hipparcos catalogue would eventually be based, the breadth of its scientific case was becaming more evident. And the importance and urgency of acquiring radial velocity measurements began to be discussed more widely (e.g. Fehrenbach, 1985).

RADIAL VELOCITY observations for a fairly limited subset of the Hipparcos stars were eventually made. These had to compete for telescope time and instrumentation in both northern and southern hemispheres. The effort was substantial, and a brief summary is given here. Further details are given in my *Astronomical Applications of Astrometry* (2009, pp32–35).

For early-type stars in the northern hemisphere, Charles Fehrenbach began a radial velocities survey for Hipparcos candidate stars in around 1982 using telescopes at the Observatoire de Haute Provence (OHP), in southern France. By the end of the programme, about 60% of the 12 000 B5–F5 stars in the northern part of the Hipparcos survey had radial velocities.

For early-type stars in the southern hemisphere, two ESO key-programmes for B5–F5 stars were started at the end of 1988 at the 1.52-m ESO telescope at La Silla: one for early-type stars nearer than 100 pc, and one for early-type stars in OB and early-A associations.

For late-type stars in the northern hemisphere, a programme of Coravel measurements using the 1-m Swiss telescope at OHP had been ongoing since 1977. It was dedicated, independently of Hipparcos, to nearby stars, cluster stars, high proper motion stars, Cepheids, and visual binaries, many of which would eventually be included in the Hipparcos programme.

For late-type stars in the southern hemisphere, an ESO key-programme was accepted in January 1989 for approximately 20 000 stars later than F5, using Coravel-South on the 1.54-m Danish telescope at La Silla. The overall observing programme was constructed along similar lines as for the northern part, and comprised Hipparcos 'survey' stars and various high-priority Galactic study programmes, all with two observations per star. Observations, totalling some 200 ESO and 200 Danish telescope nights, continued until about mid-1995.

Restrictions on the wider access to some of these results, meant that only about 20 000 radial velocities could be made more widely available at the time of the Hipparcos catalogue release. Acquiring them had been a major effort, but they represented only a small fraction of the total number of 120 000 Hipparcos stars.

IT IS PERHAPS of some historical interest to summarise the steps that were taken in trying to set up *dedicated* radial velocity measurement facilities at the time. This background, from my own records as ESA's project scientist, and those of Catherine Turon (Observatoire de Paris–Meudon), underlines how much work was invested in attempts to acquire them, the difficulties that were encountered, and the attitudes towards the importance of radial velocities that existed at the time.

ALREADY IN 1979, Catherine Turon, who would go on to lead the consortium defining the scientific priorities and catalogue content, had estimated the total number of radial velocity measurements that would be required to support the Hipparcos programme.

In 1980, a proposal was made by Albert Bijaoui to INAG (now INSU) in France for a new telescope and associated instrumentation, Coravel–North, to be installed at the Observatoire de Haute Provence, at an estimated cost of 5 MFrancs (say, 1.7M€ today). It would improve the existing Coravel limiting magnitude from $B < 14$ to 15.5, and would extend observations to A–F-type stars. At that time, the number of Hipparcos stars with no reliable radial velocity was estimated to be some 34 000 out of some 60 000 survey stars, and some 25 000 out of 40 000 non-survey stars with $B = 9 - 13$, corresponding to around 6 years of dedicated observations, assuming 3 observations per star. This proposal, and an updated proposal in September 1981, were unsuccessful.

IN JANUARY 1982, a meeting in Marseille between Marcel Golay, Michel Mayor and Fredy Rufener from Geneva Observatory, James Lequeux, Marc Azzopardi, Louis Prévot and Yvon Georgelin from Marseille Observatory, Renaud Foy and Guy Monnet from Lyon Observatory, Albert Bijaoui from Nice Observatory, and Suzanne Grenier and Catherine Turon from Paris Observatory, continued to emphasise the importance of radial velocities being available at the time of the Hipparcos catalogue publication, estimated then as around 1991.

They decreased the target observations per star to 2, and formulated a plan for a southern hemisphere programme of 12 000 observations per year for 5–6 years, with 6000 observations per year in the northern hemisphere. Given that some observations had already been made with the OHP Coravel, and that further northern hemisphere observations could start immediately, both goals were considered achievable within 10 years.

A new southern 1.5 m telescope was proposed as a collaboration between ESO, Geneva, and INAG, at a cost of 4 MFrancs (say, 1.4M€). The proposal, presented to ESO by James Lequeux and Michel Mayor in May 1982, was also unsuccessful. Nevertheless, these requests probably helped the decision to build the Elodie spectrograph for the 1.93 m telescope at OHP.

IN 1987, two years before launch, Roger Griffin, of the Institute of Astronomy, Cambridge, took up the challenge and submitted a proposal to the UK funding body, the SERC (now STFC) to acquire radial velocities for the Hipparcos stars and, in any telescope time still available, the Tycho stars. The proposal called for two identical 1-m telescopes, one in each hemisphere, at a total cost of about 1 M£ (say, 2M€ today). As he stated, *'The Hipparcos results will be made infinitely more valuable by the provision of radial velocities to complement the transverse motions measured by the satellite.'*

The proposal was supported by many, including Adriaan Blaauw as chair of the Hipparcos Proposal Selection Committee, Catherine Turon as chair of the INCA Consortium, and me as Project Scientist. But two separate proposal submissions were, again, unsuccessful.

I also made the science case to senior ESA management. It was again rejected, on the grounds that ESA had no mandate to fund ground-based astronomy, and these sorts of supporting measurements should be made from the ground. We had reached the final impasse.

MANY PROGRAMMES were later undertaken with existing instruments. But the systematic acquisition of radial velocities for all Hipparcos stars, in time for inclusion in the published catalogue, was a vision which, while articulated by many, never came to fruition.

Many papers using the Hipparcos data, from 1997 onwards, nevertheless remarked on their absence. Would it not have been a good idea, they implied, to have acquired and published radial velocities along with the final astrometric and photometric catalogues?!

In the years since, and with Gaia in mind, major ground-based programmes have contributed substantially, with the southern-hemisphere RAVE survey, at the UK Schmidt telescope in Australia, contributing radial velocities for 450 000 stars between 2003–213.

THESE PROTRACTED EXPERIENCES were central to my efforts in getting radial velocity measurements, for as many stars as possible, included as part of Gaia during the study phase before the mission's approval in 2000.

Doing so onboard would give measurements at the same time as the astrometric and photometric observations were being made for each star. Eventually, from an instrument point of view, observing all stars to a faint limit of about 13 mag would prove to be feasible.

FAST FORWARD to today, and EDR3 includes more than 7 million stars measured by Gaia, to 0.2–3 km s^{-1}.

But how were we able to clinch the case for these measurements to be made (and funded) as part of Gaia – when this had failed so dismally for Hipparcos?

That, my friends, is a story that can be told in just one minute, but it will have to wait for some other time!

9. Gaia and GDP

WHAT CONDITIONS exist in our society that allow expensive high-technology projects like Gaia to be undertaken in the first place? Why do funding agencies support these huge efforts? What are the returns for society, not only intellectually, but more generally?

ARGUMENTS THAT can be put forward for major investments in space missions include:

• scientific: scientists often argue that the search for the underlying laws of physics, and a deeper understanding of the Universe, are strong arguments for investment and research for any advanced and civilised society;

• unexpected spin-offs: as an extension of the pursuit of 'pure research', entirely unexpected and unpredictable spin-offs can eventually deliver substantial benefits to society, with associated economic dividends;

• applied spin-offs: more calculated approaches to exploiting potential spin-offs arising from space research can be encouraged and coordinated through technology transfer programmes and business 'incubation';

• technology development: space exploration provides new and inspiring challenges, requiring the direct development of new technologies;

• industrial: related to technology and economic return is the strong industrial interest in developing large-scale facilities and capabilities for space exploration;

• political: space exploration spectacularly demonstrates national and international capabilities, and can foster international cooperation in ambitious projects on an unprecedented scale, underpinning the 'peacekeeping' aspects of international collaboration;

• societal and cultural: space exploration offers extraordinary appeal to the public worldwide. It raises public interest, inspires and develops scientific culture to new levels, and attracts young people toward scientific and technical careers;

• economic: technological developments create new possibilities for innovation and economic growth, spanning business opportunities for industry as well as access to new resources.

THE DRIVERS are evidently a complex mix of science, technology, politics, and economics. But it is the last of these, the economic consequences, that I will focus on here – I imagine that economic considerations ultimately provide the strongest motivation for politicians, and so tax-payers, to commit to their very high costs.

Scientists often argue that knowledge is important for its own sake, and can be more vague when it comes to the wider benefits for society. But such arguments are easy to ignore, and budgets for space science can readily be scaled-back in the face of other priorities.

Naturally, basic scientific research (including astronomy and astrophysics) cannot be decoupled from questions of affordability, value for society, and industrial development and return.

So while we can see that Gaia is already having a major impact on scientific knowledge, we can also ask: will it have any beneficial effect on the economies of the countries that have been involved?

I will attempt a very general view of how fundamental research feeds through to economic growth.

SPACE IS NOW widely recognised as an essential resource for our civilisation, providing key services for telecommunications, satellite navigation, environmental monitoring, meteorology, crisis management, and science. Accordingly, it is widely supported by governments and private operators.

What are some headline numbers? In 2013, when Gaia was launched, `Euroconsult` indicated a space expenditure of €200 billion worldwide, with US government spending some $40 billion. In 2020, the same organisation valued the space economy at $385 billion, with commercial revenues totaling over $310 billion.

Across Europe in 2013, the space sector was generating civil and military sales of some €5.5 billion, and employing around 33 000 people. Governments of Japan, China, France, Germany, Italy, India and the EU each invested more than $1 billion in space activities.

Many more details along these lines are given annually, for example, by `Euroconsult`, and by `Eurospace`, the trade association of the European space industry.

In most European countries, space programmes are contracted out to national industries. And since the 1960s, economists have tried to measure the economic impact of these programmes with a variety of tools.

As one example, the Bureau d'Economie Théorique et Appliquée (BETA) of the University of Strasbourg attempted to evaluate the indirect industrial ('spinoff') effects of projects carried out by the European Space Agency (Cohendet 1999). One of their goals was to obtain quantitative figures that could be used to evaluate the effectiveness of any particular programme in order to justify their public-sector financial commitments.

Along similar lines, analysis by the OECD has estimated the economic return on investments in space across Europe (e.g. OECD Space Forum, 2011).

Skipping over numerous caveats, they gave the following figures for 2009: in Norway, an investment of NOK 1 million provided a return of some NOK 4.7 million. In Denmark, €1 million of Danish contributions to ESA generated a turnover of €3.7 million. In the UK, the space industry's value-added multiplier was estimated to be 1.91. In the US commercial space transportation industry, the factor was 4.9. At the upper end of the scale, NASA claimed a 7:1 return for every US dollar spent in the space programme.

A study by the Open University (2019) assessed the economic impact of the European Exploration Envelope Programme (E3P) missions. It found that each euro invested in E3P industry creates €3 in immediate impact, while each job contracted to industry creates two additional jobs in the space and broader industry.

All these studies go in the direction of trying to quantify the argument that national investment in space yields economic as well as societal and cultural benefits.

But now let me turn to the more specific question: is there a link between *basic research* and prosperity? Here, the distinction between basic research and applied 'research and development' is essentially a distinction between discovering the laws of nature, and harnessing them for practical purposes.

A more considered definition is given in the OECD's *Frascati Manual for Research and Development Statistics* (2002, p30). Basic research is defined there as: *'experimental or theoretical work undertaken primarily to acquire new knowledge of the underlying foundation of phenomena and observable facts, without any particular application or use in view'*. Applied research is defined as: *'directed primarily towards a specific practical aim or objective'*.

It is not difficult to find numbers used in arguments for funding basic research. For example, the then President of the Max Planck Society, Peter Gruss, argued in the *Max Planck Magazine*, in 2012: *'80% of economic growth in the industrialised countries results from the development of new technologies. And it is research, after all, that contributes vital ideas for new technologies'*.

Where do such numbers come from? While more traditional economic theories seem to ignore the role of basic research in economic prosperity, attempts have been made more recently to quantify their link.

In the class of economic theories characterised by 'endogenous growth', technological progress is generated by the accumulation of knowledge. An important result of these models is that growth is strongly dependent on spending on *basic* research, and indeed ceases without it (e.g. van Bochove, 2012).

Although much has been written on the subject, there is no simple answer. The absence of a clear consensus is probably reflected in the varying percentage of gross national product that is spent on space R&D by each country, and their contribution to ESA.

Again, estimates by `Euroconsult` indicated that, in 2013, the annual expenditure on space per capita ranged from $12 (0.03% of GNP) in the UK, $17 (0.05%) in Italy, $19 (0.04%) in Germany, $44 (0.10%) in France, and all the way up to $150 (0.32%) in the US.

That basic research is likely to be an indispensable catalyst for wealth and societal advance was at the heart of the Lisbon Strategy, an action and development plan set out by the European Council in March 2000, for the economy of the EU between 2000–2010. Through various policy initiatives, its aim was to make the EU *'the most competitive and dynamic knowledge-based economy in the world capable of sustainable economic growth with more and better jobs and greater social cohesion'*.

While the connection between research funding (both pure and applied) and economic growth seems to be plausible, and arguably convincing, the details remain imperfectly understood. Nonetheless, demanding that a research proposal must be able to demonstrate a *direct link* between the research and its economic benefits, would perhaps seem misguided.

Gaia started on its path as an idea that two scientists sat down and discussed in Leiden in 1993. During its technical development, it involved thousands of scientists, engineers, and managers across dozens of European industries. Its exploitation now involves hundreds of scientists, both career professionals as well as early-stage researchers who will move on to positions in science, industry, education, and management.

Can we then argue that Gaia is bringing significant economic returns as well as enormous scientific advances? I suspect so, but I have no quantitive evidence to support such a claim!

These considerations were part of my contribution to a study undertaken in 2014 by the European Academies Science Advisory Council (EASAC), chaired by Thierry Courvoisier, and which were included in its final report *'Strategic Considerations of Human versus Robotic Exploration'*.

10. Catalogue data releases

GAIA WAS launched from Kourou, French Guiana, on 19 December 2013, and after a 6-month commissioning phase, started routine operations in July 2014.

Gaia continuously scans the sky following a carefully specified scanning 'law', providing a fairly uniform scanning of the celestial sphere, as well as a robust 'separation' of the astrometric parameters of each star.

There is no pre-defined observing programme. Instead, objects brighter than a specified threshold are detected as they enter the instrument's fields of view, then followed as they cross the array of CCDs performing the astrometric, photometric, and radial velocity measurements. Typically, all stars brighter than about 21 mag are observed, including variable stars and moving objects. Just a handful are too bright to be observed.

Fundamental to the determination of absolute trigonometric parallaxes, Gaia has two widely separated fields of view, 106° apart. The angle between them is rigidly maintained, and the two fields of view are superimposed in the combined focal plane.

IN THE SIMPLEST case, a star's location is characterised by its position on the sky at some reference epoch (e.g. the mid-point of the observations), given by two angular coordinates (RA and dec). Its space motion, projected on the plane of the sky, is characterised by its angular motion in both coordinates, usually expressed in milli-arcseconds (milli-arcsec or mas) per year.

A star's finite distance results in an apparent change in its position as the Earth moves in its annual orbit around the Sun. From this parallax angle, typically measured in milli-arcsec (or mas) and denoted ϖ, its distance, in parcsec, is determined as $d = 1/\varpi$.

Gaia can only measure radial velocities (and thus all 6 quantities specifying its position and motion in Euclidean space) for stars brighter than about 13 mag.

AS MORE measurements are accumulated over time, the accuracy on these various quantities improves. At the same time, observations over more than a year start to get a better measure of their proper motions, and observations over more than 18–24 months start to get a much better separation between the effects of the star's proper motion through space, and its parallax. For more complex binary or multiple stars, and especially in the case of orbital motion, more than 5 or 6 quantities are required to describe the star's motion, and more observations are required to permit their estimation.

Successive catalogue releases comprise more observations, and as the temporal baseline increases, the accuracy of the astrometric parameters improves as a result of these two effects. Each successive catalogue supersedes the previous, so earlier data releases become redundant. In addition to orbital effects (due to binarity or planetary companions), more observations also give better handles on stellar variability, on the motion of solar system objects, and on systematic effects in the data (in particular the global parallax zero-point).

A summary of the three data releases to date are given in the table, and outlined below. My goal here is to summarise some key features useful in understanding the scientific use of the data. Many more details of each release are described under the given url's.

GAIA DATA RELEASE 1 (DR1) is based on observations collected between 25 July 2014 and 16 September 2015, and was released on 14 September 2016.

It contains source identifier, positions (α, δ) and G magnitudes for sources with acceptable standard errors, the five-parameter astrometric solution (position, proper motion, and parallax) for stars in common between the Tycho 2 Catalogue (epoch around 1990) and Gaia (epoch around 2015), based on the Tycho–Gaia Astrometric Solution (TGAS).

DR1 was incomplete because of limitations in sky coverage, and data processing constraints. Many bright stars, $G < 7$, were missing, as were many fainter stars, sources close to bright objects, high proper motion stars, and stars in areas with very high surface densities (above some 400 000 deg^{-2}).

An A&A special issue in 2016 contained 16 papers describing the various aspects of the DR1 data release.

	Gaia DR1	Gaia DR2	Gaia EDR3	accuracy (indicative)
Observations:				
– time period	Jul 2014–Sep 2015	Jul 2014–May 2016	Jul 2014–May 2017	
– observations duration	14 months	22 months	34 months	
– reference epoch	J2015.0	J2015.5	J2016.0	
– catalogue release date	14 September 2016	25 April 2018	3 December 2020	
– url: www.cosmos.esa.int/web/	`gaia/dr1`	`gaia/dr2`	`gaia/earlydr3`	see hyperlinks
Astrometry:				
– total number (3–21 mag)	1,142,679,769	1,692,919,135	1,811,709,771	0.01–1 mas
– 5-parameter solutions	2,057,050	1,331,909,727	585,416,709	
– 6-parameter solutions	2,057,050	–	882,328,109	
– 2-parameter solutions	1,140,622,719	361,009,408	343,964,953	
Photometry:				
– mean G magnitude	1,142,679,769	1,692,919,135	1,806,254,432	0.3–6 mmag
– mean G_{BP} photometry	–	1,381,964,755	1,542,033,472	1–100 mmag
– mean G_{RP} photometry	–	1,383,551,713	1,554,997,939	0.6–50 mmag
Radial velocities (4–13 mag)	–	7,224,631	7,209,831 (DR2)	0.2–3 km s^{-1}
Other:				
– variable sources	3,194	550,737	as DR2	
– known asteroids with epoch data	–	14,099	as DR2	
– Gaia-CRF sources	2,191	556,869	as DR2	
– effective temperatures	–	161,497,595	as DR2	
– extinction and reddening	–	87,733,672	as DR2	
– radius and luminosity	–	76,956,778	as DR2	

Gaia Data Release 2 (DR2) was based on observations between 25 July 2014 and 23 May 2016, and was released on 25 April 2018.

As described under the relevant ESA Gaia www pages (and where further details of various caveats are given), DR2 contains:

five-parameter astrometric solution (α, δ, μ_α, μ_δ, ϖ) for 1.3 billion sources in the range $G = 3-21$. Parallax uncertainties are up to 0.04 mas for $G < 15$, around 0.1 mas for $G = 17$ and around 0.7 mas at $G = 20$. Uncertainties in the proper motion components are up to 0.06 mas yr^{-1} (G <15 mag), 0.2 mas yr^{-1} ($G = 17$ mag) and 1.2 mas yr^{-1} ($G = 20$ mag). DR2 parallaxes and proper motions are based only on Gaia data, and no longer depend on Tycho-2/TGAS.

two-parameter astrometric solution 361 million additional sources for which a two-parameter solution is available: the positions on the sky (α, δ) combined with the mean G magnitude. These have a positional uncertainty at $G = 20$ of about 2 mas (at J2015.5).

radial velocities Median radial velocities (i.e., the median value over the epochs) for 7.2 million stars with a mean $G = 4-13$ and $T_{\rm eff} \sim 3550-6900$ K. This leads to a full six-parameter solution: positions and motions on the sky with parallaxes and radial velocities, all combined with mean G magnitudes. The overall radial velocity precision at the bright end is $\sim 200-300$ m s^{-1}, while at the faint end it is ~ 1.2 km s^{-1} for $T_{\rm eff} = 4750$ K and ~ 2.5 km s^{-1} for $T_{\rm eff} = 6500$ K.

G magnitudes for more than 1.69 billion sources, with precisions varying from around 1 mmag at $G < 13$ to ~20 mmag at $G = 20$. The photometric system for the G band in Gaia DR2 is different from the photometric system in Gaia DR1.

GBP and GRP magnitudes for more than 1.38 billion sources, with precisions varying from a few milli-mag at $G < 13$ to 200 mmag at $G = 20$. Passband definitions are given for G, $G_{\rm BP}$ and $G_{\rm RP}$.

epoch astrometry for 14 099 known solar system objects based on more than 1.5 million CCD observations. 96% of the along-scan residuals are in the range –5 to 5 mas, and 52% are in the range –1 to 1 mas. The transit observations are part of Gaia DR2 and have also been delivered to the Minor Planet Center (MPC).

effective temperatures for 161 million sources brighter than 17 mag with $T_{\rm eff} = 3000$–$10\,000$ K. For a subset of 87 million sources, the line-of-sight extinction A_G and reddening E(BP–RP) are also given, and for a part of this subset (76 million sources) the luminosity and radius are also available.

classification for more than 550 000 variable sources consisting of Cepheids, RR Lyrae, Mira and Semi-Regular candidates as well as high-amplitude Delta Scu, BY Dra, and SX Phe candidates and short time scale phenomena.

A series of 25 papers in A& A Volume 216 (2018) describe the details of the Gaia DR2 data release.

GAIA DATA RELEASE 3 is split into two parts: the early release, Gaia Early Data Release 3 (Gaia EDR3), and the full Gaia Data Release 3 (Gaia DR3).

Gaia EDR3 was released on 3 December 2020, and the full Gaia Data Release 3 (Gaia DR3) is planned for the first half of 2022 [postscript: actually on 13 June 2022].

EDR3, as well as DR3, are based on observations between 25 July 2014 and 28 May 2017, i.e. a period of 34 months (compared with 14 months for DR1, and 22 months for DR2). The reference epoch for each release is given in the table.

AS DESCRIBED under the relevant ESA Gaia www pages (and where further details of various caveats are given), EDR3 contains:

five-parameter astrometric solution (α, δ, parallax, and proper motions) for 585 million sources, with a limiting magnitude of about $G \approx 21$ and a bright limit of about $G \approx 3$.

six-parameter astrometric solution a 6-parameter solution for a further 882 million sources includes a 'pseudo-colour', for sources lacking high-quality colour information.

two-parameter astrometric solution for around 344 million additional sources.

G magnitudes for around 1.806 billion sources.

G_{BP} and G_{RP} magnitudes for around 1.542 billion and 1.555 billion sources, respectively. The photometric system for the G, G_{BP}, and G_{RP} bands in Gaia EDR3 is different from the photometric system used in Gaia DR2 and Gaia DR1. Passband definitions are given for G, G_{BP}, and G_{RP}.

celestial reference frame sources for about 1.614 million celestial reference frame sources (Gaia–CRF3).

cross-matches between Gaia EDR3, and Gaia DR2, Hipparcos-2, Tycho-2 + TDSC merged, 2MASS PSC (merged with 2MASX), SDSS DR13, Pan-STARRS1 DR1, SkyMapper DR1, GSC 2.3, APASS DR9, RAVE DR5, allWISE, and URAT-1.

SOME FURTHER details of the typical accuracies, limitations, and caveats follow (see also some example plots on the following page):

• survey completeness: EDR3 is largely complete between $G = 12 - 17$. The source list for the release is incomplete at the bright end, and has an ill-defined faint magnitude limit, which depends on celestial position.

The combination of the 'scanning law' sky coverage, and some filtering on data, leads to some regions of the sky displaying source density fluctuations reflecting this scanning law pattern. In addition, small gaps exist in the source distribution, for example close to bright stars.

• coordinate system: positions and proper motions are referred to the ICRS (International Celestial Reference System), using TCB (barycentric coordinate time) as the fundamental time coordinate.

• astrometry: position uncertainties are 0.01–0.02 mas ($G < 15$), 0.05 mas ($G = 17$), 0.4 mas ($G = 20$), and 1.0 mas ($G = 21$). Parallax uncertainties are 0.02–0.03 mas ($G < 15$), 0.07 mas ($G = 17$), 0.5 mas ($G = 20$), and 1.3 mas ($G = 21$). Proper motion uncertainties are 0.02–0.03 mas yr^{-1} ($G < 15$), 0.07 mas yr^{-1} ($G = 17$), 0.5 mas yr^{-1} ($G = 20$), and 1.4 mas yr^{-1} ($G = 21$).

• parallax zero point: the parallax zero point deduced from extragalactic sources is about $-17\,\mu$as. The uncertainties for the 6-parameter solutions are slightly larger than for 5-parameter solutions. Uncertainties for the 2-parameter solution (i.e. position only) are 1–3 mas.

• photometry: G-band standard errors are around 0.3 mmag ($G < 13$), 1 mmag ($G = 17$), and 6 mmag ($G = 20$). G_{BP} standard errors are around 0.9 mmag ($G < 13$), 12 mmag ($G = 17$), and 108 mmag ($G = 20$). G_{RP} standard errors are around 0.6 mmag ($G < 13$), 6 mmag ($G = 17$), and 52 mmag ($G = 20$).

• radial velocities: Gaia EDR3 (as Gaia DR2) contains median radial velocities for 7.21 million stars with $G \sim 4 - 13$ and $T_{\rm eff}$ in the range 3550–6900 K. The precision of the radial velocities is 200–300 m s^{-1} at the bright end. At the faint end, accuracies are around 1.2 km s^{-1} for $T_{\rm eff} = 4750$ K and around 3.5 km s^{-1} for $T_{\rm eff} = 6500$ K.

• astrophysical parameters: some were published in DR2; no new updates were provided with Gaia EDR3.

• variable stars: some were classified in DR2; no new updates were provided with Gaia EDR3.

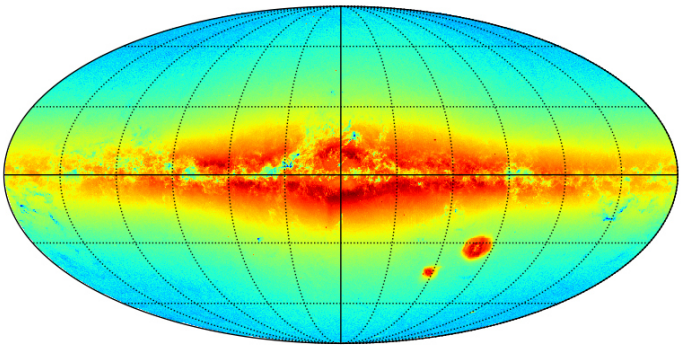
• solar system objects: information on the currently available asteroids was provided with Gaia DR2. Orbits will become available with the full Gaia DR3 release.

A SERIES OF papers in A&A (2021) describe the details of the Gaia EDR3 data release. These cover a summary of the contents and survey properties, as well as nine papers describing technical details of the data processing and calibration.

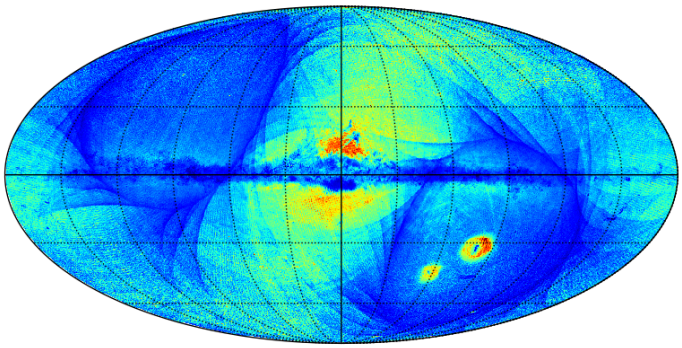
In addition, the data release was accompanied by four papers on selected science topics, viz. the Gaia catalogue of nearby stars, the structure and properties of the Magellanic Clouds, the Galactic anticentre, and the acceleration of the solar system from Gaia astrometry.

Some example statistical diagrams, given on the following page, are from `gea.esac.esa.int/archive/documentation/GEDR3`, where further details (and colour legends) can be found.

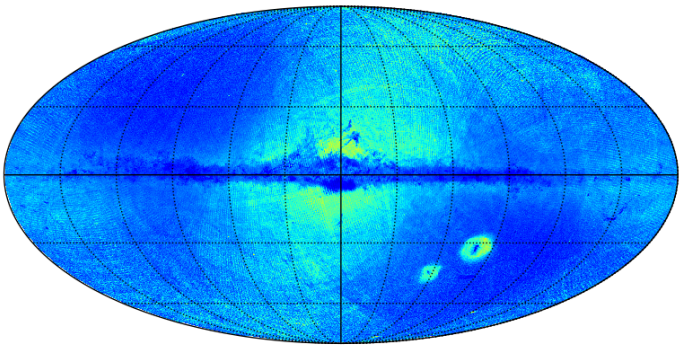
Further details of Gaia DR3 will be given separately when available [postscript: see essay #76, 13 June 2022].



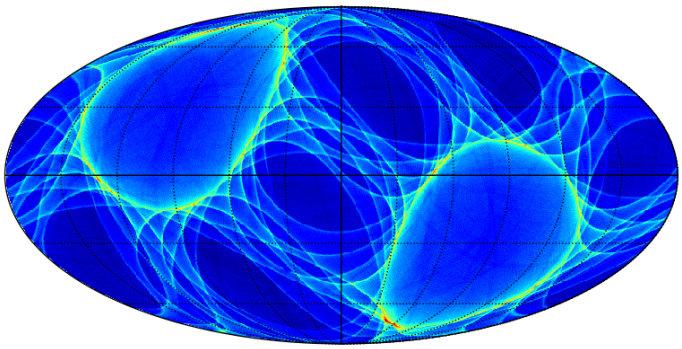
Density counts for sources with a 5-parameter solution from EDR3 (Galactic coordinates). Densities range from around 100 (light blue) to some 10 000 (dark red) stars per square degree.



Median right ascension error for sources with a 5-parameter solution from EDR3 (Galactic coordinates). These range from below 50 micro-arcseconds (dark blue) to about 300 micro-arcseconds (dark red).



Median parallax error for sources with a 5-parameter solution from EDR3 (Galactic coordinates). These range from below 50 micro-arcseconds (dark blue) to about 300 micro-arcseconds (yellow).



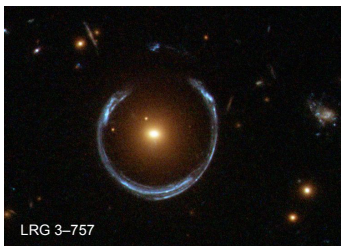
Median of number of good astrometric observations for sources with a 5-parameter solution from EDR3 (Galactic coordinates). These range from about 200 (dark blue) to more than 1000 (red).

11. Astrometric microlensing

THE FIRST observations confirming 'light bending' as described by general relativity were made by F. W. Dyson and A. S. Eddington using the 1919 solar eclipse seen from Brazil. More compelling confirmation of light bending by the Sun included the 1973 solar eclipse, and the full-sky Hipparcos observations from 1997.

The possibility that gravitational lensing by a nearby star could result in two distinct images of a background star was pointed out by Eddington in 1920. In 1936, Einstein commented *'Of course, there is no hope of observing this phenomenon directly. First, we shall scarcely ever approach closely enough to such a central line. Second, [the angles] will defy the resolving power of our instruments'*.

Further theoretical work led to the discovery of the first case of 'strong lensing', a double image of the quasar QSO 0957+561 (Walsh et al., 1979). More than a hundred such galaxy-lensed systems are known today.



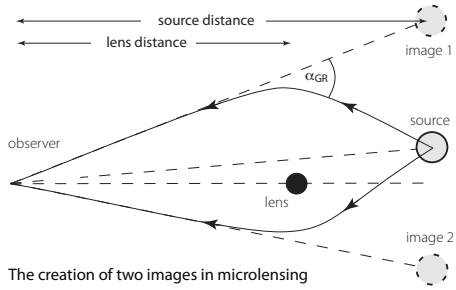
HST: ESA/NASA

Distorted, arc-like images of galaxies were reported by Lynds & Petrosian (1986). Mainly through later Hubble Space Telescope observations, many examples are now known. A first incomplete 'Einstein ring', resulting from almost perfect alignment, was reported by Hewitt et al. (1988), and dozens of more-or-less complete Einstein rings have been discovered since, including LRG 3–757.

HOW DO THESE CURIOUS structures arise? In general relativity, matter distorts spacetime, and the path of electromagnetic radiation is deflected as a result. With almost strict alignment, light rays from a distant background object (the source) are bent by the gravity of a foreground object (the lens) to create images which are distorted (and possibly multiple), and which may be highly focused and hence significantly amplified.

But these all rely on the chance alignment of a background source, an intervening lens, and an observer.

It is termed *strong lensing* if the effects are seen at an individual object level, or as *weak lensing* if it is only observed in a statistical sense.



The creation of two images in microlensing

Strong lensing can be further divided, somewhat subjectively (depending on the telescope resolution), into *macrolensing* (resulting either in multiple resolved images, or in 'arcs' in which the source is both sheared and magnified) and *microlensing* (in which discrete multiple images are essentially unresolved).

Relative motion between source, lens and observer leads to brightness changes, over hours or days for stellar alignments, or even over years for quasars. At peak amplification, the brightness of a background star might increase by several magnitudes over a few days.

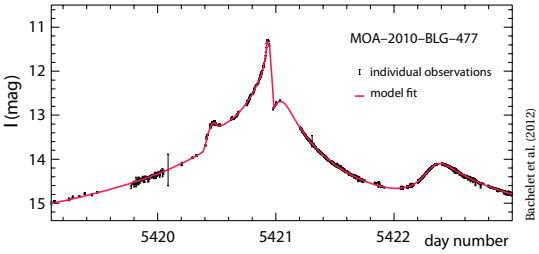
If the foreground lensing object is itself of complex structure, whether a cluster of galaxies, or a star orbited by one or more planets, then the background source may show a more complex light curve resulting from the time-varying magnification as the alignment changes.

THE EARLIEST microlensing surveys, in the 1980s–1990s, were motivated by the search for dark matter in galaxy halos, as probed by distant quasars. The observational challenges of observing these very rare and unpredictable events are enormous. Only since 1993, when massive observing programmes surveying millions of stars got underway, was microlensing observed by the EROS, OGLE, MACHO, DUO, and MOA projects.

To date, several thousand microlensing events have now been detected in the Galaxy. By the early 2000s, these surveys had excluded the possibility that the dark matter of our Galaxy's halo contained a significant contribution from massive objects of stellar mass.

With the constraints on dark matter largely resolved, observations turned to the search for exoplanets. The first, a planet three times the mass of Jupiter, was reported in 2004, and one just five times the mass of Earth was discovered in 2006. A two-planet system, in which orbital motion was measured during the lensing event, was observed in 2008.

By the end of 2020, more than 100 exoplanets had been discovered from the amplified light during the chance alignment of the system with a distant star.



Gaia is contributing to the study of microlensing events in fundamentally new ways. One is by providing the proper motion of the foreground object in these lensing events, which can be used to characterise the lens star.

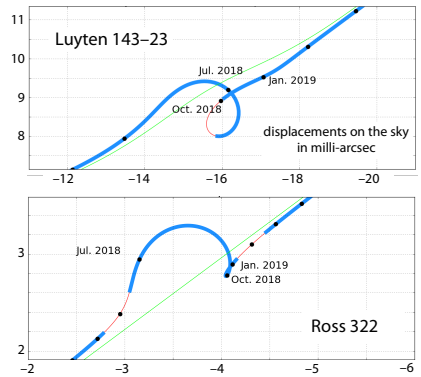
A very different application is to use the proper motions of stars in the vicinity of a rapidly moving star, to predict which might pass close to it on the sky months or years in advance, and to prepare for such lensed and magnified events in the future. Here, a number of papers have been published, using the Gaia DR2 data.

In addition to this *photometric* manifestation of microlensing, time-varying magnification of the unresolved microlensed images should also lead to a small time-varying shift of their photocentre, of up to around 1 milliarcsec. This tiny movement of a star image would be undetectable from the ground, but should be observable by Gaia, leading to mass estimates of the lens star accurate to 1%. Many studies of the effect and its detectability were made before the launch of Gaia.

The shift was first measured by Sahu et al. (2017), with Hubble. They measured a 2 milli-arcsec shift in the position of a background star as the nearby (52 pc) white dwarf Stein 2051B passed in front. Models yielded a white dwarf of 0.67 ± 0.05 solar mass – and further confirmation of the physics of degenerate matter.

The Gaia DR2 data was used by Klüter et al. (2018) to predict astrometric microlensing events by foreground stars with high proper motions, which will pass by a background star in the coming years. From a list of 148 000 stars with proper motions larger than 150 mas yr^{-1}, they searched for background stars close to their paths, calculating the dates and separations of closest approaches, and calculating the expected astrometric shifts and magnifications of the predicted events.

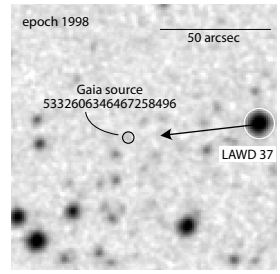
They detected ongoing events by two high proper motion stars. Luyten 143–23 had a predicted closest separation of 108 mas in July 2018 with a shift of 1.7 mas. It will pass by another star in March 2021 with a closest separation of 280 mas, and an expected shift of 0.7 mas. Ross 322 had a predicted separation of 125 mas in August 2018, and an expected shift of 0.7 mas.



Although the first of the Luyten fly-bys was not observed as part of the Gaia sky scanning, the other two were. Results are awaited!

Mustill et al. (2018) searched for potential lenses within 100 pc, using parallaxes and proper motions of the lenses and background sources, then calculating peak magnifications and displacements. They found seven possible events that will occur before 2035. Of particular interest is a 14.9 mag star, which will lens a 13.9 mag background star in early 2030.

All 1.7 billion stars in GDR2 were searched by Bramich (2018), who predicted 76 microlensing events between July 2014 and July 2026. Nine of these will be caused by the white dwarf LAWD 37, and another five by the white dwarf Stein 2051 B (measured by Sahu with the Hubble Space Telescope, and also predicted by McGill et al. (2018). Other estimates have been made by McGill et al. (2020).



The motion of LAWD 37

It is too early to estimate how many events will finally be measurable by Gaia. Meanwhile the method also has applicability to more exotic lens types.

Ofek (2018) identified two candidate events of Gaia DR2 stars involving lensing by a foreground pulsar (with known proper motion) in which the shift of the background star will exceed 10 micro-arcsec.

Rybicki et al. (2018) investigated the impact of combining Gaia astrometry from space with precise, high cadence OGLE photometry from the ground. For the archival event OGLE3-ULENS-PAR-02, which is likely a black hole, they predict that at the end of the nominal 5 yr of the Gaia mission, for the events brighter than 15.5 mag at the baseline, caused by objects heavier than 10 solar masses, it will be possible to derive lens masses with accuracies of around 10%.

12. Multiple-planet mandalas

THE IDEA THAT THE Earth was at the centre of the Universe, and that the Sun and planets were all in orbit around it, was central to the views of the ancient Greeks. But in this framework, the motions of the planets (being the Greek for 'wanderers'), was impossible to fathom.

Only with the adoption of the Copernican view, that the Sun was at the centre of the solar system, and that the planets (including Earth) moved in orbit around it, did their motions become possible to comprehend.

The very complex apparent motions of the planets, as seen from Earth, could eventually be explained by the combination of the Earth moving in an elliptical orbit around the Sun, and the other planets orbiting with slightly different eccentricities. This is because elliptical orbits, with the Sun at one focus, are 'permitted' types of orbit under Newton's inverse square law of gravity.

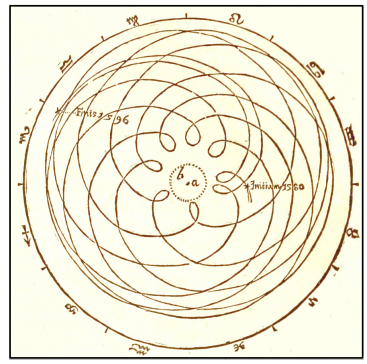
MORE PRECISELY, planets orbit the *centre of mass* of our solar system, rather than orbiting the centre of mass of the Sun itself. In many case, this (often very small) difference may have no obvious effect. But it will become relevant with high-accuracy measurements.

Let us first conisder an analogy to get a clear picture of what this might mean. Imagine two ice skaters, holding hands, and twirling around each other. If one is particularly bulky, and their partner very slight, it may almost appear that the latter is circling the former. But if our two skaters are of equal weight, it should be obvious that both circle around their common centre – their centre of mass (their 'barycentre', in scientific speak).

The same is true of a system of one or more planets orbiting a star. If the star is very massive and the planets are very small, it may seem that the latter are strictly orbiting the star. But imagine the planets becoming more and more massive, and it again might be more obvious that everything – the planets and the star itself – are orbiting their overall centre of mass.

IN 1609 Johannes Kepler published his *Astronomia Nova* (A New Astronomy), including his analysis of the orbit of Mars from 10 years of observations.

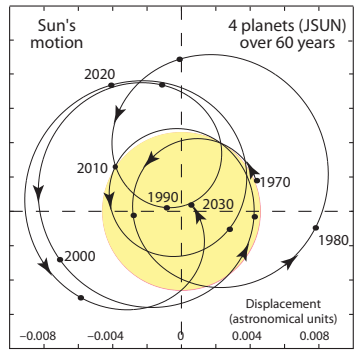
In late 1604, and after various attempts at describing its orbit around the Sun, he at last hit upon the idea of an ellipse, which he had previously assumed to be too simple a solution for earlier astronomers to have overlooked. Finding that an elliptical orbit fitted the Mars data, Kepler concluded that all planets move in ellipses, with the Sun at one focus (his first law of planetary motion). His diagram of the orbit of Mars, seen from Earth, appeared in *Astronomia Nova*, and confirms the remarkable accuracy of his observations.



Kepler's study of the orbit of Mars

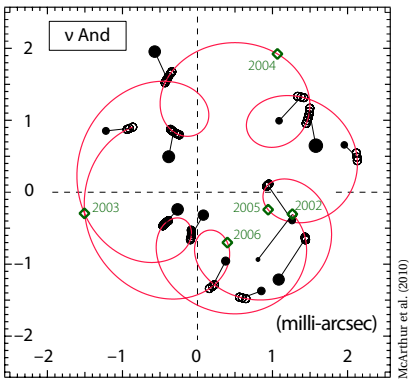
Our Sun's own path over decades reflects the combined gravitational effects of all solar system objects.

Indeed, as already noted by Isaac Newton, the actual motion of the Sun about the solar system barycentre is rather complex *'since that centre of gravity is continually at rest, the Sun, according to the various positions of the planets, must continuously move every way, but will never recede far from that centre.'*



Looking down on the plane of the solar system, this figure shows the effect of the four most massive planets (Jupiter, Saturn, Uranus, and Neptune) over 60 years, from 1970–2030. The Sun itself is tugged continuously, smoothly but somewhat erratically due to the different planetary periods. We can see that the Sun often moves by more than its own radius over long periods of time.

THIS GRAVITATIONAL TUGGING on the motion of a host star will exist for any star orbited by planets. If the orbit is viewed edge-on, the effect will be most noticeable in the Doppler shift (radial velocity) of the star's spectrum. Viewed face-on to the orbital plane, the star will move in the same way as we have seen for the Sun.



The orbit of υ And, from Hubble Space Telescope

A single orbiting planet moving in an elliptical orbit will cause the host star to move in a scaled-down elliptical orbit, mirroring the planet's motion.

Multiple planets lead to a more complex star motion over time, dependent on the number of planets, their masses, and their orbital periods.

The bright star *υ* And was one of the first known multi-planet systems, discovered from long-term monitoring of it radial velocity, and now known to be orbited by at least three planets. Observed with the Hubble Space Telescope a number of times over four years, McArthur et al. (2010) were able to reconstruct the orbit of the host star (shown here as the red curve) from the various individual positional observations (solid dots).

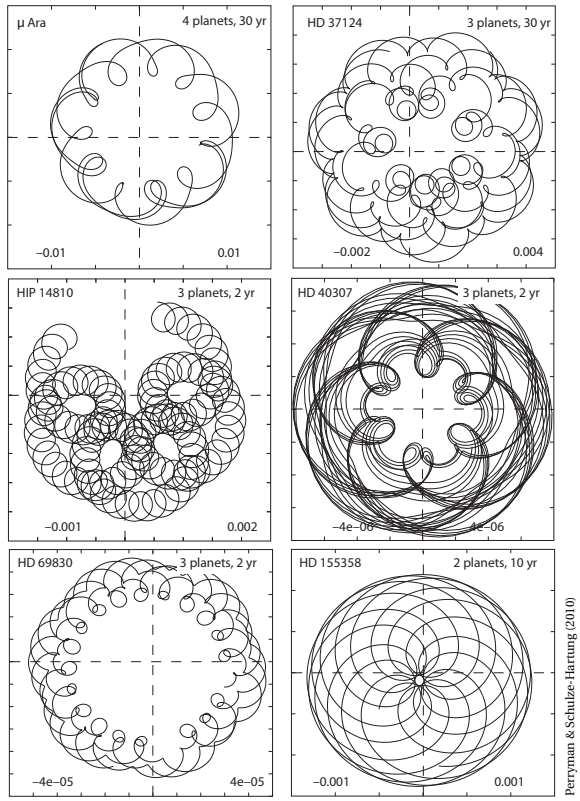
THE TYPE OF positional accuracy needed to make this kind of astrometric observation has not been routinely available so far. But this is the type of observation that is being made, in enormous numbers, by Gaia.

To give an indication of the expected yield, the orbits of just a handful of systems have been measured by the Hubble Space Telescope. By 2020, Gaia will have already acquired around 100 positional observations for perhaps 30 000 exoplanet systems. The results are not available for study, because much more work must be done to calibrate, validate, and interpret the results before they can be made available. Perhaps they will be available around 2025.

But meanwhile, we can predict what the results will look like. We can rather easily take the information that is known about each planetary system, in particular the mass of the host star, and the masses and orbits of the planets that have so far been discovered around them.

Together these will define the path on the sky that the host star will follow over time. Gaia will measure some 100–200 data points on these curves over a total interval of 5–10 years, and in many cases the smooth path of the star can then be reconstructed.

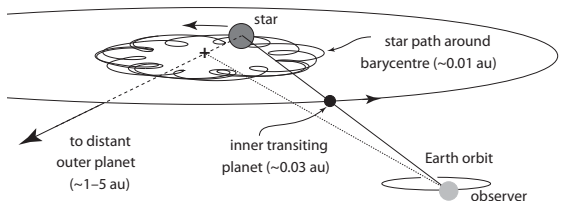
Some examples of the complex star paths predicted for real multi-planet systems are shown in the figure. These types of beautiful curve, drawn out by Nature, have been named *planet mandalas*, after the Sanskrit for circle.



Astrometric orbits expected for some known planetary systems

THIS KIND OF DIAGRAM allows us to visualise the motion of the host star in a multi-planet system. Viewed edge-on, we can see how the transit times of an orbiting planet will be not be regularly spaced at exactly the planet's orbital period, but instead modulated by the star's motion as affected by other planets in the system.

These sorts of 'transit timing variations' are seen frequently in the accurate Kepler transit data, and allow a wealth of new information about the system to be derived, including the mass and orbital periods of any non-transiting planets that might orbit the same system.



13. The distance to the Pleiades

MEASURING DISTANCES across the vast scale of our Galaxy and beyond has long been a central problem in astronomy. Only for objects within a few tens of parsecs has direct distance determination, through parallax measurements, been possible from the ground.

Beyond that, too far away for distances to be determined directly, estimates have had to rely on a distance scale 'ladder', constructed from a sequence of indirect and often uncertain measurements relating the closest objects to increasingly distant ones.

Open clusters have long been a crucial step in this sequence of distance estimates.

OPEN CLUSTERS are groups of several hundred stars that were formed from the same giant molecular cloud and have roughly the same age. They are still being formed in our own Galaxy, at an estimated rate of one every few thousand years.

More than a thousand have been discovered within our Galaxy, and many more are thought to exist. A few, such as the Hyades (at about 45 pc), the Pleiades (at about 130 pc), and the Alpha Persei cluster (at about 175 pc), are visible with the naked eye.

Open clusters are key objects in the study of stellar evolution. Because the cluster members are of similar age and chemical composition, their properties (such as distance, age, metallicity, extinction, and velocity) are more easily determined than they are for isolated stars.

ALTHOUGH LOOSELY bound by mutual gravitational attraction when they formed, open clusters slowly disperse as gas (and therefore mass) is stripped by the radiation pressure of their hot young stars. They are further disrupted by close encounters with other cluster stars, and other external structures, as they orbit the Galaxy. Open clusters may survive as recognisable clumps for a few hundred million years, with the most massive clusters surviving for a few billion years.

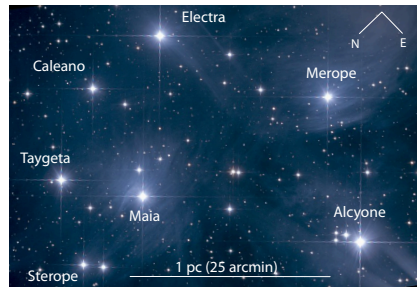
As stars slowly escape the gravitational field of the cluster, many will still be moving through space on a roughly similar path around the Galaxy, in what is known as a moving cluster or moving group.

DISTANCES TO THE nearest open clusters, notably the Hyades and the Pleiades, have long been used to calibrate the distances to others more remote, notably by matching their main sequences on the Hertzsprung–Russell diagram. And clusters containing luminous Cepheid variables can be used to calibrate the Cepheid period–luminosity relationship, allowing them to be used as standard candles out to more distant galaxies in the Local Group.

Determining the distances to these clusters has its own long history. For the Hyades this gathered pace with the work of Lewis Boss (1908) whose opened his study with the words: *'The phenomenon of neighbouring stars moving athwart the sky with motions of the same order of magnitude and in sensibly parallel directions has been noticed many years ago. It has been demonstrated that the greater part of the stars in the Pleiades are moving in this manner.'* And over the next 30 years, other big names tackled the problem, including Jacobus Kapteyn, Willem de Sitter, Henry Plummer, and Ejnar Hertzsprung.

Fast forward 60 years (and dozens of papers), and Hipparcos allowed the distance to the Hyades to be established more securely, with the cluster centre being estimated at around 46 parsec (Perryman et al., 1998).

THE PLEIADES, in the constellation Taurus, has been recognised as a group of stars since antiquity, and has long played an important role in establishing the distance scale.



The cluster contains more than 1000 members, with a total mass of about 800 solar masses. Its light is dominated by young, hot blue stars, some dozen of which can be seen with the naked eye. The cluster has a 'core radius' of about 2–3 pc, and a 'tidal radius' of about 13 pc .

UNTIL THE Hipparcos measurements of its trigonometric parallax were published, various alternative and somewhat less-direct methods had all converged on a Pleiades distance of about 130–135 pc.

But, in 1999, the results from Hipparcos yielded an unexpected result. The technique that should have provided the most accurate results to date, viz. measuring stellar parallaxes from space, gave a mean distance of only 118 pc, 7% 'shorter' than the previous consensus.

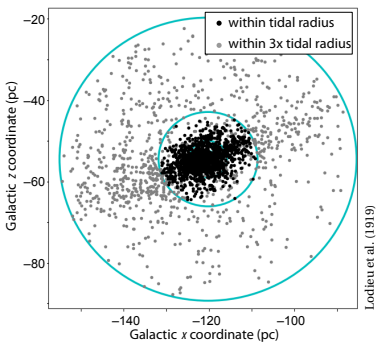
Later work, amongst them optical interferometric observations of the binary star Atlas, main-sequence fitting in the infrared, Hubble Space Telescope Fine Guidance Sensor observations, VLBI observations, and others, consistently argued that the Hipparcos distance for the Pleiades must be erroneous.

But the problem did not disappear: a detailed reanalysis of the Hipparcos data a decade later led to only a small revision, from 118 to 120 pc (van Leeuwen, 2009).

WHAT DO the Gaia results have to say on the subject? The first Data Release, DR1 in 2016, included the Tycho–Gaia Astrometric Solution (TGAS), a subset of about 2 million stars incorporating the Hipparcos and Tycho-2 positions centred at 1991.25. Their parallaxes for 152 Pleiades members gave a mean distance of 133.7 ± 0.6 pc, agreeing with other non-Hipparcos values of the distance (Gaia Collaboration et al., 2017).

And their results for 19 open clusters, ranging from the Hyades (at just under 47 pc) to IC2422 (at nearly 440 pc) gives a good agreement with a similar study using the Hipparcos data – with the one exception of the Pleiades cluster, which remains unexplained.

THE SAME prescription was used by Abramson (2018) to estimate the cluster's distance using the new Gaia DR2 data. From 1594 cluster stars, an order of magnitude more than available for the DR1 analysis, he derived a mean distance of 136.2 ± 5.0 pc.



Stars escaping from the Pleiades

A more rigorous treatment of cluster members in DR2 allowed Lodieu et al. (2019a) to identify 1248 members, yielding a mean distance of 135.15 ± 0.43 pc. They estimated the cluster's age, of 132 ± 26 Myr, from its single white dwarf LB 1497. And they detected a stream of stars escaping from the cluster, extending up to 40 pc from its centre.

A separate analysis of DR2 found 1454 cluster members, and a mean distance of 136.0 ± 0.1 pc (Gao, 2019).

Recent Pleiades distances (Hipparcos and Gaia highlighted)

Year	Distance (pc)	Technique	First author
1999	118.3 ± 3.5	Hipparcos	van Leeuwen
2004	132.0 ± 4.0	binary star (Atlas)	Zwaalen
2005	133.8 ± 3.0	infrared main-sequence	Percival
2005	134.6 ± 3.1	HST–FGS	Soderblom
2007	122.2 ± 2.0	Hipparcos re-reduction	van Leeuwen
2009	120.2 ± 1.9	Hipparcos re-reduction	van Leeuwen
2014	136.2 ± 1.2	VLBI	Melis
2017	133.7 ± 0.6	Gaia Data Release 1	van Leeuwen
2018	136.2 ± 5.0	Gaia Data Release 2	Abramson
2019	135.2 ± 0.4	Gaia Data Release 2	Lodieu
2019	136.0 ± 0.1	Gaia Data Release 2	Gao

TAKEN TOGETHER, it is now clear that all methods, whether ground- or space-based, converge on one agreed mean distance to the Pleiades cluster, of about 136 pc, with the collective exception of the various Hipparcos estimates, which placed it significantly, but erroneously, some 10–15 pc nearer.

The papers cited here have much more to say about the Pleiades cluster than just its distance. Gaia reveals, for the first time, the cluster's depth, allowing each star to be pinpointed in space with respect to its centre. It allows us to recognise stars that are slowly escaping from the cluster, gradually merging into the background population over millions of year. And it clarifies its age and stellar content from its Hertzsprung–Russell diagram.

WHILE THE PLEIADES distance provided, by far, the largest controversy in the Hipparcos results, and while it remains an important problem to resolve, I will make a couple of observations.

At 135 pc distance, the mean parallax of the Pleiades stars is around 7.4 mas. Hoping to establish the distance with a 10% error was always going to be at the limit of the Hipparcos measurement accuracies. Compounding the problem, the stars are also much closer together on the sky than the satellite was ever designed to probe.

The Hipparcos distance of the Pleiades was an 'outlier' compared to the other cluster distances. And even before Gaia, the cosmic distance ladder could be weighted to other nearby clusters where a better consensus existed.

Today, the Gaia parallaxes have been pushed to such great distances that the Hyades and Pleiades, for many decades key steps in unravelling the cosmological distance scale, are no longer relevant in this endeavour – a century of ingenuity and experimental perseverance have been consigned, at a stroke, to historical curiosity.

Open clusters will remain of great importance for dynamical and evolutionary studies of stellar systems, and it would be good to see the Hipparcos parallax of the Pleiades more definitively resolved. Meanwhile, it has surely made its final appearance in the lengthy and proud history of the distance scale in astronomy.

14. Testing modified gravity

SINCE ITS FORMULATION AND experimental confirmation in the early years of the 20th century, Einstein's theory of general relativity has been widely accepted as providing the best known description of gravity, on all spatial scales. However, various observations, starting with the flat rotation curves observed in most spiral galaxies, but now embracing the existence of large-scale structure in the Universe, demand an additional non-visible 'dark matter' component to fit the data.

At the present time, and despite much experimental effort, there has been no decisive detection of dark matter. This leaves open the possibility that some modified theory of gravity might explain these perplexing observations without this hypothetical dark matter.

MODIFIED NEWTONIAN DYNAMICS (MOND) is a theory originally proposed by Milgrom (1983) which attempts to account for these long-range gravitational effects without invoking dark matter.

Subsequent years have seen several developments on the theoretical side, notably the incorporation of MOND into more generalised theories of relativity. Meanwhile, the most straightforward versions of the theory have been ruled out by more rigorous observations, notably using precise timing effects in the neutron-star merger GW 170817 (Boran et al., 2018).

A convincing detection of dark matter would settle the question. But in the absence of such a dark matter detection, new tests which can discriminate between dark matter, and modified gravity, are highly desirable. One such family of tests probes the observational effects of gravity under conditions of very low accelerations.

PREVIOUS WORK on tests of MOND-like gravity over the past decade has hinted at deviations of the form expected from a MOND-like gravity. But the observations available to date have been of insufficient quality to conclude one way or another.

Gaia was expected to provide much improved prospects for such a test, based on the orbital behaviour of a number of very wide-separation binary stars.

For separations larger than about 5000 times the Sun–Earth distance (i.e. 5000 astronomical units), the stars in such a system have sufficiently small orbital accelerations – below about $10^{-10}\,\mathrm{m\,s^{-2}}$ – to provide a direct probe of MOND-like theories. This minuscule acceleration can be compared to that experienced by a body at the Earth's surface, of about $9.8\,\mathrm{m\,s^{-2}}$.

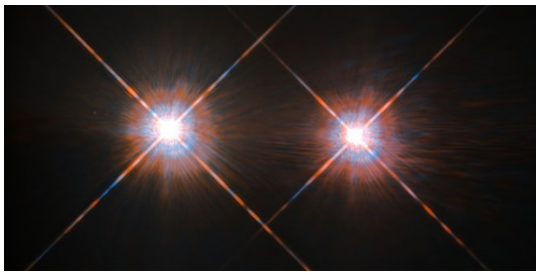
Several studies have already made use of the Gaia DR1 and DR2 data to make a start on the problem (amongst them El-Badry, 2019; Hernandez et al., 2019; and Pittordis & Sutherland, 2019).

BINARY, AND occasionally triple or even higher multiplicity star systems, form – over a range of separations – in the swirling gas clouds of dense regions of the interstellar medium. These can be density enhancements triggered by the passage of our Galaxy's rotating spiral density waves. If a binary forms with a small separation, their orbits can slowly spiral inwards until they eventually coalesce. Wider separation binaries can be slowly torn apart by external gravitational forces.

Systems with extremely short orbital periods, of days or even hours, are well known and widely studied. And there are various ideas of how very wide binaries can form. But there is no clear picture of how far apart binaries can survive as a gravitationally bound pair, i.e. while still maintaining a stable orbit around each other.

In our solar neighbourhood, Alpha Centauri A and B orbit each other every 80 years. They are separated by about 11 times the distance between the Sun and Earth (i.e. 11 astronomical units) at their closest approach. Many binary star systems are known with a much wider separation, yet still remaining gravitationally bound.

BACK IN 1937, Armenian astronomer Victor Ambartsumian calculated that a very wide binary, with its very weak gravitational bond, rarely breaks apart due to a single close encounter with another star, but rather as a result of numerous distant passages that each gently pull on the binary until it slowly evolves from being bound to being unbound.



Alpha Centauri A and Alpha Centauri B

Thus, an ultra-wide binary with a separation of 0.5 parsec (1.6 light-years, or 100 000 astronomical units) is likely to break up within about 100 million years. A binary with a separation of 0.1 pc (0.3 light-years) might survive for more than a billion years. At these enormous separations, two stars will be widely separated on the sky, with very long orbital periods, but sharing an almost identical space motion over millennia.

How then can two stars be recognised as a physical binary? Close binaries will generally be unusually close together on the sky, much closer than the average star density on the sky would imply. Careful monitoring of the their space motions or radial velocities over years or decades can hope to reveal their orbital motion, thus confirming their gravitational coupling.

But how can a *widely separated* binary be distinguished from two completed unrelated stars? Any orbital motion of a slowly orbiting wide-separation binary would require extremely accurate measurements to recognise. In other words, the wider a binary is, the more difficult it is to identify – and this has been a major barrier to discovering wide binaries in the past.

GAIA IS IN THE PROCESS of identifying many thousands of very wide and ultra-wide binaries from their highly accurate space motions. Their properties will help to determine the most likely mechanism responsible for their formation, and their proper classification will allow for the sorts of tests necessary to confirm, or discard, MOND-like theories of gravity.

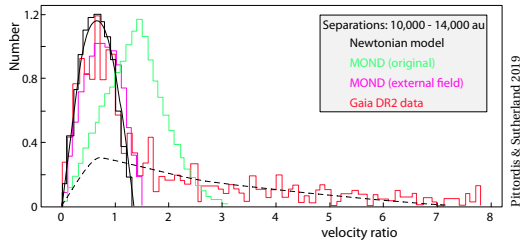
RETURNING, THEN, TO the tests of gravity, it turns out that the original MOND-like models result in relative orbital velocities which are significantly different to those predicted by Newtonian models, specifically they can allow bound binaries with relative velocities well above the Newtonian 'ceiling'.

But MOND models including what is called an 'external field effect' (which are also preferred in present theories) give predicted relative velocities much closer to Newtonian, but do still show subtle but well-defined deviations in both their true space velocities, as well as in their projected velocities on the plane of the sky, as measured via the Gaia proper motions.

A high-velocity tail, well above the Newtonian prediction, could provide evidence favouring such a modified theory of gravity. And various studies suggested that there is also an optimal window of projected separation, between about 5–20 000 au, for the practical application of such a test.

Using the Gaia DR2 catalogue, Pittordis & Sutherland (2019) selected stars within 200 pc of the Sun, and bright enough ($G < 16$ mag) to give a good accuracy on the astrometry. They then selected pairs of stars with (projected) separations up to 40 000 astronomical units, using both the parallax and proper motion measurements to sift out nearly 25 000 plausible physical pairs.

They then calculated their likely orbits according to the various models (Newtonian, and MOND with and without an 'external field effect'), characterising them through some appropriate velocity ratio metric.



Gaia wide binaries compared with theoretical models

The figure shown here is just part of their findings, covering only the subset of binaries with projected separations between 10 000–14 000 au. From the main peak of the observed histogram (in red), they found that the original MOND model (green) provides a very poor match to the observed data, while the inclusion of the 'external field effect' (purple) works much better.

The biggest surprise was a very long tail in the velocity ratio, across all of the separations observed. This long tail makes it impossible to decide between a Newtonian model, or a MOND model with an external field.

Pittordis & Sutherland (2019) found that this tail can be explained by pairs of stars which were born in the same open cluster, but which are currently undergoing a chance close 'flyby'. Clarke (2020) suggested that it can be attributed to a population of hidden triple systems.

In a similar study using Gaia DR2 data on 81 wide binaries, Hernandez et al. (2019) also found results consistent with Newtonian predictions below 7000 au, but inconsistent with it at larger separations.

WHILE THE PRINCIPLES of this sort of test have been demonstrated, the Gaia DR2 data are insufficient to rule on the reality or otherwise of a modified MOND-type gravitational field. The Gaia DR3 data will allow further advances. Meanwhile, it is clear that the Gaia data will have much to say about the formation – and eventual demise – of very wide-separation binary stars.

15. The Enceladus stream

THE PAST two decades have witnessed a major advance in our understanding of galaxy formation. Present theories, supported by extensive numerical simulations, argue that large galaxies, such as our own Milky Way, have been built up from a series of mergers with smaller galaxies over the past several billions of years.

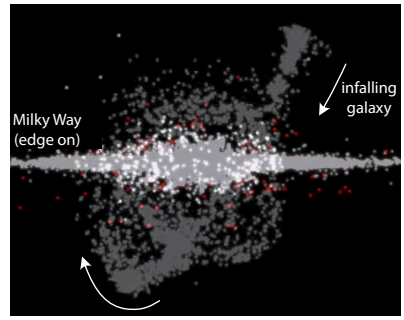
Studies of the structure of our Galaxy identify it as comprising a massive spherical central bulge, and an extended rotating stellar disk in which a few spiral arms (major sites of ongoing star formation) are embedded. On more careful examination, the disk comprises two rather distinct populations: the dominant 'thin disk', some 300 pc in vertical extent, and a more extended 'thick disk'. All of these sit within a vast, more diffuse, and largely spherical halo. This halo is dominated by unseen and still mysterious 'dark matter'. Over cosmic history, these dark matter halos have been the centres of galaxy formation, and of star formation within them.

CONTROLLED BY long-range gravitational forces, disk stars rotate around the Galactic centre with an orbital period of some 250 million years. They also 'bounce' up and down about the Galaxy's mid-plane with a period of 80 Myr as they rotate. Stars making up the visible component of the halo are generally 'metal-poor' (astronomer-speak for low in elements heavier than H and He), these stars having formed when the Universe was itself very young. Halo stars move on much more extended, circular orbits. If they happen to pass through our solar neighbourhood on these vast orbits, they will be moving at relatively high speeds as a result.

Barnard's Star is one such fast-moving, metal-poor 'halo-like' star, passing swiftly through our solar neighbourhood. Some 10–12 billion years old, compared to our Sun's 4.5 Gyr, it is amongst the oldest stars in the Milky Way. It is moving through space at about 90 km s^{-1} and, because of its proximity at only 1.8 pc (6 light-years), has the largest of all known proper motions. This large angular motion across the sky, at around 10.3 arcsecond per year, leads to its displacement of about a quarter of a degree over a human lifetime, i.e. roughly half the angular diameter of the full Moon.

MODELS OF structure formation suggest that our Galaxy's inner stellar halo should be dominated by the debris of just a few massive progenitor galaxies merging with our own early on in its formation history. This explanation finds convincing support with the new Gaia data, where there is compelling fossil evidence for one such merger event which took place around 10 Gyr ago, and whose debris in the solar neighbourhood reveals much about this enormous and disruptive event.

Before looking at the details, let us take a look at the bigger picture. How is it possible to piece together evidence for such an ancient event? To visualise the principles, the figure shows an edge-on schematic of our Galaxy's disk, and a less massive galaxy falling onto it from the upper right.

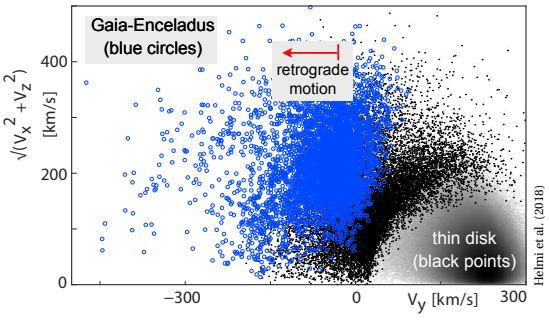


The lower mass incoming galaxy is captured by the gravity of the more massive galaxy, and leads to a stream of stars which becomes progressively more torn apart as it encircles and merges with the cannibalising galaxy. Over the subsequent few billion years, over several orbits, the disrupted stellar stream becomes progressively more mixed into the stars of the more massive host. But it is not difficult to imagine that these infalling stars might still be recognisable as interlopers if their parent galaxy had distinctive chemical signature.

It is perhaps less obvious, but nevertheless crucial to the recent discoveries, that these stars are likely to retain some common and inherited features of their orbital motion around the more massive galaxy which has entrapped them. Specifically, although after a few billion years, and after several disruptive orbits, they end up being well-spread throughout space, their angular momentum around our Galaxy's centre retains some memory of their original orbital motion.

EARLIER STUDIES WHICH considered both the stellar chemistry (i.e. the elements which dominate their spectra) and the dynamical motions of the nearby halo stars together have hinted at the presence of such infalling stellar streams and clumps, and of correlations between the stars' chemical abundances and their orbital parameters. Indeed, the publication of the Hipparcos star catalogue in 1997 allowed the discovery of one such 'disrupted event' identified from this type of correlated space motion (Helmi et al., 1999).

More recently, analysis of data from the Sloan Digital Sky Survey, and from Gaia DR2, has revealed the presence of two distinct sequences in the colour–magnitude diagram, and of a prominent kinematic structure, all in the nearby halo. These may well be traces of an important accretion event experienced by the Galaxy. Interestingly, this kinematic feature is slightly 'retrograde', that is, the stars within it are moving around our Galaxy in the opposite sense to the bulk of the disk and halo stars.

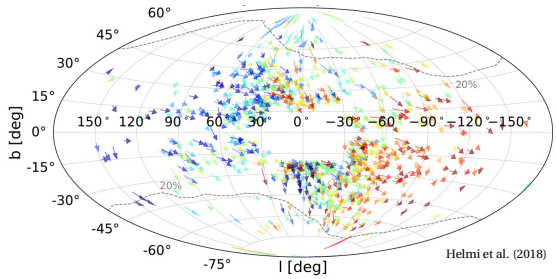


The Gaia–Enceladus stream in velocity space

The type of structure observed in the joint SDSS and Gaia DR2 data nicely confirms predictions from cosmological simulations, namely that this type of substructure in our solar neighbourhood is most apparent among the fastest moving stars, typically reflecting more recent accretion events (Koppelman et al., 2018).

THE GAIA DR2 DATA was then used to study the kinematics, chemistry, age and spatial distribution of stars in a relatively large volume around the Sun. As we have seen, this volume samples two major Galactic components, the disk and the stellar halo.

This study, by Helmi et al. (2018) and reported in the Journal *Nature*, demonstrated that the inner halo is dominated by debris from an object which at infall was slightly more massive than the Small Magellanic Cloud, and which the authors referred to as Gaia–Enceladus. In Greek mythology Enceladus was one of the Giants (Titans), the offspring of Gaia (representing the Earth), and Uranus (representing the Sky), buried under Mount Etna and held to be responsible for earthquakes in the region (the specification as Gaia–Enceladus avoids confusion with the Saturnian moon of the same name).



Gaia–Enceladus: proper motions, coloured by radial velocity

The stars originating from the Gaia–Enceladus accretion event cover nearly the entire sky, and their motions reveal the presence of streams and slightly retrograde and elongated trajectories. Hundreds of RR Lyrae stars and thirteen globular clusters following a consistent age–metallicity relation can be associated to the merger on the basis of their orbits.

With an estimated 4:1 mass-ratio between the young Milky Way and Gaia–Enceladus, the merger would have led to the 'dynamical heating' of the precursor of the Galactic thick disk, increasing their space velocities, and therefore increasing their scale height with respect to the Galaxy mid-plane. It seems highly plausible that the merger between Gaia–Enceladus and the Milky Way contributed to the formation of our Galaxy's thick disk component some 10 Gyr ago. Most probably, this was the last significant merger that our Galaxy experienced.

THE FINDINGS are in line with simulations of galaxy formation, which predict that the inner stellar halo should be dominated by debris from just a few massive progenitors. But the agreement goes further.

Amongst these very large-scale cosmological simulations, the EAGLE project has been shown to produce a realistic population of galaxies reproducing a broad range of observed galaxy properties. The largest of the EAGLE simulations, L100N15043, has a cubic volume of 100 Mpc in size, and includes the effects of both baryonic and dark matter. Its huge volume of simulated space–time provides numerous Milky Way-type galaxies, and with a wide range of merger histories.

Amongst these mergers, Bignone et al. (2019) identified one with remarkably similar properties to the Gaia–Enceladus event, also occurring around 9 Gyr ago. These specific simulations result in merger debris on a slightly retrograde orbit (as found for Gaia–Enceladus), bursts of star formation in the early disk, the formation of a dynamically heated thick disk (as seen in our Milky Way), and with a large fraction of the debris deposited at large heights above the Galactic disk, corresponding to our Milky Way's stellar halo.

All-in-all, a most remarkable triumph of state-of-the-art space observations combined with state-of-the-art cosmological simulations!

16. Quasars, as seen by Gaia

QUASARS ARE the highly luminous nuclei of distant active galaxies. Their light is dominated by emission from an accretion disk as matter falls towards a central supermassive black hole, millions to billions times the mass of the Sun. The radio source 3C 273 was the first whose redshift, $z = 0.158$, was determined in 1963.

There are more than a million quasars known today, many discovered from the Sloan Digital Sky Survey. The most distant are at redshifts above $z = 7.5$, meaning that they date from a 'mere' 700 Myr after the Big Bang. Quasar activity was more common in the distant past, peaking at around 10 billion years ago.

THE MOST IMPORTANT role that quasars have in Gaia is in their direct manifestation of the global reference system. Their enormous cosmological distances means that their individual proper motions are effectively zero.

More than 500 000 are being observed by Gaia. Reasonably well distributed across the sky, they serve to define a highly accurate inertial reference system – a subject that I cover in more detail elsewhere.

MANY QUASARS are also of individual interest, for example because of their use as probes of cosmological evolution and structure formation. Many are radio sources, some show outflowing 'jets' that appear to be superluminal due to relativistic effects and line-of-sight orientation, and a few occur in physical groups.

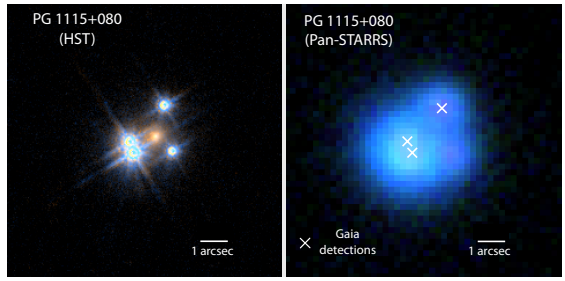
Several hundred show multiple images as a result of gravitational lensing, and these can be used for detailed studies both of the source itself, and of the lensing galaxies. Source variability results in measurable time delays between the separate images, and these can be used to determine the Hubble constant, today reaching a precision of just a few percent (e.g. Bonvin et al., 2016).

On-board sampling of Gaia's CCDs provides source images with an ultimate resolution of about 0.1–0.2 arcsec in the scanning direction (Gaia's astrometric accuracy is much better than this because the image centroid can be determined much more accurately than this effective resolution). This angular resolution is much better than can be routinely measured from the ground.

Accordingly, Gaia should be able to discover lensed quasars with image separations below about 1 arcsec. Small-separation lenses are the most common, and can be used to probe high-redshift/low-mass lensing galaxies. But they are more difficult to find from the ground.

Lemon et al. (2017) demonstrated this discovery potential with Gaia DR1, showing that Gaia correctly identified the three brightest images in the previously-known 5-image system PG 1115+080. Indeed, out of 49 known lensed quasars with image separations below 2 arcsec, they recovered 8 from Gaia DR1, discovered four new lensed systems with sub-arcsec separation, and estimated that Gaia would eventually be able to discover some 1400 with image separations above 0.5 arcsec.

A further 24 gravitationally lensed quasars were similarly discovered and characterised using the DR1 data by Lemon et al. (2018), and an additional 22 using the Gaia DR2 data by Lemon et al. (2019).



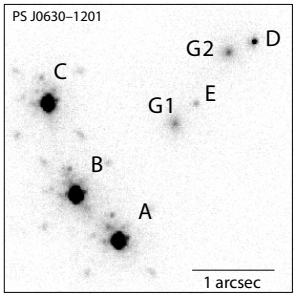
Lemon et al. (2017)

A DIFFERENT APPROACH was taken by Krone-Martins et al. (2018) with Gaia DR2. From the various compilations of known or possible quasars, they constructed a starting list of 3 112 975 objects, of which 1 839 143 have a Gaia DR2 counterpart within 0.5 arcsec of that starting position. They then matched these with other Gaia DR2 detections within a radius of 6 arcsec.

Applying other astrometric and photometric tests, as well as physical modelling, they finally extracted two new high-reliability quadruple-lens candidates, GraL 113100–441959 and GraL 203802–400815 (GraL designating a gravitational lens system). Five previously known lensed systems were also re-discovered.

Further examination of Gaia DR2 by Ducourant et al. (2018) showed that out of 481 known multiply imaged quasars at that time, 206 had at least one image in Gaia DR2. Among the 44 known quadruply-imaged systems, 29 had at least one image in DR2. For twelve of these, all 4 components were found in DR2, while eight have 4 components, eight have 2, and one has only 1.

Their physical modelling of the quadruple system HE 0435–1223 shows that models are much better constrained when using Gaia astrometry, in particular for the relative positions of the background quasar and the lensing object. It follows that more detailed modelling will benefit from Gaia's sub-milliarcsec astrometry.



Ostrovski et al. (2018)

Gaia DR1 data was used in the discovery and modelling of the five-image quasar PS J0630–1201, at $z = 3.34$ (Ostrovski et al., 2018). Four of the images, ABCD, lie in a canonical 'cusp' configuration, of which A, B, and C were detected and flagged by Gaia, with G magnitudes of 19.95, 19.76, and 19.61 respectively. In addition, Keck near-infrared imaging revealed two lensing galaxies, G1 and G2, and an additional point source E. The relatively bright fifth image raises the possibility of measuring a total of ten time delays, all predicted to lie in the range of 1–245 days.

Of course, brightness variations, whether caused by physical effects intrinsic to the quasar (notably source variability), or extrinsic (specifically gravitational lensing), all have an effect on the apparent position of the quasar, and can affect linking of the Gaia reference frame to an inertial one. Details of these and related effects have been given by Bachchan et al. (2016), and many other studies have been carried out subsequently.

THE GAIA DATA is being widely used to re-visit earlier compilations of quasars, and possible quasars, using the new high-accuracy proper motions or colours to validate or refute such candidates.

This has been carried out using Gaia DR2 applied to the 190 000 quasar candidates in the Kilo-Degree Survey Data Release 3 (KiDS DR3) by Nakoneczny et al. (2019).

Cross-matching with Gaia DR2 similarly assisted in the fifth release of the Large Quasar Astrometric Catalogue (LQAC–5), resulting in a list of 592 809 objects with 398 697 Gaia counterparts (Souchay et al., 2019).

Gaia DR2 data was also used in the finalisation of the Sloan Digital Sky Survey IV quasar catalogue from Data Release 16 of the extended Baryon Oscillation Spectroscopic Survey (eBOSS). Their 'quasar-only' subset contains 750 414 quasars, and is estimated to be 99.8% complete, with 0.3–1.3% contamination (Lyke et al., 2020).

THE HUGE quantity of data acquired by Gaia each day, around 40–50 Gbytes, and the vast numbers of sources observed each day, means that the data processing on ground has to rely on semi-automated methods of object classification and parameter estimation.

An insight into results expected from Gaia DR3 was given by Delchambre (2018). He used the blue (B_P) and red (R_P) spectra from the satellite to determine, through supervised machine-learning, the astrophysical parameters of quasars, including estimates of their redshift, their continuum slope, and the total equivalent width of their emission lines.

More details of the classification using Gaia DR2 data alone are given by Bailer-Jones et al. (2019).

PRE-GAIA, extremely luminous quasars were not easy to discover, because high-redshift candidates are heavily outnumbered by nearby stars, these contaminants being typically of low mass and temperature. As a result, spectroscopic follow-up was often biased against the brightest and most interesting candidates.

Using Gaia astrometry and photometry, the main contaminants can be recognised and rejected, and true quasars identified as red objects (in $B_p - R_p$) at very large distances, i.e. with proper motions consistent with zero.

Wolf et al. (2018) used the Gaia DR2 proper motions in this way to facilitate the discovery of SMSS J215728.21–360215.1, at $G = 18.286$ mag and $z = 4.75$. Seen by Gaia as an isolated single source, and thus unlikely to be strongly gravitationally lensed, they concluded that this is the quasar with the highest unlensed ultraviolet–optical luminosity known to date.

DAMPED LYMAN-α absorbers are a class of quasar with intervening absorption-line systems along our line-of-sight. They provide important information on the cosmic chemical evolution of galaxies. Finding them involves searching for objects that are reddened by metal-rich and dusty foreground absorbers.

Geier et al. (2019) used Gaia DR2 astrometry with existing optical and infrared photometry to discover a $z = 2.60$ quasar strongly reddened by dust in a heavilty damped Lyman-α absorber at $z = 2.226$. Another similar example discovery is given by Fynbo et al. (2020).

BY THE END OF 2020 nearly 100 scientific publications have used the Gaia DR1 or DR2 data to examine the reliability of earlier quasar surveys, to identify other particularly interesting objects, to isolate and classify numerous gravitationally lensed systems, and to quantify various effects important in further refining the extragalactic reference frame link.

Work on quasars using the Gaia data is clearly set to enter a new and vigorous phase over the coming years as the improving quantity and quality of the Gaia astrometry, photometry, and imaging data becomes available.

17. Solar siblings

MOST STARS are believed to have been born in molecular clouds, as members of larger star clusters.

There are hints that our Sun was born in such a cluster, of perhaps a thousand other stars. The solar system's relatively sharp outer boundary, at about 30 au, suggests that the disk of gas and dust from which it formed was truncated by interactions with other cluster members.

Another clue is the large eccentricities and inclinations of the Kuiper belt objects. This suggests that a stellar encounter occurred early in the solar system's history, an event more likely in a denser stellar environment.

But was it, instead, born in isolation? If it was part of a larger cluster, where are the other members now?

ASTRONOMERS ARE searching for possible 'solar siblings', but also for 'solar twins' and 'solar analogues'.

Solar twins are defined as stars being essentially identical to the Sun in key all astrophysical parameters: mass, age, luminosity, chemical composition, temperature, surface gravity, magnetic field, and so on. Solar twins may be stars most likely to host planetary systems similar to our own, and may be best-suited to host life forms based on carbon chemistry and water oceans.

Solar analogues are more loosely defined, as stars that looked in the past, or will look in the future, very similar to the Sun. They can therefore provide a perspective of the Sun at some other point in its evolution.

A SYSTEMATIC SEARCH for solar twins started with the work of Hardorp (1978), who surveyed the near-ultraviolet spectra of 77 solar-type stars, but found no G2 dwarf stars matching the properties of the Sun.

From the late 1990s, searches used Hipparcos data to provide precise distances, and therefore accurate stellar luminosities, finding that HIP 79672 (18 Sco) and HIP 78399 are amongst the most promising solar twins.

Of stars with planets on the various Doppler planet-search programmes, only HD 186427 was found to have properties close to those of the Sun. Careful searches have nevertheless identified, in total, only around a dozen nearby stars as plausible solar twins.

IN CONTRAST to solar twins, which are defined as having largely identical spectra regardless of their origin, solar siblings do not need to be Sun-like in terms of their key properties, such as their effective temperature, mass, or luminosity. But if they formed at the same time as the Sun, and from the same gas cloud, they must have identical ages and chemical compositions.

The search for solar siblings also aims to gain a better understanding of the conditions under which life developed on Earth. Was this, for example, somehow related to whether the Sun was born in a cluster of other stars?

And this touches on the more controversial idea of 'panspermia', in which 'spores of life' are ejected and transmitted through space, protected within small rocky bodies on their travels. If the Sun was born in a cluster environment, solar siblings might be especially interesting targets in the search for life.

OF RELEVANCE to practical searches, solar siblings should share similar *Galactic kinematics* as the Sun. In other words, in the absence of some more energetic ejection from the birth cluster, they should be on similar orbits around the Galaxy as our Sun.

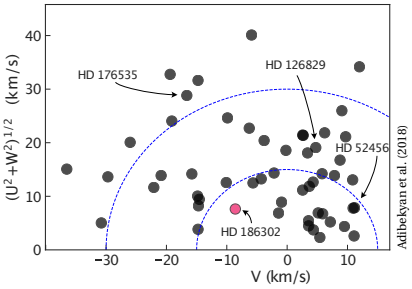
Given accurate space motions of the Sun and other nearby stars, and given a good model of the Galaxy's mass distribution (and hence gravitational potential), it should then be possible to 'reverse' their orbits, and calculate their original common birthplace.

The most optimistic estimates pre-Gaia have suggested that perhaps 10–60 still exist within 100 pc of the Sun (Valtonen et al., 2015). And while various solar siblings have been proposed in the past, with some of these subsequently contradicted by later work, no reasonably unambiguous candidates have been identified.

Martínez-Barbosa et al. (2016) concluded that our knowledge of the Galactic potential, e.g. concerning spiral arms and molecular clouds, is still too uncertain for the unambiguous identification of solar siblings based on astrometry and radial velocities alone. Accurate stellar ages and precise chemical abundances would be essential in making the searches more efficient.

TWO MAJOR EFFORTS have used the Gaia DR2 to advance the task of identifying possible solar sibling candidates. Both demonstrate how the problem of pinpointing possible candidates is strongly influenced by the quality of the astrometric data, by the uncertainty in the present knowledge of the mass distribution in the Galaxy (i.e. the Galactic potential), and perhaps even more so, by present inaccuracies in determining stellar abundances and, most acutely, stellar ages.

Adibekyan et al. (2018) used a database of 17 000 star spectra (AMBRE), and selected 55 with metallicities closest to the Sun's. Restrictions based on other chemical abundances resulted in 12 solar sibling candidates with chemistries very close to solar, and finally leaving just four candidates whose stellar ages are close to that of the Sun. For two of these, even the measured carbon isotope ratios are compatible with the solar value.

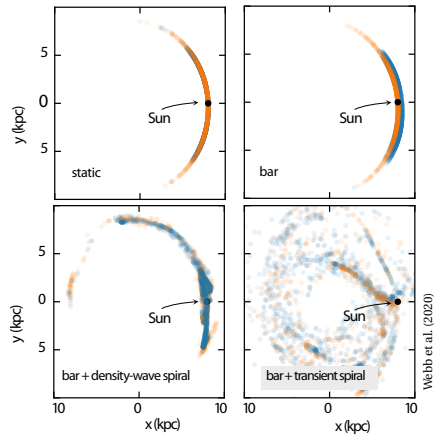


Given the difficulties of propagating the space velocities of these sibling candidates, and the Sun, backwards in time within the uncertainties of the Galactic potential, they simply used the Gaia astrometry to calculate the position of their candidates in the so-called 'Toomre diagram'. As shown here, this represents each star's combined vertical and radial kinetic velocities as a function of their rotational velocity.

They concluded that HD 186302 is the most precisely characterised and the most probable of their candidates, in terms of its chemical composition, its age, and its orbital dynamics.

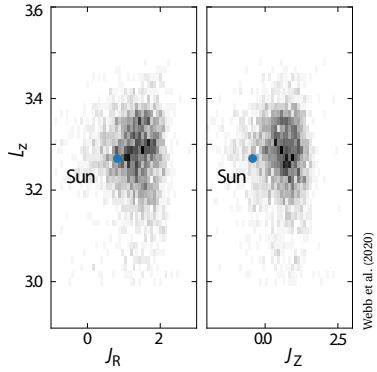
THE MORE RECENT study by Webb et al. (2020) further concentrated on the kinematical aspects of the search. Their starting sample was the SDSS–APOGEE DR14 catalogue, comprising more than 19 000 stars with high-quality spectra that have solar values of [Fe/H] within their measurement uncertainties. Combining this with the Gaia DR2 astrometry provides the largest dataset of stars with measured abundances, as well as full three-dimensional positions and space velocities.

They started with a model star cluster within a static Galaxy containing a bulge, a disk, and a halo. They assume that it 'dissolves' quickly, so that escaping stars will have similar velocity distributions to models where the Sun was born in an unbound association, since most stars will then be unaffected by intra-cluster interactions. They then introduce subsequent time-dependent effects by adding either a rotating central bar, or a rotating central bar with two different types of spiral arms.



They then integrated the present Galactocentric position of the Sun backwards in time over 5 billion years, in each potential, to determine its location at birth. The model cluster is then initialised at this location, then the cluster itself is evolved forward again for 5 Gyr. Orbital 'actions' for each star are calculated all the way along these evolving orbits.

Without entering into the details of these procedures, all we need to know here is that the three orbital quantities referred to as 'actions' (J_R, L_z, J_z) provide a powerful tool for characterising the orbits of stars in the Galaxy's gravitational potential, essentially describing the amount of oscillation of the star along its orbit in the Galactocentric directions (R, ϕ, z).



When these quantities are compared to the values for the Sun, they could pick out those stars which are closest to the Sun in terms of their orbital motions over the past 5 billion years.

Although their detailed conclusions are sensitive to the assumed details of the Milky Way's bar, spiral arms, and giant molecular cloud populations, they ended up with a list of just over 100 primary candidates, and they provide the values of the orbital actions that any true solar sibling is likely to possess. Interestingly, several of the candidates previously suggested by Martínez-Barbosa et al. (2016) also meet the orbital criteria estimated by Webb et al. (2020). And, in contrast with some previous work, they were able to exclude M67 as the Sun's birth cluster.

Their top candidate, Solar Sibling 1, lies at a distance of 360 ± 80 parsec. It has a dynamical history very close to that of the Sun, even in the absence of strong interactions with the bar, spiral arms, or giant molecular clouds.

GAIA, IT SEEMS, has allowed the discovery of the first of our Sun's birth cluster siblings. Presumably others will soon be found. And I imagine that this new field of study will develop in a variety of interesting ways.

18. The origin of OB associations

ALL STARS ARE believed to have been born in dense molecular gas clouds. Clusters and associations appear to have formed within their massive dense cores.

What is the difference between clusters and associations? Did all stars form in clusters, some of which spread out to form looser associations and moving groups, before 'dissolving' completely? Gaia has added some very significant new insight to these questions.

BUT FIRST WE will need some definitions. Lada & Lada (1991) defined open clusters and associations as groups of stars of the same physical type whose surface density significantly exceeds that of the field for stars of the same physical type.

They considered *clusters* to be physically related groups of 10 or more stars whose stellar density, of around $1 M_{\odot}\,\mathrm{pc}^{-3}$, would render it stable against tidal disruption by the Galaxy, as well as by passing interstellar clouds. *Associations* are loose groups of 10 or more physically related stars whose stellar space density is considerably below the tidal stability limit. Moving groups are, loosely, the remnants of such structures.

A typical cluster contain a few tens to several thousands stars within a typical radius of 1–10 pc. Constituent stars progressively dissolve back into the field over time through a variety of mechanisms, notably gravitational perturbations from the disk or passing interstellar gas clouds, and supplemented by cluster ejection through gravitational encounters within the cluster. Typical lifetimes are of order 100 Myr, with the least tightly bound surviving for only a few million years, and the richest for as much as a billion years.

In practice, OB associations are characterised by low stellar densities and a large spatial extent, and survive as a recognisable group only for a short time, of order 25 Myr. They are made recognisable not necessarily by their general overdensity with respect to field stars, but by their overdensity of luminous O and B stars.

Their large masses and high luminosities imply that the constituent stars are young and short lived, and are therefore associated with sites of recent star formation.

FROM THE THEORETICAL side, the 'monolithic formation scenario' holds that most (if not all) stars form in gravitationally-bound clusters. In this picture, the gravitationally-unbound OB associations found in the solar neighbourhood and beyond must have been significantly more compact at the time of their formation, and must have subsequently expanded into the configurations we see today. The process suggested to initiate expansion was the expulsion of residual gas from embedded clusters through stellar feedback.

Over the past few years, this view of monolithic formation has seemed somewhat in contradiction to observations of present-day star formation, which appears to proceed over a wide range of environments, including both large-scale hierarchical structures and isolated young stellar objects.

THere are a couple of dozen clusters, moving groups, and OB associations within about 150 pc of the Sun. Associations include those of Scorpius–Centaurus (with subgroups including Lower Centaurus Crux and Upper Scorpius), Taurus–Auriga, and Hercules–Lyra. And with the exception of the more distant α UMa (Dubhe) and η UMa (Alkaid), all the stars in the Plough/Big Dipper constellation are part of the Ursa Major association. OB associations have also been found in the Large Magellanic Cloud and the Andromeda Galaxy.

The large extent of nearby OB associations on the sky has traditionally prevented accurate kinematic membership determination for any but the brightest stars.

Many studies of OB associations were made with the Hipparcos data, with one of the most extensive being by de Zeeuw et al. (1999). They made a comprehensive census of the stellar content of OB associations within 1 kpc from the Sun, based on 9150 Hipparcos candidate members. It was part of a long-term project to study the formation, structure, and evolution of nearby young stellar groups. Hipparcos provided a major improvement in the kinematic detection of these structures, resulting in improved astrometric members for 12 young stellar associations out to a distance of 650 pc.

THE SCIENTIFIC CASE for Gaia, back in 2000, included the goal of detecting and studying associations to much larger distances. For those nearer than 2000 pc, reliable member selection should be possible down to the lowest stellar masses. This would include objects that are still on their way to the main sequence, amongst them objects bright in X-rays.

The first studies of OB associations using the early Gaia data already suggested disagreement with the monolithic formation scenario, in which associations must be expanding from their early origins as gravitationally bound clusters. The problem was that none of the associations exhibited any significant evidence of an expanding velocity field. Indeed, their bulk kinematic properties were far more consistent with randomised velocity fields than expanding ones.

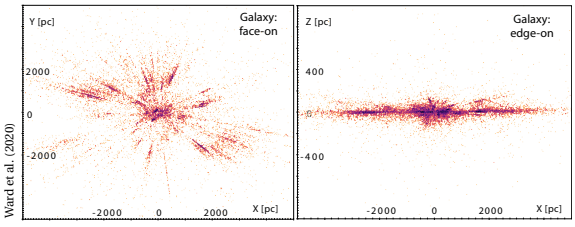
Specifically the kinematics of 18 nearby OB associations using the first Gaia data release found little evidence for systematic expansion (Melnik & Dambis, 2017). The same was true for 18 other associations (Ward & Kruijssen, 2018), while Wright & Mamajek (2018) used Gaia DR1 to show that the Sco–Cen OB association was most likely formed in a highly sub-structured state with multiple small-scale star formation events rather than a single, monolithic burst of star formation.

A subsequent study of 28 associations with DR2 found that the majority of associations are not undergoing expansion (Melnik & Dambis, 2020). In a possible counter-example, Cantat-Gaudin et al. (2019) found that, while the Vela OB2 association *is* expanding, the Gaia data suggests that this expansion began before the stars in Vela OB2 were formed, and therefore probably the result of a supernova-driven shock.

A MUCH LARGER study was carried out with the Gaia DR2 data by Ward et al. (2020) and the numbers of associations, and numbers of objects available for study in each, are worth emphasising.

Their starting point was the Galactic OB star catalogue, GALOBSTARS, containing some 16 000 OB-star candidates. Further selection according to Gaia colours resulted in a set of 11 844 OB stars from DR2.

In the resulting distribution of the selected OB stars out to 3000–4000 pc from the Sun, shown below, the apparent elongation of the associations in the radial direction is primarily due to the present uncertainty in distance of those stars.

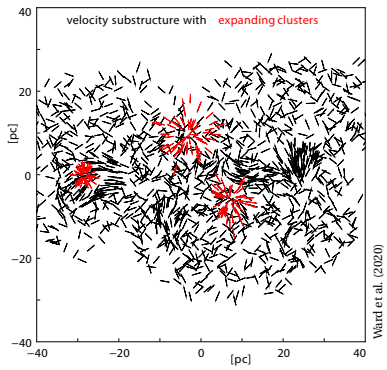


Use of a cluster-finding algorithm resulted in the discovery of 109 OB associations out to about 3000 pc. With typically 10–50 OB stars in each of these, they then searched for additional, lower mass members from the wider DR2 catalogue. This resulted in a typical total of about 6000 stars in each association.

The main part of their analysis then involved determining the key kinematic properties of these associations, such as their velocity dispersions. They could then quantify the extent to which they are undergoing expansion in excess of what could be expected from a random velocity distribution.

They concluded that a simple monolithic cluster that subsequently underwent gas-expulsion driven expansion can be firmly ruled out as an origin of the associations. But their kinematic properties are also found to be inconsistent with purely random velocity fields.

Only with a combination of small-scale localised expansion events, shown here in red, along with positional substructure and a randomised but sub-structured velocity field were they able to reproduce the kinematic properties of these OB associations.



Velocity maps based on a Gaia-observed OB association

THE GAIA DATA convincingly rule out simple models in which Galactic OB associations are formed from the expansion of previously compact clusters. They also contradict the picture that most stars form in clusters, of which a large fraction is subsequently unbound by gas expulsion-driven expansion.

The Gaia results are far more consistent with a scale-free, hierarchical picture of star formation, in which stars are formed across a continuous density distribution throughout molecular clouds, rather than exclusively within clusters, and in which OB associations are formed *in situ* as relatively large-scale and gravitationally-unbound structures.

While localised expansion of individual substructures within associations does appear to be an important component of their kinematic properties, this expansion is not the primary driver of their large-scale structural evolution.

19. How many exoplanets?

THE DISCOVERY OF the first exoplanets (i.e. planets beyond our own solar system) was a major landmark in astronomy. The first discoveries, in 1995, were proof that planets indeed existed around other stars, and that they might even be rather common.

The principle of detecting exoplanets by observing their transit across the face of their host star dates back to at least the mid-19th century, when the prolific Irish scientific writer Dionysius Lardner suggested it as one of five explanations for periodic variable stars in his 1851 *Handbook of Natural Philosophy and Astronomy*.

The idea is simple enough: careful observations of a star's brightness should show periodic dips if an orbiting planet transits the star. But it is hugely challenging in practice: for a period of 1 year (as Earth), transits will be rare, and observations over a long time would be needed to establish its periodicity as a result of multiple transits.

The problems are compounded by the need for suitable orbital alignment with the line-of-sight to the Earth, and for the tiny drops in brightness expected in the majority of realistic cases.

IN 1938, David Belorizky from Marseille Observatory argued that *'stellar photometry with 0.01 mag precision will provide the means of discovering the existence of other planetary systems'*.

In 1952 Otto Struve remarked: *'It is not unreasonable that a planet might exist at a distance of 1/50 astronomical unit. Its period around a star of solar mass would then be about 1 day.'* This is somewhat surprising because, at the time, the shortest orbital period of any known planet was just 88 days, for Mercury. Why did Struve speculate about planets with an orbital period of only 1 day? But his remark was prescient, for the first planet discovered, 51 Peg b, had an orbital period of just 4 days.

Even before the detection of the first exoplanet in 1995, and before the first transiting exoplanet was seen in 1999 (HD 209458 b), the method was considered to be one of the most promising means of detecting planets of Jupiter mass, with the detection of Earth-class planets quickly seen as being within its capabilities.

In the past two decades this method of detecting and characterising planets has flourished, through numerous impressive ground-based monitoring programmes, and through dedicated space telescopes, notably CoRoT, Kepler, and TESS. As of February 2021, and of more than 4300 exoplanets now known, the transit method has discovered around 3300, the majority of these being with NASA's Kepler satellite, operated between 2009–18.

As of the same date, 824 planets had been discovered from their host star's 'radial velocity' (or Doppler shift), which oscillates back and forth along the line-of-sight as a planet orbits around it.

A further 106 have been discovered by the difficult technique of 'gravitational microlensing'. And just one very massive planet has been discovered by astrometry.

In 2018, Exoplanet Encyclopedia compiler Jean Schneider, of the Observatoire de Paris (Meudon), was justified in stating that *'Nowadays the most powerful method to detect extrasolar planets is the transit method.*

ASTROMETRY WAS ALSO recognised as being a possible way of detecting exoplanets. At the same time that Lardner was mentioning the transit technique, Captain W. S. Jacob, at the East India Company's Madras Observatory, had been measuring the astrometric orbit of 70 Oph. In 1855, he reported that anomalies in the orbit made it 'highly probable' that the system contained a planetary body. He was wrong.

The method has continued with a checkered history to this day. Claimed detections, all later proven to be false, included those of Thomas See, in his measurements of binary star orbits at the Lowell 24-inch telescope near Mexico City in the 1890s; and Erik Holmberg, who suspected that Proxima Centauri was orbited by a planet (in 1938), and similarly 70 Oph (in 1943).

In 1943 Kaj Strand, later Scientific Director of the US Naval Observatory, 1963–77, reported his results for 61 Cyg: *'The only solution which will satisfy the observed motions gives the remarkably small mass of…16 times that of Jupiter…Thus planetary motion has been found outside the solar system.'* Strand was also wrong.

Lengthy disputes surrounded extensive ground-based observations of Barnard's star, for which two planetary mass bodies with periods of 12 and 20 years were proposed by Peter van der Kamp in 1963. All of these, and a few others, were later shown to be false.

But these efforts, though unsuccessful, should be given due credit: the measurements were challenging because of the tiny angles to be measured, and were plagued with difficulties as a result. But it nevertheless shows how, already more than a century ago, some understood the issues, and made efforts to find exoplanets.

IN THE EARLIEST DISCUSSIONS of space astrometry that I have found, from 1964, Paul Couteau and Jean Claude Pecker, of the Nice Observatory, considered the search for planetary systems. The Hipparcos space mission was originally proposed by Pierre Lacroute in the late 1960s, adopted by the European Space Agency in 1980, and operated between 1989–1993. Exoplanet science didn't figure in its original scientific objectives, although a search for Jupiter-like companions to nearby stars using the Hipparcos data was later suggested by Wilhelm Gliese, of nearby star compiler fame, in 1982.

Following the announcement of the first three exoplanets discovered by radial velocities (47 UMa, 70 Vir, and 51 Peg) in 1995, I used the pre-publication Hipparcos data to place upper limits on their masses (Perryman et al., 1996). Orbits for a few others were later measured by Hipparcos and the Hubble Space Telescope.

IDEAS FOR Gaia began to take shape around this time. A first design, Roemer, was proposed by Danish astronomer Erik Høg and Swedish astronomer Lennart Lindegren. It was developed in a form to measure all billion stars to 20 mag, with photometry and radial velocities, by Lennart Lindegren and myself. In subsequent years, large teams worked to submit a detailed proposal to ESA's advisory committees, and it was adopted by ESA's Science Programme Committee in 2000.

More than 100 scientists contributed to the enormous scientific case for Gaia, and these specialist contributions were assembled by Tim de Zeeuw of the Leiden Observatory and myself, running to 100 pages of the overall 360-page scientific and technical proposal.

The case for exoplanets occupied five of those pages back in 2000. With a young researcher, Ana Colorado, I made the first estimates of the numbers that Gaia would detect based on a rather simple model: counting the number of stars of suitable spectral type out to 200 pc, taking into account the estimates of planet occurrences known at the time, and using estimates of the accuracies achievable by Gaia as a function of star magnitude.

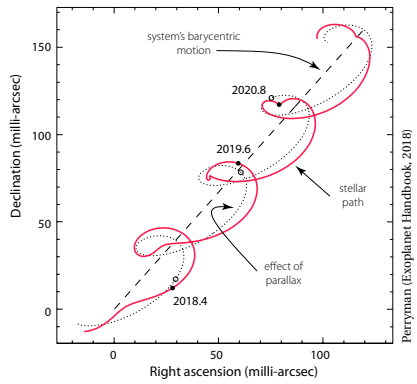
We concluded that 10 000–50 000 Jupiter-like planets should be detectable, and discussed the implications for the deeper studies of these (and multi-planet) systems.

Incidentally, in my role as ESA's project scientist for Gaia, we argued at the time of selection, in 2000, that it could be developed and launched in 2012. This impressively challenging mission was duly launched in 2013.

MORE CAREFUL ESTIMATES of the numbers of planets detectable by Gaia were made before launch, using improved models of the instrument and its sky scanning, improved estimates of planetary occurrences, and improved models of data analysis and orbit fitting.

Amongst these, detailed simulations including orbit fitting were made by Casertano et al. (2008) who concluded that some 8000 giant exoplanets (of 1–3 times the mass of Jupiter, orbiting F, G, and K-type stars) should be detectable by Gaia astrometry.

Later estimates extended the host stars to a broader range of spectral types, distances, and magnitudes, using the best estimates of exoplanet frequency distributions, and detailed simulations of the Gaia instrument and its scanning of the sky (Perryman et al., 2014a).



Simulated star–planet motion on the sky

We found that some 21 000 Jupiter-mass long-period planets should be discoverable out to distances of 500 pc for the nominal 5-yr mission (including some 1000–1500 around M dwarfs out to 100 pc), rising to some 70 000 for a 10-yr mission. The planets that Gaia should discover in large numbers – Jupiter-like planets, in Jupiter-like orbits – will not be the sort of habitable planets that are so keenly targeted by exoplanet scientists today. But they will clearly signal planetary systems that are, perhaps, very much like our own, with a Jupiter-like sentinel orbiting far out from the star capturing potentially hazardous objects left over from the system's formation.

Amongst these should be 25–50 transiting planets with periods in the range 2–3 years, perfectly placed for detailed studies of their atmospheres through follow-up transit measurements from the ground.

THE MOST POWERFUL method of detecting extrasolar planets today is without doubt the transit method, as emphasised by Jean Schneider. But planet detection through astrometry, recognised as a remarkable opportunity for Gaia already 25 years ago, and more than two decades in preparation and execution, may prove to be the most powerful in just a few years from now!

20. The Hyades star cluster

THE HYADES STAR CLUSTER is the nearest open cluster to the Sun, at around 45 pc. An arresting sight in the dark night sky, and extending over about 20 degrees, it has been prominent in the development of astronomy over the past hundred years because of its central role in establishing the cosmic distance scale.

Some background on the nature of open star clusters, and their role in the definition of the cosmic distance 'ladder', is given under my text on *'The Distance to the Pleiades'*. Let me just recall that these groups of several hundred stars, formed from the same giant molecular cloud and with roughly the same age, are key objects in the study of stellar evolution because the cluster members are of similar age and chemical composition.

The nearest, notably the Hyades (at about 45 pc) and the Pleiades (at 135 pc), have long been used to calibrate the distances to others more remote, either by matching their main sequences in the Hertzsprung–Russell diagram, or by exploiting the changing geometrical perspective as the entire cluster moves through space.

Since the pioneering work of Lewis Boss (1908) more than a century ago, the Hyades has been the subject of dozens of innovative studies which have refined its distance, and explored the nature of its members through the study of its Hertzsprung–Russell diagram.

BEFORE PUBLICATION of the Hipparcos catalogue in 1997, many studies, using different approaches, had reached on a reasonable consensus in terms of its mean distance. With the Hipparcos parallaxes accurate to around 1 milli-arcsec, distances to individual cluster stars were measured with an accuracy of about 20%. And its mean distance was pinned down, geometrically, as 46.34 ± 0.27 pc for the 100 or so objects within 10 pc of the cluster centre (Perryman et al., 1998).

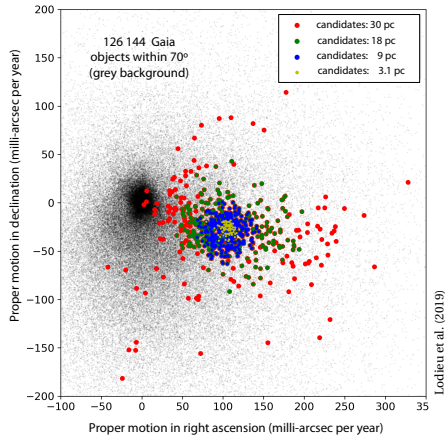
This result was close to, but much better determined than, that derived from high-precision radial velocity studies, somewhat larger than that from ground-based trigonometric determinations, and slightly smaller than those found from the most up-to-date studies of the cluster's convergent point.

The Hipparcos studies were in part limited by the pre-selection of stars demanded by the instrument itself, with 240 candidates having been pre-selected, before launch, over some 30–40 degrees on the sky. Their distances and space motions resulted in a total of 140–200 members, depending on the criteria adopted.

Further from the cluster centre, Hipparcos showed a gradual merging between definite cluster members and field stars, both spatially and kinematically. The individual proper motions were fully consistent with a uniform space motion with an internal velocity dispersion of about 0.3 km s^{-1}. The cluster zero-age main sequence could be accurately modelled, based on the estimated helium abundance and isochrone modelling, yielding a cluster age of 625 ± 50 million years.

OBSERVING ALL STARS down to around 20–21 mag, with a much improved accuracy by a factor 10–100, and with its simultaneous high-accuracy photometry, great advances were expected with Gaia, and this has indeed proved to be the case.

Early studies using Gaia DR1, with the TGAS subset, already identified 251 new candidate members (Reino et al., 2018), but deeper insights began with DR2.



Hyades: proper motions (grey) and member positions (colour)

PRE-GAIA, a total of some 800 probable members were known, including a handful of white dwarfs, and a dozen or so brown dwarfs.

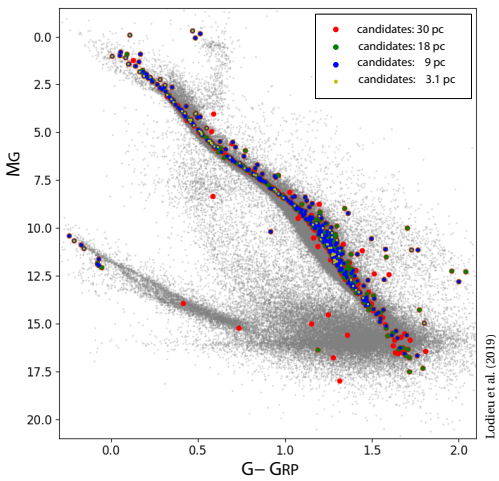
The Gaia DR2 analysis by Lodieu et al. (2019b) started with all objects within 70° of the nominal cluster centre, and with a parallax greater than 10 mas (i.e. within 100 pc), resulting in 126 144 objects. An iterative selection based on distances and space motions resulted in 1764 sharing the cluster's mean proper motion. This more than doubles the number of previously known members built up over more than a century.

Further analysis was restricted to objects which they found to be within the core radius of 3.1 pc, the tidal radius of 9 pc, the halo radius of 18 pc, and extending out to a distance limit of 30 pc. This resulted in 85, 381, 568, and 710 sources respectively, while they rejected some 200 objects previously considered as members.

From these stars within the cluster's tidal radius, they derived a mean cluster distance of 47.03 ± 0.20 pc, and a mean space velocity of $46.38 \pm 0.12\,\mathrm{km\,s^{-1}}$. They could trace the 3d distribution of stars within the cluster, and could estimate the effects of mass loss over the cluster's lifetime. They could identify clear mass segregation, with the lower mass stars being on average further away from the centre.

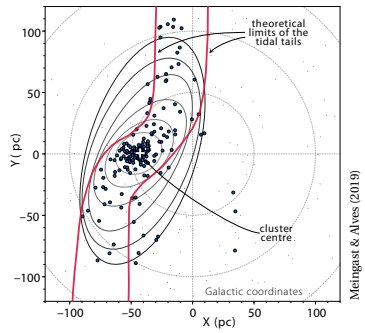
Supplemented by infrared photometry, they derived the luminosity and mass functions over the range $G = 3-26$ mag, corresponding to masses of $2.5-0.04\,M_\odot$, also showing that their shapes vary as a function of the distances from the cluster centre.

A STUDY OF THE Hertzsprung–Russell diagram of a number of open clusters and globular clusters using the Gaia DR2 data demonstrates the narrow and extremely well-defined main sequence of many of these systems, including the Hyades (Gaia Collaboration et al., 2018a). A detailed discussion of the various age estimates for the cluster is taken up by Lodieu et al. (2019b).



TWO SEPARATE studies have reported the discovery of remarkable extensions of the cluster's tidal tails using the DR2 data (Röser et al., 2019; Meingast & Alves, 2019). These tails are highly flattened in the Galactic plane, being only about 25 pc in thickness, while extending more than 100 degrees across the sky.

The leading tail, of nearly 300 stars, extends 170 pc from the cluster centre, while the trailing tail with more than 200 stars extends to 70 pc.



This top-down view of the Galactic plane reveals a distinct S-shape to the tails, similar to those observed for some globular clusters, and in excellent agreement with earlier theoretical predictions of their shape, and of their velocity dispersions.

AT THE LOWEST end of the main sequence, some 16 brown dwarfs were known pre-Gaia.

Pérez-Garrido et al. (2018) used the Gaia DR2 data, with 2MASS and WISE infrared photometry, to identify a total of 37 objects whose distances and proper motions are compatible with Hyades membership. Of these, 21 were new. And while Gaia is essentially complete to the hydrogen-burning limit for the Hyades cluster, it is still likely to be incomplete at lower masses.

FOR THE NINE known Hyades white dwarfs, Lodieu et al. (2019b) used the Gaia photometry to derive an age of 640^{+67}_{-49} Myr from evolutionary models, in agreement with other methods, including that derived from the lithium-depletion boundary method.

The Hyades white dwarfs also provide a co-eval sample with known distances which can be used to estimate their cooling times and masses. Salaris & Bedin (2018) used the Gaia DR2 astrometry to conclude that a cluster age of 690 Myr yields a slope of the initial–final mass relation closer to other white dwarf determinations.

White dwarf masses can also be derived from their gravitational redshift. The approach is possible in the case of the Hyades because the stellar radial velocities can be deduced from the astrometric data, independently of spectroscopy. Pasquini et al. (2019) found that masses derived from their gravitational redshift estimates are systematically smaller, by $0.02 - 0.05\,M_\odot$, than those derived from other methods. Though small, this is a significant finding for white dwarf models.

Confirmed Hyades members from these Gaia DR2 studies are also allowing the properties of nearly 300 X-ray sources discovered with ROSAT, Chandra, and XMM–Newton to be derived (Freund et al., 2020).

21. Measuring exoplanet radii

THE FIRST EXOPLANETS, by which we mean planets beyond our own solar system, were discovered only in 1995. But by the time of the first Gaia data release, DR1, in September 2016, more than 3000 were known.

Many of the first exoplanet discoveries were identified through radial velocity measurements of their host star. This is possible because the velocity of the star oscillates backwards and forwards (along the line-of-sight to the Earth) as the planet orbits around it.

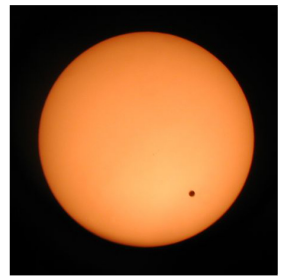
To be clear, all planets, whether those in our solar system or in others, do not orbit the centre of the *star*, but rather the *centre of mass* of the entire planetary system. This means that a star hosting one or more planets orbits this centre of gravity as well, albeit with a relatively small orbital amplitude often lying within the star's own physical boundary. So the existence of a planet orbiting a star shows itself as a tiny change in radial velocity of the star, and with the same period as the orbiting planet. All this is possible even though the planet itself is quite invisible from Earth.

MANY OF THE planets known by 2016 were actually discovered by the 'transit method'. NASA's Kepler satellite, launched in 2009, has been foremost in discovery numbers, although other instruments (such as the HAT and WASP telescopes on the ground, along with the French CoRoT satellite launched in 2006, and the NASA TESS satellite launched in 2018) have all contributed.

Like all transit searches, Kepler could only detect planets whose orbits lie *edge-on* to the sight-line from the Earth, on the alert for periodic dips caused by a planet happening to transit across the stellar disk. If the planet's orbit is inclined to this sight-line by more than a degree or so, it will escape detection by this method.

The Kepler mission transformed the field of exoplanet research, not only because of the very significant numbers of exoplanets discovered (more than 3000), but even more so because of the richness and unexpected nature of the planetary system architectures, and the deep insights into a very broad range of physical phenomena that have been gained from their study.

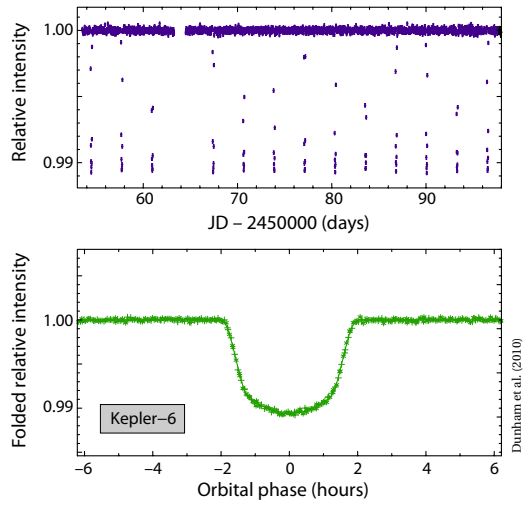
For nearly four years, the Kepler satellite locked on to one region of the sky, and monitored the brightness of some 150 000 stars, with measurements of each star every few seconds. The principles are the same as those used by Earth-based observers to watch the spectacular transit of Venus across the face of our Sun, in June 2004 and June 2012. The orbits of Venus and Earth are both slightly elliptical, and both slightly inclined to the mean orbital plane of all the solar system planets, such that Venusian transits occur only every 100 years or so, seen from Earth.



Creative Commons: Gestrgangleri

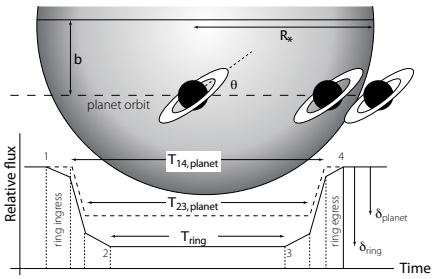
Transit of Venus, 8 June 2004

Watch any star for long enough (and with a sensitive instrument), and if it is orbited by planets, and if these planets orbit edge-on to our line-of-sight, regular dips will occur at the period of the planet's orbit. Multiple planetary systems will show a family of such dips (at different periods) if they too orbit close enough to the line of sight. This shows the data for Kepler–6 over 50 days (top). When 'folded' at the planet's period of 3.2347 days, the transit shows up clearly in the light curve (bottom).



EXOPLANET TRANSITS, observed with high accuracy, with suitable timing accuracy, and preferably in multiple spectral bands, can yield astonishing insights into the planetary system. The brightness drop as the planet transits is determined by the ratio of the planet's projected area to that of its host star, so if we can estimate the star's radius (for example, from theoretical models for a given star type) we can deduce the planet's radius. If we can measure the star's radial velocity amplitude, we can infer the planet's mass. And given its mass, and its radius, we can estimate its density.

From light curves in different spectral bands we can probe the planet's atmosphere. If there are multiple planets, we can assess the system's stability, and any interactions between them. Such observations, supported by theoretical work, provide information on planetary tides, on their atmospheric conditions, on their origin, and even on the presence of planetary moons.



ACCURATE PLANETARY radii are of great importance for many reasons. Together with the planet's mass, it provides an estimate of its average density, allowing a broad categorisation of its nature, i.e. whether the planet is a gas giant (like Jupiter or Saturn), an ice giant (like Uranus or Neptune), or a rocky planet (like Earth or Mars). Combined with information about their periods and host star properties, this leads to great insights into both the physics of planetary atmospheres and interiors, and the physics of planet formation and evolution.

But there is a serious problem, which is that the mass and radius of a transiting planet cannot be measured directly. Rather, it depends, through the transit measurements, on the assumed mass and radius of its host star. Although we cannot determine these directly either, it turns out that transit measurements do provide a direct estimate of the *density* of the star!

So, how have estimates of the mass and radius of the host star been obtained up to now? Here, the theory of stellar evolution comes to the rescue. It is known that a star of a given temperature, chemical composition, and density cannot have an arbitrary mass and radius. Rather, these three parameters together fix the luminosity and age of the star, and in turn its radius and mass. While not direct, this approach has typically provided good estimates of the properties of the system.

But we can go further: measurement of the star's flux over a broad spectral range yields the star's total (or bolometric) flux, and in turn the star's angular radius, via its 'effective temperature'. It requires just one more step to realise that an accurate distance to the star allows the star's *angular* radius to be converted to a *linear* radius, and for the star's linear radius to be used in the planetary transit measurements to yield the planet's radius.

The path just described is certainly somewhat tortuous, but can be easily summarised: the accurate distance to a star, measured by Gaia, allows the radius of a transiting exoplanets to be determined directly!

JUST SUCH A piece of work, entitled *'Accurate empirical radii and masses of planets and their host stars with Gaia parallaxes'*, was submitted to the *Astronomical Journal* by Stassun et al. (2017).

The authors published the radii of 116 stars that host transiting planets, determined using only direct observables – the bolometric flux at Earth, the effective temperature, and the parallax provided by the Gaia first data release. Although the typical uncertainties on the planet and star radii and masses, of around 10–20%, are generally a little larger than previously published (model-dependent) accuracies, they have the important advantage of being, for the first time, purely empirical.

Although the application to planetary radii is particularly forceful, their method (based on flux, temperature, and parallax measurements) yields the radius of a star whether it is orbited by a planet or not. Indeed, Stassun et al. (2017) also estimated stellar radii for more than 350 000 stars whose Gaia DR1 parallaxes were measured to better than 10%. These, along with even better Gaia distance accuracies in the future, will be of great value across many areas of stellar structure and evolution.

AS FAR AS I could make out, the work by Stassun et al. (2017) was the first refereed publication, outside of the Gaia scientific teams, to make use of the Gaia DR1 parallaxes. The timing merits a comment: the Gaia DR1 data was released on 14 September 2016. And their paper, making use of the published parallaxes, was submitted to the *Astronomical Journal* later that day!

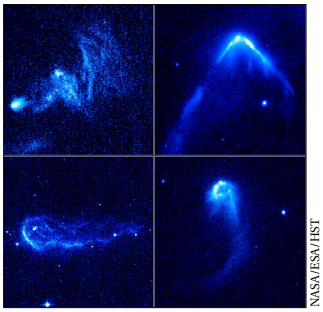
THE CONTRIBUTION of Gaia to the determination of exoplanet radii provides an excellent example of science working together and advancing across different domains. The field of exoplanetary science has exploded over the past 25 years, driven by fundamental questions such as the origin of our own solar system, and the possible existence of life elsewhere in the Universe.

Who would have foreseen, 20 years ago when Gaia was accepted by ESA, that it would find immediate application in this area of exoplanet science? And which exoplanetologists, at the birth of the field in the 1990s, could have imagined that they would be demanding accurate star distances as key ingredients for their models?

22. Hypervelocity stars

HYPERVELOCITY STARS are a rare and exotic type of star, racing through our Galaxy with velocities of 500–1000 km s^{-1} or more. How did they acquire these enormous speeds, how and where were they formed, where are they going, and what can they tell us about the structure and origin of our Galaxy?

TYPICAL STARS IN OUR Galactic precinct move, somewhat randomly, at around 20–30 km s^{-1} with respect to the bulk rotation of our Galaxy's disk. Their velocities arise from conditions of their birth in star clusters and star associations, and these are later modified by interactions with other mass structures of the Galaxy. 'Runaway stars', moving through the Galaxy disk with much higher velocities, of 100 km s^{-1} or more, have been known since the 1960s.


NASA/ESA/HST

Bow waves from runaway stars

They acquired these velocities from binary supernova ejections, or dynamical ejections from star clusters, in which extreme gravitational encounters hurl one of the interacting stars outwards at enormous speeds. Their motions often point away from known star clusters or stellar associations, providing clear signposts as to their birthplaces. They often create 'bow shocks' as they plough through the interstellar medium.

HYPERVELOCITY STARS have even more extreme velocities, velocities which cannot be reached even with these violent ejection mechanisms. They require some formation mechanism to provide them with their huge velocities which is even more extreme.

Black holes provide the answer. Black holes are objects that have undergone catastrophic gravitational collapse, leaving behind spheroidal regions of space from which nothing can escape. The most massive, perhaps millions (or even billions) times the mass of the Sun, are referred to as 'supermassive black holes'. A range of modern observations indicate that almost every large galaxy has a supermassive black hole at its centre.

Evidence for a massive black hole at the centre of our own Galaxy started to emerge in the 1930s. Radio observations by Karl Jansky suggested a highly unusual object at its centre, now known as Sagittarius A* (or Sgr A*). A large body of theoretical and observational evidence for its existence has been gathered since. Some of the most compelling are high-angular resolution optical images, published in 2009 from 16 years of data with the European Southern Observatory's Very Large Telescope, revealing a number of stars orbiting this central object. One of these, star S2, orbits the black hole in just 16 years, and was moving at 7650 km s^{-1}, or 2.5% the speed of light, at its closest approach. The mass of this central supermassive black hole is estimated to be a little over 4 million times that of our Sun.

HYPERVELOCITY STARS are an extreme type of runaway star which originate near a supermassive black hole. In this so-called Hills ejection mechanism, predicted by Jack Hills in 1988, the black hole acts as a gravitational slingshot. Simulations show that stars can be ejected from the deep potential well of a massive black hole, either as a result of scattering with another star, or through tidal breakup of a binary star system.

Stars could be ejected at 1000–2000 km s^{-1} or more, greatly exceeding speeds that could arise from either binary supernova or star cluster mechanisms. The most extreme velocity objects would not be gravitationally bound to our Galaxy, and will escape the confines of even its vast boundaries after some 300 Myr.

THE FIRST SUCH hypervelocity star, SDSS J0907, was discovered, serendipitously from the Sloan Digital Sky Survey, only in 2005 (Brown et al., 2005). Spectroscopy showed that it is a late B-type main-sequence star, with a radial velocity of more than 800 km s^{-1}. The discovery paper placed the star at a distance of about 55 kpc, some 30 kpc above the Galactic disk, some 60 kpc

from the Galactic centre, and with a motion through space consistent with it having been ejected from the Galactic centre. If so, it should have a proper motion of about 0.3 mas yr^{-1}, tiny compared with the accuracies available at the time, but which should be measurable, and so able to confirm its birthplace, by Gaia.

FINDING OTHER hypervelocity stars would open other avenues of study of our Galaxy's structure. They would throw light on their production rate at the Galactic centre, providing a better understanding of the conditions – and stellar populations – which exist there.

As a result of the enormous distances travelled on their journey through the Galaxy halo, in different directions and now at a vast range of distances from their origin, they probe the Galaxy's gravitational potential, making them possible tracers of the matter distribution in the Milky Way (Gnedin et al., 2005; Sesana et al., 2006).

Astrometric measurements with Gaia can also help pin down the Sun's position and velocity through space: if the wrong assumptions are made, the star would *appear* to originate slightly displaced from the Galactic centre (Hattori et al., 2018).

Hypervelocity stars will, however, be rare: estimates of the ejection rate from Sgr A*, of about 10^{-4} yr^{-1}, suggest that out of the Milky Way's 10^{11} stars, there should only be one hypervelocity star within 1 kpc of the Sun.

FOLLOWING THE DISCOVERY of the first hypervelocity star, a wide-field spectroscopic survey was set up at the 6.5-m Multi-Mirror Telescope (MMT), to search for others of similar spectral type and mass ($2.5-4M_\odot$). Covering over 12 000 square degrees of the northern sky, and completed in 2014, a few dozen bound and unbound candidates were discovered from their extreme radial velocities, at distances varying between 50 kpc and 100 kpc from the Galactic centre. Other surveys have been made with the Chinese LAMOST telescope (Huang et al., 2017), and a number followed-up with the Hubble Space Telescope (Brown et al., 2015).

SINCE THE DR2 RESULTS were made available in April 2018, Gaia is providing proper motions and distances of previously known hypervelocity star candidates with unprecedented accuracies, in turn helping to pinpoint their origin, while also searching the entire sky for other examples of these very rare objects.

Although the picture is still far from being complete, some of the previous candidates are indeed confirmed as having a high probability of originating close to the Galaxy centre. Amongst these are the object known as HVS 22, with a velocity of some 1500 km s^{-1}.

But others do not appear to have originated from the Galactic centre, hence suggesting that other extreme ejection mechanisms might also be at work.

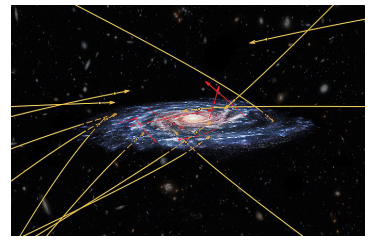
Amongst these, HVS3, appears to be coming from the centre of the Large Magellanic Cloud, and moving with a velocity of more than 800 km s^{-1} (Irrgang et al., 2018; Erkal et al., 2019). This large velocity, consistent with the Hills mechanism, provides strong direct evidence that the Large Magellanic Cloud itself harbours a massive black hole of at least $4\times10^3-10^4 M_\odot$.

NOTWITHSTANDING THEIR rarity, detailed modeling suggests that Gaia could find several hundred new hypervelocity stars within some tens of kpc from the Sun (Marchetti et al., 2018).

Amongst new discoveries from the Gaia DR2 data release, a number are believed to originate from the Galactic centre, while others are not (Li et al., 2018; Du et al., 2019; Marchetti et al., 2019). The present haul includes three hypervelocity white dwarfs, conjectured to be the companions to primary white dwarfs that exploded as Type Ia supernovae (Shen et al., 2018). The orbit of at least one of these can be traced back to a faint old supernova remnant, strengthening the idea that some hypervelocity stars may originate from extreme binary supernova ejection events (Ruffini & Casey, 2019).

Are there still other exotic ejection mechanisms, not restricted to the Galactic centre, that might be generating at least some of these hypervelocity objects?

Amongst ideas recently put forward are the existence of (less massive) intermediate-mass black holes within the Galactic disk, resulting in similar but less extreme gravitational slingshot-type ejections (Fragione & Gualandris, 2019).



ESA and Tommaso Marchetti

And perhaps yet others resulted from tidal interactions when smaller infalling galaxies were 'swallowed' and merged with our own Galaxy much earlier in its formation history (Boubert et al., 2020).

THE GAIA RESULTS on hypervelocity stars could provide a significant ingredient for refining cosmological models. In the 'concordance ΛCDM model', galaxies are embedded within extended halo structures, largely made of some 'non-viscous' dark matter, visible only through its gravitational effects. Over cosmic time, haloes grow in mass and size through hierarchical clustering, starting from the initial perturbations of a slightly inhomogeneous matter density field, but with resulting shapes and masses depending on the model details.

The motion of hypervelocity stars as they race outwards through the Galaxy halo may provide a unique probe of our Galaxy's actual gravitational potential.

23. The Maunder Minimum

A particularly cold period in Europe, which caused great hardship, was a 70-year interval from around 1645–1715, today known as the 'Maunder Minimum'. It was so named by US astronomer Jack Eddy (1976).

Eddy drew on the 19th century works of Edward Maunder and Gustav Spörer to link this extended period, some 2°C colder than average, with a seemingly unrelated natural phenomena: during this period of time, sunspots on the face of our Sun all but disappeared.

Before looking at this further (and what it might have to do with Gaia!), let us first look at some links that are known to exist between astronomical phenomena and the Earth's climate. These are on top of the recent anthropogenic contributions to climate change, which are now uncontested, and other less predictable effects, ranging from El Niño to volcanic eruptions.

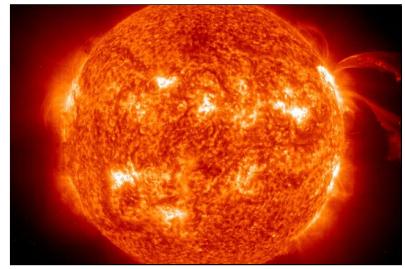
The complex orbital motion of the Earth has three dominant components, sometimes referred to as the Milankovitch cycles: the ellipticity of its orbit and the inclination of its spin axis (which give rise to seasonal effects, as well as variations over tens of thousands of years), and the long-term 'wobble' of its spin axis (the physical phenomena of precession and nutation).

Together these create short- and long-term variations of the Sun's radiation reaching the Earth's surface. And they can explain in large part the episodic nature of the Earth's glacial and interglacial periods over the last two million years. Shorter-term disturbances of the Earth's rotation, termed 'polar motion', are influenced by less-predictable effects such as ocean currents, wind systems, and even motions in the Earth's molten core.

On geological time-scales, other mechanisms may link our Galactic environment with very long-term climate variability. These include encounters with interstellar clouds (through increasing solar luminosity from accretion, or through shrinking of the Sun's heliosphere); perturbations of the Oort Cloud due to a passing star, and the injection of comets into the inner solar system; and even 'ice house' periods synchronised with spiral arm crossings by our Sun and its solar system.

The Sun throws in a number of important effects. It spins on its axis once in 27 days, a uniform rotation first detected from the motion of sunspots across its surface. The fluid motions inside the rotating Sun lead, through complex processes still imprecisely understood, to a north–south magnetic field which flips every 11 years, and a related quasi-periodicity in the appearance of cooler surface regions manifesting as sunspots.

Sunspots have been observed and recorded continuously since the early 17th century. We now know that these and other related phenomena (notably, changes in solar radiation and the ejection of solar material, as well as solar flares and coronal loops) all exhibit a synchronised fluctuation, from active to quiet and back to active again, over this 11-year period.



SoHO/ESA/NASA

The Sun's active surface from the SOHO satellite

These various phenomena, known collectively as 'solar activity', all arise from the Sun's rotation, and the continuous regeneration of its magnetic field.

A link between the variability in solar irradiance and Earth's climate has been suspected for more than 200 years, since the work of William Herschel.

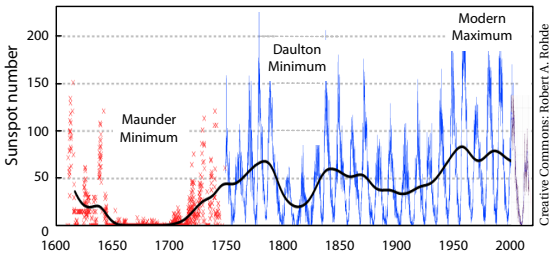
One of the many questions we can ask is: was the cold period of the 70-year Maunder Minimum linked to the (unexplained) disappearance of sunspots? In other words, does the complex variation of solar activity have a significant effect on Earth's climate?

In his 1976 study, Eddy gathered data from many sources: historical observations of the Sun going back to the telescope observations of Galileo and others of the 17th and early 18th centuries, historical reports of the aurora borealis observed from Europe and the New World, and sunspots seen with the unaided eye at sunrise and sunset in dynastic records from the Orient.

Varying levels of the carbon isotope ^{14}C are observed in tree rings. Dated through dendrochronology, ^{14}C can actually be used as an indicator of solar activity. Based on its varying levels, Eddy found evidence for other similar periods of a much quieter Sun in the distant past, including an even longer 90-year span, from about 1460 until 1550, which he named the Spörer Minimum.

Both the Maunder and Spörer minima fell during the coldest parts of what climate scientists refer to as the 'Little Ice Age', a period extending from the 16th to the 19th centuries. This has suggested a meaningful connection between the longer term behaviour of the Sun and of the Earth's mean surface temperature.

Contrasting warmer periods have also existed, notably the 'Medieval Maximum' around 1000 CE which was accompanied by the migration of the Vikings.



Four hundred years of sunspot observations

THROUGH WHAT mechanism might these relatively small changes in solar activity affect our climate? One idea is that, when the Sun is active, an increased solar wind more effectively reduces the Galactic cosmic ray flux which reaches Earth, producing more ^{14}C. In turn, cosmic rays hitting the atmosphere lead to the ionisation of aerosols in the troposphere, and these could be responsible for the condensation of water droplets and variations in Earth's cloud cover. These links are hypothesised but unproven, and climatic variations as a result of volcanic activity has also been suggested.

Nevertheless, the known northern hemisphere climatic responses to the Maunder Minimum have been successfully reproduced in global climate model simulations, and these simulations support the idea of solar irradiance variability being capable of altering major modes of dynamical variability in the troposphere.

Meanwhile, recent analysis of cosmogenic radioisotope records from Greenland ice cores, as well as the ^{14}C results from dendrochronology, suggest that the current period of relatively high solar irradiance may perhaps end sometime later this century.

At the same time, and as inferred from the various types of geophysical data, the so-called solar 'grand minimum' events (periods over which several solar cycles exhibit lower than average activity for decades or centuries) appear to be more randomised than periodic.

IN THE USUAL SPIRIT of scientific endeavour, we would like to be able to test the hypothesis that these solar 'grand minima' affect Earth's climate. And we would like to understand how often they occur for our Sun. Is the present cycle of solar activity unusual or transitory, and are Maunder-type minima common phenomena in other stars like the Sun?

Going further back in time to search for these dependencies has reached the limits of what appears possible on Earth. But an ingenious approach has tackled the problem using the data from Gaia's forerunner Hipparcos: instead of trying to estimate the prevalence of our Sun's grand minima over, say, 100–1000 years, we can instead make a survey of the magnetic activity of Sun-like stars in our solar neighbourhood.

If we can choose stars whose properties are very close to those of the Sun, we can then try to argue that the fraction of solar-type stars showing evidence of very low activity corresponds to an estimate of the fraction of the Sun's own lifetime spent in a grand minimum state.

WHAT CONSTITUTES a Sun-like star? Ideally, we would like to seek out non-binary stars which are essentially identical to the Sun in all its astrophysical properties: mass, age, luminosity, chemistry, temperature, surface gravity, magnetic field, rotation velocity, microturbulence, and chromospheric activity.

Some early surveys of Sun-like stars with low activity, using chromospheric emission in the lines of Ca H and K, suggested that perhaps 30% of Sun-like stars appeared to be in these low-activity states (Baliunas & Jastrow, 1990). Henry et al. (1996) observed more than 800 stars within 50 pc, and concluded that only some 10% were inactive. If the observations are considered to be a sequence of snapshots of the Sun during its life, then the Sun will spend about 10% of the remainder of its main-sequence life in Maunder minimum-type phases.

Wright (2004) used accurate Hipparcos parallaxes for stars within 60 pc to figure out their location with respect to the main sequence which, from stellar evolutionary models, yields their age. He showed that nearly all stars previously classified as Maunder-minimum candidates were actually evolved stars, and that it was not obvious how to identify such a state from a single observation. However, long-term monitoring programmes suggest that Sun-like stars exhibit a variety of activity behaviour, including solar-like cycles and flat-activity states. If the Sun's Maunder Minimum was a transition between the two, then continued monitoring should detect some stars switching between them.

SOME 15 very inactive stars were already ranked in priority order as Maunder Minimum candidate stars by Lubin et al. (2012). Their high-accuracy Gaia data are ready to serve as proxies for probing the climatic variations on Earth.

24. Occultations of Europa and Triton

A STELLAR OCCULTATION occurs when a solar system object, such as an asteroid or a planetary moon, passes in front of a star as seen from the Earth, causing a temporary drop in the observed brightness of the star.

This brightness drop can be used to determine the occulting object's position, along with its size and shape, often with kilometre precision. And it can probe other properties of the occulting object, such as the presence of an atmosphere, structures around it such as rings or moons, or even the measurement of its topographic features, such as mountains or valleys.

The first occultation of a star by an asteroid, Juno, was recorded in 1958. But until the Hipparcos results became available in 1997, the limited accuracy in the knowledge of star positions made it difficult to predict future occultations with any confidence. With Hipparcos positions, some 30 such events have been observed every year since. Amongst them, occultations of asteroids, satellites of asteroids, Centaurs and Kuiper Belt Objects, and even Pluto, have all been observed in this way.

OF THE FOUR Galilean moons, only a handful of occultations have ever been observed: of Ganymede (in 1911, 1972, and 2016) and of Io (in 1971), with none of Europa or Callisto. Their rarity is largely because of the difficulty in predicting potential occultations due to the uncertainties in the positions of background stars.

The first recorded occultation of one planet by another was of Jupiter by Mars in 1170 CE, observed by the monk Gervase of Canterbury in the west, and Chinese astronomers in the east. A recent computation of past and future planet–planet events, based on their accurate orbits, found two in the 19th century, none in the 20th, and five in the 21st century, all involving one of Mercury and Venus, with one of the planets beyond Earth.

Occultations of one solar system moon by another moon can also occur. Infrared imaging of an occultation of Io by Europa in 2015 yielded 2-km resolution maps of Io's volcano Loki Patera. Even triple transits of Jupiter's Galilean moons occur once or twice a decade: the crossings of Callisto, Io and Europa were observed by the Hubble Space Telescope in January 2015.

IT IS STRAIGHTFORWARD to appreciate how the results from Gaia are transforming this field. The Hipparcos catalogue contained nearly 120 000 stars distributed fairly uniformly over the sky, or about 3 stars per square degree, all brighter than about 10–11 mag. In reality, this is a rather sparse coverage of the sky, and it means that the chance of any particular target object passing in front of one of them is rather small. Planning this type of occultation observation, of course, requires predicting such an alignment weeks or months in advance.

A further complication is that since all stars are moving due to their proper motions, knowing a star position 30 years after the epoch at which the Hipparcos measurements were centred (about 1990) means that we must also have an accurate knowledge of their proper motions in order to predict their *current* positions.

The Gaia satellite is measuring every object brighter than about 20 mag, and with a much greater accuracy – in both positions and proper motions – than even Hipparcos. In particular, since relatively bright stars are required for this type of occultation measurement, Gaia is providing a catalogue of some two million or more stars brighter than about 10–12 mag, or some 50 stars per square degree on average. This dense grid of highly accurate star positions means that there are greatly improved prospects of finding – and predicting – a suitably bright occulting star.

THE FIRST OF my two examples of the use of the Gaia data for occultation measurements are for Jupiter's Galilean moons. Between 2019–2020, Jupiter was projected against a very dense star region, in the vicinity of the Galactic centre, a configuration that will not occur again until 2031. The high background star density means that the probability of a stellar occultation by the Jovian moons increases dramatically.

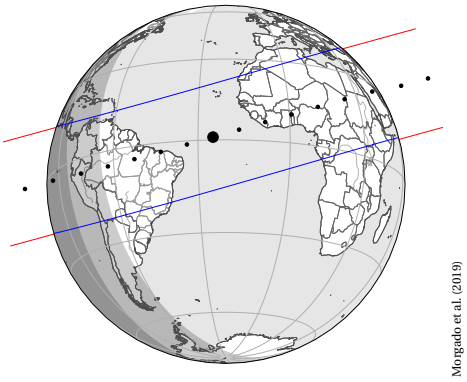
And this provides an even better opportunity to observe stellar occultations by the Galilean moons, determine their positions, improve their orbits, and measure their shapes independently of satellite probes, thus helping (for example) in the study of tides generated by gravitational forces by Jupiter itself.

In this context, it is worth emphasising that accurate orbits, of both Jupiter and its satellites, are essential to the task of preparing space missions targeting the Jovian system. In the near future, the European Space Agency's JUICE mission (the JUpiter ICy moons Explorer) is targeting launch in June 2022, and NASA's Europa Clipper mission, focusing on its moon Europa, in 2024.

ADVANCES IN THIS technique exploiting the Gaia data were reported by Morgado et al. (2019), who made the first such observation of Europa on 31 March 2017.

They predicted that Europa would occult a 9.5 magnitude star at 06:44 UTC on that day. The shadow path, several thousand km wide, would cross South America with a velocity of 17.78 km s^{-1}. The prediction was made using the state-of-the-art description of Europa's orbit, itself referred to the orbit of Jupiter, provided by the Jet Propulsion Laboratory (JPL). The star's position was obtained from the best-available Gaia catalogue (DR1) at the time, updated using its proper motion and parallax for the predicted occultation time.

Three stations in Chile and Brazil observed the occultation, and measured a brightness drop corresponding to the chord of the satellite that each observed. Taken together, they gave estimates of Europa's major and minor ellipsoidal axes, 1562.0 ± 3.6 km and 1560.4 ± 5.7 km respectively. These values, and the body's resulting oblateness, are in good agreement with those from the Galileo mission images. Topographic features of Europa are predicted to be at a level of only some hundreds of metres, below the resolution of these observations.

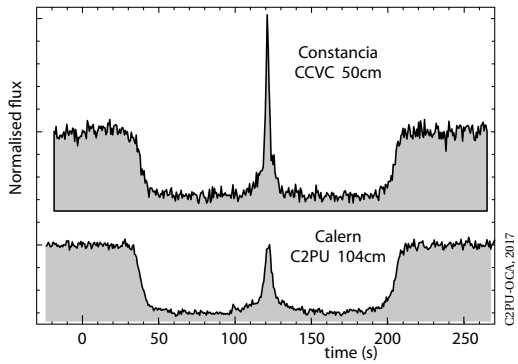


Predicted occultation of Io, on 2 April 2021

Various occultations of bright Gaia stars by Jupiter's Galilean moons have been predicted for the future. This example shows the predicted occultation by Io, of a 5.8 magnitude star, expected on 2 April 2021 10:24 UTC. The blue lines show the satellite's expected size, and the blue dashed lines correspond to an uncertainty of 20 milli-arcsec in its position. Black dots show the centre of the body's shadow at 1 minute intervals.

A FEW MONTHS AFTER these observations, on 5 October 2017, the largest of Neptune's satellites, Triton, passed in front of a 12.6 magnitude star.

A major goal of this particular event was to examine Triton's atmosphere. During the occultation of a body possessing an atmosphere, stellar rays will be refracted and focused by it, creating a flash which can be detected by an observer very close to the line of 'centrality'. This was first observed for Triton by Sicardy et al. (1990). With a width of about 100 km, this band subtends an angle of about 5 milli-arcsec at the distance of Triton.



Occultation of Triton and its central flash, 5 October 2017

To allow Triton's position on the sky to be determined accurately enough for these observations, a preparatory observational campaign was carried out at the Observatório do Pico dos Dias (Brazil), between 15–23 September 2017. This provided the positions of Triton as it completed one complete orbit around Neptune.

Together with pre-release Gaia DR2 star positions, this gave Triton's position to 3 milli-arcsec at the time of the event, corresponding to 60 km on the Earth, and to 8 seconds in the event's predicted time.

On 5 October 2017, as predicted, Triton's shadow swept across Europe and North Africa, and a few minutes later, the US East Coast. More than a hundred observing stations were ready to observe the event, which had a maximum duration of three minutes.

The event was successfully witnessed at more than 70 sites, 25 of them detecting the central flash. Its shape and spectral dependence are sensitive to the atmosphere's detailed structure, and to the presence of hazes near its surface. The data are still being analysed.

The next occultation by Triton will be by an 11 magnitude star on 6 October 2022, and will be visible from India, China and Japan.

OTHER OCCULTATIONS using the Gaia positions have since been made, amongst them a handful of trans-Neptunian objects, the Centaur Chariklo, and Saturn's irregularly orbiting satellite Phoebe. And many more of these coveted events can now be planned, for example as coordinated by the `Lucky Star project`.

25. The origin of Oumuamua

THE MENTAL PICTURE of our solar system that I had in my mind as a young boy was, of course, absurdly simplistic. Nine planets orbited our Sun. Some were big and gaseous or icy, others were rocky, some were surrounded by moons, and there was also a ring of asteroids and the occasional comet.

Half a century on, my view is very different. I see it as magnificent in its scale, in its architecture, and in its beauty. I see it as bewilderingly and endlessly complex, rich in its complex chemistry, pervaded by a startling variety of remarkable physical phenomena, and whose origin and formation we are only just beginning to comprehend. Everywhere we look, and the closer we look, it seems ever more bizarre, ever more complex, ever more finely balanced, and ever more mysterious.

Major advances in our knowledge have been gained only rather recently: for example, the energy source of the Sun was understood, as nuclear fusion, only in 1927. The vast complexity of the planets and their satellites has been won through deep space missions, observing them with flybys, orbiters, and landers.

And the seeming infinity of its smaller bodies and their complex motions – asteroids, near-Earth asteroids, Kuiper belt objects, trans-Neptunian objects, comets and their deep-space reservoir of the Oort Cloud – continue to surprise both observers and theoreticians.

THE DISCOVERY of the first planets beyond our solar system in the 1990s, and the remarkable advances in discovery numbers, their highly detailed observations, and supporting theories and numerical modelling, has transformed our understanding of the formation of planetary systems in general, and of our solar system in particular. From these multidisciplinary and concentrated efforts we now have a much better 'big picture' of its formation and subsequent evolution.

Four and a half billion years ago, gas and dust in the interstellar medium collapsed into a flattened rotating disk, our Sun forming at the centre, and the planets progressively growing out of the material in the disk as these assembled into ever bigger agglomerations. Closer to the Sun, where temperatures were higher, the rocky planets and their satellites grew out of the less-volatile debris, their gravity competing for any residual material in their orbital dominions. Further out, where temperatures were much lower, the more volatile ices and gases gradually condensed and slowly swept up the vast remaining reservoirs of gas to form the outer giant planets.

Left over from the final assembly of the planets was a vast range of debris still circling the Sun: rocky asteroids closer to it, the small icy lumps of the Oort cloud comets far out in a vast spherical cloud.

Gravitational forces between the bigger bodies and the smaller debris particles acted as slingshots, propelling the latter (and sometimes the former) into perturbed and often unusual orbits. Some of these smaller bodies would have been hurled out at very high velocities, often high enough to escape the gravitational embrace of the host system.

IF THIS BROAD PICTURE is correct – and the enormous weight of today's vast range of observational, theoretical, and modelling efforts make it compellingly so – one specific consequence seems inevitable.

Detailed models of the gravitational scattering that takes place during the later stages of planetary formation, in our solar system and in others, implies that interstellar space should have extrasolar planetesimals and comets passing through it, flung out as a by-product of the star-formation and planetary-formation process.

Some of these, from nearby stellar systems, will end up passing through our solar system. These *interstellar vagabonds* should be recognisable by their extreme hyperbolic orbits, contrasting with the elliptical (bound) orbits of the vast majority of solar system objects.

SERIOUS SEARCHES FOR SUCH objects started around 1990. The usual way of discovering objects in the solar system is from repeated deep imaging of some part of the sky, repeated days or weeks later to see if any objects might be moving. The changing position of a moving object can be measured repeatedly, and its orbit eventually determined and refined.

A handful of possible candidates, including the objects denoted C/2007 W1 and C/1853 E1 discovered around 2010–15, were later rejected on the basis of improved orbits which ruled out hyperbolic trajectories, and categorised them instead as bound to the solar system.

THEORETICAL WORK over the past 20–30 years has attempted to estimate the number of interstellar vagabonds that might be passing through our solar system today. Assuming that each star system ejects 10^{13} km-sized planetesimals, the local space density would be only one object in every 1000 cubic astronomical units. One might pass within 5 au of the Sun every year or so. So not only would they be extremely rare, but they would also be extremely difficult to spot: a 1-km size object at 5 au would be only ~24 mag. It is not surprising that none had been found.

Later estimates used improved models of debris disks, planetesimal scattering, expected size distributions, and assuming the low reflectivity of inactive comets, but with similar conclusions.

THE ASTRONOMY WORLD, and many beyond, were captivated by the announcement of a real such interstellar traveller in October 2017. Named Oumuamua, from the Hawaaian for 'scout', it was discovered with the wide-field survey instrument Pan–STARRS, some 40 days after its closest approach to the Sun.

Still within the orbit of Mercury, and very faint at only 22 mag, it was quickly confirmed as a minor body some 100 metres in size, on a hyperbolic orbit with a very high eccentricity (at $e = 1.192$, the highest known), and with an incoming velocity of around 26 km s^{-1}.


Artist's impression of Oumuamua

ESO/M. Kornmesser/PA

By the end of 2017, only three months after its discovery, more than 30 groups had studied its orbit, and its shape, rotation and surface properties.

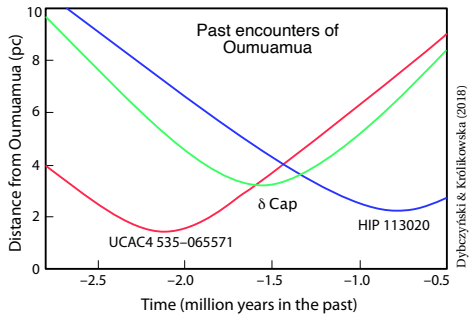
One of the big surprises was its strongly elongated shape (estimated at 5:1, or even 10:1) deduced from photometric measurements. An explanation was quickly forthcoming: in its travel through interstellar space for perhaps a million years, and free from larger impactors, many high-velocity impacts from micron-sized interstellar dust grains, would be energetic enough to repeatedly dislodge splinters and sculpt its elongated shape.

Some speculated that it was an 'alien' source, and radio telescope observations provided upper limits to any transmitter power. Some even suggested a fly-by mission, to be launched within 5–10 yr, and to chase it down before its departure from the solar system.

THERE HAS been much speculation about the origin of this interstellar traveller. But the challenge in pinpointing its origin comes down to our limited knowledge of star positions and their space motions. Traveling through interstellar space at 26 km s^{-1}, it would take a million years to cross 25 parsecs. Throughout that time, all the stars in our solar neighbourhood are themselves moving, and it will have been further deflected by close star encounters on its journey.

Its origin can only be deduced by following its motion backwards in time, taking into account the overall gravitational potential of our Galaxy, and all the individual perturbations from stars it has passed along its way.

In one of the pre-Gaia data studies, Dybczyński & Królikowska (2018) started with 201 763 nearby stars, finding just 109 that Oumuamua would have sped by within 3.5 parsecs. Only seven had an encounter closer than 1 parsec, with most of these occurring over the past 50–100 000 years. The closest, HIP 3757, had a fly-by distance of only 0.04 parsec, and happened 118 000 years ago. It passed by the second closest, GJ 4274, within 0.4 parsec, a mere 23 000 years ago.



THESE ARE UNLIKELY to be the origin of Oumuamua, because the fly-bys occurred with large relative velocities, of around 50–100 km s^{-1}, implying a similarly large ejection velocity. Searching further back in time, and examining more than 200 000 candidate stars, they found only four candidate progenitors. Their most promising, HIP 113020 (GJ 876), is known to host a four-planet system. The encounter occurred 790 000 years ago, with a relative velocity of just 3.9 km s^{-1}.

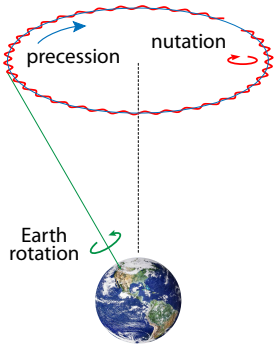
More detailed models along similar lines, but using the Gaia DR2 data, are discussed by Bailer-Jones et al. (2018a). Their closest encounter, at 0.60 pc some 1 Myr ago, was with the same M-dwarf, HIP 3757, with a relative velocity of 24.7 km s^{-1}. They found a more distant encounter, at 1.6 pc and 3.8 Myr ago, but with a lower encounter velocity of 10.7 km s^{-1}, with the G5 dwarf HD 292249. They found only two other stars with encounter distances and velocities intermediate to these.

The origin of Oumuamua remains a mystery, but even more definitive studies will be possible with the Gaia DR3 data. And beyond!

26. Polar motion

THE EARTH'S ROTATION AXIS is not perpendicular to the plane of the planetary orbits, but inclined to it by about 23.°5. Neither does it remain fixed in space, but instead moves as a result of various external and internal forces. The overall spin axis motion is conventionally decomposed, according to their periods, into three components: precession, nutation, and polar motion.

Precession and nutation are caused by the gravitational forces of the Sun and Moon, and even other bodies in the solar system, on the Earth's equatorial bulge. This causes the spin axis to precess, tracing out a circle with a period of about 25 700 years. Hipparchus discovered precession around 130 BC, by comparing his observations to those of the earlier Babylonians.



Nutation describes the smaller oscillations of the rotation axis, which are dominated by the 5° tilt of the Moon's orbit to the ecliptic plane. It has a period of 18.6 years and an amplitude of 17 arcsec, and was discovered by James Bradley in 1728.

More irregular movements of the Earth's spin axis (and its poles) are termed 'polar motion'. These short-term effects were predicted by Swiss mathematician Leonhard Euler in 1765. Using a rigid model of the Earth, he predicted a 10-month oscillation period.

OBSERVATIONAL PROOF of these predictions was obtained in the mid-1880s, when American Seth Carlo Chandler discovered that the dominant term has a 14-month period, and an amplitude of 0.3 arcsec. The difference between Euler's predicted period and the actual 'Chandler wobble' is now known to be due to elasticity of Earth's mantle, and the mobility of the oceans.

All of these effects had to be taken into account when interpreting positional measurements made from the ground, especially when using instruments which interpreted right ascension in terms of the transit times of star images across the meridian line.

FOLLOWING THE DISCOVERY of the short-term motion of the Earth's spin axis, an international programme was set up to monitor its detailed motion. From 1900, the International Latitude Service, established by the International Association of Geodesy, set up a network of astronomical stations.

This started with six observing sites using visual zenith telescopes, at Carloforte, Cincinnati, Gaithersburg, Mizusawa, Tschardjui and Ukiah.

The moving Tropic of Cancer, Mexico

Others joined from the 1930s, and from 1955, some 60–90 stations were being coordinated by the Bureau International de l'Heure (BIH). The ILS was renamed the International Polar Motion Service (IPMS) in 1962, in turn replaced by the International Earth Rotation Service (IERS) in 1988.

The International Astronomical Union (IAU) set up a 'working group' to study Earth rotation in 1978, and developed an international programme to *'monitor Earth-rotation and inter-compare the techniques'* of observation and analysis (MERIT). With the inclusion of VLBI, and both satellite and lunar laser ranging, observations from about 60 ground-based instruments were subsequently coordinated by the Shanghai Observatory.

THESE OBSERVATIONS formed a somewhat inhomogeneous set, with systematics of 0.1–0.2 arcsec. But they provided the only information on the irregularities of the Earth's orientation before the advent of space techniques. Several reductions of the ILS data were undertaken in the past, including a monthly publication of the pole coordinates by the now-defunct IPMS.

Today, optical astrometry from the ground is no longer competitive for measuring these ever-changing effects of polar motion. Instead, modern measurements are now made by a combination of very-long-baseline radio interferometry (VLBI) and GPS, along with lunar laser ranging and satellite laser ranging techniques.

SINCE THE MOVE to space with Hipparcos, the effects of the Earth's precession, nutation and polar motion ceased to be relevant for the construction of these advanced star catalogues. Separating the observing platform from the Earth freed it from the enormous complications introduced by the Earth's irregular rotation.

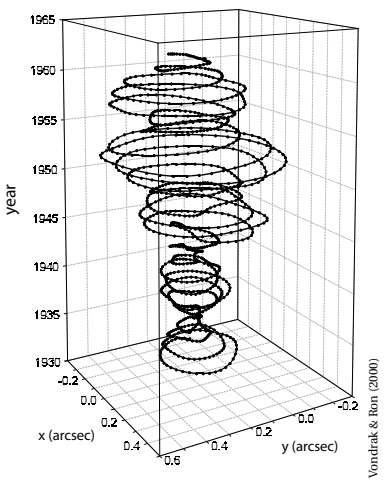
However, a fascinating by-product of the Hipparcos mission was that the catalogue could be used as an accurate reference system extending back in time over more than a hundred years, using the proper motion of each star to propagate their positions backwards to the epoch of the historical Earth rotation observations. Remarkably, Hipparcos provided a framework within which the historical optical observations could be interpolated to yield information on polar motion in the past.

The key point is that, with positions and proper motions accurate to 1–2 milli-arcsec, the accumulated position errors, even over 100 years, would still be at levels of only 0.1–0.2 arcsec. Observations going back to the 1900s could therefore be re-interpreted within a much more accurate reference system than was available then.

COMMISSION 19 of the International Astronomical Union, as the body responsible for monitoring the Earth's rotation, duly created a dedicated group to collect the past observations of polar motion, and to prepare for their analysis in the Hipparcos reference frame.

Their historical data set was eventually based on 4 million observations of the Earth's rotation, acquired with 47 instruments at 30 observatories between 1899–1992. The first results, reduced to a preliminary pre-release Hipparcos catalogue H37, were reported in 1997, with more complete solutions reported in 2000.

The results offer some intriguing insight into the Earth's motion, and indeed into its changing structure, over the past 100 years.



Polar motion 1931–1962 from Hipparcos

One part of this huge historical data set, extending from 1931–1962 and taken from Vondrák et al. (2000), is shown here as a 3d representation.

Time increases along the vertical axis, and the two components of the motion of the Earth's pole are shown as the other two coordinates.

Significant and surprisingly rapid changes and modulations are immediately apparent.

Before looking at what this means for the Earth's rotation, some other effects have to be accounted for. Over the past decades, as observations and theories have improved, other physical phenomena have been recognised that affect the observations of individual observatories. And these have to be corrected before attempting to interpret the motion of the bulk Earth.

Amongst these 'corrections' are several of geophysical origin, including plate tectonic motions, and variations of the local verticals as a result of oceanic tides.

For the former, for example, detailed models allow the plate tectonic motions, at around 0.01–0.05 arcsec per century, to be listed for each observatory. These sorts of corrections are particularly important for observatories lying close to plate boundaries.

THESE HISTORICAL observations reveal some remarkable phenomena at work. The most prominent periodic terms are the 14-month period Chandler wobble. Since these are superimposed on significant annual variations, they lead to a prominent 'beat period', of about 6 years, which modulates their amplitude.

The wobble's amplitude has varied since its discovery, reaching its largest size in 1910 and fluctuating noticeably from one decade to the next. Excursions during the 1950s (seen here) were also particularly prominent.

THEORY TELLS US that the Chandler wobble should die down in a matter of decades. So it has long been realised that there must be driving forces that continually re-excite it. Many possibilities have been considered and debated, including changes in the mass distribution and angular momentum of the Earth's outer core, atmosphere, oceans, or crust.

As one example, simulations show that, from 1985 to 1996, the Chandler wobble was excited by a combination of oceanic and atmospheric processes, dominated by pressure fluctuations on the ocean floors. These, in turn, arise from oceanic circulation caused by variations in temperature, salinity, and wind. And a steady drift, of about 20 m since 1900, has been attributed to the redistribution of water mass as the Greenland ice sheet melts, and to the resulting isostatic rebound, i.e. the slow rise of land formerly burdened with ice sheets or glaciers.

Other contributions to these decadal variations arise from torques between the core and mantle caused by the uneven motions at the core–mantle boundary. Even major earthquakes can cause abrupt polar motion by altering the volume distribution of the Earth's solid mass. And geomagnetic 'jerks', rapid changes of the Earth's geomagnetic field, also appear to contribute.

WILL THE huge leap in Gaia accuracies allow further insights into the interior of our Earth (e.g. Gross, 2019), or even of other planets like Mars, where similar polar motion, with an amplitude of just 10 cm, has recently been observed? I will watch with interest!

27. The Celestial Reference Frame

CLASSICAL ASTROMETRY from the ground, up until the Hipparcos catalogue publication in 1997, could only measure positions – and parallaxes – with respect to other stars nearby on the sky. Even the $6^\circ \times 6^\circ$ Schmidt plates used for the grand photographic sky surveys of the second half of the 20th century, and later HST in space, could only make these relative measurements.

Piecing the measurements together, to form the best global reference network, always left local distortions which varied across the sky. Even the best star positions were found to have systematic errors of 0.2–0.3 arcsec once the Hipparcos reference frame became available.

BUT WHAT WERE these measurements referenced to in the first place? What defines the origin of a stellar reference system? Although the details are intricate, the principles are straightforward, and analogous to the geographical framework of longitude and latitude used to define locations on the Earth's surface. In this equatorial coordinate system, astronomers agree on an origin for 'right ascension' (the equivalent of longitude) and for 'declination' (the equivalent of latitude).

The origin of right ascension was chosen long ago, by Hipparchus around 130 BCE. This 'First Point of Aries', or 'vernal equinox', is one of the two points on the celestial sphere at which the celestial equator (the imaginary circle in the same plane as Earth's equator) crosses the ecliptic (Earth's orbital plane around the Sun). In the same way, declination is defined with respect to the Earth's equator, north and south from 0 to $\pm 90^\circ$.

THE PROBLEM gets more complicated because the Earth's spin axis is not inertially fixed in space, but rotates slowly westward about the poles of the ecliptic, completing one sweep in 26 000 years. This 'precession' causes the equatorial coordinates of celestial objects to change continuously, by about 1° in right ascension over 70 years. The problem is further compounded by the shorter term effects of 'nutation' and 'polar motion'.

This led to the choice of reference systems which were revised, every few decades, by adjusting the epoch at which the Earth's coordinate system was specified. Thus, over the past 200 years, astronomers have used reference systems which were successively specified by the Besselian epochs B1875, B1900, and B1950, and more recently the Julian epoch J2000. Within any such system, the star position itself also changes (due to its proper motion) according to when it was measured.

As position measurements improved, the complex motion of the Earth introduced effects which were increasingly difficult to explain, and to account for.

These wobbling terms include not only the Sun and Moon's gravitational torques of precession and nutation, but a whole host of complex effects responsible for polar motion: some internal to the Earth, others forced by climatic and seasonal changes due to oceans, tectonic plate motions, and many others.

This led, in turn, to efforts to construct a 'dynamical reference system', linked to the observed motion of solar system bodies, whose orbits around the Sun should be largely decoupled from the motion of the Earth.

By the 1990s, radio VLBI measurements became possible, for a few dozen radio stars and quasars, at higher accuracies than were possible using optical measurements from the Earth. In consequence, the celestial reference system adopted by the International Astronomical Union moved to one defined at radio frequencies and, in particular, one tied to distant quasars which better represented the ideas of an inertial reference system, decoupled from the Earth's wobbling motion.

BY MAKING POSITIONAL measurements from space, Hipparcos and Gaia achieve vastly improved accuracies from above the Earth's perturbing atmosphere. At the same time, measurements from a space platform means that they were, at a stroke, freed from the hugely complicating effects of the Earth's spin-axis motion.

A further central technique used by both Hipparcos and Gaia is their two widely-separated fields of view on the sky, which are superimposed in a common focal plane. As set out by Pierre Lacroute in his first ideas for space astrometry in 1968, and further developed by Lennart Lindegren in the 1970s, a carefully chosen angle between the two – one which is not a simple rational

fraction of 360° (Hipparcos used 59°, Gaia used 106°), leads to an extremely rigid reference frame over the entire sky. The consequences are far reaching, in that the parallax of every star is 'absolute'. In other words, a star's parallax is no longer defined relative to that of another; each has the same 'offset', or zero-point, as every other.

TWO PROBLEMS remain. The first is to establish the zero-point of this parallax scale. The second is related, and particularly awkward: it turns out that, as the satellite scans the sky, any tiny changes in the angle between the two viewing directions (specifically, if phased with the sixth harmonic of the spin frequency) the resulting effect is indistinguishable from a common offset in the parallax zero point.

Unfortunately, this is precisely the effect that results from the Sun's changing illumination acting on the spinning satellite. Many details of the instrument design, both for Hipparcos and for Gaia, were driven by efforts to decrease this dependency, but the fact remains: any tiny changes in this angle can propagate through to a tiny shift in the zero-point of the totality of parallaxes.

AS THE DEFINITION of the Hipparcos observing programme took shape in the early 1980s, plans were put in place to include stars that could be used, once the catalogue was finalised, to link the rigid reference frame defined by its 120 000 stars to an extragalactic 'inertial' reference framework, and in the process estimate and correct for any tiny offset in the parallax zero-point.

The goal was, in other words, to determine the global orientation and rotation (or spin) of the coordinate frame defined by the Hipparcos positions with respect to extragalactic sources. The big difficulty was that, because of its limiting magnitude of about 12 mag, only one quasar, 3C 273, could be included in the observing programme, and even that was so faint that it contributed very little to the final link.

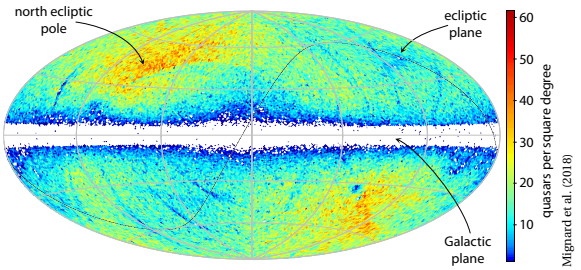
The effort required to establish this link was substantial. The contributions of several groups over a number of years, and using a variety of less direct techniques, were essential (Kovalevsky et al., 1997).

These indirect methods included interferometric observations of radio stars by radio interferometry (VLBI, MERLIN and VLA); observations of quasars relative to Hipparcos stars using CCDs, photographic plates, and Hubble Space Telescope; photographic programmes to determine stellar proper motions with respect to extragalactic objects; and a comparison of Earth orientation parameters obtained by VLBI and others.

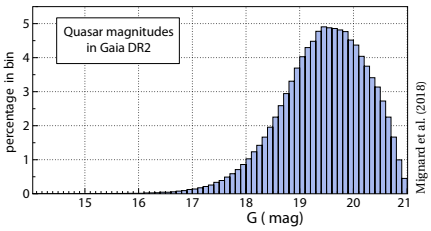
Combined and suitably weighted, the coordinate axes of the published catalogue were finally believed to be aligned with the extragalactic radio frame to within ±0.6 mas at the mid-catalogue epoch J1991.25. And it was estimated to be 'non-rotating' with respect to distant extragalactic objects to within ±0.25 mas/yr.

THE PROBLEM IS as central to Gaia as it was for Hipparcos, and the accuracy for the link correspondingly more demanding. But there is one very big difference: Gaia's limiting magnitude, at 20–21 mag, allows very large numbers of quasars, all across the sky, to be observed by the instrument itself. And this means that the problem can be tackled much more directly.

The challenge was considered in depth during the mission's feasibility study before its selection in 2000. Studies then indicated that some 500 000 quasars would be observable directly by Gaia, with a mean density on the sky of about 25 per square degree. Issues of sky uniformity, colour dependency, and possible small structural changes in position were all considered.



The second release of Gaia GDR2 contains the positions of 556 869 quasars, extending to $G = 21$ mag, and defining a kinematically non-rotating reference frame in the optical (Gaia Collaboration et al., 2018c). A subset have accurate VLBI positions allowing the reference frame axes to be aligned with the International Celestial Reference System (ICRF) radio frame.



Median positional uncertainties are 0.12 mas for $G < 18$, and 0.5 mas at $G = 20$. Large-scale systematics are in the range 20–30 μas. The optical positions for a subset of 2820 sources in common with the ICRF show very good overall agreement with the radio positions.

SO ALREADY IN 2018, based on less than 40% of the data from the nominal 5-year Gaia mission, we have the first realisation of a global, non-rotating optical reference frame that meets the ICRS prescriptions, being built only (and directly) on extragalactic sources. Its accuracy matches that of the current radio frame of the ICRF – but with a much higher density of sources.

And such an accurate reference frame may have cosmological implications previously considered unimportant and unmeasurable, such as detecting the tumbling of a triaxial dark matter halo (Perryman et al., 2014b).

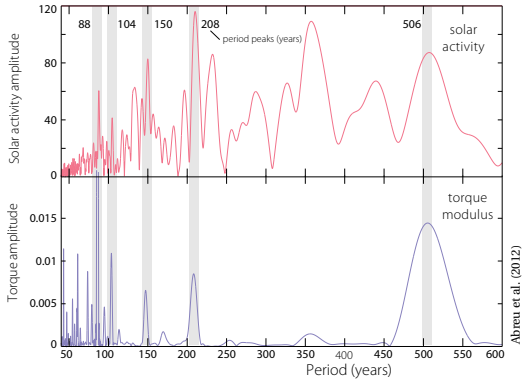
28. Solar activity – and dark matter?

THE ROTATION of the Sun is central to the two main hypotheses which try to explain the 11-year solar activity cycle. Present ideas are that the activity cycle is related either to a turbulent dynamo operating in or below the Sun's convection envelope, or to a large-scale oscillations of a fossil magnetic field in its radiative core.

However, the precise nature of the solar dynamo, and the details of the associated solar activity (such as the details of the sun spot cycles, or the prolonged Maunder-type solar minima) remain unexplained.

MEANWHILE, VARIOUS investigations (since the 1850s) have long hinted at some sort of link between the Sun's motion around the centre of mass of the solar system, and various solar variability indices.

One example is shown here. It compares the periodicity of variations in solar activity (top) with the corresponding changes in a measurement of planetary torque due to the motions of the most massive planets in the solar system (bottom).

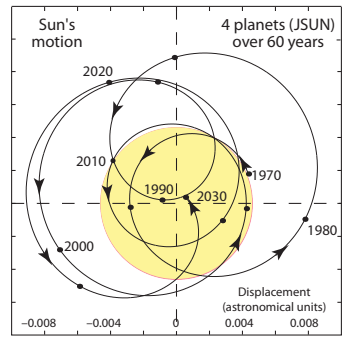


In more detail, since the 1960s, acceleration in the Sun's own orbital motion, due to the motion of its orbiting planets, has been linked to phenomena such as the Wolf sun spot number counts climatic changes, the 80–90 year Gleissberg cycles, the prolonged Maunder-type solar minima, short-term variations in solar luminosity, sun spot extrema, the 2400-yr cycle inferred from ^{14}C in tree-rings, hemispheric sun spot asymmetries, oscillations in long-term sun spot clustering, violations of the Gnevishev–Ohl sun spot rule, cosmogenic radionuclide correlations over 9400 years, variations in total solar irradiance since 1978, and strong planetary-like cycles at radio frequencies. Strangely, claimed effects even extend to dust storms on Mars, and river discharges.

THIS BARRAGE of technical terms is simply to underline that, over the past 50 years, many have argued that a link might exist. But while the figure shown here, and others, might suggest that some of the peaks do coincide, others do not. This confusing picture has led some to argue that a connection must exist, while others have rejected these claims as statistically unconvincing.

Importantly, there is also no known physical process which might connect the motion of the planets with physical effects occurring on the Sun. As a result, many scientists who might look at this question would probably dismiss it as being unworthy of further investigation.

A SPECIFIC CURIOSITY of the Sun's barycentric motion is evident in the next figure. Around 1990 (and before that, in 1811 and 1632) the Sun had a *retrograde* (backwards) barycentric motion, i.e. its angular momentum with respect to the centre of mass was negative.



Various attempts have been made to identify a coupling mechanism between solar rotation and this retrograde motion, e.g. invoking tidal forcing, or some sort of effects on the 'tachocline', the transition region between the radiative interior and the differentially rotating outer convective zone. These ideas have been contested by others. In any case, a physical picture relating the Sun's rotation and its orbital motion remains unclear.

LET ME summarise the story so far. On the one hand, a full theory of the solar dynamo, one which explains and predicts the time variations of solar activity, does not yet exist. On the other, some studies over the past few decades has hinted at a possible connection between solar activity and the motions of the planets.

Again, let me stress, we have no ideas as to why the motions of the planets might affect solar activity.

But in science, what may appear as crazy ideas, departing from current wisdom, do occasionally turn out to have substance, setting foundations for new theories with even greater explanatory and predictive power.

A FIRST STEP might therefore be to see whether we can test the idea that planetary motions might affect solar activity – whether we believe it or not, and whether we have a theory for it or not.

In this spirit, Perryman & Schulze-Hartung (2011) showed that this idea could indeed be tested in other exoplanetary systems. Systems showing large changes in the star's acceleration due its orbiting planets should also show related signatures in the activity indices of the host star, if this hypothesis is correct.

The exoplanetary systems HD 168443 and HD 74156, for example, both have periods when the changes in orbital angular momentum of their host stars exceed that of the Sun by 5 orders of magnitude. They could offer an independent test of any link between a star's barycentric motion and stellar activity. But, to my knowledge, no such tests have been made so far.

THIS STORY now takes another turn. In December 2019, I was contacted by a scientist working at CERN, Konstantin Zioutas, who drew my attention to his recent paper in the journal *Physics of the Dark Universe*, with the title *'The Sun and its planets as detectors for invisible matter'* (Bertolucci et al., 2017) and was keen to hear my opinion of it.

Their work had looked at the incidence of solar flares over several decades (both what are called M flares and X flares), and found a significant relationship between this type of solar activity, and the positions of Mercury, Venus, and the Moon. And they tied this behaviour to the possible effects of dark matter.

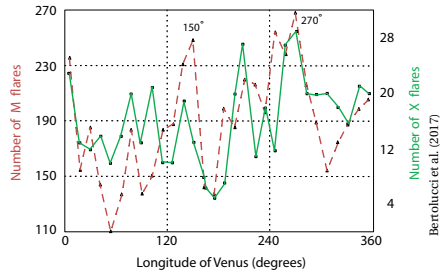
The detection and characterisation of dark matter is, of course, one of the central challenges in modern physics. The strongest evidence for it comes from large-scale gravitational effects, but more direct searches for it have so far provided no convincing evidence.

While observations, and cosmological theories, suggest that the dark matter halo in the Galaxy is distributed rather uniformly, there is also evidence for ancient stellar streams in the outer regions of our Galaxy which represent smaller galaxies captured by it billions of years ago. These captured galaxies presumably also contained entrained dark matter.

THEIR IDEA was as follows. For certain dark matter particle masses, streaming towards the Sun with velocities $10^{-4} - 10^{-3}$ times the speed of light, the particles would be gravitationally focused by a planet. If there are preferred directions in the dark matter streams, then more pronounced solar activity would be expected at certain planetary heliocentric longitudes.

Their work focused on the two inner planets, Mercury and Venus, given their relatively short orbital periods compared to a solar cycle of about 11 years. They found, since 1976, over four solar cycles, statistically significant solar activity signals when one or more planets have heliocentric longitudes between 230 – 300°.

Their hypothesis, then, is that the activity of the Sun is triggered by the influx of invisible massive matter, and that this matter has some preferred direction or stream, which gets gravitationally lensed by the planets.



UNABLE TO SEE any basic flaw in their arguments, I suggested two tests that could be made, again by appealing to other planet-hosting star systems.

The first would be to look at the stellar activity of nearby exoplanet systems, to see if it is associated with regular alignments of its planets with some preferred direction in space. And if the hypothesised dark matter stream is related to large-scale 'infall' structure in the Galaxy, then flare activity in those systems should be correlated with the same Galactocentric longitude.

The second appeals to an examination of Gaia data. Let us hypothesise that the various solar variability indicators are indeed revealing the presence of infalling dark matter, and that the dark matter flow correlates with the path of some ancient stellar stream. Then there could be evidence for correlated space motions of ancient low-metallicity stars following this same dark matter path.

I don't have enough understanding of both sides of the problem, dark matter physics and stellar halo streams, to decide whether the hypothesis has some obvious flaw. I sent an outline of the problem, and my suggested tests, to a number of colleagues who perhaps had the tools at hand to examine it more carefully. I await further insights into the question with interest!

—

Postscript, 10 Mar 2022: a study by Edmonds (2022) argues that the inclusion of Planet 9 in the Sun's barycentric motion improves the correlation with solar activity cycles.

29. White dwarf surveys

WHITE DWARFS are the endpoint of stellar evolution for some 97% of all stars, specifically those below about 8 solar masses. Most are dominated by C or C/O cores, which result from the exhaustion of nuclear fusion for these stars. They have a narrow mass distribution peaking around 0.58 solar masses. White dwarfs of lower mass are expected to have He cores.

No longer held up by nuclear fusion, they have collapsed to a small and very dense state, with a mass comparable to that of the Sun, but with a size closer to that of the Earth. They have very low luminosities, attributable to the slow release of their residual thermal energy.

White dwarfs are of great importance across many fields of study: for theories of star formation and evolution, of degenerate matter at extremely high density, for distance scale determination, and for understanding planet survival beyond a star's main-sequence lifetime.

THEIR CLASSIFICATION consists of an initial D, followed by a letter describing the dominant spectral feature. Thus DA dwarfs have atmospheres dominated by H I, DB by He I, DC have a continuous spectrum, DO are dominated by He II, DQ by carbon, and DZ by metal lines.

They span a vast temperature range, with the DA class ranging from 170 000 K to 4500 K, falling slowly, over billions of years, as the white dwarf cools. Consequently, their temperature provides a direct age indicator. They overlap instability regions like those seen around the main sequence, including DA dwarfs (and the very common pulsating ZZ Ceti stars), DB dwarfs, and DO dwarfs (including pulsating GW Vir stars).

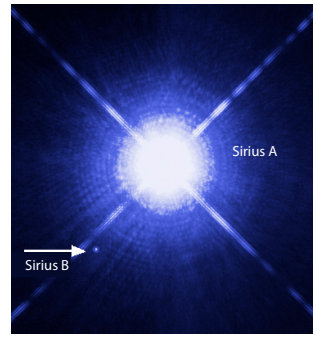
WHITE DWARFS are very common, and accordingly very numerous in the solar neighbourhood. But their very low luminosities means that any survey completeness falls rapidly with increasing distance, even within 20–50 pc. Since all reasonably bright dwarfs are also relatively nearby, their parallax distances measured from the ground were already of reasonably high relative accuracy, even in advance of the Hipparcos mission.

Catalogue of white dwarfs have been maintained by George McCook & Edward Sion since 1987, when the known count stood at 1279. The fourth edition in 1999 listed 2249, and the 2016 on-line version 14 000. The Sloan SDSS DR7 White Dwarf Catalogue lists 20 000 objects, with 13 000 DA and 1000 DB spectral types.

WHITE DWARFS PROVIDE an important insight into the behaviour of matter at extreme densities. No longer supported by nuclear fusion, they consist of a 'degenerate electron gas' at densities of $10^6 - 10^8$ gm cm^{-3}. This results in a relation between mass and radius first derived by Subrahmanyan Chandrasekhar in 1931, and later refined to included different chemical composition (He, C, Mg, Si, S, and Fe), and models with C or C/O cores and different configurations of H and/or He layers.

The mass–radius relation remains a largely theoretical construct, with observational confirmation still resting on only a handful of objects with accurately-known masses and radii. But it is a central assumption in studies of their mass distribution and luminosity function, and for a range of related applications including distances to globular clusters, ages of the Galactic disk and halo from white dwarf cooling sequences, dark matter investigations, and establishing limits on any variations of fundamental physical constants, notably $\dot{G}/G$.

The occurrence of white dwarfs in visual binaries provides one way to determine their masses and radii: masses from radial velocities and modelled orbits through Kepler's third law, while radii can be derived from their effective temperatures if their distance is known.



Hubble Space Telescope

At just 2.6 pc distance, Sirius is the prototype of the important class of Sirius-like binaries, comprising a main sequence star and a white dwarf secondary, with an orbital period of 50 years.

SINCE CHANDRASEKHAR'S work almost a century ago, it has been clear that progress in this field requires a much larger sample of white dwarfs with much better estimates of their masses and radii.

Hipparcos was able to make only a modest contribution, because their low luminosity, combined with the relatively bright limit of the satellite observations of 10–12 mag, meant that it could observe only 22 of the nearest. Of these, 11 were field white dwarfs, 4 were in visual binaries, and 7 were in common proper motion systems. The majority were of spectral type DA, but with one each of the spectral types DB, DC, DQ, and DZ.

Pre-Gaia, then, a total of some 20 000 white dwarfs were known, with around 200 within 20 pc. Pre-launch models suggested that Gaia could increase the numbers tenfold to perhaps 200 000 objects, with completeness to around 20 mag, and to distances of at least 100 pc.

ALREADY BY THE end of 2020, more than 100 scientific papers have examined different aspects of white dwarf science based on the Gaia results, using the advances of Gaia astrometry or photometry to place new constraints on previously-known white dwarfs.

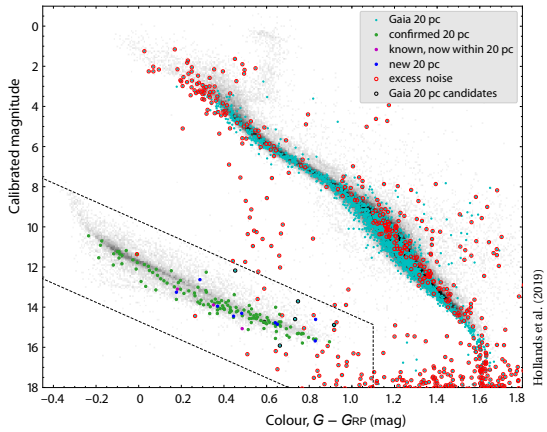
Amongst these are studies of the mass–radius relation, the discovery of new white dwarfs in binary and triple systems (including the first *triple* white dwarf system), the discovery of new white dwarfs with dusty debris disks (related to mature planetary systems), new candidates in the Hyades and Praesepe clusters as well as in the Hercules stellar stream, searches for white dwarfs in the Galactic disk, in the halo, and in other star clusters, studies of the kinematics of the Galaxy disk in the solar neighbourhood, and evidence for massive white dwarfs resulting from merger events.

BUT HERE, I will just take a look at some broad statistics from early analyses of the astrometric and photometric data of Gaia DR2. Although much better quality data is to come, we can already see that Gaia is making a major contribution to this important field.

Jiménez-Esteban et al. (2018) found 73 221 candidates from their position in the Hertzsprung–Russell diagram. Of these, 8555 are within 100 pc, yielding the largest and most complete volume-limited sample to date. Dominated by cool (< 8000 K) objects, they found 8343 C/O-core and 212 O/Ne-core candidates, and an overall space density of $4.9 \pm 0.4 \times 10^{-3}$ pc^{-3}.

Noteworthy features include a bifurcation in the Hertzsprung–Russell diagram not predicted by current theories, and a significant number of massive ($0.8 M_\odot$) white dwarfs whose origin remains uncertain.

Out to larger distances and fainter magnitudes, Gentile Fusillo et al. (2019) identified 260 000 candidates to 21 mag, estimating 85% completeness for $G < 20$ mag and $T_{\rm eff} > 7000$ K, at Galactic latitudes above 20°.



The Gaia white dwarf sample within 20 pc

Out to 20 pc from the Sun, Hollands et al. (2018) used positions in the colour–magnitude diagram to identify139 systems, nine of which are new, with the closest at only 13.05 pc. They estimated the local white dwarf space-density to be $4.49 \pm 0.38 \times 10^{-3}$ pc^{-3}.

Kim et al. (2020) compiled a catalogue of 531 candidates having large transverse motions relative to the Sun (above 200 km s^{-1}), and therefore likely to be members of the local Galactic halo population.

A NUMBER OF particularly interesting objects already feature amongst this remarkable haul.

Tremblay et al. (2020) used spectroscopy of 230 new candidates out to 40 pc to confirm 191 as real. Amongst these are 89 DA, 76 DC, and 2 DQ white dwarfs. Amongst their 14 new DZ (metal-rich) white dwarfs is the first ultra-cool object with metal lines. Three show at least four different metal species. One is strong in Fe and Ni, features now taken to indicate the recent accretion of a planetesimal-type body with core-Earth composition.

About 150 extremely low-mass white dwarfs, with $M < 0.3 M_\odot$, were known before Gaia. The Universe is not old enough for these to have formed as single stars, but rather imply a common-envelope binary, or following mass-overflow in a multiple system. Most will merge over a few billion years, each final merger being a strong source of gravitational waves. Recent theories have predicted a much larger space density of these objects, and Pelisoli & Vos (2019) duly used the Gaia DR2 data to derive a much-enlarged sample of 5762 extremely low-mass candidates, with $M < 0.3 M_\odot$.

Vincent et al. (2020) combined the 260 000 white dwarf candidates found from DR2 with ground-based photometry to measure the temperatures and masses for all white dwarfs in the northern hemisphere within 100 pc. From a sample of ZZ Ceti candidates within the instability strip, 90 were observed with high-speed photometry to reveal 38 new ZZ Ceti stars, including two very rare ultra-massive pulsators.

30. The motion of globular clusters

THE SPACE MOTIONS of globular clusters in our Galaxy, and of the dwarf spheroidal galaxies in orbit around it, depend on our Galaxy's gravitational potential, and therefore its mass distribution. Together with knowledge of their chemistry and ages, these provide strong constraints on theories of formation of our Galaxy, including when and how the halo and disk actually formed.

Studies of the motions of stars *within* globular clusters, based on their proper motions, extend back to the days of photographic plate measurements more than a century ago. Establishing and understanding the nature of their Galactic orbits has been a problem tackled since the 1950s. By the 1980s, studies had revealed systematic orbital motion for more than 60 clusters, velocity dispersions increasing with distance from the Galactic centre, and a Galactic rotation curve revealing the existence of a massive Galaxy halo extending out to at least 30 kpc.

In the 1990s, Lynden-Bell & Lynden-Bell (1995) identified streams of clusters tracing out the orbits of satellites that had long since merged with the Galaxy, including streams associated with the Magellanic Clouds, Fornax, and the Sagittarius dwarf. And they listed 22 clusters whose bulk proper motions were predicted to exceed 1 mas yr^{-1}.

THE TOTAL MASS of our Galaxy itself can be determined, through Newton's laws, from the orbits of these distance objects. When Gaia was under consideration by ESA in 2000, the best mass estimates were based on the known motions of 27 systems beyond 20 kpc.

Estimates at that time concluded that the total mass of our Galaxy is $2.3^{+3.9}_{-1.6} \times 10^{12} M_\odot$, while the mass within 50 kpc is $5.5^{+0.1}_{-1.1} \times 10^{11} M_\odot$. These uncertainties meant that the mass, and extent, of the Milky Way halo was one of the least well-known of all Galactic parameters.

Further advances in understanding cluster orbits have been limited by incomplete knowledge of their space velocities. Specifically, although their distances and radial velocities could be estimated, a complete orbit determination also requires the two components of the transverse motion on the sky, along with the gravitational potential in which the clusters move.

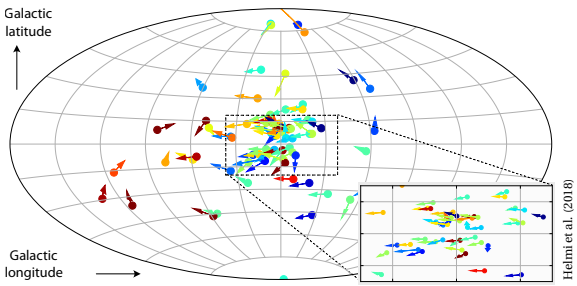
NO STARS in globular clusters were bright enough to be observed by Hipparcos. But in the mid-1980s, during the preparation of the Hipparcos input catalogue (the list of stars defining its observing programme), careful effort was devoted to including a number of bright reference stars lying close to a small number of selected globular clusters, specifically to provide an inertial reference frame for studies of their space motions.

In the absence of nearby Hipparcos stars, other studies have had to try to rely on background quasars and distant galaxies to define an 'absolute' reference frame.

With the Hipparcos results in 1997, some progress in understanding and explaining their orbits was duly made in the case of 38 halo clusters (Dinescu et al., 1999), and four low-latitude inner Galaxy clusters (Dinescu et al., 2003), They found, for example, space velocities all smaller than the escape velocity of the Galactic bulge, concluding that all were confined to the bulge.

WITH THE developments over the past 20 years, what were the insights expected from Gaia?

It was anticipated that Gaia's high-accuracy astrometry of globular clusters would shed further light on: the mass of our Galaxy; their formation and evolution (e.g. which formed *in situ* and which were accreted); the effect of external tides and of physical processes within the cluster; the existence of remnant streams; whether they formed in mini-halos, or are devoid of any dark matter; and whether they host intermediate mass black holes.



Space motions for 75 globular clusters observed by Gaia (colours are related to their radial velocities)

A FIRST INDICATION of what the Gaia data will mean for globular cluster studies was made by Gaia Collaboration et al. (2018b). They used the Gaia DR2 astrometry, parallaxes and proper motions, to examine the space motions of 75 Galactic globular clusters. This represents about half of those known in our Galaxy, and focused on the most nearby clusters, within about 12–13 kpc. Radial velocities, measured by Gaia, are available for 57.

Selecting according to distance and proper motions, Helmi et al. could identify stars out to each cluster's tidal radius, and down to a limit of about 20 mag.

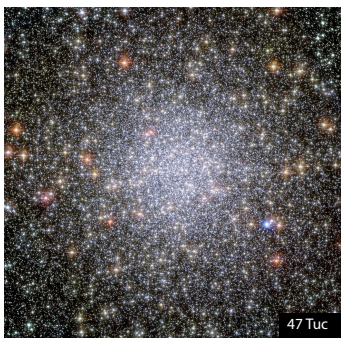
The sheer numbers of individual stars that could be measured by Gaia in each of these clusters is worthy of mention, especially when contrasted with the inability to measure even a single star in any globular cluster by ESA's trailblazing astrometry mission, Hipparcos.

With Gaia, *several thousand* star members are detected and measured in each cluster. Some 20–30 000 are identified in a number, and more than 60 000 are counted in the case of the second nearest globular cluster known (after ω Cen), NGC 104. Also known as 47 Tucanae, or 47 Tuc, this is some 4000 pc distant, around 40 pc in diameter, and also visible with the naked eye.

Mean globular cluster distances can then be derived. For example, for the nearest, 47 Tucanae, the mean parallax is 0.1959 ± 0.0002 milli-arcsec, corresponding to a distance of 5105 ± 5 pc. Remarkable!

THE BULK PROPER MOTIONS derived for each cluster are some one or two orders of magnitude larger than their parallaxes, and thus the measurements are typically both robust and significant. Their location in Galactic coordinates, and their space motions, could then be derived for all 75 clusters (see figure).

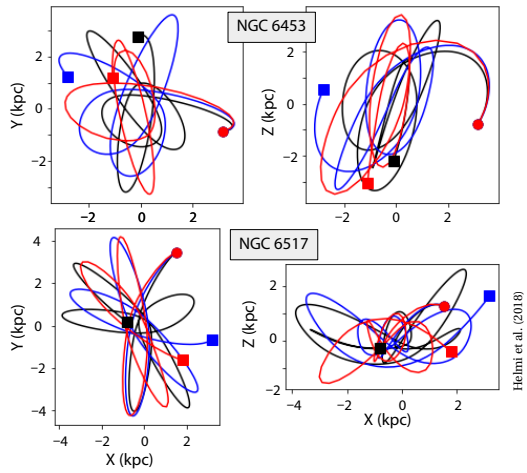
The outstanding quality of the DR2 data, together with the absolute reference frame (i.e. free of expansion and rotation) in which the proper motions are defined, has allowed the clear detection of rotation in five of their 75 globular clusters. For three of them (NGC 104, NGC 5139, and NGC 7078), this was already known, and rotation was also detected in NGC 5904 and NGC 6656.



NASA, ESA, and the Hubble Heritage Collaboration

Other interesting velocity structures can be seen in many. NGC 3201, for example, shows a very pronounced 'perspective contraction', a geometrical effect due to its large radial motion and its relatively large parallax.

In NGC 6397, considered to have been subject to core collapse, there is evidence of an expanding outer halo.



Galactic orbits of two globular clusters (face-on and edge-on) (different colours are for different Galaxy models)

THEY THEN used these positions and space motions, along with various state-of-the-art models of the Galaxy's mass distribution (for example comprising a stellar bulge, star and gas disks, and a dark matter halo), allowing Helmi et al. to follow their orbits backwards in time around the Galaxy over the past 250 million years.

The rather small differences in these predicted orbits over short timescales implies that the post-Gaia ability to predict their past orbits is impressively good. These predictions might therefore be used to search for the 'tidal tails' of some of the clusters, i.e. stars slowly lost from the cluster as it ploughs its way around the Galaxy.

Future studies of the orbital properties of the more distant globular clusters will also allow their relation to the present-day dwarf galaxy population to be examined. And if some of the globular clusters observed with Gaia are associated remnants of long-gone accreted galaxies, their stellar debris might be discovered and characterised using data from DR2 (and beyond).

WHILE FURTHER analysis, and much better data from Gaia, are both still forthcoming, it is now firmly expected that the Gaia data will eventually allow determining the mass distribution of our Milky Way Galaxy far out into the extremities of its dark matter halo.

In more technical terms, the data should allows resolving the degeneracy between the slope of the mass density profile and the orbital anisotropy. Its unknown anisotropy has been the limiting factor in attempts to pin down our Galaxy's mass to date, but Gaia DR2 has shown that this should soon be an observable quantity.

It is worth stressing that, compared to the pre-Gaia era, errors have been reduced by ~100. Gaia Collaboration et al. (2018b) concluded: *'The measurements for these clusters are of outstanding quality, with the formal and systematic uncertainties being effectively negligible'.*

31. The motion of dwarf spheroidals

DWARF SPHEROIDAL galaxies, or dSph, are small, low-luminosity galaxies comprising an old stellar population with very little dust. In contrast to dwarf elliptical galaxies, they are roughly spheroidal in shape. Some two dozen are known as companions to either the Milky Way or to the Andromeda Galaxy (M31). They are named after the constellation in which they are found.

The first known, Sculptor and Fornax, were discovered by Harlow Shapley in 1938, where he described them as *'unlike any known stellar organisation'*. But despite weighing in at around 10^7 solar masses, further discoveries were challenged by their low luminosities and low surface brightnesses.

ESO/Digitized Sky Survey 2

By the late 1990s, their rarity seemed to be in conflict with the ΛCDM (Lambda cold dark matter) cosmological model, which predicted that massive galaxies like the Milky Way should be surrounded by many dark matter dominated satellite halos.

This conflict eased with the discovery of around a dozen very faint Local Group dwarfs from the Sloan Digital Sky Survey around 2000, and a similar number discovered by the Dark Energy Survey around 2015.

THE DISTINCTION between dwarf spheroidals and globular clusters is not always sharp: one discriminant may be the presence of a significant amount of dark matter in the former, and its absence in the latter.

As typified by Sextans and Hercules, their orbits, structure, and internal dynamics, often appear to be affected by the gravitational forces of the galaxy (either the Milky Way or M31) that they are orbiting.

Better knowledge of their structures and orbits would have numerous implications, ranging from the scale of the formation of the smallest galaxies in the Universe to constraints and challenges for cosmological models, to the effect of the environment on their dynamical and chemical evolution, and to constraints on the form of the hot gaseous halo of the Milky Way.

THESE KINDS OF specific studies were keenly anticipated in the scientific case for Gaia presented to the ESA advisory committees at the time of its selection in 2000, when just eight dwarf satellite galaxies were known. As we stated there:

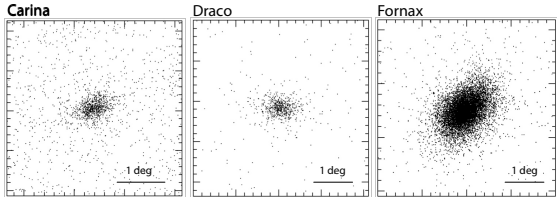
> "These dwarf spheroidal galaxies provide key dynamical tracers of the outer mass distribution of the Milky Way, at larger distances than any other available tracer. For the nearer dwarfs, especially Ursa Minor, Gaia will allow internal dynamical studies… Ursa Minor is unique among the dwarf spheroidal galaxies in showing marginal evidence for minor axis rotation, an indicator of possible triaxiality, tidal perturbation by the Milky Way, or non-isothermality in the dark matter… 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."

IN THEIR STUDY of the motions of globular clusters and dwarf spheroidal galaxies with Gaia DR2, Gaia Collaboration et al. (2018b) examined the nine 'classical' dwarf spheroidals as examples of what can be achieved.

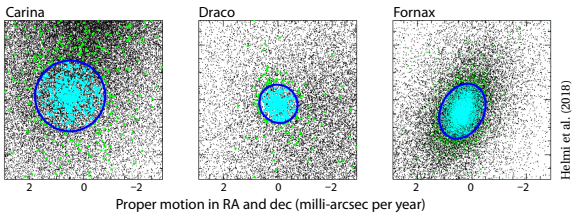
Their selection of stars as candidate galaxy members proceeded by selecting Gaia objects within a one or two degree field satisfying the two most basic criteria: from the Gaia astrometry, as lying within 2σ of the system's mean proper motion. And from the Gaia photometry, as stars populating the red giant branch and blue horizontal branch of these old stellar populations.

This possibility of selection according to proper motion reveals, in many cases, asymmetries in the distribution of the stars on the sky.

Thus, in the examples shown here, there is an indication of tidal streams in the case of Carina, and there are spatial asymmetries in the case of Fornax.



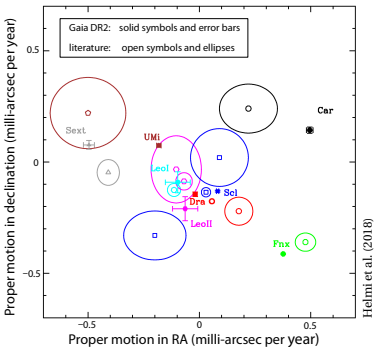
Helmi et al. (2018)



In the maps of proper motion, where the same three example galaxies are shown here, stars surviving the astrometric and photometric selection criteria are shown as cyan dots. They clearly clump much more strongly in the diagrams than the likely non-members (shown as black points).

The extension in proper motion space is, however, considered most likely due to the errors of the present proper motions. The blue ellipses correspond to a 3σ dispersion around the mean values. Green points correspond to stars that fall within the photometric selection criterion, but outside the astrometric cut-off.

THESE DWARF spheroidal galaxies are typically very distant, ranging from about 26 kpc in the case of Sagittarius, to around 250 kpc in the case of Leo I, well beyond the Magellanic Clouds. Their bulk proper motions are consequently very small, rarely reaching 0.5 milli-arcsec per year. The best previous determinations of these tiny motions have mostly been made possible from the Hubble Space Telescope.



The agreement is reasonable, but the Gaia errors are much reduced. This is especially true for the galaxies for which more than a few hundred (and up to several thousand) members have been identified, such as Carina, Draco, and Fornax. Compared with HST observations, Gaia has the advantages of covering the entire galaxy, with the proper motions being in an absolute reference frame. The proper motion of the 'ultra-faint dwarf' Bootes I is also determined for the first time.

WHAT I FOUND MOST fascinating about these early results was the Galactic orbits that Gaia has been able to illuminate for these dwarf spheroidal galaxies.

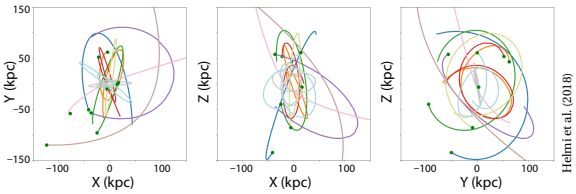
Gaia Collaboration et al. (2018b) used their positions and space motions, with various state-of-the-art models of the Galaxy's mass distribution (for example comprising a stellar bulge, star and gas disks, and a dark matter halo), to follow their orbits backwards in time around the Galaxy over the past 250 million years.

Gaia DR2 distances to dwarf spheroidal galaxies

Name	X [kpc]	Y [kpc]	Z [kpc]
Fornax	$-33.1^{+2.6}_{-2.7}$	$-51.1^{+4.1}_{-4.2}$	$-134.5^{+10.8}_{-11.0}$
Draco	$4.0^{+0.3}_{-0.3}$	$62.6^{+5.2}_{-4.5}$	$43.5^{+3.6}_{-3.1}$
Carina	$-16.7^{+0.9}_{-0.9}$	$-95.7^{+5.0}_{-5.3}$	$-39.7^{+2.1}_{-2.2}$
Ursa Minor	$-13.9^{+0.5}_{-0.6}$	$52.1^{+2.1}_{-2.0}$	$53.6^{+2.2}_{-2.0}$
Sextans	$-28.4^{+1.4}_{-1.3}$	$-57.0^{+2.8}_{-2.5}$	$57.9^{+2.6}_{-2.8}$
Leo I	$-115.5^{+7.6}_{-7.2}$	$-119.6^{+7.9}_{-7.4}$	$192.0^{+11.9}_{-12.6}$
Leo II	$-69.0^{+3.9}_{-3.8}$	$-58.3^{+3.3}_{-3.2}$	$215.2^{+11.9}_{-12.3}$
Sagittarius	$25.2^{+2.0}_{-1.8}$	$2.5^{+0.2}_{-0.2}$	$-6.4^{+0.5}_{-0.5}$
Sculptor	$3.1^{+0.2}_{-0.2}$	$-9.8^{+0.7}_{-0.7}$	$-85.4^{+5.7}_{-6.1}$
Bootes I	$22.7^{+1.1}_{-1.0}$	$-0.76^{+0.03}_{-0.04}$	$61.0^{+2.8}_{-2.7}$
LMC	$7.1^{+0.3}_{-0.3}$	$-41.0^{+2.0}_{-2.0}$	$-27.8^{+1.4}_{-1.4}$
SMC	$23.3^{+0.9}_{-0.9}$	$-38.1^{+1.5}_{-1.5}$	$-44.1^{+1.7}_{-1.7}$

Detailed inspection shows that Draco and Ursa Minor have very similar orbits, and possibly constitute a physically connected group. However, the orbital planes of most others are different, with that of Sagittarius being orthogonal to those of Draco and Ursa Minor.

Most, it turns out, are on (slightly) prograde orbits, while Fornax is retrograde, qualitatively similar to what has been found for globular clusters. However, their orbital eccentricities are very different. Few have very eccentric orbits, with Carina even somewhat circular.



ALL THIS leads to two important conclusions. First, there is only a weak similarity, if any, between the orbits of globular clusters and dwarf spheroidals. Second, their eccentricity distribution is inconsistent with the findings of recent cosmological simulations, where they are predicted to be on rather radial orbits.

It has been suggested that the Milky Way dwarf satellites lie in a plane. Gaia Collaboration et al. (2018b) find, instead, that their orbits tend to be almost perpendicular to the Galactic plane, but spanning a range of orientations. This implies that even though the orientation of the average orbit plane may be similar, they may rotate in the opposite sense. Sculptor and Sagittarius move in planes that are nearly perpendicular to each other, and to the Galactic disk.

This ordered complexity might indicate some collective infall from a preferred direction, perhaps a 'cosmic web filament' aligned with the z-axis. But it appears to exclude a single event underlying their origin.

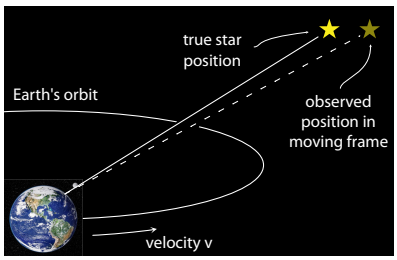
32. Aberration and Galactic rotation

STELLAR ABERRATION is the term used to describe the apparent displacement of celestial objects from their 'true' positions, as a simple consequence of the velocity of the observer.

To picture the phenomenon, let us start with a very simple analogy from daily life. Imagine that we are standing in the rain, with rain falling straight down on us from above. If we walk briskly, in any direction, the rain will appear to be coming from ahead of us. This is a simple example of the 'vectorial addition of velocities', and is experienced in many everyday situations. For example, riding a bicycle on a windless day, an 'apparent' wind will always be blowing in our face.

THE SAME EFFECT is well-known in astronomy, and is considered here only in its non-relativistic form. For an observer on Earth, moving through space with velocity v, then a star's position will *appear* to be displaced in the direction of the observer's motion by an angle v/c, where c is the speed of light.

There is an important point to make at the outset. If the observer's motion through space was constant, unchanging over time, then the position of all stars would be shifted by the same amount. And we would not be able to able to differentiate between their 'true' positions, and their apparent positions.



But the Earth's velocity does change, over months, as it moves in its annual orbit around the Sun. Knowing the Earth's orbital radius we can calculate the speed of the Earth, around the Sun, to be about $v = 30\,\mathrm{km\,s^{-1}}$. The speed of light is around $c = 300\,000\,\mathrm{km\,s^{-1}}$, so we find that v/c is about 0.0001. This quantifies the angular displacement in radians, which is about 20 arcsec. A star's apparent position can therefore be displaced by up to 20 arcsec from its 'true' position, depending on its location on the sky.

We can easily measure the effect because the 20 arcsec shift is in one direction when the Earth is at one point in its orbit, but in the opposite direction 6 months later. The effect is large compared to star positional measurements of an arcsec or better, let alone at the level of the milli- or micro-arcsec accuracies of Hipparcos or Gaia.

The aberration due to the Earth's orbital motion around the Sun is referred to as 'annual aberration'.

There is also a similar effect, albeit smaller in amplitude and on a shorter time-scale, due to the Earth's spinning motion about its rotation axis. This contribution is referred to as 'diurnal aberration'. We will see that there are others!

STELLAR ABERRATION, due to the observer's *velocity* through space, is distinct from the effect of parallax, which is due to the observer's changing *position*. But the two are entwined in the history of astronomy, specifically in the protracted efforts to measure the first stellar parallax during the 17th, 18th and 19th centuries.

Even before 1600, astronomers were in agreement that the crucial evidence needed to detect the Earth's motion around the Sun, and so to confirm the Copernican hypothesis of a Sun-centred solar system, was the measurement of trigonometric parallax. A major breakthrough came with Edmond Halley's discovery of the first stellar proper motions in 1718. And as instrumental accuracies reached the levels of a few arcseconds, the Reverend James Bradley, England's third Astronomer Royal, was immersed in his own efforts to measure parallax, targeting the bright star Gamma Draconis.

While Bradley failed to measure parallax, he did detect a small systematic shift in his star positions, which he eventually correctly attributed as resulting from the addition of the velocity of light to the Earth's velocity in its orbit around the Sun. From his angular shifts, he estimate the speed of light at $295\,000\,\mathrm{km\,s^{-1}}$. Announced in 1729, it has been described as one of the most significant discoveries in the history of astronomy.

Fast forward to the present day, and the effects of the observer's motion, whether on Earth or from a space platform like Hipparcos and Gaia, are well understood.

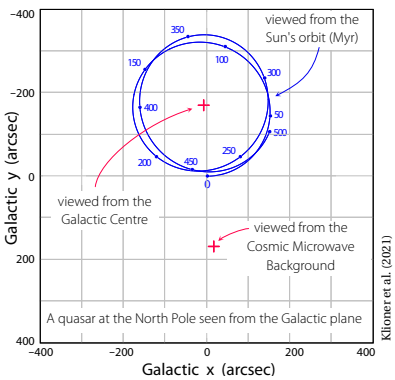
Even at it's L2 orbit location, 1.5 million km from Earth, Gaia's position and velocity in space are determined to better than 1 km, and some 0.2 m s^{-1} respectively. With this knowledge, and in the framework of special relativity, the effects of stellar aberration can be accurately predicted. And they are fully accounted for in the data analysis as part of the Gaia star catalogue preparation.

BUT ANOTHER COMPLICATION now raises its head. If the velocity of the observer is changing with time, i.e. if the observer is accelerating, the star displacements also change with time. If the measurements are accurate enough, and if the changes with time are fast enough, a systematic pattern of apparent proper motions, in the direction of the acceleration, would become apparent.

Such a possibility was already noted almost two centuries ago by John Pond, UK's sixth Astronomer Royal. But in the context of the structure of the Universe as it was known at the time, there was no particular reason to postulate an acceleration. And the idea was essentially discarded as being unmeasurable in any case.

WHAT EFFECTS could cause an acceleration of the solar system? One is the fact that, in the enormity of space and time, the Sun is moving in a (roughly) circular orbit around the Galaxy, with an orbital period of 250 million years. On an even more epic scale, our Galaxy itself is in motion within the Local Group of galaxies, and with respect to the cosmological reference frame defined by the Cosmic Microwave Background.

If stars in our Galaxy provided the only reference frame available, an acceleration term would probably be impossible to separate from other effects. But with its inertial reference frame materialised by 500 000 distant quasars, it turns out that the Gaia data are indeed accurate enough to discern the *acceleration* of the solar system due to its 250 million year orbit around the Galaxy.



Klioner et al. (2021)

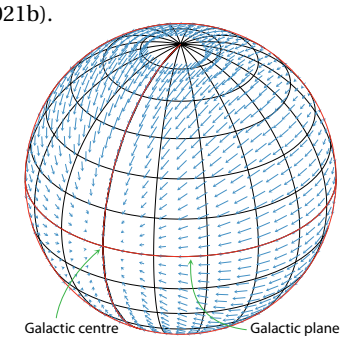
Models of the mass distribution of the Galaxy allow the effect to be predicted, here over 500 million years (in blue), or two Galactic rotations.

And from the Cosmic Microwave Background as measured by the Planck mission, we can predict the effect of our Galaxy's motion with respect to the Local Group of galaxies. For the latter, positional shifts are predicted to be around 200 arcsec. But in the absence of its variation over time, the effect is unmeasurable.

THE POSSIBILITY of discerning the first of these, viz. the effects of Galactic rotation, or 'secular aberration', has been pursued since the early 1980s in the framework of high-accuracy radio VLBI observations. But with the systematic displacement of the best-placed quasars being only around 100 μas, the effect is not much above the error of individual VLBI positions.

The latest studies, based on 39 radio sources from almost 40 years of VLBI observations, give an acceleration of 5.83 ± 0.23 μas yr^{-1} in the direction $\alpha = 270.^{\circ}2$, $\delta = -20.^{\circ}2$ (Charlot et al., 2020).

IN THE CONTEXT of Gaia, the problem was first recognised as being of relevance by Bastian (1995), and it was duly included as one of the targeted mission goals (Perryman et al., 2001). A quarter of a century later, a detailed assessment has been made with Gaia EDR3 by Gaia Collaboration et al. (2021b).



Klioner et al. (2021)

For the acceleration due the Sun's orbit, they used the distance and motion of Sagittarius A*, the black hole at its centre, to predict an acceleration of 6.98 km s^{-1} Myr^{-1} towards the Galactic centre, and an *expected* pattern of quasar motions shown here.

And they estimated the contributions from other terms, such as our Galaxy's central bar, the Sun's motion with respect to the Galactic plane, and the contribution of individual extragalactic objects in the Local Group.

TO DERIVE THEIR acceleration, Gaia Collaboration et al. (2021b) started with 1 614 173 EDR3 sources identified as 'quasar-like', teasing out the acceleration from an examination of the proper motion vector field.

They found a Galactocentric acceleration of 5.05 ± 0.35 μas yr^{-1} in the proper motions. This is a thousand times smaller than the Galactic rotation and shear effects of our Galaxy's stellar population, which is around 5–10 mas yr^{-1}. In units of more relevance to the scale size of the Galaxy, this corresponds to 7.33 ± 0.51 km s^{-1} Myr^{-1}. And in units more recognisable in terms of the acceleration due to gravity at the Earth's surface, it amounts to $2.32 \pm 0.16 \times 10^{-10}$ m s^{-2}.

THIS FIRST DETECTION of secular aberration in the optical agrees with the theoretical expectations from Galactic dynamics. Further improvements will be possible with future Gaia data releases. It is even possible that the 'secular extragalactic parallax', caused by the motion of the solar system with respect to the rest frame of the Cosmic Microwave Background, could be discerned.

33. Nearby stars

A DETAILED understanding of the nearby stellar population is central to many areas of astronomy. While 'nearby' is a vague term, it is often taken to mean the spherical region of space out to (say) 10, 20, or 50 pc from the Sun. Compared with the scale of our Galaxy, in which the Sun sits some 8000 pc from the Galactic centre, this region is dominated by stars of our Galaxy's disk.

Surveys of this nearby region provide sturdy foundations for defining our Galaxy's stellar luminosity distribution, the local mass density (in both stars and gas), their velocity distribution, the distribution and occurrence of binary and multiple stars, and the occurrence and nature of many other types of objects which comprise our Galaxy, including brown dwarfs, white dwarfs, and even exoplanets. Clearly, the best region of space to sample and to study is that closest to us, where the measurement accuracies are highest.

EVEN TODAY, it remains an enormous challenge to compile a complete census of stars in our immediate solar neighbourhood, even out to distances of, say, 10–20 pc. The pioneering ground-based parallax surveys of the early 1900s were successful in identifying nearby bright stars, but problems persist especially for the lowest luminosity stars, where a complete survey for low-mass stars and brown dwarfs even out to 10–20 pc has proven to be impossible.

Surveys searching for high-proper motion stars in the 1970–80s were successful in detecting potentially nearby stars which were then added to parallax measurement programmes, but they resulted in samples biassed towards high-velocity halo objects. For this reason, early nearby star compilations used spectroscopic *and* photometric distance estimates to try to identify more nearby candidates. The advent of accurate all-sky multicolour surveys in the past two decades has further facilitated the search for nearby, low-luminosity stars.

ONE OF THE FIRST ATTEMPTS to compile a census of stars in the solar neighbourhood, largely based on trigonometric parallaxes, was led by British Astronomer Royal Richard Woolley, and published as the *'Catalogue of Stars within Twenty-Five Parsecs of the Sun'* in 1970. Another, the Catalogue of Nearby Stars (CNS), originally led by Walter Fricke, has been updated and maintained by Heidelberg astronomers for more than 60 years.

CNS1, published in 1950, contained 915 single stars and multiple systems within 20 pc, with parallax errors of about 10 mas. CNS2, in 1969, enlarged the distance limit to 22.5 pc, and contained 1049 stars and multiple systems within 20 pc. CNS3, in 1991, extended the census to some 1700 stars nearer than 25 pc.

CNS4, in 1997, included data from the Hipparcos Catalogue, and accordingly provided the most comprehensive inventory of the solar neighbourhood up to a distance of 25 pc from the Sun at that time. But Hipparcos could only observe pre-selected stars, contained in its 'Input Catalogue'. This extended to about 11–12 mag, but with completeness only to about 9 mag.

Other ground-based surveys since then, amongst them RECONS and SUPERBLINK, have searched for faint nearby stars (such as M dwarfs) from infrared measurements or proper motion surveys.

Pre-Gaia compilations knew of some 5000 stellar systems within 25 pc, and a northern hemisphere survey has identified around 100 000 M dwarfs within 100 pc.

FOLLOWING ITS LAUNCH in 2013, Gaia is now 7 years into a potential 10-year mission, and is observing essentially every object on the sky brighter than about 21 mag. Early Data Release 3 (EDR3), published in December 2020, covered (nearly) the first 3 years of mission data, and lists nearly 2 billion stars with parallax accuracies better than about 1 mas, even for the faintest stars.

This implies that if a star is brighter than about 21 mag, and it lies within 100 pc of the Sun, it will be observed and identified as such. And with an accuracy on its distance of better (and often much better!) than 10%.

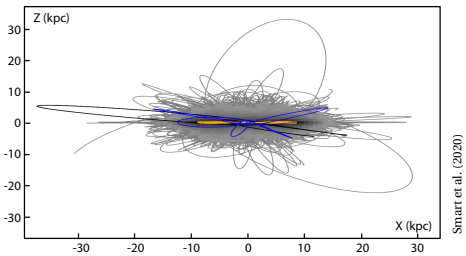
A first detailed assessment of what this means for our knowledge of stars within 100 pc was published in December 2020 (Gaia Collaboration et al., 2021d), from which the following gives a flavour.

THIS FIRST Gaia Catalogue of Nearby Star, or GCNS, contains an unprecedented 331 312 objects within 100 pc, a factor 100 more than its ground-based forerunners. It includes stars as faint as spectral type M9, i.e. with masses down to about 1/10 that of the Sun.

Within our immediate 10 pc horizon, and including a handful of stars too bright for Gaia (amongst them Sirius, Fomalhaut, Vega, Procyon, Altair, and Mizar), we now know of 383 stars, all with accurate distances. This includes five companion stars with distances measured for the first time, but not counting a few known unresolved binary systems (notably Procyon, η Cas, and ξ UMa). A few very low-luminosity T/Y brown dwarfs are also known, but too faint to be observed by Gaia.

This enormous census is providing rich new insights into our Galaxy's structure and dynamics. We can estimate the Sun's distance from the disk's mid-plane to be about 4 pc, and characterise the disk thickness for various stellar spectral types. Space velocities, in staggering numbers, allow detailed characterisation of individual and bulk motions in our region of the Galactic disk.

Galactic orbits, integrated over 1 Gyr, show that the most common disk stars follow circular orbits in the Galactic plane, while rarer halo stars visiting our neighbourhood have higher eccentricities and inclinations.



Galactic orbits of the GCNS stars over 1 Gyr (edge-on view)

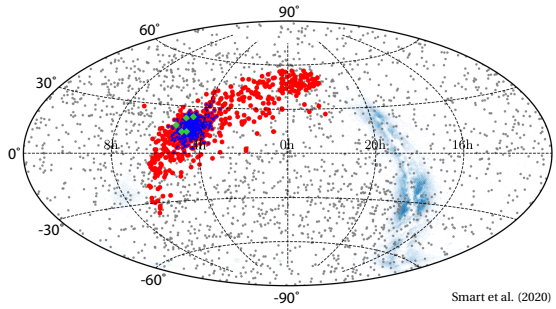
The Sun's height above the Galaxy mid-plane, its vertical velocity with respect to the plane, and the total mass of the Galaxy disk together determine the gravitationally-determined motion of the Sun up and down through its mid-plane, and the maximum height reached above and below it. Our Sun performs this oscillation with a period of about 80 million years. Knowing this allows the Sun's orbit through the Galaxy, and in particular through the Galaxy's spiral arms, to be reconstructed backwards in time over hundreds of millions of years. Some studies have linked this to 'ice house' periods in the Earth's climate over geological times (e.g. Gies & Helsel, 2005).

Of great importance in astronomy is the luminosity function (the numbers per cubic parsec) of the various star populations, such as main-sequence stars, giant stars, and white dwarfs. These estimates enter our models of star formation, and of the star formation history of our Galaxy. The Gaia Nearby Star Catalogue allows all of these to be characterised in unprecedented detail.

THE 100 PC SAMPLE contains two well-known open clusters, the Hyades (at a distance of about 47 pc, or 150 light-years) and Coma Berenices (at about 86 pc). The Hyades is an arresting sight in the dark night sky, with bright cluster members forming a loose enhancement in the distribution of bright stars over an area of about 10 degrees. The Pleiades (the Seven Sisters) is another prominent star cluster visible to the naked eye, but slightly more distant (at about 110 pc). Both the Hyades and Coma Ber clusters stand out in the GCNS as density concentrations in space, as well as in their velocities.

The Hyades members, of which some 500 are prominent in the Gaia census, show a dense concentration forming the cluster's core, and two prominent 'tidal tails', where cluster stars are slowly escaping its gravitational confines, over millions of years, as they are tugged by the pull of the Galactic centre. Studies of the Hyades with the earlier Gaia DR2 data suggest that the cluster is close to final dissolution, with only some 30 Myr of its existence remaining (Oh & Evans, 2020).

Several other stellar streams and stellar superclusters are clearly visible in the Gaia survey, including the Gaia Enceladus stellar stream (Helmi et al., 2018).



Hyades cluster members (blue) and its tidal tail (red)

STUDIES OF both intermediate-separation and wide-separation binary star systems have already been made with the first two Gaia data releases (DR1 and DR2). More than 16 000 resolved binaries are evident in the new GCNS sample, with some 10% of F, G, and K spectral types clearly seen to be wide binaries.

Wide-separation binaries have very low 'binding energies', meaning that they are, gravitationally, only loosely bound. They can therefore be used not only in models of star formation and of the dynamical evolution history of the Galaxy, but also as probes of the mass distribution and number density of potentially disruptive 'dark' objects in the Milky Way.

WHITE DWARFS stand out in the diagram of star colours plotted against their absolute magnitudes. Based on the Gaia data alone, more than 20 000 stars have a high probability of being white dwarfs, of which more than 2500 were previously unknown.

34. Perspective acceleration

ASTROMETRY CONCERNS the positions of objects on the sky, along with their projected motions in the plane of the sky due to their motion through space (their proper motion), and their apparent motion as the Earth orbits the Sun (their parallax).

Classical astrometry ignores a star's radial velocity, i.e. its space motion *along* the line-of-sight. This is because a star's radial velocity generally has no effect on its position on the sky. Indeed, not only does it have no effect on angular positional measurements, but conversely neither can its radial velocity be determined from these angular measurements. It is irrelevant in typical astrometric surveys, and it is ignored.

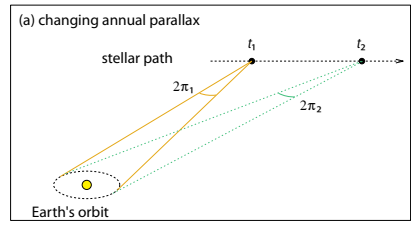
Knowing a star's radial velocity is nonetheless important for fully defining its complete space motion. The full space motions of stars are, in turn, essential in understanding kinematics and dynamics, both for individual stars, as well as for groups or populations.

Radial velocities are determined by measuring the Doppler shift of the stellar spectral lines. Extremely high accuracies can be reached in these velocity measurements: typical large-scale stellar surveys may reach accuracies of around $1\,\mathrm{km\,s^{-1}}$, while today's dedicated high-precision radial velocity spectrometers used for exoplanet studies routinely measure stellar velocities, along the line-of-sight, with accuracies of about a few centimetres per second!

BUT RADIAL VELOCITIES are no longer irrelevant in very high accuracy astrometry, where it can affect the determination both of the parallax of a star, and of its proper motion. How is this possible?

This systematic change in trigonometric parallax due to the radial displacement of a star is most easily appreciated from the figure: a star moving through space, with some radial motion, has a parallax which changes with time, by a tiny amount proportional to the product of the radial velocity and the square of the parallax.

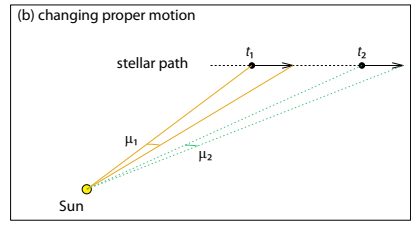
The effect is most apparent (if at all) for nearby stars with large radial velocities, but quickly diminishes for smaller radial velocities, and for larger distances.



Dravins at al. (1999)

The question was first considered by Schlesinger (1917). He concluded that the change in parallax is very small for all stars, and that detecting it would need high-accuracy measurements over years or even decades. For Barnard's star, for example, with a parallax of about 500 mas, and a radial velocity of $110\,\mathrm{km\,s^{-1}}$, the expected parallax change is a minuscule 34 microarcsec per year. It may just be measurable by Gaia.

THE SECOND METHOD is somewhat similar, and exploits the fact that a single star moves with uniform velocity through space.



Dravins et al. (1999)

For a fixed space velocity, the angular velocity (or proper motion) varies inversely with the distance to the object. However, the tangential velocity changes due to the varying angle between the line-of-sight and the direction of its space velocity. The two effects result in a changing proper motion with time, which is interpreted as an apparent (or 'perspective') acceleration of the star's motion on the sky.

This apparent acceleration turns out to be proportional to the product of the star's parallax, its proper motion, and its radial velocity. It is always a tiny effect, but largest for nearby stars with a high proper motion.

Whether a changing proper motion is due to a *real* acceleration, e.g. if it is part of an orbital binary system, or due only to some unrecognised *perspective* acceleration, can only be clarified by further orbital measurements or by measurements of the star's radial velocity.

HISTORICALLY, the phenomenon was first described by Seeliger (1900). It was used by Ristenpart (1902) in an attempt to determine a change in proper motion for Groombridge 1830, by Lundmark & Luyten (1922) for Barnard's star, and it was proposed by Russell & Atkinson (1931) as a way of confirming the hypothesised gravitational redshift, of several hundred km s^{-1}, predicted for the white dwarf Van Maanen 2.

In the 1970–1980s, photographic observations over several decades were used to measure this perspective acceleration for the white dwarf Van Maanen 2, for Groombridge 1830, and for Barnard's star.

Dravins et al. (1999) combined the Hipparcos results with those of the historical Astrographic Catalogue to determine astrometric radial velocities for 16 stars with large parallax–proper motion products. These included the fast moving Barnard's star, Kapteyn's star, as well as Groombridge 1830 and 61 Cygni.

Kürster et al. (2003) observed Barnard's star for more than five years with an accurate spectrograph at the ESO Very Large Telescope. They measured a secular acceleration fully consistent with the predicted value, of 4.50 m s^{-1} yr^{-1}, based on the Hipparcos proper motion and parallax combined with the known radial velocity of −110.506 km s^{-1}.

THE EFFORT INVESTED in this sort of task goes beyond attempting to measure the effect for its own sake. In practice, Doppler measurements represent the combination of the true velocity of the stellar centre of mass, combined with surface effects on the star such as atmospheric dynamics and gravitational redshifts.

The accurate determination of stellar radial velocities from geometric principles, i.e. without using spectroscopy or invoking the Doppler principle, can be used, in principle at least, both to examine these stellar phenomena, but also to establish fundamental radial velocity standards amongst the nearby stars.

AN EXTENSION of these principles applies to open clusters. Since all cluster stars share the same (average) velocity vector, apart from a (small) random velocity dispersion, the cluster's apparent size changes as it moves in the radial direction. This relative change, revealed by the proper motion vectors towards the cluster apex, corresponds to the relative change in distance. Since the individual stellar distances are known from parallaxes, their radial velocities can be estimated.

The method provides an estimate of the space velocity and internal velocity dispersion of a cluster using astrometric data only, along with improved parallaxes.

It was developed and applied to the Hipparcos data for the Hyades cluster by Lindegren et al. (2000) and to the Ursa Major, Coma Berenices, Pleiades, and Praesepe clusters by Madsen et al. (2002) with a number of detailed insights into the cluster members.

Predicted shifts (Δ in milliarcsec) for Gaia GDR2

Star name	Hipparcos catalogue	rad. vel. [km s^{-1}]	Δ [mas]
Barnard's star	87937	−110.51	1.975
Kapteyn's star	24186	245.19	1.694
Van Maanen 2	3829	263.00	0.573
61 Cyg A	104214	−65.74	0.313
61 Cyg B	104217	−64.07	0.297
Groombridge 1830	57939	−98.35	0.239
α Cen C (Proxima)	70890	−22.40	0.208
ϵ Ind	108870	−40.00	0.163
Ross 47	26857	105.83	0.144
ϵ Eri	15510	87.40	0.141

AT THE ACCURACY of Hipparcos, around 1 milli-arcec, perspective acceleration was of marginal importance, and then only for the nearest stars with the largest radial velocities and proper motions. Nonetheless, it was accounted for in the 21 cases for which the accumulated positional effect over two years exceeds 0.1 milli-arcec.

WHAT THEN of Gaia? For which stars is it most important, and what steps have been taken to include its effects? For the first Gaia data release, GDR1 in 2016, the astrometric accuracies were limited. The effect was simply ignored, by assuming zero radial velocity for all objects (Lindegren et al. 2016).

For the second data release in 2018, GDR2, it was taken into account for just 53 nearby Hipparcos catalogue objects, by using the values of radial velocities taken from the existing literature (Lindegren et al. 2018). The 10 stars with the highest predicted shifts over the 22 months of the Gaia data entering the construction of GDR2 are listed in the table, the largest being for Barnard's star, at 1.975 mas.

For the third data release, EDR3 in 2020, the effect was taken into account, whenever possible, using radial-velocity data from Gaia's own radial-velocity spectrometer (Lindegren et al. 2020). For a small number of nearby stars (mainly white dwarfs), this was complemented with radial velocities from the literature.

As an extension of its classical application to determine the distances to moving clusters, very clear perspective contraction has already been observed for the globular cluster NGC 3201, due to its very high radial velocity and relatively large parallax (Helmi et al. 2018).

GAIA'S ACCURACIES will continue to improve over the coming years. And although perspective acceleration is not a dominant effect in the data, Gaia nonetheless provides a secure observational footing for the theoretical description of the effect presented more than a century ago. And it opens the possibility of some detailed investigations into the physics of stars for which the effect can be measured.

35. Stellar flybys

INTERSTELLAR SPACE is really very empty. The nearest stars to us are at a distance of a little more than 1 parsec, or about four light-years. This means that their light, travelling at 300 000 km per second, would take four years to reach us. And if one of them were to be heading directly at us, at a typical speed through space of say 20 km s^{-1}, it would take 50 000 years to arrive.

This is not long in geological terms. But all the stars are moving, certainly with some common motion around the centre of our Galaxy, but with random motions too. So, we could ask, how long would it take for a typical star to collide with another in our region of the Galaxy? Mathematically, with respect to the other stars in the solar neighbourhood, this is termed the 'mean-free time' between collisions.

Estimates put this at around 10^{13} years, or 10 000 billion years – a thousand times the age of the Universe, which is a mere 13 billion years old. Stars in the solar neighbourhood therefore behave like a 'collisionless gas'. They zip around, but they will never collide.

NONETHELESS, noticeable effects due to nearby star passages and our Sun can occur. Indeed, one predicted consequence of a close approach between another star and our own solar system is this: the gravitational pull of the passing star can perturb the somewhat delicate equilibrium of the Oort cloud comets, with the possibility of an increased impact hazard on Earth.

Why might this be so? Comets are small icy bodies, believed to have formed far out in the solar system during the early stages of its formation, 4.5 billion years ago. Often described as 'dirty snowballs', they comprise frozen ices such as water, methane and carbon monoxide, along with dust grains and small rocky particles. They range from a few 100 m to tens of km in size.

When seen in the inner solar system, comets often have very eccentric orbits, with periods around the Sun of several years to several million years. Approaching the Sun, they sublimate and become 'active', producing a visible atmosphere or coma (due to solar radiation), and sometimes also a tail (due to the solar wind).

Many of the long-period comets, those with periods of 200 years or more, are inferred to have come from a region called the 'Oort cloud'. This is believed to be a roughly spherical region of predominantly icy planetesimals, far from the Sun, where temperatures are a chilly –260 C, just a few degrees above absolute zero.

Although it cannot be observed directly, its existence was predicted by Dutch astronomer Jan Oort in 1950. He argued that, given the instability of cometary orbits and their loss of volatiles during perihelion passage, a continuous replenishment source must exist. Current models suggest that this distant reservoir comprises some 10^{12} icy clumps, of typical size 1 km, and with a total mass of just a few times that of the Earth.

Extending out to perhaps 200 000 astronomical units (around 1 pc), far beyond the Kuiper belt, the Oort cloud occupies the indistinct outer limits of our solar system, where the pull of our Sun's gravity eventually loses out to more distant forces. Some indication of its vast scale is that Voyager 1, the fastest and most distant space probe, will only reach it 300 years from now, and will take another 30 000 years to pass through it, sailing past these isolated bodies spaced tens of millions of km apart.

Occasionally, these distant icy fragments can be nudged towards the Sun by gravitational perturbations caused by our Galaxy's tidal field… or by passing stars.

CAN WE HOPE TO find any evidence of stars that have come close to the Sun in the past? One approach is to see whether there might be a correlation between ancient stellar flybys, and increased cratering events, on the Earth, or elsewhere in the solar system.

A compilation of impact cratering events is maintained at the Earth Impact Database, and there have been many attempts to examine whether such impact events are connected to large-scale extinctions on Earth (e.g. Firestone, 2021). In addition to the largest-known Vredefort and Sudbury craters at around 2 Ga, these impact structures include the Kara-Kul at about 5 Ma, the Popigai and Chesapeake Bay events at 35 Ma, the Chicxulub event at 65 Ma, and the Morokweng at 145 Ma.

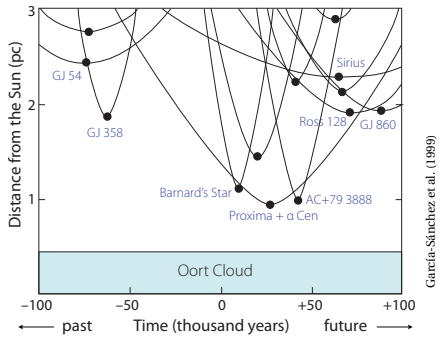
The Moon provides an outstanding archive of impact cratering in the solar system over the past 4.5 Gyr. It preserves this record much better than the larger terrestrial planets, which have largely lost their ancient crusts through geological reprocessing and hydrospheric or atmospheric weathering. Considerable bombardment evidence also occurs widely elsewhere, including on Mercury, Venus, and Mars, as well as on numerous minor solar system bodies, including Ceres and Pluto.

WITHOUT GOING further into this extensive topic of cratering events, the main question here is: could stellar flybys have been responsible for any periods of excess cratering recorded over geological time?

It is not difficult to imagine that if we have a measure of how the Sun is moving through space, and of the distances and motions of the stars around us, we can hope to track their paths backwards in time, and see whether any close stellar encounters happened in the past, and whether others might occur in the foreseeable future.

SEVERAL ATTEMPTS have been made in this endeavour, extending over the past and into future, and making use of distances and space motions from the Hipparcos catalogue (available since 1997) combined with radial velocities to give their full space motions.

García-Sánchez et al. (1999) found that one star, Gliese 710, will have a closest approach of less than 0.4 pc some 1.4 million years in the future, while several others come within 1 pc during a ±10 Myr interval. But their dynamical simulations showed that none of these passing stars perturb the Oort cloud sufficiently to create a *substantial* increase in the long-period comet flux in the vicinity of the Earth's orbit.



Past and future flybys, from Hipparcos

Other pre-Gaia work suggested that the closest past encounter was the low-mass binary WISE J0720–0846 (Scholz's star), at about 50 000 au (i.e. within the Oort cloud) some 70 000 yr ago (Mamajek et al., 2015). And tracing back the orbit of the Sun suggests that the estimated rate of close encounters, those within 400 000 au, lies in the range 20–60 per million years, or every 20–50 000 years (Martínez-Barbosa et al., 2017).

THE DISRUPTIVE effect of these passages actually depends not only on the flyby distance, but also on the total mass of the stellar system, and on its relative velocity compared to that of the Sun. This means that a smaller flyby velocity might result in a much larger perturbation for some more distant passages than for much closer ones – essentially the potentially disruptive gravitational pull is acting over much a longer time.

Because of its extreme accuracy, its faint limiting magnitude, and its survey completeness, Gaia was long expected to find many new perturbing stars, including nearby M dwarfs with low relative velocities.

BETWEEN THE FIRST Gaia data release in 2016, and the end of 2020, more than a dozen papers have examined various aspects of this 'flyby' question.

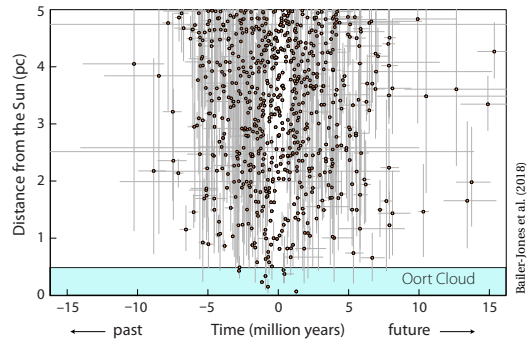
The first data release, Gaia DR1, was already used to refine the flyby of GJ 710, finding that it will pass even closer than the pre-Gaia estimates, at 13 000 au from the Sun in 1.35 Myr from now, and that it will consequently inject a larger flux of Oort cloud comets toward the inner solar system (Berski & Dybczyński, 2016).

Bobylev & Bajkova (2017) selected 216 000 stars from GDR1 with known radial velocities. They found several with encounters closer than 1 pc, including GJ 710. Bailer-Jones (2018) started with an even larger sample of 320 000 stars, then calculated their orbits within the Galaxy, finding 16 stars with a flyby distance of less than 2 pc, the closest again being GJ 710.

Gaia DR2 allowed Bailer-Jones et al. (2018b) to construct a sample of 7.2 million stars with further improved trajectories. They found 694 stars with flybys within 5 pc, all occurring within ±15 Myr. Of these, 26 may pass within 1 pc, and 7 within 0.5 pc, while GJ 710 has the largest perturbing effect on the Oort cloud.

They estimated that only 15% of flybys inside 5 pc, and within ±5 Myr, have been identified, mainly due to the absence of radial velocities. Their models suggest one encounter within 1 pc every 50 000 years.

WITH THE IMPROVED Gaia data to come, and with the progress being made in impact cratering studies, more exciting advances are sure to lie ahead.



694 past and future flybys within 5 pc, from Gaia

36. Science alerts

As the Gaia satellite scans the sky, it detects and observes all objects brighter than a given threshold. This avoids the use of a pre-defined observing programme, and it ensures that all objects bright enough at the specific time of their observation – whether regular or irregular variables, or moving objects within the solar system – are detected and observed.

An important type of object, which could never appear in a pre-defined observing programme, and which this sort of onboard detection was specifically designed to include, are the class of 'transients' sources. These are objects which can suddenly, and usually unexpectedly, increase in brightness for a number of reasons, and which therefore suddenly become measurable. With this kind of onboard detection capability, Gaia could hope to discover new transients, and contribute astrometric and photometric measurements as the source changes in brightness.

The most obvious objects in this class are supernovae, the explosive brightening of an otherwise unremarkable progenitor star. Amongst a number of dedicated supernovae search programmes, the ASAS–SN (All Sky Automated Survey for SuperNovae) operates 20 robotic telescopes which can survey the entire sky once a day. In the past two decades, a few hundred supernovae are discovered each year.

But there are many other sorts of transient events, including Galactic novae, cataclysmic variables, stellar flares, comets, gravitational microlensing events, and even tidal disruption events.

Simulations made during the early studies in the 1990s showed that Gaia could detect supernovae out to distances of 500 Mpc, or to redshifts of about 0.1, corresponding to perhaps 100 000 detections over its 5-year nominal lifetime (Høg et al., 1999). Although the satellite observations would be too sparse to produce the all-important light-curve from Gaia data alone, alerts to ground-based observers with rapid follow-up monitoring could perhaps provide light curves for perhaps 50 000. Estimates of the number of gravitational microlensing events were made around the same time.

A group at the Cambridge Institute of Astronomy is leading the processing of the satellite photometric data. A sub-group, led by Simon Hodgkin, has taken responsibility for the handling of these Gaia 'alerts'.

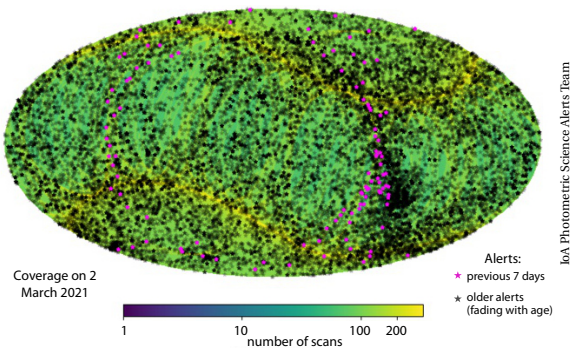
The repeated, high-precision measurements which form the basis of the high-precision measurements of stellar positions, are also ideal to look for variations in brightness as well. The group runs a dedicated data processing pipeline to look for transient events in the Gaia data, picking up 'new' sources where nothing had been detected previously, or sudden dramatic changes in brightness of previously detected stars.

These Alerts are made public immediately after the data processing and alert identification, typically just 2–3 days after the observation by the satellite.

Their `www alert pages` include all the data Gaia has collected for each source, including detected and historic G-band magnitudes, the light curves and the low-resolution Gaia (B_P/R_P) spectra.

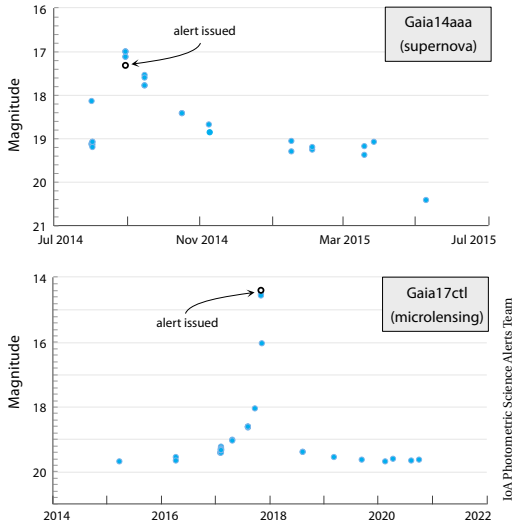
Dedicated apps allow the discoveries to be followed up by ground-based astronomy facilities. Both professional and amateur astronomers, and even groups of school children, are now involved.

Following the convention used for supernova discoveries (the prefix SN, followed by the year of discovery, suffixed with a one or two-letter designation), the Gaia discoveries are designated GaiaYYaaa, GaiaYYaab, GaiaYYaac,... where YY encodes the discovery year.



AFTER SATELLITE commissioning in mid-2014, the first reported discovery, Gaia14aaa, was a supernova of Type Ia, observed on 30 August, and reported on 12 September. It had an observed magnitude of 17.32, compared with a historic magnitude of 19.22 ± 0.42.

By the end of 2020, with 6 years of observations, nearly 15 000 events had been issued via the Cambridge alerts page, around 50 brighter than $G = 12$, and some also detected by other monitoring programmes.



Although some two thirds of these events have not (so far) been classified, around 5000 have. Some 500 are listed as cataclysmic variables, 100 as active galactic nuclei, more than 600 as variable quasars, 1500 as supernovae Type Ia, and 500 as supernovae Type II.

NEARLY 50 gravitational microlensing events have been discovered as part of the Gaia alerts pipeline, mostly in the Galactic plane where lensing cross-sections are highest. Some of these were also detected by OGLE. One of the brightest to date, Gaia 17ctl, showed a brightness increase of 4.5 mag at peak amplification.

The subject of microlensing, both photometric and astrometric, is a broader one for Gaia than simply event detection, and is picked up elsewhere. It touches on questions of exoplanet discovery and characterisation, as well as astrometric microlensing event prediction due to future close alignments as a result of rapidly moving nearby stars.

ONE OF THE AREAS that the Gaia alerts may impact in future is the detection of 'tidal disruption events'. These were predicted, 50 years ago, to occur when a star approaches sufficiently close to a supermassive black hole, and a fraction of the star's mass can be captured into an accretion disk around the black hole. This should result in a temporary flare of electromagnetic radiation as matter in the disk is consumed by the black hole.

Possible candidates were first discovered by the X-ray satellite Rosat in 1990, while more convincing examples have only surfaced in the last 3–4 years, with observations from ASAS, WISE, and TESS.

Although the Gaia transients include 100 classified as originating in active galactic nuclei, none are explicitly confirmed as such tidal disruption events. But analysis by Kostrzewa-Rutkowska et al. (2018), using different detection criteria, yielded nearly 500 nuclear transients, only five of which had been published by the Cambridge group, raising the prospects that this new type of transient may be more routinely detectable in the future.

THE INFORMATION on the variability of the Gaia alert events is extracted from the source's average magnitude over the 45 second crossing time of the full astrometric field of view, and compared with previous magnitudes measured over previous scans. The 45-second average results from the star images passing across 10 individual CCD detectors, each with a sampling period of 4.5 seconds. Whether this faster time sampling can contribute significantly to studies of 'fast' transient sources has been studied by Wevers et al. (2018).

With their finding that transient brightness variations down to an amplitude of 0.3 mag could be probed on time-scales ranging from 15 s to several hours, they identified four candidate fast transients within an area of just 20 square degrees. Two were tentatively classified as flares on M-dwarf stars, one as a flare on a giant star, and one potentially a flare on a solar-type star.

ANOTHER EXAMPLE of the potential relevance of the Gaia alerts comes from the emerging field of gravitational wave astronomy. The first such detection was announced by LIGO in February 2016. Since then, and until the end of 2020, 20 such events have been discovered by LIGO and VIRGO. In just one of these, GW 170817 (a binary neutron star coalescence), an electromagnetic counterpart was discovered, firstly in the optical, and a few days later in the X-ray and radio.

Studies of the Gaia data by Kostrzewa-Rutkowska et al. (2020) have shown that, from the current gravitational wave observing run (O3, which began in April 2019), about 16–25 per cent should fall in sky regions observed by Gaia 7–10 days after the event. They suggest that their specific detection algorithm would provide about 20 candidates per day over the whole sky.

THIS FOCUS ON photometric 'transient alerts' is quite distinct from the vastly broader topic of stellar variability, and the huge impact that Gaia will have on the quantity and quality of data on both regular and irregular variables, as well as on stellar rotation.

Neither has it touched on the detection of moving solar system objects and their orbit reconstruction, nor on the photometric detection of planetary transits.

37. Ultra-wide binaries

BINARY STARS, as well as triple or even higher multiplicity star systems, are common. They form, in the swirling gas clouds of dense regions of the interstellar medium, over a very wide range of separations.

Binaries with very close separations can eventually spiral in and merge, while those formed with wider separations, and less weakly bound, can eventually be broken apart, either due to close stellar encounters (for example in star clusters), or as a result of numerous distant passages that incrementally pull on the binary, and it slowly evolves from being bound to being unbound.

Indeed, the wider their separation, the more easily they are disrupted, whether by passing stars, molecular clouds, or the Galaxy's spiral arms. Indicative survival times are of order a billion years at 0.1 pc separation (20 000 au, or 0.3 light-years), and perhaps around 100 million years at 0.5 pc (100 000 au, or 1.6 light-years).

ALTHOUGH THERE is no precise definition, we can refer to those with separations of 100–2000 au as 'wide' binaries, and those above about 30 000 au as 'ultra-wide'. At these enormous separations, the two stars will be very widely separated on the sky, by a degree or more, with very long orbital periods, but sharing an almost identical space motion over millennia.

How can two stars be recognised as a physical pair? Close binaries will often be unusually close together on the sky, much closer than the average star density. Monitoring their space motions or radial velocities over years or decades can reveal their orbital motion, and so confirm their gravitational connection.

But how can a wide or ultra-wide binary be distinguished from two completed unrelated stars? The orbital motion of a very wide-separation binary requires extremely accurate measurements to recognise. Indeed, the wider a binary is, the more difficult it is to identify – and this has been a major barrier to discovering and studying wide binaries in the past.

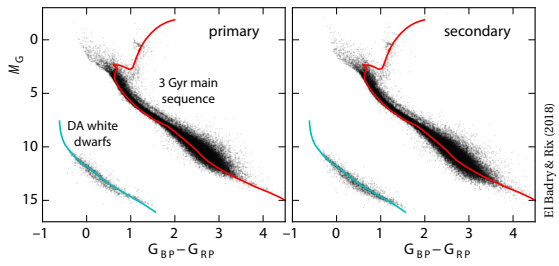
As an example of the situation pre-Gaia, a programme aimed at identifying wide binaries in the Galactic halo found less than 100 candidate pairs from the Sloan Digital Sky Survey (Coronado et al., 2018).

GAIA IS IDENTIFYING many thousands of very wide and ultra-wide binaries from their highly accurate space motions. Essentially, a very widely separated binary can be recognised if both components share identical, or very similar, distances and space motions.

The resulting discoveries are allowing great progress in understanding their formation, evolution, and the various processes that contribute to their eventual disruption. I will look here at some advances enabled by Gaia DR1 and DR2 across a rather wide range of topics.

THE PARALLAXES AND proper motions of Gaia DR2 allowed El-Badry & Rix (2018) to compile a catalogue of more than 50 000 wide binaries, in the separation range 50–50 000 au, and lying within 200 pc of the Sun.

While its sheer size is impressive, more interesting is the make-up of the resulting binary population. More than 50 000 are pairs of main-sequence stars, with a typical age of around 3 Gyr, while more than 3000 are pairs comprising a main-sequence primary star and a white dwarf secondary. And nearly 400 systems consist of white dwarf–white dwarf pairs.



In terms of numbers versus separation, a marked difference between the three populations is evident. The main sequence–main sequence binaries are reasonably consistent with a single power-law of slope over separations 500–50 000 au, while the main sequence–white dwarf, and white dwarf–white dwarf binaries, show distinct breaks at 3000 and 1500 au, respectively.

These distributions can be explained if the white dwarfs receive a 'kick' of about 0.75 km s^{-1} during their formation, presumably due to asymmetric mass-loss.

STARS BORN together should be chemically homogeneous, and wide binary systems provide an opportunity to test this assumption. For 25 systems comprising main-sequence stars of similar spectral type in Gaia DR2, Hawkins et al. (2020) obtained high-resolution spectra to derive chemical abundances for various elemental classes, including iron-peak and α-elements.

Twenty pairs (80%) were found to be homogeneous in [Fe/H], as well as in all other elemental abundances. That true physical pairs are more chemically homogeneous than random pairs of similar spectral type, provides another route to characterising very wide binaries.

GAIA DR2 has been used to examine the eccentricity distributions of wide binaries. This is of interest because their orbits provide a fossil record of their formation and early dynamical evolution. Specifically, the direction and speed of their relative motions contain statistical information on their eccentricities, which is typically inaccessible due to their long orbital periods.

Tokovinin (2020) used the 50 000 wide binary catalogue of El-Badry & Rix, 2018 to show that the eccentricity distribution is close to that expected from dynamical interactions within their natal cluster environment.

He found that pairs with projected separations below 200 au are more circular, a result expected where forces due to tides or gas friction have been operating. He found that wide pairs, above 1000 au, have an excess of very eccentric orbits, along the lines predicted by stellar ejections from unstable triple systems.

AIMING TO examine how the wide binary populations observed in star-forming regions or OB associations compare with that in the field, Deacon & Kraus (2020) used the Gaia DR2 data to show that the wide binary population in the Alpha Per, Pleiades, Praesepe, and Hyades open clusters is significantly lower than the wide binary fraction observed in the field population.

However for two groups of young stars that likely originated in looser associations (young moving groups and the Pisces–Eridanus stream), the wide binary fraction was similar to, or even above, that of the field.

THE OCCURRENCE of binary stars at the largest separations has long been recognised as an important potential probe of the possible existence of Massive Compact Halo Objects (MACHOs) in the Galactic halo.

Tian et al. (2020) constructed samples of ultrawide binaries in the solar neighborhood, with separations 0.01–1 pc, using Gaia DR2 to define kinematic populations according to their tangential velocities, i.e., disk-like ($V_\perp < 40\,\mathrm{km\,s^{-1}}$), intermediate, and halo-like ($V_\perp > 85\,\mathrm{km\,s^{-1}}$) binaries, with thousands in each sample.

Fitting power laws to the separation distribution, they found that its slope at 300–10 000 au is the same for all subpopulations, while it steepens at larger separations, an effect increasing with age.

They argued that the trends are contrary to that expected if the steepening at wide separations was due to gravitational perturbations by molecular clouds or stars (which would preferentially disrupt disk binaries), or to MACHOs (which would require a population inconsistent with other constraints). Instead, they could model the distribution of wide binaries as having been formed in dissolving star clusters, with the steepening at larger separations due to the finite size of the birth clusters.

ULTRA-WIDE SYSTEMS provide an interesting laboratory for testing the predictions of general relativity, and comparing them with the predictions of modified theories, like MOND, which have been put forward as an alternative to 'dark matter'. At separations beyond about 5000 au, the stars in such a system have sufficiently small orbital accelerations that they can be used as a direct probe of binary orbits in low-gravity regimes.

Several such tests have been carried out with the Gaia data, and the statistical properties do show some anomalies. Others have suggested that these can be attributed to hidden triple systems. This subject is treated separately elsewhere.

THE IMPORTANCE of binaries in the formation of planetary nebulae, which form at the end of life of intermediate-mass stars, is well established. Binarity can explain the frequent occurrence of asymmetrical nebulae, and the formation of bipolar lobes. But establishing the incidence of binarity has not been easy.

González-Santamaría et al. (2020) used the extremely precise measurement of parallaxes and proper motions in Gaia DR2 to search for wide binary companions, out to around 20 000 au from the central stars of 211 planetary nebulae.

They found wide binary companions for eight of the sample, all with projected separations less than 15 000 au. They concluded that binarity plays a role, but that the Gaia DR3 release would be needed to expand the search to objects located closer to the central star.

INTHE FIELD of exoplanetary science, there is an ongoing debate as to whether the host stars of massive closely-orbiting 'hot Jupiter' planets are more likely to be found in wide binaries with separations above 100 au. A search for comoving, very wide companions with separations 1000–10 000 au was made by Hwang et al. (2020). They found that only around 10% of hot Jupiter hosts have companions at this sort of separation.

Their preliminary conclusion is that the formation of 'hot Jupiter' planets is not particularly sensitive to this sort of binarity, but improved Gaia data will clarify the picture further.

38. The Magellanic Clouds

THE MAGELLANIC CLOUDS are two 'nearby' irregular dwarf galaxies, visible to the unaided eye in the dark skies of the southern hemisphere. They became known in Europe following Portuguese voyages in the 16th century, and in particular through Ferdinand Magellan's circumnavigation of the world of 1519–1522.

Compared to our own Galaxy's diameter of about 100 000 ly (light-years; divide by 3.26 if you prefer your distances in pc), the Large Magellanic Cloud (LMC) has a diameter of 14 000 ly, and lies 160 000 ly away. The Small Magellanic Cloud (SMC) has a diameter of 7 000 ly, and lies about 200 000 ly away. The two are separated by 20° on the sky, 75 000 ly apart in distance. Only the smaller Sagittarius Dwarf Elliptical Galaxy (discovered in 1994), and the Canis Major Dwarf Galaxy (discovered in 2003) are our more proximate neighbours.

Both probably have large dark matter halos. The LMC is now believed to be the fourth most massive of over 50 galaxies comprising the 'Local Group'. Observations and theory suggest that the Magellanic Clouds have both been distorted by tidal interaction with the Milky Way. Their gravity has, in turn, affected the Milky Way, distorting the outer parts of our Galaxy's disk.

WHETHER THE LMC and SMC are bound as orbital companions to our Milky Way remains uncertain. If they are, their orbital period is at least 4 billion years. The other possibility is that they are on a 'first approach', and we are witnessing the start of a galactic merger that may overlap with the Milky Way's expected merger with the Andromeda galaxy sometime in the future.

In radio images sensitive to neutral hydrogen, the LMC reveals a clear spiral structure. Streams of neutral gas connect them both to the Milky Way (the Magellanic Stream) and to each other (the Magellanic Bridge). They are both gas-rich, with a higher fraction of their mass in hydrogen and helium compared to the Milky Way, and both are also more metal-poor than the Milky Way.

They both host nebulae and young stellar populations, with stars ranging from the very young to the very old, telling of a long stellar formation history.

DETERMINING THE DISTANCE to the Large Magellanic Cloud in particular has been a long-standing challenge in astronomy, representing an important early step in establishing the overall distance scale in the Universe, and hence also the associated Hubble constant.

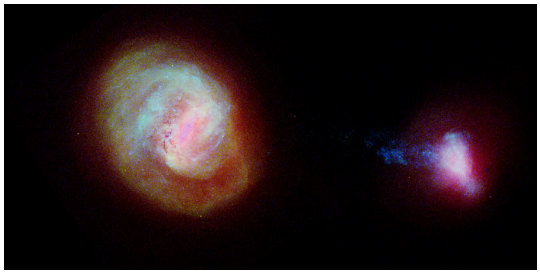
However, its stars are far too distant for any direct geometrical (parallax) determination, even for Hipparcos. Seen from a distance of 160 000 ly (about 50 000 pc), a star's parallax is a minuscule 20 micro-arcsec, a factor 50 smaller than the angles accessible to Hipparcos!

Angular accuracy aside, Hipparcos could observe only a few LMC and SMC stars because of its magnitude limit of around 11–12 mag, as well as very strict crowding constraints related to the functioning of its on-board detection system. A careful pre-selection of stars, satisfying these various constraints, resulted in just 36 LMC and 11 SMC objects in the observing programme of Hipparcos, all with final accuracies of around 1.5–2 milli-arcsec.

PRE-GAIA, distance estimates to the LMC were based on a consensus of various methods, each providing individual steps in a somewhat shaky distance 'ladder'. These drew together estimates from Population I stars (Cepheids, red clump giants, Mira variables, and eclipsing binaries), Population II stars (subdwarf fitting to globular clusters, horizontal branch stars, and RR Lyrae stars), as well as white dwarf sequencing, and estimates based on the bright supernova, SN 1987A.

As I wrote in 'Astronomical Applications of Astronomy' in 2009: *'Further in the future, one-step trigonometric parallax estimates to individual objects in the LMC will become accessible to Gaia. At that point, the tortuous complexity underlying the various distance estimates used to date will be relegated to historical curiosity, and still deeper insights into the physical nature of the 'standard candles' used to date will commence.*

I WANTED TO RECALL the Hipparcos results on the Magellanic Clouds because the step-change in what was state-of-the-art in space science just 20 years ago, and the Gaia results of today, are indeed spectacular.



Luri et al. (2021) and ESA/Gaia/DPAC

Gaia view of the LMC and SMC, colour denoting populations

Gaia Early Data Release 3 (EDR3) was made available on 3 December 2020. It is based on the 34-month of satellite data obtained between July 2014 and May 2017. With respect to the earlier DR2 it includes proper motions improved by a factor 2, along with much improved photometry.

The Magellanic Clouds are close to the distance limits at which even the Gaia data accuracy starts to break down. Nonetheless, the enormous improvement in accuracy and numbers of stars represents a paradigm shift in understanding their structure and kinematics. A first assessment has been published by Gaia Collaboration et al. (2021c), from which the following is a summary.

THEIR SELECTION of stars as likely members of the Large and Small Magellanic Clouds used carefully chosen criteria based on each star's position, parallax, and proper motion. This was necessary both to remove foreground contamination from our Milky Way star populations, and also to allow a proper separation of the stars belonging to each. Further division according to star colour allowed the classification of stars into groups of similar approximate age and evolutionary phase.

Starting from a coarse selection of more than 27 million Gaia objects within a 20° radius for the LMC, and more than 4 million objects within a 11° radius for the SMC, further refinement according to proper motion and parallax led to a total of 11 156 431 objects belonging to the LMC, and 1 728 303 belonging to the SMC.

The fact that Gaia measures high-accuracy photometry in a number of spectral bands (and at the same time as the astrometry) is crucial: it allows the construction of the diagnostic 'colour–magnitude diagram'. Like the Hertzsprung–Russell diagram, this can be used to categorise each star according to its evolutionary phase.

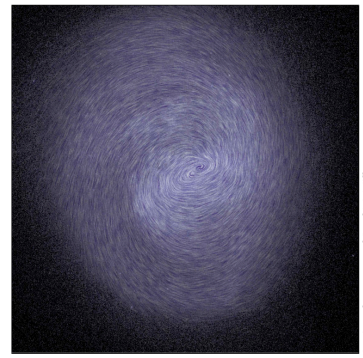
Accordingly, the members of the LMC and SMC can be further subdivided by age, notably as very young main sequence (ages < 50 Myr), young (50–400 Myr), and intermediate-age (up to 1–2 Gyr) main-sequence populations. In addition, we can identify stars of the red giant branch, the asymptotic giant branch (including long-period variables), RR Lyrae stars, classical Cepheids, and red clump stars.

ALL OF THIS allows the ordered and random motions for multiple stellar evolutionary phases to be separated for a galaxy disk outside the Milky Way for the first time, and allows the spatial structure and motions in the central region, the bar, and the disk to be examined.

THE LARGE Magellanic Cloud is taken as the prototype of the class of barred Magellanic spiral galaxies, characterised by an off-centre bar and one prominent spiral arm. The dynamical interactions between the Large and Small Magellanic Clouds are probably responsible for this and its other spiral features.

Given the huge numbers of stars in each of the LMC and SMC, future modelling is expected to reveal much more about their three-dimensional structure, their orientation with respect to the line-of-sight, and their precise distances. Meanwhile, the stellar density maps from Gaia reveal more concentrated and clumpier distributions for younger stars in the bar and inner spiral structure than the older disk stars.

Smoothed maps of the proper motion field, revealing the bulk stellar motions, shows a clear ordered rotation of the LMC, while the SMC is more chaotic. And we see that younger stars rotate faster than the older ones.



Luri et al. (2021) and Paul McMillan

Rotation of the LMC viualised by Gaia

One of the most prominent features in the outskirts of these and other interacting galaxies is the existence of a bridge between them. This is due to tidal forces that strip gas and stars from the least to the most massive galaxy. The relative position of the Milky Way with respect to the LMC and SMC places us in the privileged position of witnessing the close encounter between them, and of studying the Magellanic Bridge.

USING TWO different evolutionary phases (the young stars, and the red clump samples), Gaia Collaboration et al. (2021c) could trace the density and velocity flow from the SMC towards the LMC following the Magellanic Bridge. The Gaia data shows that it appears to wrap around the LMC, connecting with the young southern arm-like structure. Additionally, the outskirts of both Magellanic Clouds reveal other well-known features, such as the north and south tidal arms of the LMC and the northern enhancement in density of the SMC.

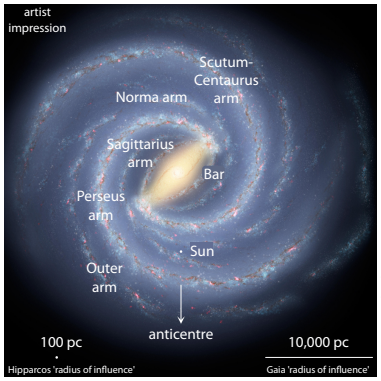
The next major data release scheduled for 2022, and further releases beyond that, will shed much further light on these majestic neighbours that encode so many details of star formation and galaxy evolution.

39. The Galactic anticentre

THE BROAD features of our Galaxy are today well established. Our Sun sits close to the mid-plane, and 8 000 pc out, in a rotating spiral galaxy comprising a flattened disk, a central spherical bulge and an inner bar. All are embedded in a spherical halo comprising normal matter as well as invisible dark matter of unknown form.

Our Galaxy has been shaped over its 10 billion year history by the accretion of other smaller galaxies, with evidence for major mergers in the distant past, as well as others still slowly ongoing today. Probing the details of these various populations, and understanding their history, is one of the major goals of modern astronomy. And in this, Gaia is proving to be something of a revolution.

OBSERVATIONS TOWARDS the Galactic centre probe its densest and most complex regions, with multiple populations superimposed along our line-of-sight to the centre: the thin and thick disks, star clusters, spiral arms, a bar-like inner structure, and the central bulge.



NASA/JPL-Caltech

The opposite direction, 180° away on the sky, is termed the Galactic anticentre. In this direction, star densities are lower, interstellar extinction is lower (meaning that observations can more easily probe to larger distances), and stars of the disk and halo dominate.

Nonetheless, the structural and dynamical phenomena on display, including the remnants of ancient and recently disrupted stellar systems of extragalactic origin, are proving to be an important window on its dynamics and past history.

It has not been easy to understand how these various features have arisen. For example, the prominent spiral arms, seen in 60% of all galaxies, arise from some combination of density waves, and self-propagating star formation, but they remain incompletely understood.

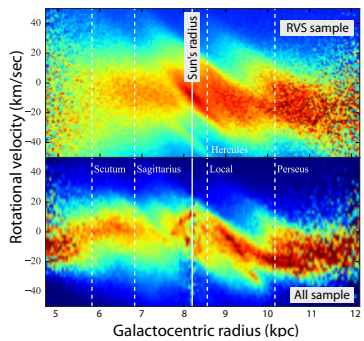
THE BAR towards the centre of our Galaxy was recognised only in the 1960s, while the fact that the flattened disk actually comprises two separate 'populations' – a thin disk, and a thick disk – was discovered only in the 1980s. The picture has become even more confusing over the past 20 years, with various other complex structures and dynamical features still being uncovered.

Gaia is providing huge numbers of stars with well-defined distances and velocities, allowing many of these features to be mapped in much greater detail, such that better models and theories of them can be developed.

SINCE THE late 1950s, astronomers have recognised that the disk is not flat, but warped – slightly curved up on one side, and down on the other. One idea was that this resulted from an irregular-shaped dark matter halo. But velocities of 12 million giants stars, from Gaia DR2, suggest instead that it is a ripple-like effect due to the Sagittarius dwarf galaxy, which orbits the Milky Way, and which has probably ploughed through the Galaxy's disk several times in the past (Poggio et al., 2020).

Amongst other features suspected before the Gaia survey are vertical asymmetries in the star counts linked to vertical bending and breathing waves, and large-scale substructure and velocity patterns in the disk.

Gaia DR2 has already clarified many of these. From orbital velocities around the Galaxy, between 5–12 kpc from the centre, Kawata et al. (2018) found several prominent diagonal 'ridges' in two large samples: of 861 680 radial velocity stars, and 1 049 340 brighter than 15.2 mag.



Kawata et al. (2018)

In another example, Bennett & Bovy (2019) used Gaia DR2 to confirm that the local disk is undergoing a wave-like oscillation, and they established a dynamical model of this perturbed local vertical structure.

THE ASTROMETRIC MEASUREMENTS from EDR3 have allowed further progress, reported in some detail by Gaia Collaboration et al. (2021a). Distant regions of the Galaxy in all directions, especially in the direction of the anticentre, can now be explored using positions and velocities of unprecedented quality. We can now probe structures and motions to distances of 15 000 pc or more from the Galactic centre, out to the very outskirts of our Galaxy's disk. And spatial and velocity structures can be rigorously classified as a function of age and radius.

IN THE OUTER DISK, beyond 12 000 pc, the velocity field is seen even more prominently – and with further structure – than with Gaia DR2. Here, velocities are dominated by an upwards warping motion of 5 km s^{-1}, and attributed either to the passage of the Sagittarius dwarf galaxy, or to our ancient collapsing disk that never achieved dynamical equilibrium.

The ridge-like features, already seen in the circular velocities of the Gaia DR2 data, are now detected in EDR3 up to 14 000 pc from the Galactic centre. Two additional ridges are apparent, still part of the disk's circular rotation, but now extending out to 16–18 000 pc. The precise nature and origin of all of this detailed velocity structure is not yet known.

IN THE GAIA DR2 data, two distinct populations of redder and bluer stars were evident in the Hertzsprung–Russell diagram of the subset of stars with large tangential velocities near the Sun. These high-velocity stars are nearby members of the stellar halo population.

Gaia Collaboration et al. (2018a) suggested that the bluer stars are an accreted population arising from the ancient merger of the Gaia 'Enceladus' galaxy, while the redder stars are a distinct thick disk population that was present at the time of the Enceladus merger.

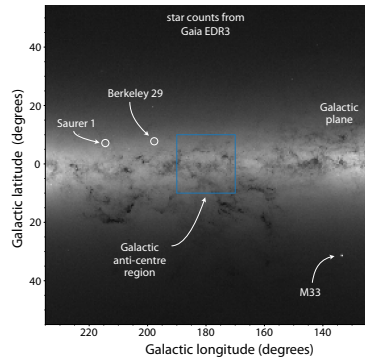
The Gaia EDR3 data extends the distance out to which the Gaia Enceladus debris merger can be traced, to distances of 17 000 pc or more from the Galactic centre. The new data also show that most of the (local) halo is made up of debris from this single accretion event.

OTHER DENSITY STRUCTURES are seen towards the edge of the disk in the anticentre direction. The deep Sloan Digital Sky Survey had already hinted at the existence of a 100°-wide structure in their star count maps in 2002. Now known as Monoceros, and some 10 000 pc distance from the Sun, later studies have confirmed its existence, and its large extension on the sky. Together with the Anticentre Stream, and the Triangulum–Andromeda 'overdensities', they are all part of a complex and sub-structured outer disk.

The earlier idea that these could be the remains of an accreted dwarf galaxy seem less likely today, in part because there is no obvious progenitor, and in part because their stars are so similar to the rest of the disk.

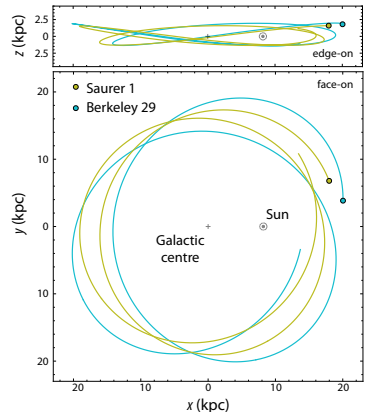
OPEN CLUSTERS can also be used to trace the global structure and evolution of the disk, and a major advance in this area already took place using Gaia DR2. Berkeley 29 and Saurer 1 both lie in the anticentre direction and, with ages of several Gyr, are among the oldest Galactic open clusters known.

Their unusual location, 20 000 pc from the centre, and more than 1000 pc above the mid-plane, led several authors to question whether they are really associated with the disk, or whether they had an extragalactic origin. The difficulty of interpretation was simply because, at these very large distances, small proper motion errors translate into very large uncertainties in space velocities.



Using the vastly better Gaia EDR3, and the much improved ability to assign cluster membership, Gaia Collaboration et al. (2021a) showed that the two clusters are indeed in disk-like orbits.

But their distant location raises questions as to their origin: does the disk really extend so far out, or were these clusters delivered there by other means, such as radial migration, interaction with a passing galaxy, or born from material expelled from the disk?



Orbits of Saurer 1 and Berkeley 29

These clusters may be small in size, but they provide important clues about the nature of the outer disk.

THE TRANSFORMATION brought by Gaia is hard to overstate. Where space science of the late 20th century gave astronomers a catalogue of 120 000 stars, and a 'radius of influence' of around 100 pc, Gaia provides a census of 2 000 000 000 stars, extending this radius of influence out to the very edges of our Galaxy's disk.

As Antoja et al. (2021) expressed it: *'The Gaia EDR3 data, together with the advantage of having astrometry and photometry from the same mission, have allowed us to extend the horizon for exploration towards the very end of the disk, to travel to the past to explore its ancient components, and to detect its small constituents and phase space features with much greater resolution.'*

40. The distance of Omega Centauri

GLOBULAR CLUSTERS are tightly bound spherical groups of up to a million or more stars. In contrast to the younger open clusters, which are found mainly in the disk, globular clusters are amongst the oldest populations, contain many more stars, and are more representative of the overall Galaxy population, occurring in the disk, in the bulge, and most prominently in the halo.

There are more than 150 globular clusters known in our own Milky Way, some 20% within a few kpc of its centre, and others extending out to distances of 30–40 kpc. The nearest, M4, is at about 2.2 kpc. Most galaxies of sufficient mass in the Local Group and beyond have their own systems of globular clusters.

Ages of the oldest are comparable to that of the Universe itself, as derived from its expansion rate, suggesting that at least some were formed early in its formation. Others appear to be remnants of the Galaxy's early accretion phase, captured from smaller galaxies during mergers or collisions. And although they represent a negligible fraction of the light and mass of the stellar halo, they are important tracers of our Galaxy's age and dynamics.

THE PROPERTIES of even the nearest globular clusters are such that Hipparcos could make no direct contribution to determining their distances or ages: they lie well beyond the horizon of its parallax measurements, even their brightest stars are too faint, and the average sky density of Hipparcos targets, around 2.5 per square degree, would in any case imply coverage of only one or two cluster stars – even if they had been accessible.

But Hipparcos made important *indirect* contributions to their distances and ages. The distances, and therefore luminosities, of certain nearby stars could be measured and calibrated. Then if these stellar types are also found in globular clusters, their apparent luminosities can be inferred, and used to estimate their distances.

At the same time, globular cluster distances fix the luminosities of stars on the cluster's main sequence, and hence the cluster's age from main-sequence fitting or, more effectively, the main-sequence 'turn-off point' from theoretical models of stellar evolution.

Around 1994, some estimates of the Hubble constant, including those from HST observations of Cepheids out to the Virgo cluster, indicated a value as high as $H_0 = 80 \pm 17\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, and a resulting Universe expansion age of 8 Gyr. By 2001, further observations gave a revised $H_0 = 72 \pm 8\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, and an expansion age of around 12 Gyr (Freedman et al., 2001).

Before the Hipparcos results in 1997, an awkward problem was the 'age paradox': the ages of the oldest globular clusters were estimated to be around 15 Gyr, or even older in the case of NGC 6541.

The disturbing implication for cosmology was that some globular cluster ages appeared to exceed the 'expansion age' of the Universe. Something was wrong.

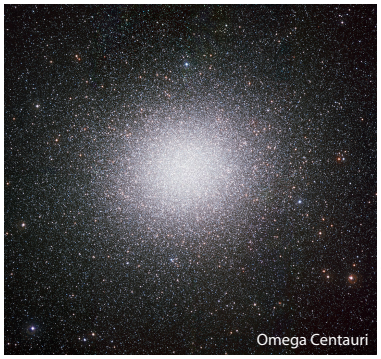
A POSSIBLE SOLUTION to the paradox came from Hipparcos parallaxes for the nearest Cepheids (Feast & Catchpole, 1997). The inference that the Large Magellanic Cloud Cepheids were 10% further than previously estimated, and thus brighter, led to the conclusion that the globular clusters were more distant than previously thought, that their luminosities were larger, and that their ages were younger than previously thought.

The Hipparcos contribution to the ages of globular clusters rested on the observation of nearby subdwarfs, i.e. metal-poor halo stars which happen to be passing close to the Sun at this time. Since they also occur in globular clusters, careful luminosity calibration as a function of metallicity allows them to be used in globular cluster distance determinations through main-sequence fitting.

Independent distance estimates can also be made using the calibrated luminosities of RR Lyrae variables, since these very luminous stars also occur in both the field and in globular clusters. However, even the nearest RR Lyrae are at the limits of the Hipparcos parallaxes.

Whichever way teams tried to tackle the problem, major advances in determining globular cluster distances, ages, and indeed dynamics would have to wait for the arrival of direct trigonometric parallaxes able to reach distances of several *thousands* of parsecs.

OMEGA CENTAURI (ω Cen, NGC 5139) is a globular cluster in the constellation of Centaurus, first identified as non-stellar by Edmond Halley in 1677, and one of the few such systems visible to the naked eye.



At a distance of about 5200 pc, it is the largest, most massive globular cluster in the Milky Way. It contains some 10 million stars, with a total mass of 4 million solar masses.

With a diameter of 50 pc, it appears almost as large as the full Moon, and it is thought to have originated as the core remnant of a disrupted dwarf galaxy. But it is too distant, too crowded, and just too faint, for any of its stars to have been observed by Hipparcos.

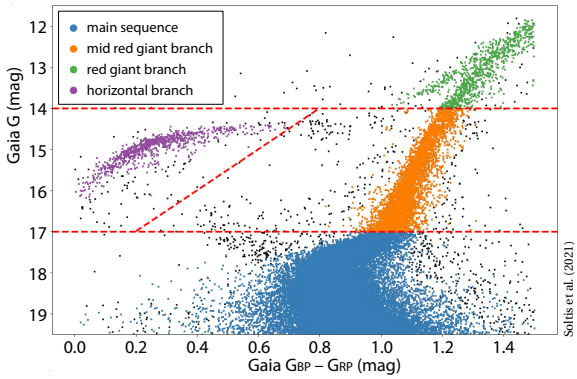
GAIA IS NOW transforming our understanding of this important stellar system. As Soltis et al. (2021) stated: *'The recent Gaia Early Data Release 3 opens a new chapter in the measurement of parallaxes, placing the precise and accurate determination of the distances to nearby Galactic globular clusters within reach. None may be more prized than that of ω Cen... A precise determination of its distance will characterise the luminosity of a broad sample of stellar types.'*

An important feature of the Hertzsprung–Russell diagram (or colour–magnitude diagram) seen in globular clusters is the 'tip of the red giant branch'. This is the point of maximum brightness for red giant branch stars, and this tip originates from the sudden start of helium fusion in low-mass stars. After this 'helium flash', stars quickly expand and dim, resulting in a rapid decline in numbers at magnitudes brighter than the tip.

Once calibrated, it provides an important distance indicator for evolved stellar populations that can be used to reach the galaxy hosts of Type Ia supernovae, and so influence determination of the Hubble constant.

Because the tip of the red giant branch is not the identity of some individual star, but rather a feature of the entire globular cluster population, its luminosity has not been easy to determine in the past. This is partly because of the impossibility, before Gaia, of determining the individual parallaxes of such distant objects. But also because, since bright red giants are relatively rare, few clusters contain enough stars to reasonably define the position of this 'tip'.

In practice, ω Cen provides the best opportunity in the Milky Way, in part because it is 'relatively' nearby, but also because it contains nearly 200 stars within a magnitude of the tip, enough for a rather good definition.



Colour–magnitude diagram of 66 467 members of ω Cen.

The numbers of stars available with Gaia is spectacular. Soltis et al. (2021) selected 178 548 from EDR3 within 45 arcmin of the cluster centre. They estimated a mean cluster proper motion of $\mu_\alpha = -3.25\,\mu$as, $\mu_\delta = -6.76\,\mu$as, in good agreement with the value found for DR2 (Baumgardt et al., 2019).

Selecting stars participating in this common space motion then resulted in 108 054 candidate members. Further restriction according to their location in the colour–magnitude diagram resulted in 66 467 members with good Gaia astrometry and 2-colour photometry.

Their resulting mean parallax of ω Cen is 0.191 ± 0.001 mas, corresponding to a distance of 5236 ± 28 pc, and in good agreement, for example, with the classical photometric distance of 5.2 kpc from Harris (1996).

WHAT DOES this mean for cosmological distance calibrations based on the tip of the red giant branch? Soltis et al. (2021) estimate that its I-band absolute magnitude is $M_I = -3.97 \pm 0.06$ mag. This is slightly fainter, by 0.07 mag, than the calibration used in recent determinations of the Hubble constant (Freedman et al., 2019).

This change raises the value of H_0 by 3.2%, so yielding a Hubble constant from the SN Ia distance ladder, based on this 'tip', of 72.1 ± 2.0 km s^{-1} Mpc^{-1}.

This is in good agreement with other local measures, including those from Cepheids. But it is significantly larger than that predicted by the Planck mission's cosmic microwave background data used to calibrate ΛCDM models (Verde et al., 2019). The 'age paradox' may have disappeared, but this 'Hubble Tension' remains.

THE DETERMINATION of the parallax of ω Cen represents a new milestone of distance measurements in astronomy.

But while the focus here has been on the cluster's distance, Gaia has also already thrown much new light on various properties of ω Cen not touched on here, including the cluster's three-dimensional structure and depth; its orbit around the Galaxy; and the tidal stellar streams torn off the cluster as it orbits the Milky Way.

41. The age of our Milky Way Galaxy

THE AGE OF THE EARTH has been pieced together from a complex story involving the geological and fossil record, radioactive dating, and evidence from the Moon and meteorites. The oldest known meteorites in the solar system are around 4.56 billion years old, and it appears that the Sun was born at around the same time.

The oldest white dwarfs in our Galaxy have been 'cooling' for about 12.7 billion years. The oldest globular clusters are about 13.4 billion years old. Distant Cepheid variable stars, which trace the expansion age of the Universe, suggest that it has been expanding, at around the current rate, for about 13.7 billion years.

And maps of the cosmic microwave background radiation, interpreted in terms of the temperatures and structures that existed when the radiation was emitted, yield ages close to 13.77 billion years (or 13.77 Gyr).

ACCORDING TO TODAY'S 'standard model' of Big Bang cosmology (Lambda cold dark matter, or ΛCDM cosmology), the Universe contains three main constituents: dark energy, cold dark matter, and ordinary matter. One of its great triumphs is its success in describing structure formation in the early Universe, and its growth over cosmological time, leading to the galaxies and clusters of galaxies that we observe today.

Observations of the stellar populations of our Galaxy, supported by these cosmological models, suggest that our Milky Way comprises both a thin and thick disk population, along with a central spherical bulge and more extended central bar. All are embedded in a vast spherical halo, itself containing extended stellar 'tidal' streams which originated from the capture of smaller galaxies much earlier in our Galaxy's history.

We now have a rather secure understanding of the age of our solar system, and of the oldest stars and stellar populations making up our Galaxy. Today, we are in a position to ask – and answer – more detailed questions about the sequence of events that made up our Galaxy's history. How old is the thin disk, and the thick disk? How old is the stellar halo? And can we discern the age of the stellar streams that are contributing to it?

THE MOST GENERAL METHOD for estimating stellar ages rests on mathematical models of stellar evolution. A star's mass and chemical composition are the most important inputs, and its luminosity and temperature provide the main observational constraints.

Detailed physical models and extensive computer calculations predict the changing state of the star over time, yielding a data grid that can be used to determine the evolutionary track of the star across the Hertzsprung–Russell diagram (charting absolute magnitude versus temperature), or the closely-related colour–magnitude diagram. The age of any particular star is estimated by comparing its physical properties with those of stars along a matching evolutionary track.

Spectroscopic observations provide both the star's temperature and its chemical composition, while the star's distance is the most crucial quantity in determining its luminosity. Gaia distances for hundreds of millions of stars out to many kiloparsec distance hold the key to unravelling their ages, and the ages of entire stellar populations throughout our Galaxy, both as a function of location and of their kinematics.

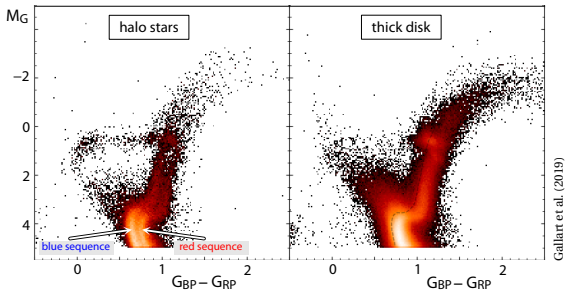
APPLIED TO OUR GALAXY'S main stellar populations, Gallart et al. (2019) used Gaia DR2 astrometry to extract two large samples of halo stars and thick disk stars. Both samples were restricted to stars within 2000 pc of the Sun, such that accurate distances and absolute magnitudes could be derived *directly* from the parallaxes.

For their halo star sample, they selected stars with large tangential velocities ($> 200\,\mathrm{km\,s^{-1}}$) relative to the Sun, such extreme space velocities being characteristic of the halo population. Their sample of a kinematically defined halo population contains about 60 000 stars.

For their thick disk sample, they employed the same distance limit, while using the distance and directional information to retain only those stars more than 1100 pc above or below the Galactic plane (and not included in the halo sample). At these distances, the majority of stars are expected to belong to the thick disk, rather than the younger thin disk component.

Their thick disk sample contains around 500 000 stars. In comparison with previous studies of the thick disk population, it is worth stressing that their sample is defined purely morphologically, rather than being based on chemical or kinematic properties.

The location of both samples in the Gaia colour–magnitude diagram, M_G versus $G_{BP} - G_{RP}$, are shown here. Again, it is worth emphasising that both axes, absolute magnitudes in the G band, and colours in the $G_{BP} - G_{RP}$ bands, are as measured by Gaia.



WHAT GAIA REVEALS is that the halo population consists of two distinct sub-groups: a 'blue' sequence and a 'red' sequence, indicating the presence of two distinct sub-populations within the halo.

At this point, a little history is probably useful. In trying to understand the 'big picture' of galaxy formation and evolution, a lengthy debate started more than 50 years ago, when two different possibilities were being considered: either that galaxies formed by a sort of 'monolithic collapse' of the available gas, or that formation and evolution was mediated by satellite accretion.

Support for the latter has been growing over the past two decades, including the discovery of stars from an ancient merger seen in the Hipparcos data (Helmi et al., 1999). Evidence for two populations within the 'hot' (high-velocity) halo stars has also been accumulating from stellar abundances (e.g. Nissen & Schuster, 2010).

Based on Gaia data, the blue sequence has recently been associated with a significant merging event, which has been named Gaia–Enceladus. This occurred early during the Milky Way formation (Helmi et al., 2018), and has been described separately in these 'essays'.

The nature of the red sequence has been less clear. One hypothesis is that it is somehow associated with the thick disk, perhaps being ancient disk or bulge stars 'heated' to halo kinematics by a galaxy merger.

THE RECENT GAIA results from Gallart et al. (2019) represent a further important advance in understanding our Galaxy's halo for the following reasons.

Comparing the halo colour–magnitude diagram with theoretical stellar evolution models shows that the red and blue sequences arise from two populations with different chemical compositions, but of *the same age*.

That they are coeval is a remarkable finding. But their models also show that these populations formed very early on in the life of the Universe. Thus, while current estimates put the age of the Universe at around 13.77 Gyr, this halo population formed with a peak age of 13.4 Gyr, and with half of the star having formed by 12.3 Gyr. Indeed, not only did they form at similar times, but they also stopped forming at similar times.

We can go still further in this cosmological sleuthing. There is, today, a well-established relation between a galaxy's mass and the fraction of 'metals' (elements heavier than H and He in astronomy-speak) in its stars. This means that the stars in the red sequence, being more metal rich, must have formed in a galaxy more massive than that in which the blue sequence formed.

AN IMPORTANT CONCLUSION appears to be inevitable: both populations were involved in a merger event, with the red sequence belonging to the main progenitor of the Milky Way, and the blue sequence belonging to a smaller accreted galaxy, the one now referred to as Gaia–Enceladus. The metallicity difference between the more massive Milky Way progenitor and the smaller accreted galaxy suggests that their mass ratio was about 4:1.

The inference is that this ancient galaxy–galaxy encounter heated some of the main progenitor stars that had been forming in a disk-like structure, increasing their space velocities (through close encounters) to the sort of extreme kinematics that lead them to be classified as halo stars. Encouragingly, the existence of this type of halo star had already been predicted by cosmological simulations of Milky Way-type galaxies based on kinematic results from Gaia DR1 (Bonaca et al., 2017).

THE CONCLUSIONS OF Gallart et al. (2019) clearly summarises the current picture of our Galaxy's formation, deduced from the Gaia DR2 data.

New stellar age distributions enabled by Gaia, aided by state-of-the-art cosmological simulations of disk galaxy formation, present a clear picture of the formation of our Galaxy: a primitive Milky Way had been forming stars during some 3 Gyr when a smaller galaxy, which had been forming stars on a similar timescale but was less chemically enriched owing to its lower mass, was accreted by it.

This merger heated a fraction of the existing stars in the main progenitor to a stellar halo-like population. A ready supply of infalling gas during the merger ensured the maintenance of a disk-like configuration, with the thick disk continuing to form stars at a substantial rate.

The measured age distributions indicate that the thick disk reached its peak star formation rate around 9 Gyr ago, or some 4.5 Gyr after the first stars formed in the Milky Way. Subsequently, around 8–6 Gyr ago, the gas settled into a thin disk that has continued to form stars up to the present day.

42. Surprises in the HR diagram

All professional astronomers, and many amateurs, will have heard of the Hertzsprung–Russell (or 'HR') diagram. Like the works of Shakespeare, perhaps, few will be familiar with all of its details, but different experts have studied every aspect of it, such that any new or unexpected features will be greeted with surprise and excitement: they tell us something about the workings of Nature that we didn't know before.

The Hertzsprung–Russell diagram shows the relationship between a star's luminosity (or absolute magnitude) versus its temperature (or colour). It was created independently around 1910 by the Dane Ejnar Hertzsprung and American Henry Norris Russell (physicists will know his name in the context of quantum mechanical 'LS coupling' or Russell–Saunders coupling).

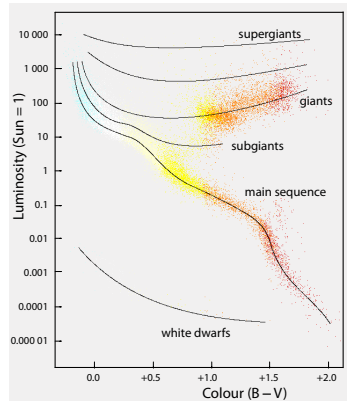
This graphical presentation of stellar properties, in which stars of higher luminosity are towards the top of the diagram, and stars with higher surface temperature (or bluer colour) are towards its left side, facilitated a major step in understanding stellar evolution.

It remains a powerful tool for interpreting, through stellar evolutionary models, the properties of individual stars, star clusters, and entire stellar populations.

As an example of its main features, and the state-of-the-art pre-Gaia, the HR diagram shown here was constructed from 22 000 stars from the Hipparcos Catalogue, supplemented by 1000 low-luminosity stars (red and white dwarfs) from the Catalogue of Nearby Stars.

The ordinary hydrogen-burning 'dwarf' stars, like the Sun, are found in the main band running from top-left to bottom-right, which is referred to as the 'main sequence'. Giant stars form their own clump on the upper-right side of the diagram. Above them lie the less common, and particularly bright, supergiants.

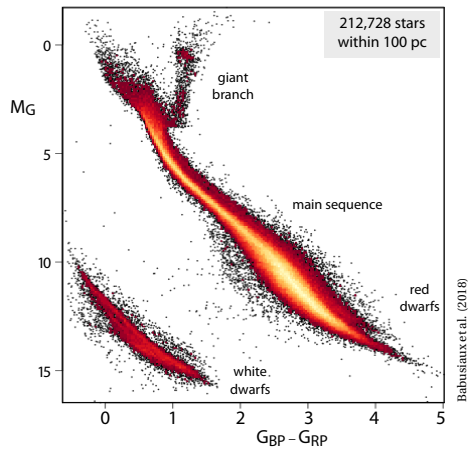
Tracing a band at the lower-left are the white dwarfs. These are the dead cores of old stars which have finally exhausted their H or He 'fuel supply' and, with no energy source remaining, simply cool, slowly over billions of years, down towards the bottom-right of the diagram.



Richard Powell (2011, Wikimedia, CC BY-SA 2.5)

With the Gaia second data release, Gaia DR2, with high-precision distance measurements to more than 1.7 billion stars, many new examples and applications of the HR diagram are appearing. The unprecedented numbers of stars now available for study paints a remarkable panoramic picture of stellar evolution.

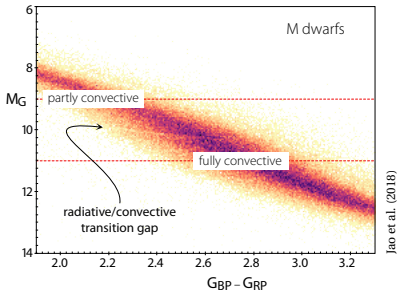
A detailed discussion, and many examples, of these HR diagrams constructed from DR2 are given by Gaia Collaboration et al. (2018a). In the following I will focus on some new features in these diagrams at the lowest luminosities: the red dwarfs and white dwarfs.



Babusiaux et al. (2018)

THE FIRST OF THESE is a distinct gap which has been discovered in the main sequence for M dwarfs, in the region towards the lower-right of the HR diagram.

While the main sequence is generally smoothly populated as a function of stellar temperature, a tiny but pronounced discontinuity in number density occurs within the sequence of the coolest, faintest red dwarfs. This gap was first reported in the Gaia data by Jao et al. (2018), confirming a similar feature already hinted at in infrared data from 2MASS.



Qualitatively, the gap arises because of the very different internal structure of stars on either side of it: low-mass M dwarfs have a fully convective interior, while more massive stars (including the Sun) have a convective envelope surrounding a radiative zone, a denser area where these convective motions are suppressed. The transition occurs at about $0.35 M_\odot$ (solar masses).

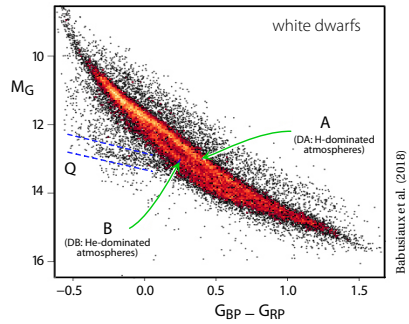
A detailed theoretical explanation was soon forthcoming (MacDonald & Gizis, 2018). The gap arises because the convective motions in the cores of low-mass stars help mix the intermediate nuclear fusion products, allowing the star to fuse hydrogen more efficiently.

Their detailed models show that the mixing of ^{3}He during the merger of the envelope and core convection zones occurs over a narrow range of masses. This successfully replicates an associated dip in the luminosity function which is responsible for the gap.

IN THE LOWER-LEFT part of the HR diagram, the white dwarf sequence shows several remarkable features, identified by Gaia Collaboration et al. (2018a).

They constructed a sample with relative parallax uncertainties better than 5%, yielding a set of 26 264 white dwarfs. The accurate parallaxes, combined with simultaneous accurate Gaia photometry, then yield accurate absolute magnitudes, which allow them to be precisely located in the Hertzsprung–Russell diagram.

Several new structures are evident. First is a clear concentration of stars distributed continuously from the upper-left to the lower-right (A), coinciding with the evolutionary tracks for the DA white dwarfs (whose envelopes are dominated by hydrogen). Just below the main band is a second, distinct concentration (B). Gaia Collaboration et al. (2018a) attribute this to the DB white dwarfs (whose atmospheres are dominated by helium).



This prominent split in the white dwarf cooling sequence between H and He white dwarfs was actually first detected in the colour–colour diagrams from the Sloan Digital Sky Survey. But the Gaia data reveal it for the first time in the HR diagram. The very narrow sequences also confirm the sharp peak of their mass distribution around $0.6 M_\odot$. Further details are given by, e.g., El-Badry et al. (2018) and Gentile Fusillo et al. (2019).

There are also a number of white dwarfs which lie above the main DA sequence, and these are attributed to white dwarfs in binary systems.

A THIRD, WEAKER concentration is evident in the figure. A rising transverse feature labelled 'Q', it is a faint but significant 'band' of stars visible between the two dashed lines. Tremblay et al. (2019) showed that this is due to core crystallisation as the white dwarfs cool.

In the early 1960s, Abrikosov, Kirzhnitz, and Salpeter independently predicted that their cores should slowly crystallise as they cool, resulting in a lattice rather than a gas. In the process, the hot plasma fluid (of nuclei and electrons) releases an associated latent heat, providing a new source of energy that delays the object's cooling.

There has only been indirect evidence for this theory to date, although the details are crucial in estimating cluster ages. Gaia reveals the crystallisation as a mass-dependent pile-up in the HR diagram as they spend time at this location while they release their latent heat.

Tremblay et al. (2019) showed that the observed position of this transverse sequence (Q) agrees with the range of absolute magnitudes and colours at which the bulk of the latent heat from crystallisation is released over the full range of white dwarf masses.

THE GAIA DATA on white dwarfs thus provides direct evidence that a first-order phase transition occurs in high-density Coulomb plasmas. Importantly, it is a theory that cannot be tested in terrestrial laboratories because of the extreme densities involved.

Meanwhile, it will be some 5 billion years until the Sun evolves into a white dwarf, and another 5 billion years before it cools enough to form a crystalline sphere.

I imagine that Hertzsprung and Russell would have been astonished at these results: the spectacular emergence of subtle physics from exquisite data.

43. Cepheid variables

CEPHEIDS ARE pulsationally unstable stars, located in a narrow region of the HR diagram, with typical periods of 1–30 days, but extending up to about 100 days.

There are two sub-classes. Classical Cepheids (or δ Cephei stars) are young high-mass core He-burning supergiants, Population I objects found in the Galactic plane, notably in spiral arms and in open clusters.

Type II Cepheids are low-mass metal-poor Population II objects found at high Galactic latitudes, in the Galactic bulge, and in globular clusters (and sub-divided by period into BL Her, W Vir, and RV Tau-type variables).

The immediate precursors of the classical Cepheids are massive young O and B main-sequence stars. As they evolve rapidly off the main sequence, they pass through a zone (termed the 'instability strip') where their outer atmospheres are unstable to periodic radial oscillations. High-mass stars pass through the instability region at higher luminosities (cooler temperatures) than lower-mass stars, resulting in a Cepheid instability strip which slants upwards and to the right in the HR diagram.

Basic physics considerations lead to the existence of a mass–luminosity relation for Cepheids, and hence also a radius–luminosity relation. However, since neither mass nor radius are easily observable for the majority of stars, the mass–luminosity relation cannot be used to predict luminosities, nor therefore distances.

THE IMPORTANCE OF Cepheid variables as distance indicators is that there exists a correlation between period and luminosity, discovered empirically by Henrietta Leavitt (1908), and subsequently explained theoretically (a historical review is given by Fernie (1969).

The relationship nevertheless shows a significant scatter about the mean line, even when corrected for reddening, due to the finite (temperature) width of the instability strip. If a colour-term is introduced, the scatter is significantly reduced.

While the Cepheid period–luminosity relation has traditionally provided to be the most accurate method to derive distances to nearby galaxies, various complications have been encountered in practice.

THERE IS AN ENORMOUS literature on Cepheid variables, and their application to the determination of the astronomical distance scale, both within the Galaxy, and beyond.

One main goal of Cepheid studies is to establish the slope and zero-point of the period–luminosity relation, such that an observed period yields the object's luminosity and thereby its distance. Until the Hipparcos results, the most accurate zero-point for the period–luminosity relation came from Cepheids in open clusters and associations through main-sequence fitting.

An important and related question is whether the period–colour and period–luminosity relations for classical Cepheids in the Galaxy, and in the Large and Small Magellanic Clouds, have the same slopes and zero-points; differences would greatly complicate the use of Cepheids for the extragalactic distance scale.

IN ADDITION TO THEIR USE as distance indicators, the fact that Cepheids can be seen to large distances, and the fact that they reflect the young population of the Galaxy, means that they also provide an important tracer of spiral arms, while their proper motions provide a powerful probe of Galactic rotation.

Pre-Hipparcos studies of Galactic rotation could only sample a small region around the Sun. The first such contribution making use of the Hipparcos data to explore a significant region of the Galactic disk was by Feast & Whitelock (1997). They used 220 Cepheids with Hipparcos astrometry to derive the Oort constants A and B from the first-order expression for Galactic rotation.

Interesting information is also encoded in the vertical distribution of Cepheids above and below the Galactic plane, and its age dependence. In a simplified picture, Cepheids with a very young age are found preferentially close to the Galactic plane, their assumed birth sites. Evolving in scale height with age as a result of their initial vertical velocity component, they reach their maximum distance and return to the plane after times depending on the local mass density, somewhere in the range of 70–100 Myr.

THE HIPPARCOS CATALOGUE contained 280 Cepheids, of which 32 are either Type II (mainly W Vir stars) or double-mode Cepheids. Of the 248 classical Cepheids, 32 are first-overtone pulsators.

The mean standard error of the 223 Hipparcos Cepheid parallaxes considered by Feast & Catchpole (1997) is about 1.5 mas. The majority are beyond about 500 pc, such that the parallaxes are typically very small, and of limited individual value. The closest is Polaris (α UMi), with $\pi = 7.56 \pm 0.48$ mas or $d = 132 \pm 8$ pc. Polaris is too bright to appear in current Gaia data releases.

THE GAIA RESULTS are transforming all of these areas of study. The high-accuracy parallaxes, combined with the multi-colour multi-epoch precision photometry, makes Gaia extremely powerful for identifying and characterising variability across the entire HR diagram.

Gaia DR1 included 599 Cepheids (and 2595 RR Lyrae stars) in the Large Magellanic Cloud region, observed at high cadence during the first 28 days in the 'ecliptic poles scanning configuration' (Clementini et al., 2016).

For Gaia DR2 (the first 22 months of the mission), a 'Specific Object Study' pipeline was used to validate and characterise Cepheids and RR Lyrae stars, originally using the period–amplitude and period–luminosity relations in the G band, and subsequently extended to G_{BP} and G_{RP} (Clementini et al., 2019; Rimoldini et al., 2019).

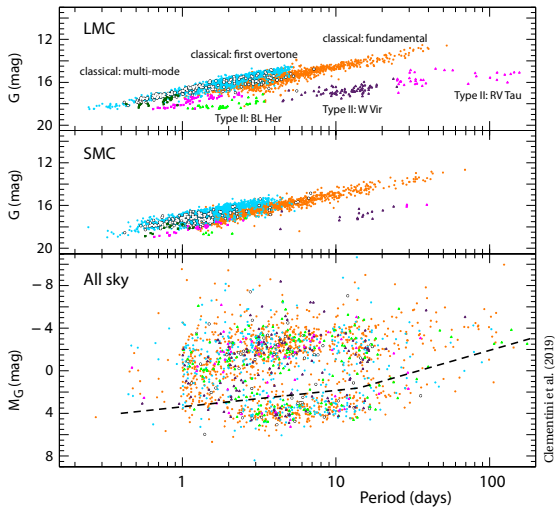
Gaia DR2 provides results, along with mean magnitudes and pulsation characteristics, for 9575 Cepheids, of which 3767 are in the LMC, 3692 are in the SMC, and 2116 are elsewhere ('all-sky'). The majority of those in the Magellanic Clouds were already known from the OGLE survey, although Gaia DR2 lists 118 new objects.

The all-sky sample includes Cepheids and RR Lyrae variables in 87 globular clusters and 14 dwarf galaxies (the Magellanic Clouds, 5 classical and 7 ultra-faint dwarfs), of which 350 Cepheids are new discoveries.

Metallicities derived from the Fourier parameters of the light curves are also given for 3738 fundamental-mode classical Cepheids with periods below 6.3 days.

IN ADDITION TO the classical and Type II Cepheids, the Gaia 'Specific Object Study' pipeline also identifies the less common double-mode Cepheids (which are observed to pulsate in two modes at the same time, usually the fundamental and first overtone), as well as the shorter-period high-mass 'anomalous Cepheids', whose evolutionary status is somewhat unclear.

THE FIGURE SHOWS the period–luminosity relation for all Cepheids identified in DR2 by Clementini et al. (2019), divided into three sky regions, and shown as a function of apparent magnitude for theLMC and SMCs, and as a function of absolute magnitude for the all-sky sample. All are uncorrected for reddening.



The colour coding, identical in all panels, is divided into the classical Cepheids (sub-divided into fundamental-mode, first-overtone, and multi-mode pulsators), and the Type II Cepheids (sub-divided by period into BL Her, W Vir, and RV Tau-type variables).

A much larger scatter is seen in the all-sky period–luminosity distribution. Clementini et al. (2019) already considered that many of the sources below the dashed line are likely to be a combination of mis-classifications, sources with very high reddening, or the consequences of a simplified treatment of binary/multiple sources.

A further more detailed analysis has subsequently been undertaken by Ripepi et al. (2019), while an independent analysis of the purity of the DR2 Cepheid sample is discussed by Molnár et al. (2018).

I WILL MAKE ONLY a brief mention of some of the other analyses that have been based on the Cepheid data from Gaia DR2, through to the end of 2020. Specific application to the estimation of the Hubble constant is taken as a separate topic elsewhere.

Kervella et al. (2019) combined the Hipparcos and Gaia DR2 positions to determine the mean proper motion of a sample of classical Cepheids, searching for proper motion anomalies caused by close-in orbiting companions. They concluded that the binary fraction of classical Cepheids is likely to be above 80%.

Other studies have used the Gaia Cepheid data to characterise our Galaxy's rotation curve (e.g. Mróz et al., 2019; Kawata et al., 2019; Ablimit et al., 2020). as well the vertical component of the velocity vector (Skowron et al., 2019b), and our Galaxy's structure more generally (Skowron et al., 2019a), which I will look at separately.

Marconi et al. (2020) derived theoretical mass-dependent 'period–Wesenheit' (reddening-free) relations in the various Gaia photometric bands, from which they derive the individual mass of each pulsator.

44. The Hubble constant from Cepheids

THE DISCOVERY, by Edwin Hubble almost a century ago, that distant galaxies are moving away from us at speeds proportional to their distance, was the first observational evidence for the expansion of the Universe. Today, Hubble's 'law' still serves as one of the key pieces of evidence supporting its 'Big Bang' origins.

Hubble's estimates of the distances to spiral galaxies (then known as spiral nebulae) were based on luminous variable Cepheid stars within them. His observations actually followed theoretical work over the preceding decade, independently by Russian physicist Alexander Friedmann and Belgian astronomer Georges Lemaître, based on the equations of general relativity, and suggesting that the Universe could be expanding.

At the time of the discovery and later development of Hubble's law, galaxy redshifts were interpreted as Doppler velocity shifts in the context of special relativity. The 'Hubble constant', H_0, was similarly formulated as the constant of proportionality relating the galaxy's distance, D, with its recession speed, v, i.e. $v = H_0 D$.

The Hubble constant is most frequently quoted in km/s per Mpc ($\mathrm{km\,s^{-1}\,Mpc^{-1}}$), i.e. giving the speed in km/s of a galaxy at a distance of 1 Mpc (although we may note that this is simply equivalent to $\mathrm{s^{-1}}$ in SI units). Its value is about $70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$.

THINGS HAVE certainly become much more complicated since the time of Hubble.

We now know, for example, that the Hubble 'constant' actually varies with time in standard cosmological models. And all observations of extremely distant objects are actually observations relating to the distant past, when the 'constant' had a different value. H_0 is, rather, the present-day value of the 'Hubble parameter' describing the expansion of the Universe with time.

Again, more generally, the Hubble parameter may be increasing or decreasing with time depending on the actual value of the deceleration parameter, q, being a dimensionless measure of the cosmic acceleration of the expansion of space. If $q = 0$, then $H = 1/t$, where t is the time since the Big Bang, i.e. the age of the Universe.

It was long thought that q was positive, indicating that the expansion is slowing down due to gravitational attraction. This would imply an age of the Universe less than $1/H$ (which is about 14 billion years). For example, a value for $q = 0.5$ (once favoured by most theorists) would give the age of the Universe as $\frac{2}{3}(\frac{1}{H})$.

A major development occurred in the 1990s, with the discovery that distant supernovae were dimmer, and therefore farther away, than previously suspected. This unexpected finding indicated that the Universe was not only expanding, but also accelerating in its expansion.

The result implied the existence of dark energy as an inevitable consequence of cosmological models, i.e. a new force pushing everything in the cosmos apart. The discovery that q is apparently negative means that the Universe could be older than $1/H$, although estimates of its age still put it very close to $1/H$.

Another complication in interpreting the Hubble law is that gravitationally interacting galaxies move relative to each other independently of the expansion of the Universe. These 'peculiar velocities' also need to be correctly accounted for in any related studies.

VARIOUS METHODS have been used to determine the Hubble constant. The story is long and involved, and things have come a long way since Edwin Hubble's own over-estimate, of about $500\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$.

What have become known as 'late Universe' methods rely on calibrated distance ladder techniques: measuring the redshifts of distant galaxies, and then determining the distances to them by some other method.

Over the past half century, many hundreds of research papers have targeted the refinement of H_0. But uncertainties in the physical assumptions used to determine these distances have resulted in varying and discordant estimates of its precise numerical value.

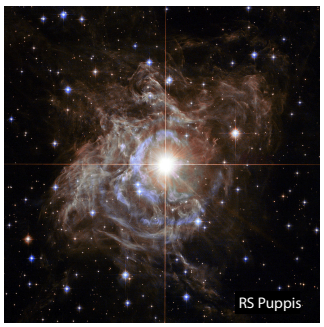
For most of the second half of the 20th century, H_0 was estimated to lie in the range 50–90 $\mathrm{km\,s^{-1}\,Mpc^{-1}}$. Over the past decade or so, data from Cepheid variables and other astrophysical 'standards' have converged on a 'late Universe' value of around $70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$.

Meanwhile, since about 2000, 'early Universe' techniques have become available. These values are based on measurements of the cosmic microwave background, viz. the 'echo' from the Big Bang that contains imprints of the Universe's fundamental properties. Initially from NASA's WMAP, and more recently from ESA's Planck mission, the most recent Planck estimates yield $H_0 = 67.66 \pm 0.42\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (Aghanim et al., 2020).

OF THE LATE UNIVERSE measurements being pursued most actively today, two are particularly powerful, and both have been advanced by Gaia.

One makes use of earlier work on red giant stars, and uses the tip of the red-giant branch as a primary distance indicator. It is based on the fact that all red giants reach the same peak brightness at the end of their lives. The latest estimates from the Hubble Space Telescope give a value of $69.8 \pm 1.9\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (Freedman et al., 2019). Combining this with recent parallax measures of the distance to the globular cluster Omega Centauri from Gaia EDR3 (discussed separately here) gives a value of $72.1 \pm 2.0\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (Soltis et al., 2021).

BUT LET US RETURN to the method based on Cepheid variables, used by Hubble a century ago, and still pertinent today given the Gaia parallax measurements.



Hubble Space Telescope (NASA/ESA)

The relation between Cepheid luminosities and pulsation periods was discovered by Henrietta Leavitt in 1908, based on variable stars in the Magellanic Clouds. Inferring the true luminosity of a Cepheid by observing its pulsation period in turn allows the distance to the star to be determined.

One example of the new distances from Gaia is for the long-period Cepheid RS Pup, shown here, where DR2 lists a geometric parallax of $0.584\,4 \pm 0.0260$ mas, corresponding to a distance of $1\,710 \pm 80$ pc. Gaia DR2 actually includes 9575 stars classified as Cepheids (Clementini et al., 2019).

FOR SOME YEARS, the strongest evidence for a high value of H_0, of around $70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, has rested on an empirical Cepheid-based calibration of the distances to galaxies hosting Type Ia supernovae.

Riess et al. (2016) based their value of 73.24 ± 1.74 on 19 such Cepheids observed with HST–WFC3, along with a more robust distance to the Large Magellanic Cloud based on late-type detached eclipsing binaries, HST observations of Cepheids in M31, and new HST-based trigonometric parallaxes for Milky Way Cepheids.

Riess et al. (2018b) added new HST-based parallaxes of a further seven long-period Milky Way Cepheids to yield a value of $73.48 \pm 1.66\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$.

Riess et al. (2018a) used HST photometry of 50 long-period, low-extinction Milky Way Cepheids observed in the same photometric system as extragalactic Cepheids in Type Ia supernova host galaxies. For the first time, Gaia parallaxes (from DR2) were included to constrain the distance scale, while simultaneously estimating the global DR2 parallax zero-point offset. Their resulting value was a rather similar $73.52 \pm 1.62\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$.

Riess et al. (2021) used 75 Milky Way Cepheids with Hubble Space Telescope photometry, now using the greatly improved Gaia EDR3 parallaxes. Applied to the calibration of Type Ia supernovae, it gave $H_0 = 73.0 \pm 1.4\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. Combined with the best complementary sources of Cepheid calibration, they found $H_0 = 73.2 \pm 1.3\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, reaching 1.8% precision, but a 4.2σ difference with the estimate from the latest Planck microwave background observations.

The inclusion of future Gaia parallaxes is expected to lead to a total uncertainty on H_0 of as little as 1.0–1.3%. As stressed by Riess et al., neither HST nor Gaia can expect to reach this goal alone: rather, this milestone will require simultaneously measuring Cepheid parallaxes to around $5\,\mu$as precision from Gaia, and measuring the brightnesses of the same objects to about 0.01 mag precision with HST within the *same* photometric systems used to measure their extragalactic counterparts.

The important point is that such differential flux measurements essentially circumvent systematic uncertainties related to instrumental zero-points and transmission functions, which otherwise impose a systematic uncertainty of 2–3% in the determination of H_0.

THE TWO VALUES presently considered to be the most reliable – the 'early Universe' estimate of $H_0 = 67.66 \pm 0.42\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ from the Planck mission, and the 'late Universe' value of $H_0 = 73.2 \pm 1.3$ from the combined HST and Gaia measurements of Cepheids – might not seem very different. But each is very precise, and the disagreement, although small, is statistically significant.

Meanwhile, various other promising methods of addressing this so-called 'Hubble tension' are also being developed, including the use of gravitational lensing, and the properties of gravitational waves.

AT STAKE is not so much the precise value of H_0 *per se*, but whether the different estimates hide some new and exotic physics, either in the understanding of stellar evolution, or in the physics of the early Universe.

Prospects for a resolution include whether we live in a local 'bubble' (e.g. Shanks et al., 2019) or whether weak gravitational lensing is affecting the measurements of Type Ia supernovae.

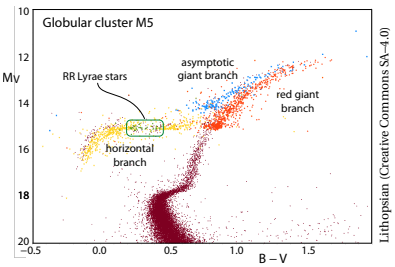
45. RR Lyrae variables

GLOBULAR CLUSTER and other Population II stars more massive than the Sun have long ago evolved into white dwarfs. In contrast, stars of approximately solar mass are common.

Their location in the Hertzsprung–Russell diagram is well accounted for theoretically. After ascending the giant branch, terminating in the helium flash, stars evolve rapidly onto the 'zero-age horizontal branch' with masses around $0.6 - 0.8\,M_\odot$, where they basically comprise a static He-burning core and a H-burning shell.

Their location along the horizontal branch depends on metallicity, from blue in metal-poor clusters, to red in metal-rich clusters, where they merge into the giant branch to form the region of the (red) clump giants.

THE RR LYRAE VARIABLES are a subset of the horizontal branch giants, occurring where the horizontal branch intersects the instability strip. RR Lyrae variables have pulsation periods of around 1 day or less.



Lithopsian (Creative Commons SA-4.0)

Like Cepheids, although less luminous, their distinctive light curves allows detection to large distances, as far as the Galactic centre in the low-absorption Baade Windows, and in crowded fields.

There are two major subgroups: the RRab, most relevant to the distance scale, are metal-poor spheroidal component stars, with asymmetric light curves, longer periods (above 0.4 day), larger amplitudes (around 0.5–1.5 mag), and pulsating in the fundamental mode.

The less numerous RRc type are old disk component stars, with more symmetric almost sinusoidal light curves, shorter periods (below 0.4 day), smaller variability amplitudes, and pulsating in the first overtone.

There are also double-mode pulsators, denoted RRd, which pulsate simultaneously in the fundamental mode and in the first overtone.

AN UNDERLYING period–luminosity relation has long given RR Lyrae stars a role as standard candles for relatively nearby targets, especially within the Milky Way and Local Group. While typically more common than the Cepheids, there are greater difficulties in accounting for the effects of metallicity, faintness, and blending. RR Lyrae stars also provide tests of evolutionary and pulsation models, and are important kinematic tracers.

Their typically large distances means that useful trigonometric parallaxes have largely been unavailable, and various other methods have been used for luminosity calibration. These include the use of RR Lyrae in Galactic globular clusters whose distances have been derived from main-sequence fitting of subdwarfs, the use of statistical parallaxes, and Baade–Wesselink determinations based on interpretation of the colour, light, and radial velocity variations during the pulsation cycle.

But my goal here is just to look at the *numbers* of RR Lyrae stars that Gaia is discovering, measuring, and characterising.

PRE-GAIA, several thousand Galactic RR Lyrae stars were known. Because of their magnitudes, just 179 were included in the Hipparcos Catalogue.

Hipparcos eventually gave useful parallaxes for only a few, and only that for the class's prototype, RR Lyrae itself, is reasonably accurate, $\pi = 4.38 \pm 0.59$ mas. Bono et al. (2002) later gave a weighted mean of the Hipparcos, HST, and pulsational parallax of 3.87 ± 0.19 mas.

THE HIGH-ACCURACY PARALLAXES from Gaia, combined with the multi-colour multi-epoch precision photometry, makes the mission extremely powerful for identifying and characterising variability across the entire Hertzsprung–Russell diagram.

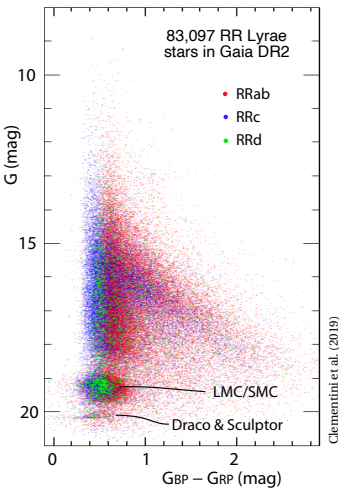
Similarly to what has previously been described here for Cepheid variables, the first Gaia data release, DR1, included 2595 RR Lyrae stars (and 599 Cepheids) in the Large Magellanic Cloud region, which was observed at high cadence during the first 28 days in the 'ecliptic poles scanning configuration' (Clementini et al., 2016).

For Gaia DR2, covering the first 22 months of the mission, a 'Specific Object Study' pipeline was used to validate and characterise Cepheids and RR Lyrae stars, originally using the period–amplitude and period–luminosity relations only in the G band, and subsequently extended to G_{BP} and G_{RP} (Clementini et al., 2019; Rimoldini et al., 2019).

Gaia DR2 accordingly provides results, along with mean magnitudes and pulsation characteristics, for 140 784 RR Lyrae stars as faint as $G = 20.7$ mag.

This huge sample includes objects in the Milky Way disk, bulge, and halo; in the Large and Small Magellanic Clouds; 1569 distributed over 87 globular clusters; and 417 distributed over 12 dwarf spheroidal galaxies (including seven ultra-faint dwarf galaxies). The largest numbers are in M3 (159), NGC 3201 (83), Sculptor (176) and Draco (176). Including some previously-known objects, not (yet) detected by Gaia, a total of 46 443 lie in the Large and Small Magellanic Clouds.

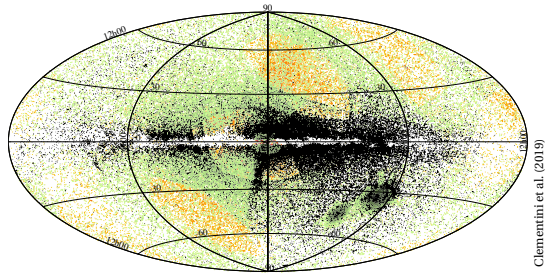
The accurate multi-epoch photometry resulted in 121 234 objects whose light curves could be modelled with at least two harmonics, and 67 681 whose light curves could be modelled with at least three harmonics. Cross-matching indicate that out of the 140 784 confirmed RR Lyrae stars in Gaia DR2, 90 564 were already known, while 50 220 are new discoveries.



For the 83 097 stars with both G_{BP} and G_{RP} photometry, the colour–magnitude diagram shows the different regions occupied by the RRab, RRc, and RRd classes, along with clumps associated with the Large and Small Magellanic Clouds, as well as the Draco and Sculptor dwarf spheroidals.

The distribution of all known RR Lyrae stars on the sky (more than 220 000 in total, shown here in Galactic coordinates) are colour-coded as: previously-known RR Lyrae stars without a counterpart in Gaia DR2 (orange); previously-known RR Lyrae stars with a counterpart in Gaia DR2 (green); and new RR Lyrae discovered by Gaia (black).

The latter two groups clearly reflect the pattern of the Gaia scanning law as it stood after the first 22 months of the mission, and suggest that many tens of thousands of additional RR Lyrae stars will be discovered by the end of the mission as the sky coverage densifies.



Let me turn briefly to how the Gaia RR Lyrae data, and especially the parallax distances, are being used. Firstly, Molnár et al. (2018) used Kepler mission photometry to conclude that the DR2 catalogue has a completeness of 70–78%.

Several groups have used the data to determine improved period–luminosity–metallicity relations, confirming the parallax quality in the process (e.g. Neeley et al., 2017; Neeley et al., 2019; Muraveva et al., 2018).

Gould & Kollmeier (2017) showed that RR Lyrae yield a meaningful zero-point of the global parallax system, comparable with quasars, despite their smaller number.

Various groups have used the distances and proper motions of several thousand stars to establish the triaxial structure and kinematics of the Galaxy halo, identifying the inner and outer halos with weak prograde and retrograde rotations respectively (Utkin et al., 2018).

Iorio et al. (2018) found no evidence of tilt or offset of the halo with respect to the Galaxy disk, while Iorio & Belokurov (2019) interpreted their density and kinematics as evidence in favour of a scenario in which the bulk of the halo was deposited in a single massive merger event.

The existence of RR Lyrae stars well beyond a system's tidal radius are providing evidence for tidal disruption and debris stripping, both for Galactic globular clusters (Kundu et al., 2019), and ultra-faint dwarf satellite galaxies (Vivas et al., 2020).

Kervella et al. (2019) combined Hipparcos and DR2 positions to find proper motion anomalies caused by close-orbiting companions, detecting 13 out of a sample of 198, and suggesting a binary fraction of at least 7%.

Ramos et al. (2020) searched for RR Lyrae associated with the Sagittarius tidal stream, finding some 6000–11 000 candidates. Similarly, Prudil et al. (2020) assessed the contributions from the Galaxy disk and from the Gaia–Enceladus stream, out of 314 RR Lyrae in the solar neighbourhood, while Du et al. (2020) studied the kinematics and spatial distribution of 15 599 RR Lyrae stars in the Milky Way bulge.

While the DR2 results are already impressive, numbers, accuracies and a vast panorama of detailed studies will continue to grow with the future Gaia data releases.

46. The iterative solution: formulation

GAIA GATHERS an enormous quantity of observations of a vast numbers of stars over several years. The goal of the data analysis on the ground is straightforward in principle: like solving a giant celestial jigsaw, the task is to find the positions and motions of each star best matching this gargantuan global set of observations.

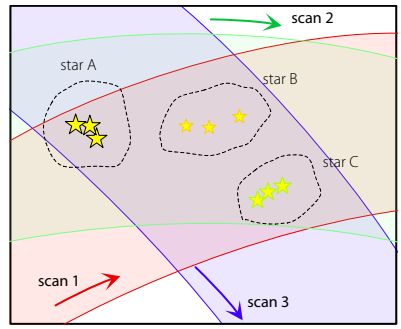
THE SCHEMATIC OPPOSITE shows three scans across a small region of sky to illustrate the concept. Depending on the scanning motion across that part of the sky at those particular times, the interval between successive scans may be several hours or several days.

Between the scans, all the stars have moved minutely, through a combination of their true motions through space, and their (apparent) parallax motions due to Earth's annual orbit around the Sun. Over months and especially over several years, and with many more scans, enough information has been collected to allow an estimate of each star's minuscule motion across the sky, along with any other motion that might affect it, such as orbital binary rotation, the effects of unseen planets, or gravitational light bending due to the Sun.

If you are looking at this problem for the first time, you may well be asking: How are the stars in the different scans matched up? What are the star positions measured with respect too? Why are there two fields of view? How are the optical aberrations of the telescope accounted for? How do any irregularities in the satellite's scanning motion affect the problem?

These considerations are all necessary for the correct and rigorous execution of the data processing (and are partly why these experiments take years to prepare and execute), but they are not central to the basic principles, and I will ignore most of them here.

AT THE HEART of the data processing carried out on the ground, then, is a global solution that matches up all the star signals generated by the CCD focal plane – several thousand every second – and solves for (a minimum of) the five astrometric parameters per star, along with all the additional unknowns describing any minute time-varying changes of the instrument.



The beauty of the problem is that the star positions, calibration and the spacecraft attitude are all tightly related, and connected by the fixed angle between the two identical telescopes simultaneously observing the sky.

And importantly, there is no satellite 'down time', in which science observations must be suspended while specific instrument calibrations are carried out; calibration is a by-product of the observations themselves.

But there is a catch: given the billions of stars, each with hundreds of observations, many thousands of calibration parameters, and with a satellite attitude sampled every second, any system of rigorous mathematical equations connecting all these unknowns is far too large and complex to solve directly.

The enormous size of the computational problem, and the experience gained through Gaia's predecessor, Hipparcos, led to the conclusion that only an iterative method might conceivably allow a solution.

INDEED, WHILE the *concept* is straightforward, the task of efficiently implementing and executing the global solution as an iterative least-squares adjustment was one of the major feasibility questions facing the Gaia project at the time of its adoption by ESA in 2000.

The mathematical solution to the problem was led by Lennart Lindegren, and described at the start of the mission by Lindegren et al. (2012), while Lindegren et al. (2016) addresses the details involved in the creation of Gaia DR1. I will look at some of the numbers involved in its numerical implementation separately.

GOING A LITTLE further into the problem, the challenge is the simultaneous estimation of a very large number of unknowns representing four distinct types of information: (a) the astrometric parameters for a subset of the observed stars, providing the astrometric reference frame; (b) the instrument attitude, representing the celestial pointing of the instrument axes in that reference frame as a function of time; (c) the geometric instrument calibration, representing the mapping from the CCD detectors to angular directions relative to the instrument axes; and (d) a few 'global' parameters describing, for example, a possible deviation of space–time from the prescriptions of general relativity.

Although the total number of stars observed by Gaia is more than two billion, only a subset are used in the astrometric core solution. This subset, of some 100 million well-behaved 'primary sources', consists of (effectively) single stars and extragalactic sources (quasars) that are sufficiently point-like and stable over time.

Nonetheless, the problem is formidable: the total number of unknowns involved is around a billion, and the solution uses some 100 billion observations extracted from some 100 000 Gbytes of raw satellite data.

AMONGST MANY details I will mention here just a few to give a flavour of the complexity.

The satellite 'attitude' specifies the telescope's orientation as it spins. The spacecraft is controlled to follow a specific 'scanning law', which provides good coverage of the entire sky, as well as maintaining a constant angle to the Sun. But the actual attitude can deviate from the nominal 'law' by up to 1 arcmin in all three axes.

The geometric instrument model defines the precise layout of the CCDs. It depends on their geometry, position and alignment in the focal-plane assembly, as well as the entire optical system including its scale, its stability, its distortions, and its other aberrations.

Gaia's high astrometric accuracy makes it necessary to use General Relativity to model the data. The formulation is based on the parametrised post-Newtonian (PPN) version of the relativistic framework adopted by the International Astronomical Union (IAU) in 2000.

Other complexities abound, including the chromaticity of the telescopes, charge transfer inefficiency of the CCDs, and attitude irregularities due to thruster noise and micro-meteoroid impacts.

THE NUMERICAL APPROACH to solving for all of these unknowns is a 'block iterative least-squares solution', the Astrometric Global Iterative Solution (AGIS).

In its simplest form, four 'blocks' are evaluated in a cyclic sequence until convergence. The blocks map to the four different kinds of unknowns mentioned previously: the source (star) update, S, in which the astrometric parameters of the primary sources are improved; the attitude update, A, in which the attitude parameters are improved; the calibration update, C, in which the calibration parameters are improved; the global update, G, in which the global parameters are improved.

The blocks must be iterated because each needs data from the three other processes. For example, when computing the astrometric parameters, the attitude, calibration and global parameters are taken from the previous iteration. These updated astrometric parameters are used the next time the A block is run. And so on!

WHILE THE BLOCK-ITERATIVE solution is intuitive and appealing in its simplicity, its implementation faced many challenges in practice: it is not obvious, mathematically, that it must converge. And if it does, it is not obvious how many iterations are required, whether the order of the blocks in each iteration matters, or even whether the converged results do, in fact, constitute a solution to the *global* minimisation problem.

Adding to the complexity is the fact that the core iterative solution also interfaces with all the other (enormous) processing tasks, amongst them the photometric analysis (including variability and 'alerts'), the treatment of double and multiple systems, the radial velocity measurements, and the object classification algorithms.

Accordingly, and in parallel with the industrial satellite development from about 2000 onwards, a Gaia 'Data Processing and Analysis Consortium' was set up with the task of developing and running a complete system to analyse all aspects of the satellite data, and so constructing the various Gaia catalogue releases.

AMONGST THE PEOPLE involved in this work (the co-authors of the 2012 papers were Uwe Lammers, David Hobbs, William O'Mullane, Uli Bastian, and José Hernández), Lennart Lindegren, of the Lund Observatory (Sweden), has made many and profound contributions to the Gaia project, and to Hipparcos before it, over a career of some 40 years.

In 2018 he was awarded the German Astronomical Society's *Instrumentation Prize* for his contributions to Gaia. And in 2020 he was awarded the Brouwer Award by the Division on Dynamical Astronomy of the American Astronomical Society, for his lifetime's contribution to astrometry.

The latter part of this citation reads *'His work has changed our understanding of the Universe at a fundamental level... The enduring legacy of his work is such that future generations of astronomers may owe their success, and even careers, to his remarkable contributions.*

47. The iterative solution: implementation

FORMULATING THE MATHEMATICAL description of the astrometric solution was one part of the challenge for the Gaia astrometric data processing. But its actual computer implementation was quite another.

As I have described here separately, the problem is formidable: both in terms of the amount of data to be treated (the total number of unknowns is around one billion, and the solution treats some 100 billion observations extracted from some 100 000 Gbytes of raw satellite data), and in terms of the way in which the iterative solution has to be executed, with its four 'blocks' (of source, attitude, calibration, and global parameters) being evaluated in a cyclic sequence until convergence.

The implementation and the data management required to make the Astrometric Global Iterative Solution (AGIS) function has been absolutely crucial to the ultimate goal, and success, of Gaia.

A KEY FIGURE in this task was William O'Mullane. While Gaia was still in its study phase, in the 1990s, O'Mullane conceived and set up the framework for running the set of iterative equations in a distributed manner. The prototype was based on the Hipparcos satellite data, which employed a similar sky-scanning.

This approach was novel in using the relatively new Java language, and in exploiting the message passing and networking capabilities to manage multiple distributed computing nodes.

Initially using an object oriented database, this was later largely replaced by more traditional solutions, although AGIS ultimately used the high-performance matrix database InterSystems Caché (which itself now uses the Gaia data as a key example on its home page!).

In 2005, O'Mullane returned to ESA with the task of developing an implementation of the global astrometric solution at scale. This led to the building up of a team of a dozen computer scientists at ESA's European Space Astronomy Centre (ESAC), outside Madrid.

Using 'agile' programming (and, in particular, 'extreme programming', XP), the group developed a system capable of producing the core astrometric solution underpinning the successive catalogue releases.

CENTRAL TO THE technical implementation was the concept of a 'data train', which performs a 'sweep' through the observational data, drawing the various algorithms behind it. The train 'picks up' an object, and passes it to all algorithms in a data-driven manner. Algorithms are called – and objects are then passed to them – allowing the system to access data efficiently, while insulating the algorithmic code from storage aspects.

As an example, while early implementations of AGIS required four passes through the billions of observations to execute the four processing blocks, a later implementation executed a single 'outer iteration' with just a single pass through the entire data set.

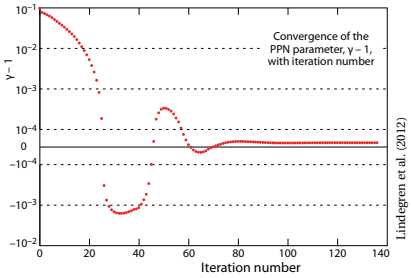
A critical constraint was to minimise disk access at all stages, also within each iteration. For example, holding the satellite attitude vectors for the entire mission in memory on each processing node was crucial, since practically every calculation used the attitude data. Likewise, calibration results from each previous iteration could also be stored in memory on each node.

Gaia processor, ESAC (José Hernandez)

LET ME FIRST LOOK at a major demonstration solution run at ESAC on simulated data, and before the satellite launch, as described by Lindegren et al. (2012).

This used an IBM cluster with 14 nodes, each node having two processors with four cores each, corresponding to 112 CPUs in total. This configuration of 14 nodes was estimated to have a total performance of 0.65 teraflop/sec (0.65×10^{12} floating point operations per second). One iteration took about 1 hr (with typically 90% CPU occupancy). The total run time for the 135 iterations executed was nearly 6 days, corresponding to a total of about 3×10^{17} floating point operations.

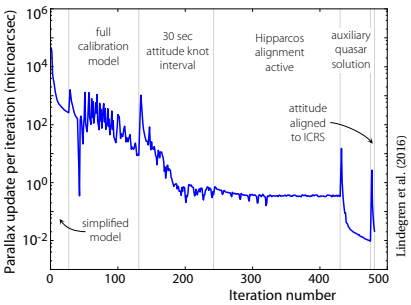
Scaled up to the projected 10^8 primary sources of a real AGIS run, this would amount to 1.5×10^{19} flop. Using a more conservative estimate of 5×10^{19} flop to account for additional features not included in that particular demonstration run, they predicted a requirement of some 60 days on a dedicated 10 teraflop/sec machine.



The figure above, from the same simulation exercise, shows how one of the 'global' terms, specifically the estimate of the parameter γ (in the PPN formulation of general relativity), eventually converges, from some initial 'assumed' value of 1.1, to a stable value close to 1.0, from around iteration 80 onwards.

THE FOLLOWING IS another example illustrating the convergence of AGIS, taken from the processing involved in the preparation of Gaia Data Release 1 (DR1, Lindegren et al., 2016). It shows how the parallax update (in microarcseconds) evolved with the iteration number. Different regimes of the underlying calibration model were adopted as the iterations proceeded.

Starting with parallax updates of around 10 milliarcsec for the initial iterations, they converged to values of around 1 microarcsec or less by around iteration 200.



THE ASTROMETRIC SOLUTION for Gaia Early Data Release 3 (EDR3) is described in detail by Lindegren et al. (2021). But I will focus here on some of the numbers related to the hardware, and to the execution times.

The astrometric solution for Gaia EDR3 was run using 32 nodes, each comprising 64 GB of RAM and 24 cores (which are Intel–Xeon CPU E5–2670 v3 running at 2.30 GHz). Together these provided a total theoretical performance of 30 teraflop/sec.

In terms of disk storage, 55 Tbytes was available for the solution for primary stars (around 40 Tbytes was used), and a further 20 Tbytes for the secondary stars.

In practice, data from the satellite is sorted before being ingested by AGIS. This AGIS-preprocessor reads and writes the data three times in order to have the main astrometric data sorted for every source. Then the data is grouped, typically in chunks of 1000 sources in the primary store and 10 000 sources (depending on the star density) in the secondary store. The AGIS-preprocessor takes about one week to run.

A total of 181 iterations were executed in generating EDR3 (as detailed in their Table 3). Each outer iteration of the 14 million primary sources (performing the source, attitude, calibration and global blocks) took 90 minutes. Although this has been further optimised (taking around 60 minutes) for the next major reduction, DR3, more primary stars will be used.

What has taken much time and resources is refining the calibration model to further improve the solution, correcting the biases observed in the preliminary runs. In practice, some 200 preliminary tests were scheduled before starting the final operational run.

THE PRIMARY STAR SOLUTION for EDR3 had to process about 6.5 billion CCD observations for the 14.3 million primary sources. The solution determined 71.5 million source parameters, together with 10.7 million attitude, 1.1 million calibration, and 2.0 million global parameters. The 'redundancy factor', being the mean number of observations per unknown, was about 76.

The secondary star solution processed nearly 78 billion field-of-view transits, generating converged solutions for 2.495 billion sources (of which 585 million were 5-parameter, 883 million 6-parameter, and 1.027 billion were 2-parameter solutions). Subsequently some of the 5- and 6-parameter solutions, and most of the two-parameter solutions, were removed because they failed to meet the rigorous acceptance criteria.

LOOKING BACK, this crucial step in the success of Gaia could easily have gone spectacularly wrong.

In the technology preparation phase in 1999, I identified this data organisation/analysis problem as one of the highest priority items for immediate further study, entering the Science Technology Document (Peacock & Scoon 1999) as Item G24. But a review panel rejected the funding request, with the one-line comment: *'Not a priority, just wait for commercial products to come'*.

It was a long process for me to reverse this uninformed dismissal of a most complex issue. After that, a 2-year industrial study, and a parallel academic study extending over several years, failed to progress this challenging implementation. Only bringing O'Mullane to ESA to work on the problem under my direct authority eventually led its successful implementation.

48. The risk of asteroid impacts

FROM TIME TO TIME, the topic of near-Earth asteroids, and their potential for impact hazards to our planet, hits the scientific and popular headlines.

I will relate an episode which played out in 2000, during the preparation of the scientific case for Gaia. This involved a number of high-profile personalities, and an important and remarkable opportunity for Gaia.

But ultimately it left me perplexed, and not a little disappointed, at the outcomes which can emerge when different people – in other words different ideas and different priorities – come together from disparate fields.

I IMAGINE THAT most who are reading this will be at least vaguely aware that some of the minor objects in the solar system, continually nudged by gravitational forces, can in principle sooner-or-later impact the Earth.

Left over in colossal numbers from the processes that shaped our solar system's formation, some of the trillions of rocky asteroids, and to a lesser extent the icy comets, could pose some sort of threat.

At the lowest masses, 'dust' and small objects rain down on us continuously. Indeed, it is estimated that $10^5 - 10^6$ kg of meteoritic material falls on Earth each day. Larger bodies are rarer, and so is their chance of impact. But their potential for damage is vastly more.

Let me go further back in time to provide some context. During the first 500 Myr of the solar system's existence, some 4–4.5 Gyr ago, Earth grew by colliding with other 'planetesimals'. The energies involved were sufficient to melt much of the growing planet, allowing dense iron melts to sink to the centre to form Earth's core.

Collision of the proto-Earth with a giant impactor, less than 100 Myr after the birth of the solar system, resulted in ejected material coalescing to form our Moon.

The subsequent decline of giant impact events, and the progressive cooling of the Earth's surface, would have allowed the formation of an initial planetary crust. Later tectonic processes destroyed all remnants of this initial crust, and the meteoritic impact sites. But traces of this 'terminal bombardment' are clearly evidenced by the remarkable cratering record of the lunar surface.

IN MORE RECENT geological history, the Earth has experienced many huge impacts, catalogued in the Earth Impact Database, which records their age and size.

For example, in addition to the massive Vredefort and Sudbury craters from 2 billion years ago, impact structures include Morokweng at around 145 million, Chicxulub at 65 million, Popigai and Chesapeake Bay at 35 million, and Kara-Kul at 5 million years ago.

An area of active research today is investigating the possible relationship between the biggest impacts, and the extinction events which are evident in the geological record. Chicxulub, notably, occurred at or close to the Cretaceous–Tertiary boundary, and may have caused the mass (dinosaur) extinction around that time.

In more recent history, the Siberian Tunguska event of 1908 has been attributed to a 50-metre diameter object which probably broke up some 6–8 km above the ground, generating a destructive blast wave and high-speed wind over more than 2000 square kilometres.

The Chelyabinsk meteor, estimated at 20-metre in size, entered Earth's atmosphere over Russia in February 2013, exploding in an air burst at a height of 30 km. Its explosion led to many hundreds of injuries, and thousands of damaged buildings in six cities across the region. Dmitry Medvedev, Prime Minister of Russia, subsequently called for a 'spaceguard' system to protect the planet from similar objects in the future.

NEAR-EARTH ASTEROIDS, or NEOs, are defined as the subset of objects whose orbits come within 1.3 au of the Sun. Only a handful were known in 1980, nearly 1000 in 2000, and more than 25 000 are known today. If their orbits cross the Earth's, and are larger than 140 m in size, they are termed 'potentially hazardous objects'. Two adopted measures, the Torino and Palermo scales, rate their impact risk and predicted consequences.

Of the various searches ongoing, the US and the European Union collaborate on the Spaceguard programme. Encouragingly, an early US Congress mandate for NASA to catalogue at least 90% of NEOs more than 1 km in diameter by 2020, was actually met by 2011.

EXCITEMENT, AND indeed concern, can grab the news headlines when a possible impactor approaches, but interest typically wanes as the immediate danger recedes. Nevertheless, various events elevated awareness and interest to unprecedented levels in the year 2000.

On the first working day of the new millennium, Lord David Sainsbury, he of the supermarket dynasty but more pertinently UK Minister for Science and Innovation at the time, unveiled a task force to assess the risks. *'The risk of an asteroid or comet causing substantial damage is extremely remote'*, he said, *'But we cannot ignore the risk, however remote, and a case can be made for monitoring the situation on an international basis'*.

Liberal Democrat MP Lembit Öpik (whose grandfather, Ernst Öpik, of Opik's law, was an Estonian astronomer who had worked at Armagh Observatory), had lobbied for the task force, and praised Sainsbury for what he said was a *'brave political move'* in launching it.

The panel comprised just three people: Dr Harry Atkinson (formerly Science and Engineering Research Council and past chair of ESA's Council), Professor David Williams (then RAS President, University College London), and Sir Crispin Tickell (former British ambassador to the United Nations). They were charged with assessing the hazards, and with suggesting how the UK should contribute to international efforts to deal with them.

IT IS DIFFICULT, today, to appreciate the widespread interest that this whole subject generated at the time. The Sainsbury panel, the warnings of scientists, the voices of those involved in experiments searching and tracking these potential impactors, and of course journalists, had brought the topic to much wider public attention. A high-profile competition was even launched to solicit ideas for detecting and deflecting them.

One voice in this drama was Russell (Rusty) Schweickart, lunar module pilot on the 1969 Apollo 9 mission. After leaving NASA in 1977, he served as California Governor's assistant for science and technology, then on California's Energy Commission. In 2002, he co-founded the B612 Foundation, aimed at 'defending Earth from asteroid impacts'. He still serves as its chair emeritus.

I met Schweickart in 2002 during one of his visits to The Netherlands, and he showed a great interest in the potential of space astrometry to assist in this task – the Hipparcos results had recently become available, and the prospects for Gaia were moving to centre stage.

THE SAINSBURY PANEL report was duly published on 18 September 2000, and it made 14 recommendations. Amongst these, their third was specific to Gaia's nascent capabilities: *We recommend that the Government draw the attention of the European Space Agency to the particular role that Gaia, one of its future missions, could play in surveying the sky for Near Earth Objects.*

MARCHING ALONGSIDE all of this activity, Gaia was competing for its place in ESA's scientific programme. As I have described elsewhere, this involved a detailed review of its scientific case by ESA's scientific advisory committees, and led to its adoption by the high-level Science Programme Committee in October 2000.

Amongst its harvest, the combination of on-board detection, faint limiting magnitude, observations at small Sun-aspect angles, and confirmation from successive field transits, showed that as well as observing all known asteroids, Gaia would discover some $10^5 - 10^6$ new objects down to diameters of 260–590 m at 1 au.

Concerning asteroid impacts, simulations showed, rather remarkably, that the predicted orbital errors based on the Gaia observations alone *100 years after the end of the mission* would be at least 30 times better than the predicted errors corresponding to the *entire set of past and future ground-based observations*.

Let me re-phrase that in somewhat over-simplified and slightly more hyperbolic language: for any big rocks out there of a size likely to inflict substantial damage to Earth, the orbits from Gaia would allow a good prediction of whether they would hit the Earth, or not, *some 100 years in advance* – time enough for some avoidance manoeuvres to be evaluated and, perhaps, enacted.

At a high-profile meeting in Paris on 11 May 2000, representatives of ESA and its Science Programme Committee, of the European Southern Observatory, of the European Science Foundation, of the European Physical Society, and of the International Astronomical Union, all endorsed the pivotal contribution that Gaia promised.

GAIA IS INDEED today measuring asteroid orbits in huge numbers, and with unprecedented accuracy.

But the '100 year' impact-warning capability applied to the accuracies which were targeted when it was adopted in 2000, viz. 10 microarcsec at 15 mag. Subsequently, in three separate 'de-scopes', Gaia's target accuracy at 15 mag dropped to 15 microarcsec in 2002, to 20 microarcsec in 2004, and to 25 microarcsec in 2006.

As Project Scientist, with full overview of the scientific and technological challenges, I agreed with the first, but not the others. They were nonetheless signed off between the Project Manager (who held the purse strings) and the ESA Director of Science, against my own recommendations and those of the community I represented.

LET ME spell out the disappointment I referred to at the start. Recommendations were made by Lord Sainsbury's panel to search for ways of predicting future impacts. Gaia provided everything demanded. But the urgency of the panel's findings soon faded, and Gaia's ability to address the problem was quietly curtailed.

When the problem resurfaces, it is worth recalling that a solution through astrometry is still at hand.

49. The rotation of our Galaxy

THE STARS making up the disk of our Galaxy rotate around the Galactic centre which, today, is generally assumed to be defined by its central massive black hole. Our Sun, about 8 kpc from the centre, participates in this general motion, moving around the Galaxy in an approximately circular orbit in about 250 million years.

It has long been known that our Galaxy, as all others like it, does not rotate like a solid body. Instead, its 'rotation curve' has an innermost region (within about 3 kpc from the centre) in almost solid-body rotation, rising to a fairly constant rotation velocity in the solar vicinity. Further out it is flat, or with a slow decline, taken to imply the presence of 'dark matter' in its outer parts.

Much about our Galaxy's origin, structure, and dynamics relies on an understanding of the detailed form of its rotation. Yet many issues complicate its measurement and interpretation, even in the vicinity of our Sun. Before looking at what Gaia has to say about Galactic rotation, we should look at these complications.

IN ANALOGY WITH THE PROBLEM faced by the ancient Greeks in comprehending the motion of the planets, we must understand how the Sun's position and velocity affect its perceived motion around the Galaxy. First, we need to know the distance to the centre of Galaxy, R_0. A difficult problem in its own right, recent results from the GRAVITY collaboration yield, for example, $R_0 = 8.12 \pm 0.03$ kpc (Abuter et al., 2018).

The 'local standard of rest' is the velocity of a hypothetical group of stars in strictly circular orbits at the solar position. Its practical definition is complicated by the wide choice of stars used to represent it, with young stars (for example) not yet being in dynamical equilibrium.

The 'solar neighbourhood' is another somewhat loose concept. It is considered to be a volume centred on the Sun much smaller than the overall size of the Galaxy, containing a statistically representative subset of its population, but with a somewhat arbitrary size dependent on the objects under investigation.

The 'solar motion' itself can be determined with respect to a range of stellar and interstellar constituents. Most frequently, this motion is estimated with respect to the local standard of rest, with $(u_\odot, v_\odot, w_\odot)$ being the difference between the Sun's velocity and that of the reference system which, by definition, moves around the Galaxy with a certain circular velocity.

This circular velocity, usually designated $\Theta(R)$, is the velocity of an object moving in a circle of radius R, in the Galactic plane and about the Galactic centre, for which centrifugal force balances the Galaxy's gravity.

OUR GALAXY'S ROTATION is often expressed in terms of the classical Oort constants, A and B, resting on the assumption of circular motion around the Galactic centre within an axisymmetric potential (the so-called Oort–Lindblad model). Physically, and in analogy with fluid dynamics, A describes the azimuthal shear of the velocity field, while B describes its vorticity.

Local values of the angular rotation rate and its local derivative can then be expressed directly in terms of local values of A and B, with its angular rotation given by $\Omega_0 = A - B$, and its local derivative by $\Omega_0' = -(A + B)$.

Complications in deriving and interpreting A and B occur if the gravitational potential is not axisymmetric, specifically in the presence of spiral density waves and the central bar, in which case the Oort constants will vary with azimuth, and the numerical values will depend on the distance of the tracers adopted.

If the assumptions of strictly circular motion and axisymmetry are relaxed, but still being restricted to motions in the plane, the velocity field can be described by an additional two (Oort) constants, namely a radial shear (C), along with a local divergence (K).

Even more generalised expressions for the velocity field in the vicinity of the Sun assume only that it can be represented by a continuous smooth flow. This was first formulated as a first-order Taylor-series expansion by Ogorodnikov (1932) and Milne (1935).

A later development dating from the 1990s describes the global set of tangential velocities in terms of vector spherical harmonics, allowing for the identification of yet more complex systematic stellar motions.

ARMED WITH THESE caveats and complications, it is worth stressing that before Hipparcos, such studies could only sample a very small region around the Sun.

Hipparcos allowed many advances, including better estimates of R_0, of the Sun's local motion with respect to the local standard of rest, of the Oort constants A and B, and of the detailed form of the rotation curve, along with higher-order velocity structures.

For example, the first contribution making use of the Hipparcos Cepheid data was that by Feast & Whitelock (1997). They used 220 Cepheids, and assumed $R_0 = 8.5$ kpc, from which they estimated $(u_0, v_0, w_0) = (9.3, 11.2, 7.6)\,\mathrm{km\,s^{-1}}$, and found $A = +14.82 \pm 0.84$ and $B = -12.37 \pm 0.64$ ($\mathrm{km\,s^{-1}\,kpc^{-1}}$), from which $\Omega_0 = A - B = +27.19 \pm 0.87$ and $\Omega_0' = -(A+B) = 2.4 \pm 1.2$.

I TOOK A FIRST LOOK at Gaia's contribution to studies of Cepheid variables in an earlier essay (#43), with their application to estimates of the Hubble constant (from EDR3) in #44. As I pointed out there, while the Hipparcos catalogue contained just 280 Cepheids, Gaia DR2 included 9575, of which 3767 are in the LMC, 3692 are in the SMC, and 2116 are elsewhere ('all-sky').

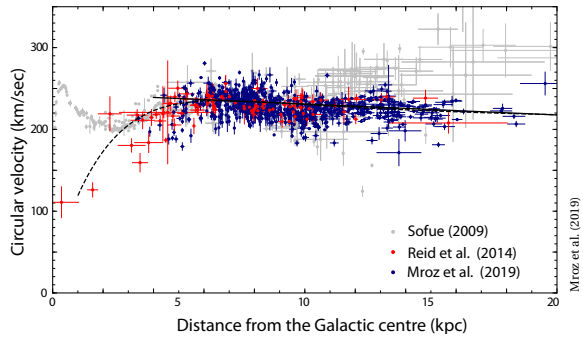
Given that Cepheids can be seen to large distances, and that they reflect the Galaxy's young population, they also provide an important tracer of Galactic rotation, as well as its spiral arms (and indeed its warp).

An examination of the rotation curve from Gaia DR1 was made by Bovy (2017). He used 304 267 main-sequence stars from the Tycho–Gaia Astrometric Solution to examine its structure out to 230 pc from the Sun. The pattern of proper motions clearly displays the effects of differential rotation. Along with the Oort constants $A = 15.3 \pm 0.4$ and $B = -11.9 \pm 0.4$, significant (non-zero) values of $C = -3.2 \pm 0.4$ and $K = -3.3 \pm 0.6$ (all in $\mathrm{km\,s^{-1}\,kpc^{-1}}$), demonstrate the importance of non-axisymmetry for the velocity field of local stars.

A NUMBER OF PAPERS have already focused on the rotation curve based exclusively on Cepheids.

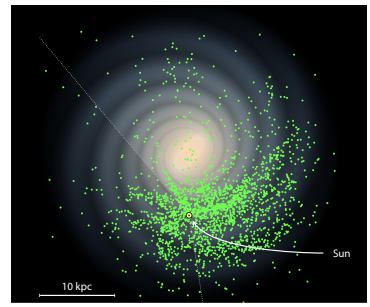
Bobylev (2017) used 260 Cepheids from DR1 to give the Sun's velocity as $(u_\odot, v_\odot, w_\odot) = (7.90, 11.73, 7.39) \pm (0.65, 0.77, 0.62)\,\mathrm{km\,s^{-1}}$, with the rotation curve (for $R_0 = 8$ kpc) described by $\Omega_0 = 28.84 \pm 0.33\,\mathrm{km\,s^{-1}\,kpc^{-1}}$, $\Omega_0' = -4.05 \pm 0.10\,\mathrm{km\,s^{-1}\,kpc^{-2}}$, and yielding a linear rotation velocity of the local standard of rest of $231 \pm 6\,\mathrm{km\,s^{-1}}$.

Mróz et al. (2019) used 773 Cepheids from Gaia DR2 to measure the rotation curve out to 20 kpc from the Galactic centre. Assuming $R_0 = 8.122 \pm 0.031$ kpc (from the GRAVITY Collaboration), they estimated the rotation speed of the Sun as $\Theta_0 = 233.6 \pm 2.8\,\mathrm{km\,s^{-1}}$. From this accurate Galactic rotation curve at distances $R > 12$ kpc, they showed that the rotation curve at Galactocentric distances $R = 4 - 20$ kpc is nearly flat, with a small decreasing gradient of $-1.34 \pm 0.21\,\mathrm{km\,s^{-1}\,kpc^{-1}}$.



Kawata et al. (2019) found reasonable agreement of the local centrifugal speed derived from 218 Galactic Cepheids in DR2 based on both an axisymmetric model, and from a simulation of a Milky Way-like galaxy containing a bar and spiral arms.

Skowron et al. (2019a) constructed a three-dimensional map of our Galaxy's young stellar population, based on the positions and distances of 2431 Cepheids from the Optical Gravitational Lensing Experiment (OGLE–IV), supplemented by distances and velocities from Gaia DR2. Their simple model of star formation in spiral arms successfully reproduces the observed Cepheid distribution.



2431 Cepheids in a 4-arm spiral model

Ablimit et al. (2020) examined 3500 Cepheids from OGLE, Gaia, and other surveys with the goal of measuring the rotation curve over Galactocentric distances of 4–19 kpc. Their analysis yields a gently declining rotation curve with a gradient of $(-1.33 \pm 0.1)\,\mathrm{km\,s^{-1}\,kpc^{-1}}$, in agreement with the findings of Mróz et al. (2019).

WHAT DOES THIS all mean? Interpreting this rotation in the framework of the Navarro–Frenk–White (NFW) model of galaxy formation, they estimate our Galaxy's (virial) mass as $(0.822 \pm 0.052) \times 10^{12} M_\odot$ within the corresponding (virial) radius of 191.84 ± 4.12 kpc, and a predicted local dark matter density of $0.33 \pm 0.03\,\mathrm{GeV\,cm^{-3}}$. They also conclude that the dark matter halo is the main contributor to the form of the Galactic rotation curve beyond distances of about 12.5–14.5 kpc.

IN ADDITION TO THIS information being extracted on the rotation curve of our Galaxy, the Gaia Cepheid data is also being used to trace out the pattern of our Galaxy's warped disk structure, as well as of its spiral arms. I will look separately at these two topics.

50. The German DIVA project

BY AROUND 1995, before the Hipparcos catalogue release, it was in the minds of various groups around the world to propose a follow-on astrometric mission.

In the US, various ideas were put forward between around 1995–2005. Amongst these were a 'point-and-stare' mission, POINTS, aiming at sub-microarcsec astrometry on a number of objects, one objective being planet detection. By early 1999, scientists at the US Naval Observatory in Washington had advanced a scanning concept, FAME, to the level of a Phase A study. And the ambitious long-baseline interferometer, SIM (part of the NASA Origins programme, and later descoped to SIM-LITE), targeted its original launch in 2005. All eventually fell by the wayside for technological or financial reasons, or through lack of wider scientific support.

In Russia, scientists were suggesting a Hipparcos clone (AIST), and there was talk of studies of other astrometric concepts (Lomonosov and Regatta–Astron). But by 1999, these ideas had also dispersed, presumably due to the country's economic situation at that time.

In Japan, scientists began to look at a series of missions – Nano-Jasmine, Small-Jasmine, and Jasmine – which would yield progress in technology and scope.

In a historical context, these various astrometry mission ideas undoubtedly merit their own more detailed consideration. But here, I will focus only on a European idea which gained more traction, DIVA, and summarise how its development and eventual demise was entwined with the progress of the ESA juggernaut, Gaia.

IN RESPONSE TO THE ESA CALL for an M3 (medium mission) proposal in 1993, Erik Høg (Copenhagen) and Lennart Lindegren (Lund) led a proposal for a European follow-on to the Hipparcos mission, which they called Roemer – along the lines of Hipparcos, but using CCDs as detector. We had looked at CCDs for use in Hipparcos in about 1982, but the suggestion (by Delft engineer M. Hammerschlag) was rejected then as being technologically immature. But by 1993, in contrast, CCDs were being used widely in astronomy, and had become the natural, if still complex and challenging, choice.

Roemer aimed at sub-milliarcsec astrometry. While of scientific value, this was not in reality a transformative advance. Roemer was rated highly by ESA's Astronomy Working Group (AWG), but eventually rejected by its more senior advisory groups in favour of Cobras/Samba (subsequently Planck) for a Phase A study in 1999.

Out of these early ideas the much more ambitious Gaia was proposed by Lennart Lindegren and Michael Perryman. It engaged much wider support, and it was eventually accepted by ESA's advisory groups in 2000.

IN 1995 Siegfried Röser, of the Astronomischen Rechen-Institut Heidelberg (ARI, now ZAH), Germany's leading institute for astrometry, began to coordinate the preparation of a mission concept, DIVA (Deutsches Interferometer für Vielkanalphotometrie und Astrometrie; viz. German interferometer for multicolour photometry and astrometry) for submission to the German Space Agency, DARA (DARA merged with the research and development activities of DLR to form the German Space Agency, DLR, in 1997).

The expectation was that DLR would issue a call for new missions before the end of 1997. The successful outcome of their Phase A study could lead to the start of Phase C/D in 1999, and a launch of DIVA in 2004.

DIVA WAS DESIGNED to fill the gap in observations between Hipparcos (100 000 stars at 1 milliarcsec accuracies at 9 mag) and Gaia (a billion stars to 20 mag, with accuracies of 10 microarcsec at 15 mag).

In a 24-month mission DIVA would measure positions, proper motions and parallaxes of all 30 million stars down to 15 mag, with accuracies at $V = 9$ mag of around 0.2 milliarcsec in positions, parallaxes, and annual proper motions, along with accurate broad-band photometry (Röser, 1999).

A key idea was that, due to the progress in technology since the time that Hipparcos was designed, DIVA would be able to surpass its performance at a fraction of its cost. Like Gaia, it started as an interferometer, with dispersed fringes providing multi-colour information.

As both mission concepts progressed over the next few years, affordability, politics, and national priorities, all played their parts in their destinies.

In addition to its near-term scientific harvest, the proponents of DIVA saw it as a guarantee of Europe's leading role in space astrometry should Gaia not be selected as a future mission by ESA, and as of great value in maintaining scientific expertise and continuity in the years leading up to launch if Gaia were selected.

Representatives of the Gaia effort saw DIVA as a confirmation of the importance of astrometry in Europe, but at the same time as a competitor both for finite financial resources (in DLR and ESA) and, as importantly, for the critical knowledge and expertise available within the European astrometry community.

In 1998, as the German delegation to ESA lobbied the ESA Director of Science, Roger-Maurice Bonnet, for assistance in funding, Bonnet expressed his view that it did not sit comfortably with him that a national programme should benefit from ESA's technical expertise, using it as a means of supporting a national programme which, they could then claim, could be developed at significantly lower cost. Bonnet countered by inviting the German delegation to instead contribute more to Gaia, with a view to facilitating an earlier launch date.

Also in the back of Bonnet's political mind was that bringing Gaia forward from its target launch of 2012, would take more wind from the sails of the ongoing US studies on FAME (itself based on many of the ideas from Hipparcos and Gaia), and on a possible linkup between FAME and DIVA which had also been mooted.

By late 1998, DLR had completed their feasibility study on DIVA, conducted by the industrial consortium Dornier/DASA. Though confident of its technical feasibility, they did not yet have a firm commitment on funds. With the constraints on Germany's funding for space science, the accepted wisdom was that if FAME were to be funded, then DIVA would be consigned to history.

A NASA press release, and an email from the FAME principal investigator Ken Seidelmann on 15 October 1999, brought the news that FAME had been approved for further study – a new entrant representing a serious threat for both DIVA and Gaia. In the fierce competition for future missions in ESA, such news was ammunition for those pursuing other priorities.

The following weeks and months were punctuated by much discussion, negotiation, rumour, and declarations of personal priorities. The Gaia Scientific Advisory Group, which I chaired, met to discuss the project's position: FAME targeted a significant advance which could weaken the political support for Gaia. And there was little enthusiasm for a trans-Atlantic collaboration. Of course if Gaia were to be approved for an early 2009 launch, FAME itself would no longer make much sense.

In this confusing atmosphere, DIVA continued its onward progress. In January 2000, Germany's DLR laid its cards on the table, favouring DIVA over Gaia from both an industrial and a science policy point of view. Influential Danish astrometrist Erik Høg stated that Gaia would be his priority, and that he could not spare effort to work on DIVA as well. The Gaia Scientific Advisory Group was also concerned by the 11 August 2000 selection data for DIVA targeted by DLR, just in advance of the planned selection date for the next ESA round.

In February 2000, ESA science programme coordinator Sergio Volonté announced that Oxford Galactic dynamicist James Binney had agreed to serve as assessor of DIVA and Gaia, as well as FAME and SIM, for the next Astronomy Working Group meeting on 9 May. Siegfried Röser informed us that the DIVA selection would be postponed until after that for Gaia. By April, the emerging wisdom was that DIVA would be too unsettling to a coordinated Gaia approach in Europe, while the FAME mission would perhaps keep the discipline vibrant without detracting European effort from Gaia.

On 22 September 2000, DLR recommended DIVA as the next space mission within the German national programme. But a financial caveat soon became known: 50% of the needed funding would have to come from other sources, for example from the federal states.

Shortly afterwards, on 12 October 2000, Gaia was in turn selected by ESA's Science Programme Committee.

For the next few months, DIVA marched on in parallel with Gaia. A major scientific workshop devoted to DIVA was held at the Max-Planck Institute for Astronomy, Heidelberg, in April 2001, and attended by about 60 scientists from the DIVA and Gaia teams. Detailed discussions examined the optical design, attitude control, radiation environment, operations and data processing. Launch was still targeted for mid-2004.

Ultimately, the survival of DIVA hinged on obtaining 15 M€ in funding from ESA. But as this request made its way through the ESA committees, the AWG on 15 November 2001 and the SSAC on 19 November, it became clear that this support would not be forthcoming.

On 11 March 2002, DLR's executive committee decided that for financial reasons the next German mission would not be carried through in the years 2003–05. Formally, this was considered as just a delay for DIVA, rather than as a cancellation.

A final high-level petition from German industry for ESA contributory funding was made at a meeting in ESTEC on 1 October 2002. But the AWG maintained its negative stance, and the rejection of ESA funding by the SPC on 5 November 2002 marked the end of DIVA.

This was a bitter blow for the DIVA team, who had made great efforts to advance space astrometry. But by pooling all European expertise, Gaia would benefit.

51. Asteroseismology – and star distances

EVEN WITH THE LARGEST of astronomy's arsenal of ground and space telescopes, most stars apart from our Sun are only observed as point sources of light. It is then perhaps surprising that we know anything at all about their inner structure, or inner workings.

Close-up studies of our Sun reveal its surface temperature and chemical composition from spectroscopy, details of surface features and convection cells from high-resolution imaging, along with time variability of its magnetic field, sunspots, flares, and coronal mass ejections from various remote-sensing instruments.

Detailed models of the inner structure of our Sun and other stars are based on numerous nuclear reaction rates which depend on the star's chemical composition and age, its density, pressure, and temperature, and adding in dependencies on their internal rotation, conditions in their radiative and convective zones, and many other detailed physical effects.

Spectroscopy of even very distant stars provides clues to their chemical composition, temperature, and surface gravity. But accurate knowledge of the distance to each star is critical in transforming *observed* properties, such as brightness and angular radius, into *intrinsic* properties, such as luminosity and physical size. A star's mass is of crucial importance for stellar models, but it is generally only measurable for certain stars in binary systems. Otherwise it is typically estimated from the position of the star in the Hertzsprung–Russell diagram.

AN IMPORTANT TOOL IN probing the internal structure of our Sun came with the advent of *helioseismology*, viz. inferring its internal structure from the 'shock' waves that can propagate through it. These waves are analogous to the seismic waves which propagate through the Earth's interior as a result of earthquakes or tsunamis.

On Earth, 'elastic waves' that can propagate in the solid crust and deeper into its interior are of two main types: 'pressure waves' or primary waves (P-waves), longitudinal waves of compression and expansion, and the slower 'shear waves' or secondary waves (S-waves) that move perpendicular to the direction of propagation.

Seismic waves provide high-resolution probes for studying the Earth's interior. One of the earliest discoveries (demonstrated by Harold Jeffreys in 1926) was that the outer core of the Earth is liquid. Since S-waves do not pass through liquids, the liquid core causes a 'shadow' on the side of the planet opposite the earthquake where no direct S-waves are observed. Processing readings from many seismometers using seismic tomography, the mantle of the Earth has been mapped with a resolution of several hundred kilometres, allowing the identification of convection cells and other large-scale features of its core and mantle.

Large earthquakes can also set the entire Earth 'ringing' like a bell. A mix of 'normal modes', with discrete frequencies and periods of around one hour, lead to resonant signatures observable even a month after a major seismic event.

THE REALISATION THAT analogous studies can be made of the Sun's internal structure came in the 1960–70s, with the discovery of tiny quasi-periodic oscillations in its intensity and line-of-sight velocity, with periods of about 5 minutes. The entirely fluid interior of the Sun leads to somewhat different oscillation modes compared to the Earth: the dominant pressure modes or P-modes, but also gravity modes (g-modes) most prominent in its radiative interior, and surface gravity modes (f-modes). These oscillations are principally caused by sound waves that are continuously driven and damped by convection near the Sun's surface.

Helioseismology observations made from the ground over several decades, with GONG and BiSON, and space observations from SoHO and SDO, has shown that the Sun has a rotation profile with a rigidly-rotating radiative (i.e. non-convective) interior zone; a thin shear layer, the 'tachocline', which separates the rigidly-rotating interior and the differentially-rotating convective envelope; a convective envelope in which the rotation rate varies both with depth and latitude; and a final shear layer just beneath the surface, in which the rotation rate slows down towards the surface.

ASTEROSEISMOLOGY extends the principles of helioseismology to more distant stars, although less information is accessible because their surfaces are unresolved. Oscillations are monitored through high-accuracy photometry over weeks or months.

Mechanisms other than convection can excite stellar oscillations, depending on a star's mass and spectral type: large-amplitude Cepheid and RR Lyrae variables are driven by a variation of radiation opacity with temperature (the 'κ-mechanism'), while tidal forces can drive oscillations in eccentric binary systems.

Over the past decade, a wealth of asteroseismic observations have been acquired on thousands of stars from space, most notably from the Kepler satellite as part of its search for exoplanet transits.

AMPLITUDES AND PHASES of the stellar oscillations are largely controlled by near-surface layers. For example, frequencies are determined by the bulk sound speed and the internal density profile. Convective motions within the differentially rotating outer convective zone modify the star's temperature, density and velocity structure. *Convective overshooting*, caused by the momentum of cool sinking material into the deeper radiative regions, alters the structure of the tachocline, the transition region between the two.

Stellar rotation further influences the details of the frequency spectrum through its variation with radius, and because of the resulting flows and instabilities which are also a function of the star's evolutionary state.

The detailed interpretation of an observed frequency spectrum proceeds via a comparison with theoretical models. For solar-type oscillations, for example, state-of-the-art asteroseismic models, based on the latest space-based observations, yield typical uncertainties of around 3–5% on stellar masses and 1–3% on stellar radii based on certain 'scaling relations'.

Uncertainties of typically 5–10% on luminosities follow from estimates of the radii from these models, and these in turn yield an asteroseismic distance, which can then be compared with accurate trigonometric distances from Gaia. As a result, the Gaia parallaxes therefore provide a powerful test of asteroseismic (and hence stellar evolution) models.

SEVERAL STUDIES using the early parallaxes from Gaia DR1 showed that the good agreement between asteroseismic and astrometric distances for solar-like oscillators (both dwarf and subgiants), which had already been demonstrated with the Hipparcos data, also hold true for the more accurate distances from the Tycho–Gaia Astrometric Solution, while the more accurate asteroseismic distances for pulsating red giants from Kepler yielded early insights into systematics in the DR1 parallaxes (e.g. De Ridder et al., 2016; Davies et al., 2017; Huber et al., 2017).

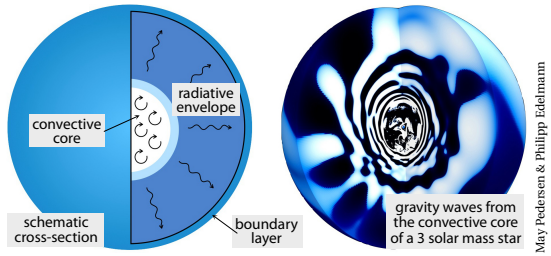
More detailed analyses became possible with the release of Gaia DR2. For example, Sahlholdt & Silva Aguirre (2018) used a sample of 93 dwarfs to show that the asteroseismic radii are 1% smaller than Gaia radii on average, possibly explained by a negative bias of 30 μas in the DR1 parallaxes. They argued that asteroseismic radii are generally accurate to within 1%, but are perhaps overestimated by 5% or more at the highest temperatures. Several other studies have further examined the implications for Gaia systematics or asteroseismic models (e.g. Hall et al., 2019; Khan et al., 2019; Zinn et al., 2019).

I WILL LOOK AT one particular study using the DR2 data in a little more detail, which illustrates the insights into the physics of stellar interiors that is being gained.

Pedersen et al. (2021) studied 26 'slowly-pulsating B stars', a class of star between 3–10 solar masses which are known to be rapidly rotating. Over this mass range, higher core temperatures lead to H–to–He fusion occurring primarily via the temperature-sensitive CNO cycle, which leads to the core being convective, while the outer envelope is ionised, transparent to ultraviolet radiation, and is consequently radiative.

The extent to which the convective core mixes with the radiative envelope in turn affects the amount of H fuel that can be accessed for core H burning, and in consequence allows the stars to live for longer, changing their evolutionary paths. But the degree of mixing, and its effect on the boundary layer and the radiative envelope, has so far remained unknown.

The Gaia distances provide strong constraints on the asteroseismic models. These show that the internal mixing among the 26 stars is far from uniform, with some having almost no mixing, while others show levels a million times higher. The mixing shows no clear dependence on the star's mass or age, but does appear to be correlated with the rotation. But this is not the only physical process at work, and improvements in the theory of the internal mixing of massive stars will follow.



THERE CAN BE LITTLE DOUBT that many new and interesting insights into stellar structure, stellar evolution, and the state-of-the-art computational models being used to interpret observations of stellar oscillations, will come from the improved astrometric solutions from Gaia in the future.

52. Interplanetary navigation

THE PAST FEW DECADES have seen many spacecraft sent out to orbit, land on, or simply fly past all of the major and many of the minor bodies of our solar system: these include landers on the Moon and Mars, flybys and orbiters of Jupiter, Saturn, Uranus and Neptune, and landers on Saturn's moon Titan in 2005, and on comet 67P/Churyumov–Gerasimenko in 2014. There have been visits to comets Giotto, Grigg–Skjellerup, Tempel 1 and Wild 2, and to asteroids Annefrank in 2002, Lutetia in 2010, Vesta in 2011, Ceres in 2015, Bennu in 2018, and Ryugu in 2019.

One of the most impressive achievements in interplanetary navigation has been NASA's New Horizons mission to the dwarf planet Pluto. Launched in 2006, it flew past Pluto nearly a decade later, on 14 July 2015, just 12 500 km above its surface. It captured images, and collected data on the atmospheres, surfaces, interiors, and environments of both Pluto and its moons.

Having completed its flyby of Pluto, New Horizons was then re-manoeuvred for a flyby of the Kuiper belt object Arrokoth (also known as Ultima Thule). This occurred on 1 January 2019, when it was more than 40 au (i.e. 40 times the Sun–Earth distance) from the Sun.

IN ADDITION TO THE ENGINEERING challenges of the spacecraft itself, and the many impulse manoeuvres needed to set it on its way and adjust its flight path *en route*, there are two basic navigational ingredients required for these remarkable space rendezvous. The first is an accurate knowledge of the target object's own position and orbital motion around the Sun. The second is an up-to-date knowledge of the spacecraft's own position and orbital motion along its pursuit path.

FOR THE FORMER, precise orbits of many solar system objects are calculated, compiled and maintained, independently by the Jet Propulsion Laboratory (JPL, California), and by the Institute for Celestial Mechanics (IMCCE, Paris). Specified at regular intervals spanning a certain period of time, these estimated positions and orbits are referred to as the object's 'ephemerides'.

Precision ephemerides of the Sun, the planets, the Moon, and other solar system objects are also used as the basis of the *Astronomical Almanac*, with its various civilian applications such as the positions of the planets, and predictions of the phases of the Moon and of civilian twilight. Ephemerides provide the positions, velocities and accelerations of each object at equally spaced intervals over a specified period. They take account of all available knowledge including their masses, the oblateness of the Sun, and relativistic corrections.

Observational data used in the fits include transit and CCD observations of planets and small bodies, lunar laser-ranging, radar-ranging, and distances measured by radio signals from interplanetary spacecraft themselves. Angular accuracies reach around 0.001 seconds of arc (arcsec) for the inner planets (1 km at the distance of Mars), and some 0.1 arcsec for the outer planets.

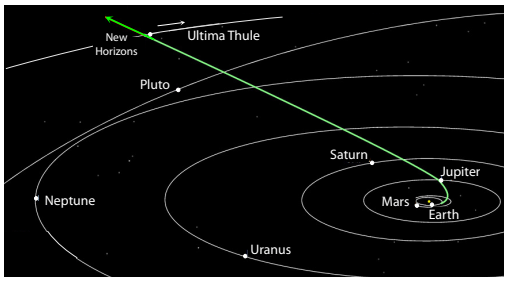
FOR THE LATTER, establishing the position and velocity of the spacecraft itself as it travels to its rendezvous makes use of various techniques. For spacecraft near to the Earth, radar tracking using ground-based antenna can be used to determine the spacecraft's instantaneous distance via the signal's time delay, and its radial velocity via the Doppler effect. Accuracies of 1 m in distance and 1 $\mathrm{mm\,s^{-1}}$ in velocity can be routinely achieved, further augmented through the use of two or more ground stations to allow triangulation. Tracking the spacecraft over time allows its orbit to be computed.

For more distant spacecraft further out in the solar system, radar tracking is supplemented by 'delta-differential one-way ranging', which uses two or more widely-spaced ground stations to interpret the spacecraft signal. The time-delay between the receipt of the signal establishes the angular position of the spacecraft, which can then be related to the known positions of distant quasars lying within a few degrees of the line-of-sight. Angular accuracies of around 10 nano-radians (2 milli-arcsec) can be achieved, corresponding to a transverse positional accuracy of 1.5 km for a spacecraft at 1 astronomical unit from the Earth.

ULTIMATELY, both the spacecraft orbit, and the target celestial object's orbit, are referenced to the position of background stars and quasars on the sky. Progress in interplanetary navigation rests on the accuracy of the best star catalogues available at that time.

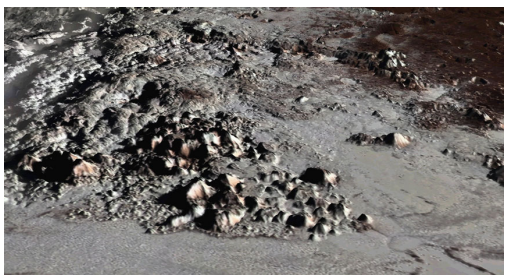
The New Horizons fly-by of Pluto was based on the JPL solar system ephemerides DE430, finalised before the Pluto encounter (Folkner et al., 2014), and itself based on the celestial reference frame materialised by the Hipparcos star positions from the late 1990s.

For New Horizons, attitude *determination*, i.e. knowing which direction it's pointing, uses star-tracking cameras, gyroscopes and accelerometers. Attitude *control* is accomplished using (hydrazine gas) thrusters. The star-tracking cameras store a map of about 3000 star positions. Every 0.1 second, this is compared to a wide-angle image of space. The spacecraft's actual orientation is determined on-board, and the hydrazine thrusters can be fired to reorient the spacecraft as required. All this allowed the spacecraft to be piloted for the fly-by of Pluto, in turn leading to the spectacular and unprecedented images of its surface, and a whole host of other scientific results enabled by the spacecraft's instruments.



Johns Hopkins University Applied Physics

ARROKOTH (formally 486958 Arrokoth; also known as 2014 MU_{69}, and as Ultima Thule) is a trans-Neptunian object within the Kuiper belt. It is a contact binary, composed of two planetesimals 21 km and 15 km in size, joined along their major axes, and with an orbital period around the Sun of about 298 years. It was discovered in 2014 by Marc Buie and colleagues using the Hubble Space Telescope as part of a search for a Kuiper belt object to be targeted for a further flyby beyond Pluto.


Paul Schenk/Lunar and Planetary Institute

Pluto's surface imaged by New Horizons in 2019: mountains rise 6 km above plains of nitrogen-ice

THE EVENTUALLY HIGHLY SUCCESSFUL flyby was based on star occultation campaigns in Argentina, Senegal, South Africa and Colombia in 2017–18, which themselves made use of the Gaia DR2 star positions provided to the New Horizons team before their general release. Images from the spacecraft's Long-Range Reconnaissance Imager, 6.5 minutes before closest approach, provided a high-spatial resolution of 30 m per pixel, along with a favourable viewing angle.

According to Science Magazine (24 Nov 2020), Marc Buie persuaded the New Horizons team to trust the new Gaia stellar framework, and a correction based on the Gaia positions was sent to the spacecraft. The article continues: When the closest flyby images came back, Arrokoth was framed perfectly. *'None of that would have happened if we hadn't had the Gaia catalogue'*, Buie says. *'It's a fundamental rewriting of how we do positional astronomy'*.


NASA/JHU-APL/SwRI/Roman Tkachenko

Ultima Thule

Scientifically, the surfaces of each lobe of Arrokoth display regions of varying brightness along with various geological features such as troughs and hills, thought to have originated from the clumping of smaller planetesimals to form its lobes. The brighter surface regions may be material that rolled down from its higher peaks, under gravity. The interior is believed to be composed of amorphous water ice and rocky material.

While moons and asteroids in the inner solar system suggest a violent collisional past, the surfaces of objects in the Kuiper Belt around Pluto and beyond reveal a more tranquil environment, due to their lower space densities and smaller orbital speeds. Pluto confirmed this through its relative lack of impact scars, while the surface of Ultima Thule supports the same ideas.

FOR DEEP SPACE NAVIGATION, beyond the solar system, spacecraft will be too distant to rely on Earth-based tracking. Traveling to the nearest stars, signals will be too weak, and light travel times will be of order years. An interstellar spacecraft will instead have to navigate autonomously, using other information to decide when to make course corrections or to activate instruments.

Radio pulsars have been considered as a solution to this future problem, employing an approach somewhat analogous to global positioning systems on Earth, but using pulsars rather than an artificial satellite constellation as navigation beacons.

Another approach to deep space navigation is via direct triangulation of stars, using the full three-dimensional positions of a set of stars relative to some well-defined reference frame. A recent feasibility study based on the accurate star positions being provided by Gaia is given by Bailer-Jones (2021).