Science with Gaia: 2022

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

Michael Perryman

Essays 53–104 (Jan – Dec 2022)

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 Many will be quickly superseded by new results.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

53. The scientific case for Gaia in 2000

AT THE START of another year, and now 22 years since Gaia was accepted by the advisory committees of the European Space Agency in 2000, and with a steady flow of scientific results now coming from the mission, I thought it could be of interest to look back at the scientific case that was prepared at that time, and which was included along with the early technical assessment, together used as the basis for the advisory structure's consideration of the mission, and its eventual selection.

WHEN A ROUND of competitive mission selection is announced by ESA, a considerable effort – generally spanning several years – will already have been devoted to preparing an overall mission concept (describing what the satellite and payload will look like in overall terms), a preliminary but authoritative technical assessment, and an outline of the overall mission organisation and management.

The formal Gaia 'Concept and Technology Study' ran between 1997–2000. What was at stake was its candidacy for the next cornerstone mission of the ESA science programme, part of ESA's long-term scientific programme arising from the recommendations of the Horizon 2000+ Survey Committee in 1994.

The resulting report, ESA–SCI(2000)4, extended to almost 400 pages. It synthesised the preparatory activities of the ESA Study Scientist (Michael Perryman), the ESA Study Manager (Oscar Pace), the scientists throughout Europe and beyond already supporting the mission's objectives, the various industrial contractors involved in preparatory technology activities, engineers in ESA's technical support directorate, and representatives of ESA's Spacecraft Operations Centre (ESOC).

AT THIS STAGE, throughout 1997–2000, I had set up the following advisory structure. I chaired the top-level Science Advisory Group (SAG), comprising just eight external scientists, together intended to cover those whose primary interest was in the overall scientific goals, those with an appropriate technical background, and at the same time covering some overall European geographical (and therefore also political) representation.

Guiding the mission through to selection in 2000, these members were K.S. de Boer (Bonn), G. Gilmore (Cambridge), E. Høg (Copenhagen), M.G. Lattanzi (Torino), L. Lindegren (Lund), X. Luri (Barcelona), F. Mignard (Grasse), and P.T. de Zeeuw (Leiden).

Within ESA, four other staff regularly supported the top-level SAG activities (Oscar Pace, Martin Hechler, Sergio Volonte, and Fabio Favata), with a further four from ESA's Space Science Department providing additional inputs. From ESA's technical directorate, 11 engineers provided specific support in the various areas of instrument design, manufacture and testing. Their expertise covered on-board data processing, radiation, thermal control, structure, antenna, communications, attitude and orbit control, CCDs, and basic angle monitoring.

Reporting to the Science Advisory Group were three 'working groups'. The Science Working Group, jointly chaired by Tim de Zeeuw and Gerry Gilmore, comprised 18 scientists. The Instrument Working Group, chaired by Lennart Lindegren, comprised 16 people; and the Photometry Working Group, chaired by Fabio Favata, comprised 18 people. A further 50 European scientists supported these various activities as 'Members at Large'.

The preparatory technical studies were undertaken through two principle study contracts issued by ESA, and again coordinated by the ESA Study Scientist and ESA Study Manager.

One of these industrial teams, Matra Marconi Space (Toulouse; now Airbus), led by Pierre Mérat and comprising around 15 engineers, focused their efforts on primary mirrors of a monolithic design. The other, Alenia (Torino), led by Stefano Cesare and comprising a similar number of support staff, focused on a (small-baseline) interferometric system.

ALREADY WITH AN OUTLINE instrument concept on the drawing board, and with a fairly comprehensive assessment of the astrometric and photometric accuracies that Gaia should achieve, we knew that the scientific breadth and reach of Gaia would be enormous, and a corresponding effort was invested in preparing and presenting the scientific case.

Within the 400-page Concept and Technology Study Report, the scientific case for Gaia ran to 100 pages. More than 60 (mainly ESA member-state) scientists contributed to the various sections, with the overall science case coordinated, and the report synthesised, by Tim de Zeeuw, Gerry Gilmore, and Michael Perryman.

THE EXECUTIVE SUMMARY presented the summary scientific case as follows, given here verbatim (while pointing out that the referenced mission accuracies targeted at the time, 10 microarcsec at 15 mag, were later descoped to around 25 microarcsec).

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

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

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

'Gaia will pinpoint exotic objects in colossal and almost unimaginable numbers: thousands of exoplanets will be discovered, and their orbits and masses determined; tens of thousands of brown dwarfs and white dwarfs will be identified; some 100 000 extragalactic supernovae will be discovered and details passed to ground-based observers for follow-up observations; solar system studies will receive a massive impetus through the detection of tens of thousands of new minor planets; inner Trojans and even new trans-Neptunian objects, including Plutinos, may also be discovered.

'Gaia will follow the bending of star light by the Sun and major planets, over the entire celestial sphere, and therefore directly observe the structure of space-time—the accuracy of its measurement of General Relativistic light bending may reveal the long-sought scalar correction to its tensor form. The PPN parameters γ and β, and the solar quadrupole moment J_2, will be determined with unprecedented precision. New constraints on the rate of change of the gravitational constant, $\dot{G}$, and on gravitational wave energy over a certain frequency range, will be obtained.'

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

IT IS INDEED SATISFYING to see these far-reaching predictions being achieved within just two decades.

In another essay here, Galactic Tracers by Design, I go more into how Gaia targeted very specific measurement goals, formulated in terms of limiting magnitude and astrometric accuracy. Gaia's limiting magnitude of around 20–21 mag, for example, was a requirement resulting from an examination of what types of stars, in different regions of our Galaxy, should be measured.

THE STUDY REPORT also included a section on specific questions that could be tackled by high-accuracy astrometry, but which remain largely inaccessible even to Gaia: rich grounds for future space missions. This included the cores of globular clusters, the Galactic centre, inhomogeneities of the Galactic gravitational potential, and geometric cosmology, viz. studying the transverse motions of distant galaxies and quasars, independently of a dynamical model of the Universe.

54. Animations, stereos and fly-throughs

ASTROMETRY, EVEN SPACE ASTROMETRY, doesn't necessarily easily lend itself to 'public outreach'.

The results of the first space mission, Hipparcos, was a catalogue of 120,000 star positions, parallaxes and proper motions. In printed form, the main catalogue ran to five volumes, more than 2000 pages of small densely packed type, superficially reminiscent of an old (printed) telephone directory. Astronomers made extensive use of the information; but in terms of visual impact, and popular appeal, Hipparcos couldn't complete with, say, the stunning images of deep space returned in glorious abundance by the Hubble Space Telescope.

And while scientists were, even in 1997 and more so today, encouraged to make the results of these ambitious space missions accessible to the wider public who, through their taxes, essentially sponsor them, the task was and remains a challenging one.

In 2009 I wrote a popular account of the Hipparcos mission under the title *'The Making of History's Star Map'*, published by Springer. A more mainstream publisher came rather close to accepting it, but eventually declined because 'two thirds of our sales in English are sold in the US, and the US public has not heard of Hipparcos'. I argued, unsuccessfully, that surely the purpose of writing a popular book was to make things known to those who had never encountered the subject before.

Curiously, I was invited instead to write a general book on the Hubble Space Telescope! The fact that I had no inside knowledge of the project, and that there were already more than 30 published titles on it, were not seen as impediments to this offer… which I declined.

AN INVITATION TO DELIVER the annual Darwin Lecture to the UK Royal Astronomical Society back in 1997 was an ideal opportunity to try for some special visualisation of the mission results. It is not easy, today, to comprehend the more primitive state of digital projection just two decades ago. So my lecture centred, precariously, around 35-mm stereo slide pairs, synchronously projected, and suitably imaged, through two identical slide projectors equipped with orthogonal polarising filters. Projected onto a large silvered screen to avoid loss of polarisation, with the audience equipped with appropriately polarised glasses, I displayed a series of star fields in their full stereoscopic glory.

The impact of Hipparcos, aided by this son-et-lumière performance, led to my Invited Discourse at the IAU General Assembly in 2000, with the stereoscopic system scaled up to hefty high-powered twin projectors, and my first efforts at including stereo videos for an audience of 1200 in Manchester's huge Bridgewater Hall. In a darkened auditorium, the effect is almost like looking up at the night sky, while resolving the depths of space. It is powerful visually, but laborious to replicate.

And an anecdote: during my preparation for this talk, which required a bulk order of cardboard-framed polarised 'glasses' from a supplier in London, my telephone enquiry was interrupted by the assistant as I was explaining my intentions. *'Let me stop you there'*, he said, *'I'm an amateur astronomer, and stars are just too far away for any stereo effect to be observable'*. After a short pause, I explained why I thought I could. *Ah!*, he said, *there's probably just one person in the world in a position to demonstrate this effect, and I'm speaking to them*!

THERE ARE VARIOUS REASONS why polarised light projection is particularly effective in displaying stereoscopic scenes. One important advantage is that the subtleties of the slightly different star colours can also be used, which aids the brain in more easily grasping the stereoscopic effect. Anaglyph (e.g. red–green) images can also work, but are considerably less forceful.

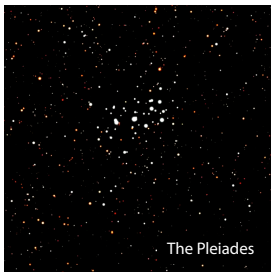
Another technique which is quite reasonable in visualising the stereoscopic effect of star fields is 'swing stereoscopy', in which two lateral projections (or a more complete video sequence), are cycled repeatedly. I will refer to a couple of examples of this below.

A further engaging technique for visualising the 3d structure of interplanetary, interstellar, or intergalactic space (and enabled by today's computational and projection technology) are simulated fly-throughs. Again, I will refer to some spectacular examples below.

As well as their stereoscopic distances, Hipparcos and Gaia also measure the *motions* of stars through space. These motions, suitably 'speeded up' and extrapolated over thousands of years, can then be compressed into just a few seconds of movie time.

We can then see things like the configuration of stars in the night sky as they will appear thousands of years in the future. As just one example, the easily recognisable constellation of Ursa Major will look very different, far in the future, compared to how it appears to us today.

To illustrate the principles, elsewhere on my www site I've included some animations based on the Hipparcos results for both the Hyades and Pleiades open star clusters. Each covers a sky region of 8° × 6°, each reconstructing the star magnitudes, colours, distances, and space motions. In the first of each series, star motions are extrapolated — over 60 000 years for the Hyades, and 150 000 years for the Pleiades.



In the second, star distances are visualised by 'swing stereoscopy' by changing the observer's effective location (and corresponding to the Earth's annual motion around the Sun).

In the third sequence, the two effects are combined to show the (amplified) annual parallax superimposed on the space motion. The continuous variations in the relative positions of the stars with time illustrates the principles used by Hipparcos and Gaia in determining (and decoupling) stellar distances and space motions.

A movie illustrating the Gaia proper motions, created by Jacqueline Faherty of the American Museum of Natural History, shows the Gaia stars (from DR2) within 200 parsec of the Sun. The animation runs over 1.5 million years, with the star motions at 25 000-year time steps compressed into 1 second of movie time:

`www.youtube.com/watch?v=qBZKXwpzdDw`

Within this vast local volume, we can see star clusters moving slowly through space. And we can see fast-moving stars racing across the field, members of our Galaxy's halo population passing through our neighbourhood. Some of her other animations are at:

`www.dropbox.com/sh/4v21ane28d6gd2x/AADJHaMG1Fu6RJGDksjBxW0Ga?dl=0`

Jackie Faherty received the American Astronomical Society's 2020 Vera Rubin Early Career Award for her work on the kinematics of very faint stars in our Galaxy, and for 'developing unique ways to engage the public and professional science teams in astrometry'. In one of ESA's 'soundbite' videos, she describes Gaia as: … *one of humanity's greatest missions… a story of vast proportions in our understanding of how stars form and evolve.*

A first detailed scientific assessment of what Gaia is contributing to our knowledge of stars within 100 pc was published in December 2020 (Smart et al., 2021), using the data from Early Data Release 3 (EDR3), which I have described in some detail in my essay #33. The authors also provided some information graphics and videos, which are collected at:

`www.cosmos.esa.int/web/gaia/edr3-gcns`

Amongst these are a fly-through of more than 300 000 stars based on this Gaia Catalogue of Nearby Stars:

`www.youtube.com/watch?v=bzQUNCleS3o`

and the Galactic orbits of a subsample of 74 281 stars:

`www.youtube.com/watch?v=k9pHGhNtyPk`

An extensive suite of Gaia star animations is the '3D Universe Simulator', comprising more than a billion objects, and made available by Stefan Jordan and colleagues in the Gaia group of the Astronomisches Rechen-Institut (ZAH, University of Heidelberg):

`zah.uni-heidelberg.de/gaia/outreach/gaiasky`

Amongst its many capabilities are the ability to navigate freely through the Galaxy, and its support for various forms of stereoscopic projection.

Amongst various other Gaia animations now publicly available, or in preparation, are a series of nice maps and interactive tools created by Kevin Jardine. These provide the possibility to fly through the catalogue, searching for some chosen star, or simply clicking on a star to find its name and properties:

`galaxymap.org/drupal/blog/1`

I will finish with two movies that are completely independent of Gaia and space astrometry, but which provide an important cosmological context.

The first is the acclaimed Millennium Simulations of Volker Springel and colleagues, from 2005:

`wwwmpa.mpa-garching.mpg.de/galform/millennium`

Based on more than 10 billion interacting particles, it shows the dark matter distribution in the Universe today. It highlights structure on scales, ranging from several billion parsecs down to substructures as small as 10 kpc. It encapsulates our understanding of how galaxies, including our own, have formed.

The second is a simulation of disk galaxy formation, from the cosmological N-body simulations TNG50, also created by Volker Springel and colleagues, which zooms in on the formation, over billions of years, of a disk galaxy very closely resembling our own:

`www.tng-project.org/media`

The point to stress here is how well these simulations, which are based on extensive physical *models* of our Universe's formation, now resemble the sort of structure that Gaia is discovering in our own Galaxy. With Gaia, astrometry is truly confronting cosmology.

55. Wow!

THE SEARCH FOR extra-terrestrial intelligence, or SETI, is motivated by the belief that intelligent life is likely to emerge under conditions similar to those on Earth. Detailed studies started with Cocconi & Morrison (1959).

Loosely connected but not implicit in such searches are various unproven and somewhat inconsistent postulates. Amongst these are the 'anthropic principle' (which suggests that no assertion can be made about the probability of intelligent life based on a sample set of one, viz the Earth); the 'mediocrity principle' (which, given the existence of life on Earth, asserts that life typically exists on Earth-like planets throughout the Universe); and the 'fine-tuning hypothesis' (which asserts that the natural conditions for intelligent life are implausibly rare).

Attempts to quantify the probability that intelligent life exists elsewhere in the Galaxy include consideration of the Drake equation (formulated by Frank Drake in 1961), and consideration of the *Fermi paradox*. Resolution of these questions may still lie far in the future.

THE FERMI paradox is the question famously posed by physicist Enrico Fermi in 1950: *'If other advanced civilisations exist, where are they?'*. Alternatively formulated as *'If alien civilisations existed, they would be here'*, or as the 'Great Silence' problem, it is a deceptively simple question that presents a challenge for theories assuming a naturalistic origin of life and intelligence. Objectively, scientists attempt to weigh the evidence for or against the existence of other intelligent civilisations.

According to current understanding, there is no evidence at present which contradicts the hypothesis that life arose on Earth due to such extraordinarily improbable events that it is unlikely to have arisen elsewhere within the observable Universe. But the converse is also true: there is no evidence at present which contradicts the hypothesis that life arose on Earth due to such ordinary and probable events that it is likely to have arisen in many other places within the observable Universe.

As the 19th century Scottish essayist Thomas Carlyle gloomily remarked on looking at the night sky: *'A sad spectacle. If they be inhabited, what a scope for misery and folly. If they be not inhabited, what a waste of space.'*

ALTHOUGH SETI is often thought of as lying outside of today's mainstream scientific research, many searches have been undertaken. Early surveys, from the 1980s, including ones supported by NASA, searched for continuous and pulsed radio or microwave signals that could have been generated by alien civilisations. Recent efforts have also searched for intense optical pulses, as could be generated by targeted high-power lasers.

While many of these searches have been discontinued, others are being actively pursued, notably by the SETI Institute (using the Allen Telescope Array at Hat Creek), and by the 'Breakthrough Listen' initiative (using the Green Bank and Parkes radio telescopes).

OVER THE YEARS there have been several false alarms and unconfirmed events. The 'SETI episode' in the 1967 discovery of pulsars is recounted by Penny (2013). Here, the Cambridge radio astronomy group became the first to confront the 'contact' problem which the SETI community faces in its 'detection and reply protocols'.

Otherwise, most of the 'interesting' signals so far detected have been transient and non-repeatable, perhaps the first points in a growing database of signals used to construct a probabilistic argument for their existence.

WHEN SCIENTISTS DISCOVER a completely new celestial phenomenon, efforts focus on attempting to interpret it in terms of known physics and natural processes. Theoreticians and observers will contribute their different expertise. Ultimately, a widely accepted explanation must satisfy the laws of physics, be verifiable, and have predictive power.

In 1967, the discovery of 'pulsars' – extremely regular and rapid radio pulses originating from specific points on the sky – could not immediately be attributed to any known celestial phenomenon. Now known to be generated by rapidly spinning neutron stars, the discoverers initially considered that the signals might have been generated by an alien civilisation.

But an enormous edifice of theory and observation of pulsars, all fitting within our detailed models of stellar evolution, can now discard such a non-physical origin.

I WILL MENTION two other examples of recently discovered celestial phenomena which probably have a purely physical explanation, but which have aroused interest as being possible alien techno-signatures, and where Gaia is contributing to a more complete picture.

My first example is the 'interstellar traveller' Oumuamua, the first object moving through our solar system known to have originated from another star system, and described in more detail in a previous essay (#25). Discovered in 2017, it is a highly elongated body some 100 metres in size, considered by some at the time to be a possible alien techno-signature.

But its existence, and its passage through our own solar system, now appears to be fully consistent with what we now know about the formation and evolution of other exoplanetary systems. Ongoing work using the enormous Gaia data base of star positions and space motions is contributing to the challenge of pinning down its likely progenitor stellar system.

MY SECOND EXAMPLE is the curious 'Boyajian's star' (KIC–8462852), discovered in the Kepler satellite data by the Planet Hunters project from its unusual light-curve (Boyajian et al., 2016). The star shows pronounced dimming by up to 20%, lasting between 5–80 days, and with an irregular cadence and unusual profile. Considerable speculation accompanied the unusual light curve, with interpretations ranging from occulting clouds of exocomets, to a 'swarm of alien megastructures' – and hence an outstanding SETI target (Wright et al., 2016).

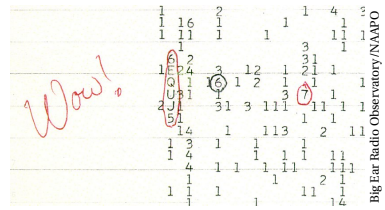
Other stars with similar light curves would help to elucidate its nature and, today, Gaia is actively contributing to this search. Twenty one further candidates from wide-field variability surveys have recently been identified as possible 'dippers' (Schmidt, 2019).

Accurate distances for all of these new candidates from the Gaia Second Data Release (DR2) in 2018 show that these objects are all located in two restricted regions of the Hertzsprung–Russell diagram: some being near the main sequence with masses near that of the Sun, and the others in the red giant region near the evolutionary track for 2 solar mass stars. Again, improved constraints will come from future Gaia data releases.

PULSARS WERE NOT the first examples of strange signals detected by radio astronomers. Indeed the technology and radio pioneer Nikola Tesla claimed to have detected interplanetary transmissions during his work in Colorado Springs in 1899.

Tesla wrote, in 1901, that: *'The changes I noted were taking place periodically, and with such a clear suggestion of number and order that they were not traceable to any cause then known to me… The feeling is constantly growing on me that I had been the first to hear the greeting of one planet to another.'*

ANOTHER EXAMPLE of Gaia's contribution to the field is in the context of the famous 'Wow!' signal of 15 August 1977, detected at Ohio State University's 'Big Ear' radio telescope. It was a strong 72-second anomalous signal at 1420 MHz – the emission frequency of neutral hydrogen, which physicists Philip Morrison and Giuseppe Cocconi speculated might be the preferred medium of extra-terrestrial communications.



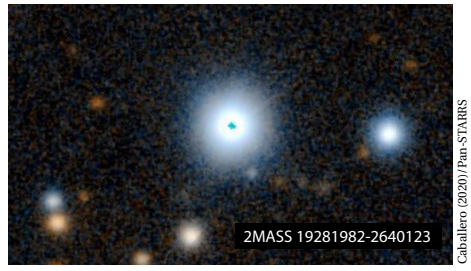
The 'Wow!' signal was so-named because Jerry Ehman, the astronomer who analysed the data print out, annotated the signal with the word 'Wow!'. The signal was never repeated, and it remains unexplained.

Observatory director John Kraus described it in a letter to astronomer Carl Sagan: *'The Wow signal is highly suggestive of extraterrestrial intelligent origin, but little more can be said until it returns for further study'*. Kraus and others searched for stars that could be the source of the signal, writing *'We checked star catalogs for any Sun-like stars in the area and found none'*.

A detailed report, *'The Big Ear Wow! Signal: what we know and don't know about it after 20 years'*, was written by Ehman in 1997 (`www.bigear.org/wow20th.htm`).

THE GAIA CATALOGUE is, of course, far more extensive, and far more detailed, than the star catalogue that Kraus had access to in the 1970s.

And amateur astronomer Alberto Caballero has now used Gaia DR2 to identify a possible stellar origin: out of 66 G and K stars in the likely sky region, one has sufficient information to infer that it closely resembles our Sun, i.e. the same temperature, radius, and luminosity. The object, 2MASS 19281982–2640123, lies in the Sagittarius constellation, at a distance of 1800 light-years (Caballero, 2022). It is not yet known whether this, or other stars in this part of the sky, are accompanied by planets.



In 2012, on the 35th anniversary of the Wow! signal, Arecibo Observatory beamed a digital stream towards Hipparcos stars HIP 34511, HIP 33277, and HIP 43587, comprising 10 000 Twitter messages solicited by National Geographic. No replies have been reported so far.

56. Gaia: interferometer or monolith?

As briefly mentioned in the 2016 A&A paper describing the Gaia mission (Gaia Collaboration et al., 2016): *'In the early phases, Gaia was spelled as GAIA, for Global Astrometric Interferometer for Astrophysics, but the spelling was later changed because the final design was non-interferometric and based on monolithic mirrors and direct imaging'.*

But the choice (interferometer versus monolithic) was not at all straightforward. And, as I will explain, Gaia owes its very existence to its interferometric origins.

In late 1983, ESA made an open call for mission proposals to the European scientific community, based on an idea for a community-driven programme, Horizon 2000, as presented by Director of Science Roger-Maurice Bonnet to ESA's Science Programme Committee. The solicitation yielded 68 proposals: including 30 in the field of astronomy, and 34 related to solar physics.

An *ad hoc* committee led by Dutch Space Research Organisation (SRON) director Johan Bleeker was set up to assess the proposals. It comprised members of ESA's Space Science Advisory Committee (SSAC), CERN, the European Science Foundation, the European Southern Observatory, and the International Astronomical Union.

In early 1984, the committee formulated plans for a series of missions divided into three categories: 'cornerstones' costing a total of two annual budgets, medium-size missions costing one annual budget, and small-size missions costing half an annual budget. In broadly indicative terms, the annual budget was assumed to be around 200 M€ from 1991 onwards.

Three such cornerstone missions were assigned to a specific field of science that competing proposals would then aim to fill, while the science objectives of medium-size missions were left open for competitive selection.

Cornerstones selected were a comet sample-return (subsequently Rosetta), X-ray spectroscopy (later XMM-Newton), and a sub-mm mission (subsequently FIRST). Cornerstone objectives not selected due to financial and technical reasons, but noted as possibilities ('green dreams') beyond Horizon 2000, included a solar probe, a Mars rover, and… an interferometry mission.

Several concepts competed to occupy this interferometric 'slot' (within the follow-on Horizon 2000+), with an emphasis on establishing interferometric *techniques* rather than being driven by specific science goals.

One interferometric concept was Darwin, proposed in 1993, and studied by ESA for many years. It aimed to image Earth-like planets around nearby stars, with (in its final configuration) four free-flying telescopes. Despite a loose collaboration with NASA on their comparable Terrestrial Planet Finder, both missions were eventually abandoned due to technical complexity and cost.

Meanwhile Rømer, proposed by Erik Høg and Lennart Lindegren in 1993 as a medium mission in Horizon 2000 (100 million stars to 15–17 mag at 0.1–1.5 mas accuracies), was reviewed by ESA's Astronomy Working Group, but lost out to the microwave background mission Cobras/Samba (subsequently Planck).

There had, in fact, been ESA study teams devoted to space interferometry since the early 1980s. Under Sergio Volonté's direction, I had coordinated some of the work around 1988, leading to a report on a 'strategy' for space interferometry. My own notes from March 1995 read: *'I was unhappy with its immediate prospects – low throughput, a relatively small number of accessible targets, complexity, UV coverage, time variability – all seemed to undermine its scientific applicability.'*

On 9 September 1993, there was a lunar interferometry meeting in ESTEC. My notes from the meeting read *'Enjoyable talking to Chris Dainty. Useful in that it jogged me into putting a proposal in to post-Horizon 2000 planning, for a project which I called GAIA, for Global Astrometric Interferometer for Astrophysics'.*

In the evening, I met with Lennart Lindegren at the Camino Real, in Leiden. My notes continue: *'This is when Gaia was born – I believed that we could exploit the present interest in interferometry with a proposal for Horizon 2000+. A letter was duly prepared and submitted to ESA by Lennart; we prepared an outline proposal, added the names of interested people, and Lennart submitted it on 12 October 1993'.*

UNDERPINNING THE GAIA interferometric concept, as formulated and proposed by Lennart Lindegren and myself in 1993, were two distinct considerations. The first was programmatic: if an astrometry mission employed interferometry, it could hope to satisfy the open cornerstone slot in the Horizon 2000+ plan.

But a key reason to consider an interferometer was related to the accuracies achievable. If λ is the wavelength and D the overall size of the instrument aperture (diameter or base length), then the characteristic angular size of features in the diffraction pattern that can be used to localise a star image is of order λ/D radians. If a total of N detected photons are available for the image location, then the theoretically achievable angular accuracy is of order $(\lambda/D) \times N^{-1/2}$ radians.

In a scanning instrument, with the astrometric accuracy determined by the star's centroid position *along-scan*, it follows that for a given mirror mass, astrometric accuracy can be gained by extending the mirror in the along-scan direction. Which is why, today, the Gaia primary mirrors are rectangular, 1.45 m × 0.5 m in size, with the longer dimension in the along-scan direction.

GIVEN THAT the baseline launch vehicle, Ariane 5, had a fairing diameter of about 4.5 m, an interesting possibility presented itself: a rigid interferometer, housed within the Ariane 5 fairing, could employ circular primary mirrors of, say, $D = 0.6$ m, with a baseline length of around 2.5 m, and an astrometric accuracy significantly superior to a much smaller diameter monolithic mirror of comparable mass.

Although later changed to a Soyuz launch vehicle, these considerations led to the original design of GAIA as a stacked pair of interferometers, and to the mission's original accuracy goals of 10 micro-arcsec at 15 mag.

Working through the numbers led to an astrometric mission substantially more performant than Rømer. And with other features including multi-colour photometry and bright star radial velocities, the Astronomy Working Group and SSAC found the resulting science case compelling. In this form, rather crucially, GAIA also satisfied the letter – if perhaps not necessarily the spirit – of the Horizon 2000 interferometer placeholder.

NOT ALL OF ESA's advisers were happy with this astrometric juggernaut which appeared out of left field. Arguments against included the fact that ESA had only just completed one astrometry mission, Hipparcos, and that it was the turn of another scientific discipline.

Others expressed the opinion that astrometry was not what the Horizon 2000 architects had in mind for an interferometry mission. There was scepticism, too, that the required technologies were attainable. But the scientific case for this high accuracy astrometric survey, extending as faint as 19 mag, duly won the day.

THE INTERFEROMETRY CONCEPT underpinning GAIA (note the capitals!) was the initial baseline. But during subsequent system studies between 1997–1999, two industrial teams (Alenia, Torino; and Astrium, Toulouse) undertook a more detailed technical evaluation.

Alenia continued to focus on, and advocate, the interferometric design. But as the studies progressed, multiple problems became apparent. One was the technological challenge of maintaining the required 10 pm optical stability over the 2.5 m baseline. Associated with this was the formidable challenge of the in-orbit spatial alignment, complicated by the multiple optical components all along the two interferometric arms.

A further difficulty followed from the smaller Airy profile: the CCD pixels would have to be reduced, along-scan, to a dimension matched to the higher spatial frequency of the star images, with problems both for manufacturing, as well as for the pixel charge capacity. This led to other penalties including a brighter limiting magnitude, and a much higher data rate to Earth.

THE ASTRIUM TEAM, in contrast, soon discarded the interferometric option, and concentrated their efforts on a monolithic primary mirror design, albeit with a penalty in achievable astrometric accuracy.

When the time came for the technical evaluation of the two industrial proposals – covering all mission aspects including technical, managerial, schedule, risk and cost – the ESA evaluation team unanimously recommended acceptance of the monolithic design. The trade-off is detailed in the Gaia Concept & Technology Report (ESA–SCI(2000)4, Appendix B).

One could argue that, by discarding an interferometric concept, the Horizon 2000+ implementation plan should have been revisited. But by this stage the scientific impact of Gaia (no longer an interferometer, and so no longer capitalised as its original acronym!) was widely appreciated and fully endorsed by ESA's scientific community as represented by its advisory committees.

My own record of this technologically and politically charged period, extending over nearly 10 years, with its many actors, and its lobbying, meetings, reports, deadlines and frustrations, runs to many pages. But this outline picks out the key points in the transition from its interferometric origins to its final monolithic design.

IN SUMMARY, interferometry was eventually dropped for objective technological and associated cost/risk reasons. But I suggest that, had Gaia not started as an interferometer, it would never have been accepted within ESA's scientific programme in 2000.

And by not capitalising on the scientific expertise carried over from Hipparcos, it might have been a very long wait until the right circumstances for the flourishing of microarcsec astrometry re-emerged.

57. Technology preparation for Gaia

WHY DOES ESA FUND the development of space science missions? Why do some of these run over budget, and over schedule? Does the fact that both Hipparcos and Gaia were launched within a year of their target launch date indicate that they were simpler missions than some others in ESA's scientific portfolio?

LET ME START WITH MY FIRST question: why does ESA facilitate and fund the development of space science missions?

Amongst ESA directorates are satellite missions with a utilitarian function, notably those within the classes of Earth observation, of telecommunications, and of satellite navigation. Missions in the science directorate instead focus on scientific research: aiming to further our understanding of the Sun, our solar system, our Galaxy, and the Universe beyond. By advancing our knowledge, they sit at the heart of our culture.

At the same time, these scientific missions push the frontiers of technology, allowing more advanced experiments, and greater advances in scientific understanding than ever before. New missions proposed and accepted by ESA's scientific advisors must push these technical and scientific boundaries. If they are too modest in their technological vision, they will be of limited appeal to national industries. But if they are too unrealistic, problems may arise during their construction, and delays and cost overruns will follow.

WHY ARE NEW technological advances a requirement when selecting new missions? Several answers can be given. New scientific questions can frequently only be answered using more advanced experiments than available before. And these will generally demand an advance in the underlying technologies.

Industries embrace the opportunity for advances in their technological capabilities because this puts them at an advantage with respect to their competitors. And the link between advances in fundamental science and benefits to national economies has been widely demonstrated (see my essay on 'Gaia and GDP', #9).

THE REASONS FOR SPACE SCIENCE MISSIONS (and indeed for any large project whether in the public or private sector) running over budget and over schedule, can be many and complex.

As examples, the project may have been overly ambitious for the technologies available. This may in turn be traced to undue optimism of the originators, to understating the challenges in order to get the project accepted, or to insufficient rigour in the project evaluation, or to an underestimate of the resources required.

Projects are often subject to unrealistic 'requirement creep' as they are implemented, where new and more ambitious goals are introduced along the way. Delays can often be traced to errors in management, including a lack of progress monitoring, changes in specifications, inadequate resources, an absence of contingency planning, and many others.

CORRECTLY ANTICIPATING the challenges presented by new requirements in technologies has always been a challenge for space missions. In the early days of space science, a project may have been accepted with only limited attention having been given to the challenges in actually implementing it.

Different project 'phases' were introduced to help quantify the requisite steps. Both ESA and NASA, for example, have long followed the idea of a project phase A being dedicated to a sort of 'feasibility study' (i.e., does the project appear broadly feasible, given what we know about current and projected state-of-the-art in technology?), leading to the mission's acceptance (or otherwise), followed by a phase B dedicated to a detailed system design. This in turn would be followed by the satellite construction and testing (phases C/D), followed by launch and operations (phases E/F).

Durations vary, but are (indicatively) around 2 years for phase B, and around 6–8 years for phase C/D. The feasibility study phase can vary more widely, anything from 2–5 years up to one or two decades or more in the case of extremely demanding technologies such as for the ESA–NASA gravitational wave mission, LISA.

RELATIVELY SMALL INDUSTRIAL teams are typical during phases A–B, perhaps involving some 2–3 industries and some 50–100 people. Costs escalate during the implementation phases C/D, when tens of companies, and some thousands of people, might be involved.

Consequently, unexpected problems encountered during phase B may have significant schedule but relatively small financial consequences, while major technological problems encountered in the implementation phase, where costs of hundreds of the associated work force might be carried along, can be substantial.

OVER THE PAST DECADE OR SO, procedures have been developed in an attempt to control these costs and schedule risks. For example, ESA today follows the approach of 'approving' a project once the scientific objectives have been endorsed and once the required technology development activities have been identified, but only finally 'adopting' the project once the technology development has reached a specified level of maturity.

Quantifying the technological maturity of each satellite component today follows a strict hierarchical 'Technology Readiness Level', where each hardware item is classified from level TRL1 (the basic principles have been reported), through TRL4 (the item has been validated in a laboratory environment), TRL6 (in which a prototype has been verified in the relevant environment), all the way up to TRL9 (in which the system has been 'flight proven' through successful mission operations). Today, ESA requires that the mission must be based on technology that can credibly achieve TRL6 by the time of the mission's *adoption*.

ACHIEVING OR demonstrating these TRLs is an interesting problem in its own right. As an example, Surrey Satellite Technology Ltd (SSTL) have employed the approach of operating satellites with a certain 'flight proven' technology serving some specific purpose (for example, attitude control), such that the mission can be largely guaranteed to function, while carrying a redundant but more innovative back-up technology to carry out the same task by way of a technology demonstrator.

As a specific example of a real schedule for an ongoing (but particularly challenging) mission, the first concept for a gravitational wave detection mission was proposed in the US in 1985, and to ESA in 1993 in response to the Call for Mission Proposals for the third 'medium size mission' (M3) within ESA's 'Horizon 2000' programme. It subsequently appeared as the fourth 'cornerstone mission' in Horizon 2000+ in 1995.

Required technologies satisfying the TRL classification system were identified in an ESA–NASA study which I chaired between 2014–16. The study report detailed a development schedule constrained to an ESA Science Programme Committee adoption in early 2025, and a launch in 2034.

GAIA STARTED OUT, in 1993, as a conceptual sketch on a piece of paper, further refined through discussions with the ESA technical directorate and, later, industry engineers.

Working closely with the appointed ESA Study Manager, Oscar Pace (who had led the Assembly, Integration, and Testing group within the ESA project team for Hipparcos) we then progressively finalised a conceptual design, and outline technological development plan.

Seven years later, as a result of a 2-year study involving our scientific and industrial study teams, and leading to the 360-page Concept and Technology Study Report (ESA–SCI(2000)4) presented to ESA's advisory groups in 2000, we had identified the various development activities needed to place the developments for Gaia on a secure technological footing. In the Executive Summary, these were listed as:

- high-performance, small-pixel (9 μm) CCDs
- high integration, low-noise detector chain
- SiC technology validation for mirrors and structure
- high data rate phased array antenna for an L2 orbit
- database/processing architecture for 20 TB of data

In more detail, identified development was foreseen for:

- three-side buttable, small pixel, high-performance CCDs for the astrometric instruments;
- large-area, highly integrated CCD-based focal plane assembly for the astrometric instruments;
- large size, high-performance CCDs for the spectrometer;
- large size, high-performance CCD for the photometer;
- highly integrated, high-speed, low-noise detection chains for the astrometric focal planes (including design, bread-boarding, and development of ASIC hybrids);
- large area SiC mirrors (manufacturing and testing);
- ultra-stable large size SiC structure and application to the payload primary structure;
- large size deployable solar array sunshield assembly;
- phased array antenna for high data rates and L2 orbits;
- complementary qualification of a FEEP based reaction control subsystem (delta-qualification for the adequate range of thrust and lifetime);
- optimised on-board data compression algorithm;
- payload data handling electronics architecture;
- ground verification calibration and required facilities;
- inch-worm actuators for the refocusing mechanism;
- data base architecture, including storage, archive and processing of the satellite data.

THE STUDY REPORT argued that the advanced technology programme could be completed in 2003, with phase B completed in mid-2005, all leading to a possible launch in 2009. When the mission was later accepted by ESA's advisory committees in 2000, and following the assignment of Bepi-Colombo as cornerstone 5, with Gaia as cornerstone 6, the target launch date was 2012.

These technology development activities were followed, and the satellite duly launched in December 2013. Whilst therefore largely adhering to its target implementation schedule, Gaia remains one of ESA's most advanced and complex satellite projects.

58. Einstein crosses

QUASARS ALLOW the successive Gaia catalogues, and in particular their parallaxes and proper motions, to be referenced directly to an extragalactic, or 'inertial', reference system. Some half a million distant quasars are observed as part of the Gaia survey, and I have outlined how they are being used to define this inertial reference system in essay #12.

Many quasars show extended structures, and time variability, which complicate the construction and definition of such a reference system. But, at the same time, these structures convey important physical information about the quasars themselves. Their intrinsic variability, for example, can be used to infer properties of the black hole accretion disks which power the intense electromagnetic radiation emanating from them.

GRAVITATIONAL LENSES, due to intervening massive galaxies and clusters along the line of sight, which result (in highly aligned cases) in multiple images of the quasar itself, are particularly rich in physical insights.

Hubble Space Telescope (NASA/ESA/STScI)

Strongly lensed quasars (i.e. those which result in multiple discrete images) provide a remarkable diagnostic for a whole range of important studies, including measuring the Hubble constant at intermediate cosmic distances; constraining the properties of dark matter; inferring the structure near the event horizon of supermassive black holes; measuring the accretion disk size; constraining the quasar broad emission line region geometry; measuring black hole spins; and testing general relativity in the strong-gravity regime.

Fortunately for the task of defining the reference system, such strongly lensed quasars are rare. But their rarity and compactness makes them difficult to find. Amongst the most keenly hunted are the quadruply-imaged 'Einstein crosses', which are some of the most constraining for detailed physical modelling, and where the time delays between the images brightness variations can provide constraints on the Hubble constant.

THE FIRST EINSTEIN CROSS to be discovered was QSO 2237+0305, a quasar at $z = 1.695$ nearly centred on a relatively nearby 15 mag spiral galaxy at $z = 0.0394$ (Huchra et al., 1985).

While crescents and rings resulting from gravitational lensing can be understood somewhat intuitively, it is not so obvious how such remarkable 4-image systems arise. Here, I would say, intuition is not such a helpful guide. Instead, we have to work through the mathematics, first noting that the details of the lensed images is governed by the *projected* surface mass density on the sky (i.e., in the lens plane). The four images cannot be due to a spherically symmetric lensing galaxy, but instead can arise when the lens is a triaxial ellipsoid.

Detailed lensing models with such a 'quadrupole moment mass asymmetry' (e.g. Chae et al., 1998) can also explain why the four images are almost circular, rather than arc-like. And they explain why the results do not contradict the 'odd number theorem' in gravitational lensing (viz. that there are always an *odd* number of lensed images): this is because one image (the fifth) lies right behind the foreground galaxy, and it is invisible because its magnification is always very small.

THE FIRST GRAVITATIONALLY lensed quasar, the 'twin quasar' QSO 0957+561A/B, was discovered with the 2.1-m telescope at Kitt Peak by Walsh et al. (1979).

Many others have since been found, often serendipitously, from radio observations with the VLA, from the Sloan Digital Sky Survey, from the WISE observatory, and others. Subsequent observations with the Hubble Space Telescope provide some of the highest-quality images.

As of December 2019, the Gravitationally Lensed Quasar Database listed 220 confirmed lensed quasars. These discoveries provide the basis for follow-up observations, time variability studies, and detailed modelling.

Today, new systematic searches of the entire sky have been made possible with Gaia. And over the past two years, the multi-national Gaia Gravitational Lenses working group (GraL) has discovered about 10% of all currently confirmed strongly lensed quasars, including 12 of the 56 known quadruply-imaged quasars.

ALTHOUGH GAIA DOES NOT produce images of the sky in the conventional sense, its effective angular resolution in the along-scan direction is nevertheless comparable to that of the Hubble Space Telescope.

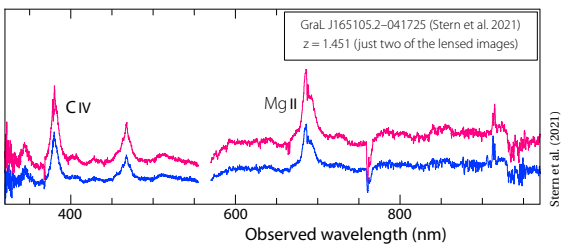
Thus, amongst the previously known quadruple-image quasars, 29 have at least one image in the Gaia DR2 data release, and 12 of these have all four images detected (Ducourant et al., 2018). These authors went on to show, for example, that the well-known quadruple HE 0435–1223 has model parameters significantly better constrained when using Gaia astrometry compared to HST astrometry, in particular for the relative positions of the background quasar, and the lens centroid.

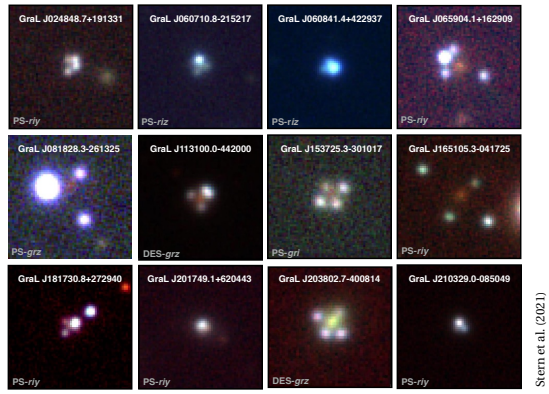
Gaia is accordingly providing the first homogeneous, magnitude-limited census of strongly lensed quasars over the entire sky, down to lensed image separations of about 0.18 arcsec. And indeed, even prior to the first data release, it was estimated that of the more than 500 000 quasars expected to be detected by Gaia down to $G < 20$ mag, some 3000 would be resolved, multiply-imaged quasars, of which only some 1600 would have image separations large enough to be resolved by ground-based atmospheric-limited observations (Finet & Surdej, 2016).

THE GRAL TEAM employs various methods to identify resolved multi-image lenses in the Gaia data. One uses supervised machine-learning to identify candidates based on their Gaia astrometry and photometry.

Another discovery technique is to use quasar light curves (e.g. from the Catalina Real-Time Transient Survey and the Zwicky Transient Facility), supplemented by Gaia astrometry and photometry. Underpinning this technique is that, for an unresolved lensed system, the observed time series is the addition of multiple identical light curves with time delays which, compared to an unlensed quasar, results in a less stochastic light curve.

In the sixth paper of their GraL series, Stern et al. (2021) identified possible quadruple-lensed systems in the Gaia DR2 data release, then used various spectroscopic observations (using the Keck telescope in Hawaii, the Palomar Hale, and the ESO NTT and the Gemini-South telescopes in Chile) to confirm or refute candidates based on their spectral similarities. In this way, they identified a total of 12 confirmed quadruply-imaged quasars, one example being shown here.



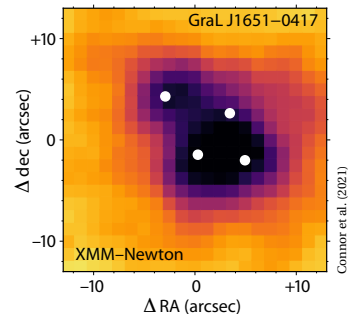


The 12 confirmed quadruply-imaged Gaia quasars

Several other groups are also using the Gaia data, along with machine learning and other supplementary observations, to find these rare systems, some of which are already noted in my earlier essay on Gaia quasars (#12). Amongst these are the discovery of a five-image lensed quasar at $z = 3.34$ using data from Gaia and Pan–STARRS1 (Ostrovski et al., 2018).

THESE NEW DISCOVERIES set the stage for various multi-wavelength follow-up activities, with some of the grand cosmological objectives mentioned in the introduction. Besides photometric monitoring, which will enable their use as cosmological probes, follow-up campaigns include high-resolution radio observations, X-ray observations, and near-infrared adaptive optics integral field unit spectroscopy to improve the source modelling.

At X-ray wavelengths, several of these quadruples have already been observed with ESA's XMM–Newton, starting in July 2020 (Connor et al., 2022). As an example, the image at 0.3–8.0 keV of GraL J1651–0417 shows exactly the same features detected by Gaia, indicated here by the white circles.



These and future X-ray observations will assist in the detailed modelling of the lensing source, since a typical quiescent lensing galaxy will not contribute a significant X-ray flux. But they may also provide estimates of black hole spins at redshifts $z > 3$, the size of the X-ray emitting region, and even X-ray spectral analysis at high redshift due to the magnified X-ray flux.

NEW PHYSICAL and cosmological insights, as well as many more new lensed quasars based on future improved Gaia data releases, will surely be arriving with us in the very near future.

59. Supernovae with Gaia

In preparing the scientific case for Gaia two decades ago, early estimates suggested that, with a faint limit of around 20 mag, the mission could detect supernovae to distances of around 500 Mpc, i.e. to redshifts $z \sim 0.1$. This would yield some 100 000 supernovae discoveries, of all types, over its 5-year nominal lifetime (Høg et al., 1999). Later pre-launch estimates pointed to numbers closer to 10 000 (Altavilla et al., 2012).

Actual Gaia discoveries are based on the photometric light-curves assembled from the satellite's repeated sky scanning. These are used to trigger ground-based observations to follow the light curves over the subsequent weeks, including some events 'caught' before maximum. Identifying possible supernovae forms part of the activities of the Alerts Working Group (essay #36).

Apart from simply adding to the number of known supernovae (and given that many other searches are ongoing), what sort of scientific results might be expected from the Gaia measurements?

Supernovae are dramatic and violent end-points of stellar evolution, and they lie at the heart of many important problems of modern astronomy. Their detailed taxonomy (they are classified according to their light curves, rather than their inferred physical origin), as well as their possible formation pathways, are both rather involved, and the underlying physics still somewhat uncertain. For this top-level perspective, I will distinguish here between just two physical classes:

(a) In *thermal runaway* events, low-mass (long-lived) stars accumulate material from a companion star (either a binary companion, or via a white dwarf–white dwarf collision). At some point, this leads to a rapid 'runaway' nuclear fusion reaction, releasing a burst of energy sufficient to disperse the star in a supernova explosion.

Type Ia supernovae fall into this class: they follow a characteristic light curve, and they present very uniform properties which make them important distance indicators ('standard candles') over intergalactic scales. Indeed, they provided the first evidence for an accelerating expansion of the Universe (Riess et al., 1998).

The wider importance of these type Ia events to cosmology is that the velocity field in the Universe is dominated by the Hubble expansion, while smaller-scale spatial structure is characterised by sheet-like regions with a high density of visible matter separated by large voids. Small-scale streaming motions of the luminous matter is not easily measured, requiring distance determinations independent of the Hubble law.

(b) In *core collapse* events, massive (short-lived) stars undergo collapse when nuclear fusion no longer sustains the stellar core against its own gravity. The prompt collapse may cause violent expulsion of its outer layers as a supernova or, if the release of gravitational potential energy is limited, the star may collapse into a black hole or neutron star with little radiated energy. According to the star's metallicity and mass (in the range $10-100 M_\odot$), different mechanisms, and different taxonomic classes (amongst them type I b/c and type II) may result.

As a result of these complex formation pathways, supernovae provide laboratories for studying extreme nuclear processes, being involved in the formation of neutron stars, black holes, millisecond pulsars and gamma-ray bursts. They are sources of gravitational waves, neutrinos, and high-energy cosmic rays.

And they, along with their precursors, are implicated in the evolutionary history and accretion rates of interacting binary systems, low mass X-ray binaries, and globular cluster X-ray sources.

Supernovae play a major role in the chemical, kinematic, and dynamical evolution of galaxies. They are the main producers of heavy elements, and are fundamental for understanding abundances and abundance patterns in galaxy clusters and in the intergalactic medium.

And supernovae rates provide a probe of the star-formation history in galaxies: due to the short lifetime of their progenitor stars, the type II (core-collapse) rate is proportional to the current star-formation rate, while type Ia rates reflect the convolution of the star-formation history with the distribution of the time difference between progenitor star formation and their explosion.

AFTER THIS PREAMBLE, I will take a brief tour of some of the Gaia studies related to supernovae, their remnants, and their progenitors, published so far.

Star parallaxes are providing improved distances of past events and their remnant structures. For the Cygnus Loop, Fesen et al. (2021) used EDR3 parallaxes of stars in or behind the supernova shell (based on their absorption lines) to set a precise distance of 725 ± 15 pc, with an uncertainty comparable to its 18 pc radius.

For the Vela complex associated with the Vela pulsar, Hottier et al. (2021) used 18 million stars over 450 sq. deg. to reveal the detailed morphology of the cavity shell. The 'last word' on the distances and kinematics of the proposed companions of Tycho's supernova using DR2 data is given by Ruiz-Lapuente et al. (2019).

THE GAIA DATA, supported by simulation models, are providing new insights into supernova progenitors. Kochanek (2022) analysed luminous stars in the vicinity of the Vela pulsar to estimate the mass of the exploding star, finding a low progenitor mass of $8 - 10\,M_\odot$.

Deep imaging of SN 1972E, the nearest type Ia supernova in more than a century, rules out models in which the companion is a He star, or a high-luminosity main-sequence star, as previously favoured in the literature.

FAST-MOVING STARS have long-been prized as forensic tools in astronomy, their space motions pointing back to the violent events that launched them. With the release of Gaia DR2, Shen et al. (2018) reported three hypervelocity white dwarfs, with velocities between 1000–3000 km s^{-1}, consistent with having been companion white dwarfs in pre-SN Ia binary systems.

van der Meij et al. (2021) used EDR3 data to show that the runaway X-ray binary HD 153919/4U 1700–37 originates in the association Sco OB1: a supernova in a compact binary can result in a high recoil velocity of the system, subsequently observable as a spectroscopic binary, a high-mass X-ray binary, and ultimately as a gravitational-wave event. Their models imply a progenitor mass $> 60 M_\odot$, most likely a neutron star, and a possible prototype for gravitational wave events such as GW 190412 (Abbott et al., 2020).

A STATISTICAL ASSESSMENT of the intrinsic velocities of neutron star progenitors was made by Yang et al. (2021). They considered 24 young (< 3 Myr) pulsars with precise parallaxes (from radio observations), and compared their space velocities with the velocities and velocity dispersions of nearby stars from Gaia DR2.

By showing that their transverse velocities are generally comparable to the velocity dispersion of stars in the local group of low-velocity pulsars, they could demonstrated that the intrinsic velocities of neutron star progenitors should be taken into account when consider their natal 'kicks' and subsequent binary evolution.

THE FRACTION OF progenitors in binaries or triples, and the fraction that survive the supernova explosion, are important for evolution models, and for predictions of the gravitational wave source populations. And these fractions can be determined by observations.

Kochanek (2021) used the Gaia EDR3 data to search 10 supernova remnants containing compact objects with proper motions for unbound binaries or triples. The Gaia data confirm the one known example of an unbound binary, HD 37424 in G180.0–01.7 related to the pulsar PSR J0538+2817, but found no other examples. Combined with previous searches for bound and unbound binaries, they found that some 70% of supernova producing neutron stars are not binaries at the time of explosion, around 15% produce bound binaries, and some 10% produce unbound binaries.

The specific conclusions in the case of HD 37424 are illustrative: at birth, the progenitor of PSR J0538+2817 was probably a $13 - 19\,M_\odot$ star and, at the time of explosion, a Roche-limited star transferring mass to HD 37424, eventually producing a type II supernova.

THE RECENT and ongoing discoveries of merging binary black holes by the LIGO–Virgo and KAGRA gravitational wave observatories have stimulated a renewed interest in understanding the origins of black hole binaries in short-period orbits, and indeed how massive stars evolve and eventually collapse to form these compact objects.

Current estimates are that there are some $10^7 - 10^9$ stellar mass black holes in our own Galaxy (e.g. Brown & Bethe, 1994). But discovering them using traditional methods such as X-ray or radio observations is notoriously difficult. As a result, these methods have resulted in only about 60 detections to date.

Detached binaries, those comprising a luminous companion orbiting a stellar mass black hole but without mass transfer, are almost impossible to detect using these well-established methods. Instead, several groups have argued that such detached systems could be detected from the orbital motion of the luminous component around its dark companion using Gaia astrometry (e.g. Barstow et al., 2014; Wiktorowicz et al., 2020).

In some of the most recent simulations, Chawla et al. (2022) estimate that the extension of Gaia to 10 years of operation, should allow for the detection of 30–300 black holes with detached luminous companions, and with orbital periods of 10 years or less. Such detections should provide new and important constraints on the many complexities of binary star evolution.

THE GAIA RESULTS to date have only touched on what will surely prove to be rich new seam of observations, facilitating the next steps in understanding the complexities of supernova progenitors, and these spectacularly cataclysmic events exploding around us.

60. Scientific project management

As in so many transformational events in life, we can look back and question the series of largely random events that led up to them. So it is that I found myself, in 1981 and aged just 26, and barely a year after completing my PhD in radio astronomy, appointed as ESA's Project Scientist for the recently adopted Hipparcos mission.

With an estimated cost of around €400M, and a conceptual design existing largely only on paper, it was the start of a series of numerous technical and organisational challenges, that I could not have anticipated. Looking back, would I have handed such a young and inexperienced scientist this sort of responsibility? I'm not sure that it would happen today, although I also suspect that, given the right skill set and character, young people can often be precisely those with the commitment, passion, energy and adaptability to succeed.

What is the skill set, or character profile, needed to drive such a project through to success? The short answer, perhaps, is to counter with the rhetorical question: *How long is a piece of string?*

In my area of expertise, space science, ESA appoints two key people, from two separate departments within the Directorate of Science, and with two distinct reporting lines within it.

The Project Manager, with a team of typically 20–25 engineers/managers reporting to them, supervises the industrial design and development of the satellite, and carries overall responsibility for the mission's cost and schedule. *Project* team members are rarely knowledgeable about the mission's *science*. Their collective success is judged on some ill-defined combination of the mission's ultimate science achievements, and its cost and schedule compared to the original targets.

The Project Scientist tends to work alone in ESA, or at most with a very small team. Calling instead upon the enormous skill set in the wider scientific community beyond ESA, but with limited financial authority to assemble or direct their activities, the task of the Project Scientist is to act as the guardian of the scientific goals of the mission at the time of its adoption by ESA.

They do this by interfacing with the scientific community representing the mission's scientific requirements on the one hand, and with the project team on the other. They communicate and translate the mission's scientific goals into technical requirements, mediating discussions between the two groups when requirements must be prioritised or relaxed, for example when cost over-runs are at stake. And they establish, coordinate, and ultimately deliver whatever the associated scientific/community supporting structure is required, for example in the scientific preparations in advance of launch, and in the preparation of the data processing facilities required to interpret the satellite data.

Throughout my time in ESA, from 1980–2009, and having been Project Scientist for two of these missions, and latterly as senior adviser on others, there was never a specification of the detailed responsibilities of this central role. It was often argued, for example, that any one mission is generally very different from any other.

At one extreme, some science missions comprise a *spacecraft*, which is managed and procured by the project team in conjunction with the industrial prime contractor, and this spacecraft system supports a set of *instruments* which are developed, built, supplied by, and eventually operated by scientific collaborations across Europe and often beyond (having been delivered to and integrated by the industrial prime contractor, according to interfaces specified by the project team).

At the other extreme, where the final mission performance is more acutely determined by the combination of spacecraft and payload performances, the payload (or experiment) may also be procured under the responsibility of the ESA project team. Examples of the former include missions like FIRST, SOHO, and Mars Express. Examples of the latter include both Hipparcos and Gaia.

With many of these responsibilities, interfaces, and final products only loosely defined, and with limited financial influence on both the project team and the scientific community, how can the Project Scientist best work to achieve some optimum balance between scientific performance, overall cost, and ultimate schedule?

ALTHOUGH PROJECT MANAGEMENT in these two areas (under the Project Manager, and Project Scientist) have many aspects in common, the boundary conditions and objectives are usually very different. In the following, I will focus on some comments relevant to the *scientific* management of these large projects.

I will argue that (1) there are skill sets that the appointed scientific leader should possess in order to best optimise the prospects of success, and (2) there are lessons that can be learned from previous large space projects, and indeed from wider sociological studies of how people and teams best work together.

A DETAILED CONSIDERATION of these aspects, along with the various caveats and nuances, would be lengthy and, like management in general, somewhat subjective. In the following, I will make some fairly simplified statements to introduce the challenges.

Firstly, scientists in the area of space science and astrophysics (and probably in physics more widely) undergo an education and training period, involving undergraduate, post-graduate, and post-doctoral studies, leaving the individual with a solid foundation in, for example, mathematics, computing, and a broad knowledge of physics, thus equipping them with the necessary foundation for a career in research.

Many of these well-qualified scientists soon start to work in wider collaborations, and often as parts of teams designing, building, and operating large and complex instruments and experiments, and taking responsibility for sub-systems or even managing the entire project. But rarely, in my experience, are these people offered advice, training, or some accumulated wisdom on how to manage projects, and how to manage people.

I mean, it's something we are either good at or not, and certainly something we can pick up along the way, right? Or we might take heed of respected Dutch astronomer Henk van de Hulst (1918–2000), who pointed out that *'doing science on your own is hard; doing it in collaboration with others is even harder!'*

I HAD THE BENEFIT, from day 1, of interfacing with the Hipparcos project team through the first ESA Hipparcos Project Manager, Franco Emiliani. Twenty years my senior, and already experienced from his work on Spacelab and others, I learned the techniques of Gantt charts and milestone trend analysis, work breakdown structures, configuration control and action item tracking.

Emiliani was interested in the scientific goals of Hipparcos, and despite our differences in age and experience (and pay grade!), we quickly formed a working practice underpinned by mutual respect, he of the goals of the scientists who would be spending decades of their working lives on the project, and me of the technical and schedule constraints that the ESA project team and the industrial contractors were working under.

Following standard practice, I set up a scientific advisory team composed of a dozen European scientists expert in the various aspects of the field: not simply representatives of the end users of the astrometric data, but those whose understanding extended to the technical aspects and to the data treatment. The team worked exceptionally well over the 17-year life of the mission, even though my early appointment of the group was in the days before I'd heard of Belbin's Teams and Apollo Teams.

AS FOR THE ATTRIBUTES of an ideal scientific project leader, they are easy enough to specify, but a challenge to identify in any one individual! They should be particularly well organised, and with familiarity with all instrument or mission aspects. As well as this 'big picture', allowing informed trade-offs when necessary, the leader should be a scientist of some repute, but willing to put an active research career on hold in order to manage properly a large and complex task.

Integrity is paramount, and an ability to take timely and frequently difficult decisions is clearly required. The ability to anticipate and confront problems before they become disastrous is particularly important. Communication skills are crucial, and taking the time to inform all team members, track actions, and liaise with funding authorities and political decision makers is essential.

Of course not all of these 'ideal attributes' will be met in any one individual in practice!

LET ME FINISH with just a few suggestions to anyone managing a large scientific project:

(1) establish a clear hierarchical advisory and decision-making structure. The team should cover all relevant project aspects, should not be more than 10–12 people, should loosely adhere to Belbin's Team principles, and should be adjusted as the project phase demands;

(2) use the regular meetings to exchange information on progress, and to agree on future 'actions' with clear objectives and fixed due dates. Expect these to be adhered to, so as to avoid a knock-on effect on overall schedule;

(3) establish clear project goals or scientific objectives from the outset, and ensure that all participants sign up to them. Beware of 'requirement creep' where new objectives are adopted as the project develops, and make realistic trade-offs between scientific aspirations and technological capability as needed;

(4) establish a top-level project plan from the start, with major milestones covering the entire project duration;

(5) view the entire project as an inter-related system, all of whose elements must be considered from the outset. As one simple example, do not wait until an instrument is designed and defined before addressing the user interfaces, and the data treatment.

It's an incomplete list, of course, but hopefully conveying the spirit of passing on accumulated wisdom.

61. Discovering variability with Gaia

HUNDREDS OF THOUSANDS of stars are, today, known to be variable. Some of this variability is truly 'intrinsic', being due to the luminosity of the star itself changing with time. In contrast, some stars classified as variable are 'extrinsic', meaning that the apparent variability occurs because something affects the stellar light on the way from the star itself to us on Earth.

Even to summarise the different *types* of variability would be lengthy: for example, `wiki/variable_star` lists around 50 types of intrinsic variables, and around 20 types of extrinsic variables. Since I want to focus here on how these are detected and classified with Gaia, let me give just a selective flavour of these variability types.

INTRINSIC VARIABLES include a huge range of pulsating star types. They change in brightness as they cyclically expand, and then contract, on periods ranging from days to weeks or months. The expansion phase is due to the blocking of the outflow of energy by gas with a high opacity, with expansion eventually halting as the density decreases, then reversing due to gravity.

In this class are the Cepheids, the RR Lyrae stars, along with Delta Scuti, SX Phoenicis, Mira, Beta Cephei, and Gamma Doradus variables. Their physics is reasonably well understood, and they vary over very different timescales, with very different amplitudes (extending up to many magnitudes), and sometimes along with multiple frequency (harmonic) components.

There are other intrinsic variables: eruptive variables show irregular or semi-regular brightness variations caused by material lost from, or accreted onto, the star. And there are 'flare' stares, along with cataclysmic or explosive variable stars including novae, dwarf novae, all the way up to the most explosive supernovae.

At much lower variability amplitudes, even our Sun (and along with it other main sequence stars) is variable. Our Sun varies over the 11-year solar cycle due to the long-term behaviour of its magnetic field. And at even lower amplitudes, the Sun oscillates in a large number of modes (with periods around 5 minutes), driven stochastically by turbulent convection in its outer layers.

EXTRINSIC VARIABLES include rotating stars which appear to vary due to surface sunspots rotating into and out of the line of sight, and other rotating stars which appear to be variable due to their underlying ellipsoidal shape. The huge class of eclipsing binaries vary because of a companion orbiting binary star which can regularly eclipse the light of its companion. Stars with orbiting planets may also show small brightness variations if their planets pass between Earth and the star, and many thousands have been discovered by NASA's Kepler satellite mission and others. Gravitational lensing provides another, rarer, form of extrinsic variability.

Detecting, classifying, and characterising such stellar variability, whether intrinsic or extrinsic, provides a vast and rapidly growing panorama for an almost endless range of studies of the physics and environment of the stars in our Galaxy and beyond.

Historically, variable stars were first discovered through naked-eye observations of individual stars, and more recently through very large-scale sky surveys (such as Pan-STARRS, ASAS, and many others) which return to the same parts of the sky again and again, and at different cadences, to detect and categorise variability on all sorts of time scales, and at all sorts of amplitudes. The most ambitious of these, under development today, is the 8.4-m Vera Rubin telescope (or LSST), which will photograph the available sky every few nights.

BECAUSE OF ITS sky scanning above the atmosphere, Gaia is contributing enormously. Its strength is not so much the *number* of observations of a given star over its possible 10-year mission, but rather its all-sky visibility, and its unprecedented photometric *accuracy*. This is in part due to its detection system, but in particular its location in space, above the Earth's atmosphere.

Gaia should also detect a reasonable number of exoplanet transits and microlensing events, although asteroseismic oscillations will lie below its detection threshold. But even within the arena of the classical pulsating stars, eruptive variables, and the ubiquitous eclipsing binary stars, Gaia's contribution is unprecedented.

GAIA SCANS THE SKY in a systematic but non-uniform 'revolving' manner, resulting in around 70 focal plane passages for an average star over 5 years; more at intermediate ecliptic latitudes, with fewer crossings at higher and lower latitudes. As the stars cross the focal plane, the images are effectively integrated, over the 9 successive astrometric CCDs, over about 40 s.

On-ground calibration then leads to (broad G-band) photometric errors, for some two billion stars, of around 0.3 mmag (for $G < 13$), 1 mmag (at $G = 17$), and 6 mmag (at $G = 20$), along with lower accuracy in two (blue and red) photometric channels (van Leeuwen et al., 2017).

Of data releases to date, Gaia DR2 (made public in April 2018) was based on observations between July 2014 and May 2016, while Gaia EDR3 (released in December 2020), included all observations to May 2017.

Through the work of the processing teams, 550 737 stars were classified as variable in Gaia DR2. While no new variability updates were provided with Gaia EDR3, many more variable stars are expected to be identified with the full Gaia DR3 release in mid-2022, and beyond.

WITHIN THE Gaia Data Analysis and Processing Consortium (DPAC), Coordination Unit 7 (CU7), led by Laurent Eyer from the Geneva Observatory, has the responsibility of detecting variable objects, classifying them, and deriving characteristic parameters for specific variability classes.

CU7 interfaces with other Coordination Units, each with specific (and often closely related) responsibilities. Thus, CU5 provides the calibrated photometric data, CU4 processes binary stars, CU6 processes the spectra and derives radial velocities, and CU8 extracts astrophysical parameters from all the available Gaia data.

In parallel to the variability processing by CU7, a Cambridge-led 'alerts team' analyses the Gaia data in near-real time to detect transient phenomena, such as supernovae, that benefit from rapid follow-up by the scientific community (Hodgkin et al., 2021).

THE MAIN ACTIVITIES of the CU7 processing are, sequentially, data import into a dedicated database, cross-matching of sources, execution of the various variability detection and classification modules, followed by export of the data to the other Coordination Units. These activities are performed iteratively, adding more observations, identifying errors, and improving the results.

Source-by-source processing allows for a parallel use of the computing cluster, in Geneva, that is gradually being expanded as the mission and data quantity progresses. In 2017, it was composed of 57 quad core nodes, totalling 456 simultaneous processes, each process having 4 GB of RAM. The incoming data from the Main Data Base of the Data Processing Centre at ESAC, Madrid, is stored in partitioned tables. The database is hosted on a single node with 40 cores and 256 GB of RAM.

THE VARIOUS STEPS in the processing of such a large and complex data set are many and varied (e.g. Eyer et al., 2017; Holl et al., 2018; Eyer et al., 2019). For example, observations tagged with the satellite on-board time are converted to the standard 'Julian epoch' time scale, and then transformed to the solar system barycentre.

The statistical processes of general variability detection aim to separate 'constant' stars from those showing any type of variability, while handling challenges such as the different number of source measurements, and the dependence of measurement error on star brightness.

The task of 'special variability detection' specifically targets short time-scale or small-amplitude periodicity, as well as variability induced by planetary transits and solar-like magnetic activity (e.g. Roelens et al., 2017).

The 'characterisation task' analyses the resulting time series. For periodic objects, variability is characterised using classical Fourier decomposition, with the process comprising various period-search methods.

The 'classification module' uses automated supervised statistical classifiers to assign probabilities that any given star is a member of one of a number of pre-defined variability types. Classification using well-established machine-learning techniques is performed using three different classifiers, known as Gaussian mixtures, Bayesian networks, and random forests.

BEFORE GAIA, the number of known variables stars was around 150 000. DR2, the latest Gaia data release providing the results of the global variability analysis (Holl et al., 2018), contains 550 737 variable sources, comprising 228 904 RR Lyrae stars and 11 438 Cepheids (Clementini et al., 2019; Rimoldini et al., 2019), 151 761 long-period variables (Mowlavi et al., 2018), 147 535 stars showing rotation modulation, 8882 Delta Scuti and SX Phoenicis stars (Rimoldini et al., 2019), and 3018 short-timescale variables (Roelens et al., 2018).

As Holl et al. (2018) conclude, the DR2 release represents just a subset of the data processed to date, and future releases will include many more variable sources. Nonetheless, DR2 already shows the very high quality of the data, and Gaia's great promise for variability studies.

SCIENTIFIC APPLICATION of these results is only just starting. As one example, and at the broadest level, Gaia Collaboration et al. (2019) have located these variables in a Galactic diagram of colour versus absolute magnitude, focusing on pulsating, eruptive, and cataclysmic variables, as well as on stars that show apparent variability due to stellar rotation and binary eclipses.

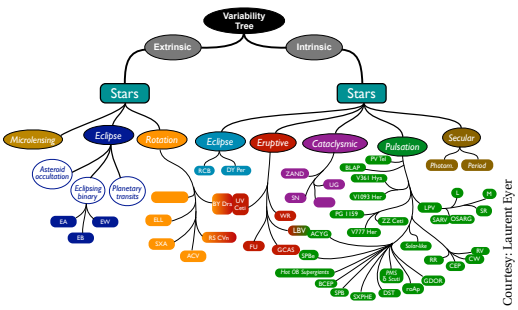
They could characterise the locations of the different variable star classes, variable object fractions, and typical variability amplitudes throughout the diagram. This demonstrates distinct regions in which variable stars occur, anticipating new insights into variability phenomena, and a greater understanding of stellar physics.

62. Variability across the HR Diagram

In my previous essay, #61, I looked at the principles and methods from which huge numbers of variable stars are being discovered and classified as part of Gaia's astrometric and photometric survey of our Galaxy.

These include both 'intrinsic variables', in which the luminosity of the star itself changes with time, as well as 'extrinsic variables', in which apparent variability occurs because something affects the stellar light on the way from the star itself to us on Earth.

Of more than two billion stars being surveyed, Data Release 2 (from just the first two years of the mission, between July 2014 and May 2016, and made public in April 2018), contains 550 737 variable sources classified by the project, comprising 228 904 RR Lyrae stars and 11 438 Cepheids; 151 761 long-period variables; 147 535 stars showing rotation modulation; 8882 Delta Scuti and SX Phoenicis stars; and 3018 short-timescale variables.



Courtesy: Laurent Eyer

A classification tree of variable stars

Most importantly, and in contrast with all other surveys, *every star* will have a measured distance, space motion, and other supplementary data, providing the information required for interpreting variability as a function of stellar luminosity, Galactic population, Galactic location, as well as age and metallicity.

Extrapolating to a possible 10-year mission, the number and variety of variable stars will provide a data set likely to occupy both observers and theoreticians with material for studies of variability, at both an individual and statistical level, for a very long time.

As just one example of why variability results from Gaia are expected to be so far-reaching are the important class of Cepheid variables, more than 11 000 of which have been characterised in Gaia DR2. The physics underlying their regular variability, the κ-mechanism, is reasonably well understood, and it depends on the changing opacity of helium in the outer layers of the star according to its ionisation state.

The strong relationship between their luminosity and pulsation period, discovered by Henrietta Swan Leavitt in 1908, established Cepheids as important indicators of Galactic and extragalactic distances, and remains central in discussions of the Hubble constant today. While uncertainties remain due to the effects of metallicity and interstellar extinction, these and others are expected to be clarified enormously by Gaia.

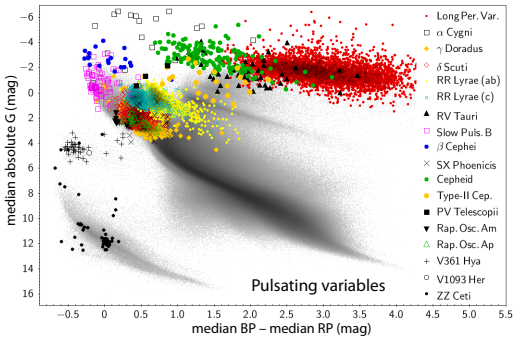
The older, and more common, RR Lyrae variables similarly result from the κ-mechanism, but their period–luminosity relation has a stronger dependence on metallicity. The Gaia distances will allow the more than 220 000 RR Lyrae variables known so far to map out the effects of metallicity throughout our Galaxy.

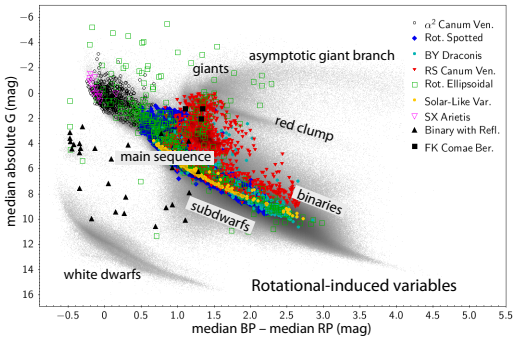
Variable stars have long been recognised as offering deep insights into stellar structure and evolution. Similarly, the Hertzsprung–Russell diagram provides a quantitative visualisation of all stages of stellar evolution. Together with its empirical analogue, the colour–magnitude diagram, it has enabled many advances in stellar astrophysics.

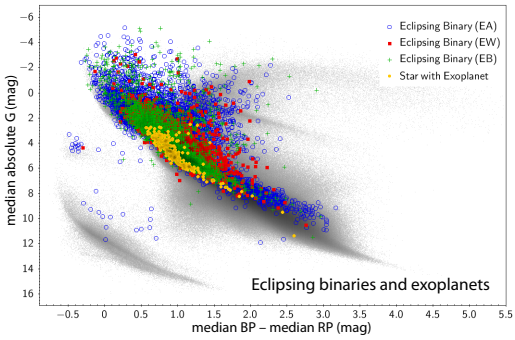
By combining the different types of variability data with the Gaia distances, the colour versus absolute magnitude for variable stars can now be reconstructed in any direction in the Milky Way, and for different populations. This builds on previous work based on the Hipparcos results (Eyer & Grenon, 1997), but on a much vaster scale.

A first investigation has been carried out by Gaia Collaboration et al. (2019), who placed these half a million variable stars in a diagram of colour versus absolute magnitude, focusing on pulsating, eruptive, and cataclysmic variables, as well as on stars that show apparent variability due to stellar rotation and binary eclipses.

In these three figures from their work, the grey background delineates key stellar populations: the main sequence (broadened due to close binary systems), the red clump (and its long tail due to interstellar extinction), the horizontal branch, the red giant and asymptotic giant branches, along with the white dwarfs, the subdwarfs, and the supergiants.



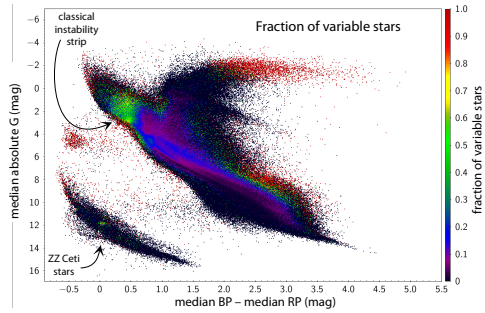




The diagrams are presently uncorrected for interstellar extinction, which blurs the boundaries (and restricts theoretical inferences) between variability classes.

Amongst Gaia objects showing rotational-induced variation are three primary categories: spotted stars, stars deformed by tidal interactions, and objects whose variability is due to light reflected by a companion, viz. binary systems with a strong reflection component in the light curve, in which the hotter component's stellar light is re-radiated from the cooler companion's surface.

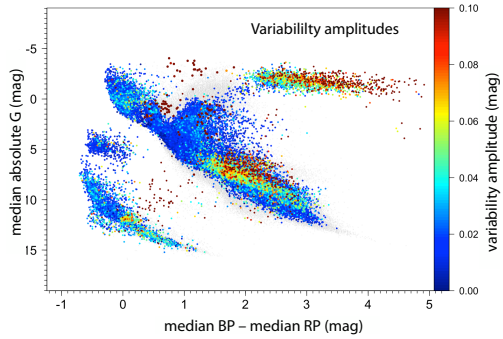
FROM THESE DATA, variable object fractions and typical variability amplitudes can be rigorously estimated throughout the diagram. Thus Eyer et al. used a sample of 13 million stars with heliocentric distances out to 1 kpc, satisfying specific astrometric and photometric criteria, and at a precision of about 5–10 mmag.



They identified variability in 9% of stars, with 50–60% in the classical instability strip being variable. For evolved stars and red giants, higher luminosity and redder colour both imply a higher probability of variability, while the red clump shows a very small fraction of variables.

The classical ZZ Ceti stars (white dwarfs featuring fast non-radial gravity-mode pulsations) are particularly concentrated in magnitude and colour, with variability in about half of the stars. This concentration is attributed to the partial ionisation of hydrogen in the outer envelopes of white dwarfs, which is only developed over a narrow range of temperature (and therefore colour).

The diagram of variability amplitudes shows a number of distinct clumps and instability regions, directly related to the various variability classes.



FINALLY, individual variable stars show both a changing absolute magnitude and a changing colour index throughout their periodic cycle. For example, pulsating stars (including long-period variables, Cepheids and RR Lyrae stars), are all bluer when brighter, showing that their brightness variations are dominated by a change in temperature, rather than in radius.

A movie of the changing loci of representative variable stars across the colour–magnitude diagram can be found at `cosmos.esa.int/web/gaia/gaiadr2_cu7`.

63. Catalogue validation

When making the successive Gaia data releases available to the world-wide scientific community, an important question is to what extent the star positions, distances, proper motions, photometry, and radial velocities can be 'trusted'? What sort of independent validation can be made before publication? Have all instrumental (and relevant astrophysical) effects been correctly accounted for? Have there been any gross errors in some part of the enormous data processing chain?

In an analysis team comprising more than 400 people, with hundreds of thousands of lines of code written and maintained over many years and in many institutes, is the same numerical value of each critical parameter consistently used throughout?

For Hipparcos in the 1980s, the relatively modest data volume and discrete processing steps, combined with the interest of the various contributing countries, permitted a parallel reduction of the entire satellite data, by two independent data analysis teams: FAST, led by Jean Kovalevsky, and NDAC, led initially by Erik Høg and, after launch, by Lennart Lindegren.

But we could not wait to compare the final astrometric parameters to identify any possible algorithmic error. Instead, the processing being comprised of a number of discrete steps, enshrined in the 3-step method, allowed each to be verified, and if necessary corrected.

The two independently produced catalogues were then combined, into a unique agreed published catalogue, through a rigorous weighting procedure developed by Andrew Murray and Frédéric Arenou.

Such a parallel reduction was out of the question for Gaia. The volume of data, its complexity, the manpower and computational resources required for even a single undertaking, ruled this out from the start.

Meanwhile, ensuring that every one of the more than 400 members of the Gaia Data Analysis and Processing Consortium would use the same values for the very broad range of natural physical and instrumental constants presented its own interesting challenge.

As examples, the (assumed) mass of the Sun, its quadrupole moment, the speed of light, and even the value of π (to 13 decimals!), are all crucial at accuracies of a microarcsec. So are many other constants of Nature. Also critical are the satellite spin rate, the CCD pixel sizes, the wavelength response of the G, B_P, and R_P systems, and hundreds of other instrumental constants, some of which may change as the mission progresses.

The solution was a centralised 'parameter data base' (Perryman et al., 2008), comprising many hundreds of relevant constants, expertly maintained by deputy project scientist Jos de Bruijne, along with structures for seamless inclusion into software codes.

Astronomical tests that the Gaia data can be subject to are many. As a simple but forceful example, Data Release 2 has transformed the field of solar system occultations. Pre-Gaia, the predicted location of occultation passages on Earth could be wrong by thousands of km, with timings off by several hours, making observations a hit-or-miss affair. Today, with EDR3 positions and proper motions, predicted timings are typically good to within a few *seconds* (Essay #24).

Other tests can be more subtle. A powerful one, applied to Hipparcos by Lindegren (1995), examines the catalogue's *negative* parallaxes. Given that a star's parallax is a geometrical measure of its distance, it may seem alarming that there should be any! But the explanation is simple: any measured quantity has a measurement error associated with it, formalised as its 'standard error'.

A star at very large distance will have a very small parallax, but the ubiquitous measurement errors may result in it 'appearing' to be negative. Indeed, it would be even more troubling if there were no negative parallaxes in the catalogue! And going further, the distribution of small and negative parallaxes results from the formal standard errors convolved with the true distribution of stellar distances. Equipped with a Galaxy model reflecting realistic star distances, the test proves to be a powerful one not only in verifying the parallax statistics, but also the fidelity of their estimated standard errors.

FOLLOWING ON from the sorts of validations for Gaia DR1 (Arenou et al., 2017), and DR2 (Arenou et al., 2018), an extensive battery of tests was made on the Gaia third data release, EDR3, by Fabricius et al. (2021). In the following, I will give an outline of some of them.

How complete is EDR3? Gaia has a practical limit for the number of simultaneous observations, dictated by the on-board memory, and telemetry capacity. Accordingly, when scanning close to the Galactic plane, and in particular across the low-extinction Baade's Window, some of the fainter observations are not sent to ground. Tests show that incompleteness sets in at around $G \sim 20$, and at around $G \sim 19$ for the crowded Baade's Window.

In globular clusters, calibrated via Hubble Space Telescope observations, completeness is around 60% for $G \sim 19$ mag even when the star density is $\sim 10^5$ stars per sq. deg. It is above 20% at $G \sim 17$ even at 10^7 per sq. deg. Not surprisingly, the inner and outer regions of globular clusters have very different levels of completeness.

FOR HIGH PROPER MOTION stars, EDR3 contains 633 sources with a proper motion above 1 arcsec yr^{-1}, and 2729 above 0.6 arcsec yr^{-1}. Such high proper motion stars are generally 'nearby', and tests have been made to verify that their parallaxes behave accordingly.

Some 8% of such high proper motion stars in the SIMBAD data base are still missing in EDR3, but it is expected that these will be picked up in later releases as the number of distinct observations increases, and as the astrometric solutions improved accordingly.

The limited sky coverage of EDR3 (with observations only between July 2014 and May 2017), combined with perturbations by close neighbours, explains the 1.5% of more than 500 million stars brighter than 19 mag without a full astrometric solution, i.e. these stars still lack an estimated parallax and proper motion.

The overall astrometric quality of EDR3 still shows certain limitations, as expected, including the residual imprints of the inhomogeneous sky coverage, again determined by the satellite scanning law. A number of large negative parallaxes can be traced to spurious solutions, again influenced by nearby stars on the sky, coupled with the still sub-optimal sky coverage.

PARALLAX ESTIMATES for distant quasars have been used to estimate the current large-scale variation of the parallax systematics, while independent catalogues of stars in the Large and Small Magellanic Clouds and the dwarf spheroidal galaxies of the Local Group, as well as catalogues of open clusters, have been used to examine the parallax zero point and associated systematics.

Meanwhile, a comparison of the proper motions from the Hipparcos Catalogue reveal a global rotation of the proper motion system well within the estimated accuracy of the Hipparcos reference frame 'spin'.

OVERALL, comparisons with the Gaia Object Generator simulation tool, or 'GOG', show that the EDR3 data are broadly as expected from our knowledge of Galactic kinematics to very faint magnitudes, with the model itself possibly the cause of residual systematics.

As described in the context of the Hipparcos Catalogue validation outlined above, and in the framework of Gaia DR1 by Arenou et al. (2017, Section 6.2.1), the 'true' dispersion of the parallaxes was estimated using a deconvolution method applied to the negative tail of the parallax distribution. Again, this method provides insights into the underestimation of the standard errors, any residual systematics, and the present contamination due to non-single sources.

As confirmed by other tests, systematic terms have decreased compared to DR2, with remaining issues largely attributable to either shortcomings in calibration, or to perturbations by non-single stars.

Dependencies of the residual parallax bias on magnitude, colour, and ecliptic latitude of the source, at the levels of a few tens of microarcseconds, have been provisionally derived by Lindegren et al. (2021).

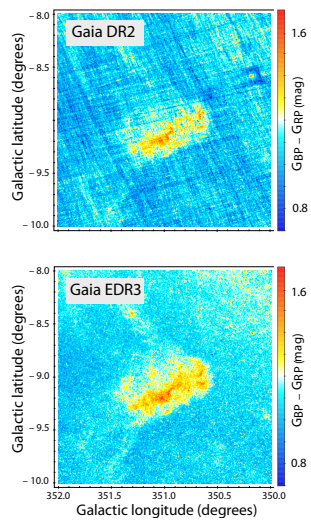
ANOTHER APPROACH to examining more subtle effects, such as the magnitude dependence of the parallax zero point, is through the use of resolved binary stars. For sufficiently long orbital periods, the proper motion of the two binary components should be similar. Potential common proper motion pairs were selected over the entire sky, being restricted to primaries up to $G < 15$ mag and to secondaries up to $G < 18$ mag. Other consistency tests were performed on the parallaxes of the two components of such wide physical binaries.

MANY VALIDATION TESTS were also made on the photometric data by Fabricius et al. (2021).

Most noticeably, artefacts in the colours due to the scan pattern in Gaia DR2, have largely disappeared in EDR3.

Their tests included internal checks of the photometric accuracy, comparisons with external catalogues, such as the Landolt standards, as well as the SDSS and Pan-STARRS1 standards, and comparisons of the $B_P - R_P$ colours with a Milky Way Galaxy model.

Photometry in crowded areas, such as globular clusters, still shows degraded quality, with the photometry in the inner regions shifted in both colour and magnitude.



64. Solar system objects

ALTHOUGH MODEST in number, the few solar system objects measured from space by Hipparcos (between 1989–1993) secured the importance and huge potential of observing solar system objects with Gaia.

For Hipparcos, all objects to be observed had to be selected pre-launch. At each measurement epoch, solar system objects had to be brighter than 12–13 mag, and smaller than 1 arcsec in angular diameter, while planetary satellites also had to be sufficiently far from their host planet to be unaffected by scattered light.

As a result, Hipparcos observed just 48 asteroids, from (1) Ceres to (704) Interamnia, and just three satellites: Jupiter's Europa, and Saturn's Titan and Iapetus. A few others, including Uranus and Neptune, and Jupiter's Ganymede and Callisto, were observed as part of the star-mapper Tycho Catalogue (Hestroffer et al., 1998).

One goal in observing these objects was to provide a dynamical reference frame based on the absence of rotation in their equations of motion. Masses, from orbit perturbations due to close encounters, notably for Massalia–Nysa (Bange, 1998), as well as photometric inferences, such as limb darkening and rotational-induced variability, were also obtained (e.g. Hestroffer, 1998).

A 7-PAGE REVIEW OF Gaia's predicted capabilities, included in the Concept & Technology study (Perryman et al. 2000), demonstrated the enormous scope of Gaia's expected contributions to solar system science.

The on-board detection of all objects brighter than 20–21 mag would yield a deep and uniform detection of minor planets and other solar system bodies (including near-Earth asteroids and Kuiper belt objects), permitting profound studies of their dynamics, structure, and taxonomy. This is important because these minor bodies retain a record of the conditions in the proto-solar nebula, and their properties therefore shed light on the formation and evolution of our solar system.

As well as known objects, Gaia was also predicted to discover upwards of 100 000 new objects, including the difficult inner Trojans, with extremely precise orbits derived from multiple observations over several years.

TO DATE, Gaia Data Release 2 (DR2), made available in April 2018, and based on observations between July 2014 and May 2016, provided limited data on around 14 000 asteroids. Early Data Release 3 (EDR3), provided no further updates on solar system objects.

Gaia DR3 will provide a further major update in mid-2022, including astrometry (and some orbits) for around 150 000 asteroids, along with an unprecedented spectral survey for more than 60 000 objects, with enormous possibilities for taxonomic and dynamical investigations.

WHILE IT IS TOO SOON to have much visibility of the impact that Gaia will have on their understanding, important 'hints' are beginning to appear.

At least for near-Earth asteroids, Gaia is providing positional accuracies close to that from radar ranging. And new alerts are being published as part of a dedicated `follow-up network` of some 100 observers around the world, leading to further improvements in orbits.

This combination of DR2 asteroid orbits, and star positions for more than 2 billion objects, is having a significant impact on stellar occultations, where robotic telescopes and international collaborations are leading to some fascinating progress (essay #60). Amongst these are early results related to the minor planet `Carnegia` in 2016; of the ringed-Centaur `Chariklo` in 2017; and of Neptune's `Triton` in 2018.

Timing uncertainties are now becoming sensitive to knowledge of the object's size, shape, and spin properties, which are, in turn, being improved by the occultations themselves (Ferreira et al., 2020; 2022).

In a preview of the DR3 results, Tanga (2021) has reported the discovery of the Yarkovsky acceleration (due to the anisotropic emission of thermal photons) amongst 20 new discoveries, and the photocentric wobble of 4337 Arecibo, the first asteroid for which a satellite has been discovered by occultations.

And Delbo et al. (2021) have shown that reflectance spectra, from B_P and R_P photometry, will probe the composition gradient of the main belt even down to `small object sizes`.

NOTWITHSTANDING their scientific importance, solar system bodies present specific challenges for the data processing, mainly because of their significant proper motions. In the remainder of this essay I'll look at how these tasks are organised within the overall Gaia Data Processing and Analysis Consortium.

The processing of satellite data leading to the successive data releases is organised broadly as a 'wheel and spoke' system: the satellite data arrives at the ESAC centre outside Madrid, where source matching, and the astrometric global iterative solution (AGIS), are executed.

A number of other 'coordination units' take these data, and subject them to various specialised tasks, generally conducted at other processing centres across Europe. Their results are returned to ESAC, and all processes are iterated, making use of the intermediate outputs of the other processing tasks as appropriate, including photometric calibration and variability analysis.

Within this framework, Coordination Unit 4 (CU4) treats three specific objects types: binary and multiple systems including exoplanets (originally led by Dimitri Pourbaix[‡], Bruxelles, but since 2021 by Frédéric Arenou, Paris–Meudon); extended objects (Christine Ducourant, Bordeaux), and solar system objects (Paolo Tanga, Nice).

The solar system objects group comprises about 25 people (mainly from Paris, Helsinki, and Bruxelles), with the computations undertaken by a team from the French national space agency (CNES, Toulouse). The paper detailing the treatment of solar system objects for DR2 had 450 authors (Gaia Collaboration et al., 2018), illustrating the extensive interconnectivity of the data analysis.

PRESENT PREDICTIONS are that, by mission end, Gaia will have observed 350 000 solar system objects at multiple epochs, many of them new. For Gaia DR2, a subset of 14 099 known objects, with known orbits, were extracted from the data stream for publication.

This involved nearly 300 000 focal-plane transits, and nearly 2 million individual CCD observations, observed over the first 22 months of the mission. Because of the scanning law, solar system objects are only observed at solar elongations between $45 - 135°$. Observations are otherwise well-distributed on the sky.

As well as having reasonably well-known orbits (facilitating object matching), these 'showcase' objects were selected to represent most main dynamical classes, including main-belt asteroids (MBA), near-Earth objects (NEO), Jupiter Trojans, and a number of trans-Neptunian objects (TNO).

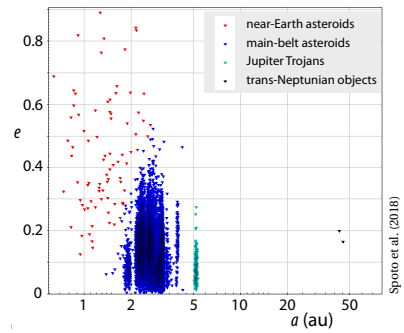
The data processing for solar system objects mostly follows the procedures in place for star observations: positional coordinates in the instrument's focal plane are converted into sky coordinates by using the transformations provided by AGIS, the astrometric global iterative solution, along with the corresponding calibrations.

COORDINATE TRANSFORMATIONS include proceeding from a non-rotating reference system (with its origin in the satellite's centre of mass) to the barycentric reference system (with its origin at the solar system barycentre), including the rigorous effects of general relativistic light bending by the Sun.

Resulting astrometry yields a single-transit accuracies below 1 milli-arcsec for objects with $G < 17.5$, and still better than 2–5 milli-arcsec for the faintest observable ($G \sim 20.5$). Although this accuracy is essentially in one-dimension (in the direction of the scan motion), the 'imprint' of the scanning law largely vanishes when several transits, in different scan directions, are combined.

Photometry for all objects, stars as well as solar system objects, is provided at focal-plane transit level by the photometric processing coordination unit. A specific complication for solar system objects is due to their large proper motion: within a focal plane transit, exposure times (4.4 s) are determined by the crossing of a single CCD. An object with an apparent motion (in the along-scan direction) of merely 13.6 milli-arcsec/s moves by one pixel during a single CCD crossing, leading to noticeable image smearing.

Orbit determination then uses a least-squares adjustment, starting with the previously known orbit elements, and employing the positions of the major solar system bodies from the JPL DE431 planetary ephemerides, to take account of gravitational terms due to the Sun, the major planets, the Moon, and Pluto, and a further sixteen massive main-belt asteroids.



THE RESULTING DISTRIBUTION of orbital eccentricity (e) versus semi-major axis (a) for the 14 099 solar system objects included in Gaia DR2, shown here, clearly distinguishes the near-Earth asteroid population from the main-belt asteroids (including the Jupiter Trojans), and the trans-Neptunian objects.

Extending to simulations representing longer data sets, Gaia Collaboration et al. (2018) have shown that Gaia should eventually lead to an improvement (with respect to the pre-Gaia orbits) of nearly a factor 10 in the orbital elements of typical main-belt asteroids for a 5-year mission, to around a factor 20 for a 10-year mission.

With this vast network of astrometric measurements of unprecedented accuracy, the Gaia orbit accuracies will, in turn, far exceed those obtained even after decades of orbit-tracking from the ground.

65. Hot Jupiters and star clustering

THE DISCOVERY OF EXOPLANETS, planets orbiting stars beyond our own solar system, confirmed in 1995, catalysed a new research field in astronomy, viz. the observation, and the theoretical and computational interpretation, of the formation and evolution of exoplanets.

Amongst many pioneering observational initiatives, and thousands of research papers, the launch of NASA's Kepler space telescope in 2009 arguably marked the next revolution in the field. In more than nine years of operation, Kepler discovered close to 3000 planets, a rich variety of planetary systems and system architectures, providing the foundation for thousands of scientific papers related to their discovery and characterisation.

Today, Gaia is contributing a major observational advance in the understanding of our Galaxy and, specifically in this context, of exoplanet systems. By determining the distances and space motions of their host stars, Gaia is precisely characterising their host star luminosities, their location in the observational Hertzsprung–Russell diagram and, via their inferred radii, the physical radii of the transiting planets themselves.

There is also the anticipated discovery, directly with Gaia, although still perhaps some years hence, of thousands of systems from their astrometric motions.

ONE OF MANY AREAS in which Gaia is contributing, and which nicely illustrates the complex foundations upon which our theories depend, is that of the origin of the perplexing class of exoplanet known as 'hot Jupiters'. These are massive gas giant planets, somewhat like Jupiter or Saturn, but which orbit surprisingly close to their host star, typically within about 0.1 astronomical units (Mercury orbits at just under 0.4 au), corresponding to extremely short periods of around 3–9 days.

In situ formation of these hot Jupiters seems implausible: in models of core accretion, both the available protoplanetary disk mass, and the opening of disk gaps, appear to preclude their growth, while in models of direct gravitational collapse due to disk instability, the gas cools too slowly for the resulting spiral structures to fragment into bound clumps.

These sorts of considerations support the idea that hot Jupiters must have arrived at their present locations from their region of formation beyond the snow line (beyond around 3 au), where solid material is inferred to be abundant in the form of condensed ices.

Assuming that they formed at much larger distances from the host star, there are then two main hypotheses as to how they arrived at their present locations: either via disk-driven inward migration, or through the generation of a highly eccentric orbit with small pericentre, followed by its tidal damping and orbital circularisation.

And there are at least three plausible mechanisms which might generate such high-eccentricity orbits: planet–planet encounters leading to orbit 'scattering', resonant orbit changes due to a distant stellar companion (the 'Lidov–Kozai mechanism'), or as a result of close stellar encounters, perhaps in a young star cluster.

This is not the place to describe these processes, and the evidence for them, in detail. Here, I will look exclusively at how Gaia is helping to clarify the role of close stellar encounters in the possible formation of these strange giant planets, so close to their host stars.

IT HAS BEEN DIFFICULT to argue convincingly that the existence of at least some of these hot Jupiters has been affected by close stellar encounters, in part because stellar groups disperse, at least in 3d space, within less than a billion years, much shorter than the ages of most known exoplanets.

An early approach to this problem, using the Gaia DR2 data, was by Winter et al. (2020). They identified old co-moving stellar groups around exoplanet host stars, and argued that system architectures have a strong dependence on local stellar clustering in position–velocity 'phase space', viz. while individual stars born in an open star cluster may long since have 'disappeared' into the field star population, evidence for their common origin may still be preserved in their space motions.

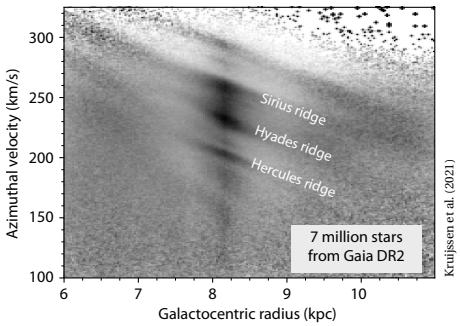
Indeed, propagating the space motions of some 'runaway' stars backwards in time, can often point back to their origin in some ancient star cluster.

After accounting for the star's age, mass, and metallicity, Winter et al. found significant differences in properties between exoplanet systems residing in phase-space density enhancements, and those in the field population. Median semi-major axes and orbit periods in the overdense regions are 0.09 au and 9.6 d, compared to 0.81 au and 154 d for planets around field stars.

They found that 'hot Jupiters' preferentially exist in phase-space overdensity regions, suggesting that their extremely 'tight' (short-period) orbits arise from environmental perturbations, rather than from inward migration or planet–planet scattering, i.e. star clustering is a key factor in determining the architectures, and indeed the atmospheric properties, of exoplanet systems.

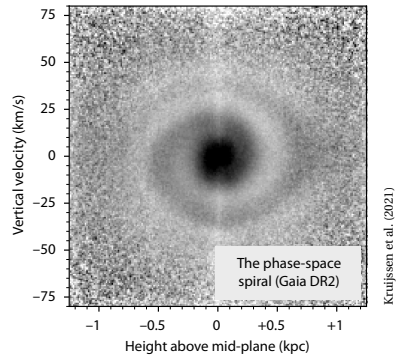
WINTER ET AL. were careful to point out that a similar sort of 'velocity clustering' can also be generated later in the planetary system's lifetime by other dynamical features within our Galaxy.

Many recent results, starting with the release of Gaia DR2 in 2018, have confirmed that the phase-space distribution of disk stars (i.e. their dependence on position *and* velocity) shows considerable sub-structure, with waves, ripples, and streams. They are generated by resonances and instabilities driven by our Galaxy's central bar and by its spiral arms, or by the passages of satellite galaxies through the Galactic disk which can drive spiral and vertical perturbations. Within these structures, star velocities are correlated over distances of several kpc.



Kruijssen et al. (2021) used the Gaia DR2 data (including the radial velocities to yield the full space motions) to show that at least some of their inferred overdensities correspond to such kpc-scale 'ripples' and 'streams' in the Galactic disk.

Another recent and relevant Gaia discovery is the so-called 'phase-space spiral', a prominent feature in the space spanned by the height above the Galactic mid-plane and the star's vertical velocity. This has been attributed to a recent crossing of our Galaxy's disk by the Sagittarius dwarf galaxy. Kruijssen et al. (2021) found that exoplanet systems associated with this phase-space spiral have a ratio of hot Jupiters to cold Jupiters that is 10 times higher than in field systems.



Another of their findings was that the ratio of hot Jupiters to cold Jupiters within any of these overdensity regions appears to increase with stellar age over timescales of billions of years, consistent with the lifetimes of these Galactic-driven overdensities. In turn, this suggests that planetary system properties are not only affected by stellar clustering in their immediate surroundings, but by processes on a Galactic scale throughout their evolution.

More details of the perturbation models are given by, e.g. Rodet et al. (2021) and Rodet & Lai (2022).

As Kruijssen et al. (2021) conclude, these new Gaia results imply that planetary systems are not only affected by stellar clustering in their immediate surroundings, but by processes on a Galactic scale throughout their evolution. It remains as a challenge for the future to establish a complete census of the physical processes at work, which will require combining studies of planet formation and evolution, star (cluster) formation, Galactic dynamics, and galaxy formation and evolution.

FURTHER PIECES of the puzzle as to how exoplanet system architectures are related to a system's passage through the Galaxy have been explored by Dai et al. (2021). They used the Gaia and Kepler data to divide main-sequence star systems into three groups according to their velocities compared to the local standard.

They found that high-velocity planet hosts (with local velocities above about 50 km s^{-1}) are associated with a higher fraction of multi-planet systems (and of lower than average eccentricity), consistent with the higher velocity stars actually experiencing fewer gravitationally perturbing events as they moved through the Galaxy. Low-velocity stars (with space velocities below about 20–30 km s^{-1}) appear to be more strongly perturbed by the effects of stellar clustering.

IN THIS FAST-MOVING FIELD, the big picture being consolidated by Gaia is of a highly complex Galaxy, in terms of both its spatial and velocity sub-structures, whose characterisation will allow a more detailed picture of exoplanet systems, their formation and evolution, and how they have been modified during their billion year journeys through our Milky Way Galaxy.

66. Exoplanet habitability: TESS and Gaia

SINCE THE FIRST confirmed discoveries in 1995, more than 4000 exoplanets (planets around stars other than our Sun) have been discovered. Planets, and indeed planetary systems, have been detected from the radial velocities of their host stars, with others from precise orbital timing or through gravitational lensing.

But the most prolific discovery method has been based on transit measurements, viz. measuring the temporary drop in the light from the star as a planet passes in front of its host star as seen from Earth. And of the more than 3000 transiting planets now known, the majority were discovered using NASA's Kepler satellite, launched in 2009 and operated until 2018.

Kepler focused on one region of the sky, $10^\circ \times 10^\circ$ in area, and discovered planetary systems covering a large range of star and planet masses and orbital periods.

A SMALL FRACTION OF THIS remarkable haul are planets of about one to a few times the mass of the Earth, variously described as 'terrestrial', 'telluric', or 'rocky'. Detailed formation models suggest that these will be composed primarily of silicate rocks or metals, i.e. Earth-like in terms of their composition. Within our solar system, the terrestrial planets are the four closest to the Sun, i.e. Mercury, Venus, Earth, and Mars.

Unlike the larger gas and ice giants (like Jupiter and Uranus), terrestrial planets have a solid surface, and are expected to comprise a central metallic core, mostly iron, with a surrounding silicate mantle, and with typical bulk densities around 5–8 $\mathrm{g\,cm^{-3}}$. They may have 'secondary' atmospheres, generated through outgassing after formation, or from cometary impacts (in contrast to the giant planets, whose 'primary' atmospheres were captured directly from the original stellar nebula).

Kepler discovered several terrestrial-type planets, amongst them bodies orbiting Kepler–20, Kepler–42, and Kepler–70. But all of these, and most of the other Kepler discoveries, are in very short-period orbits around their host stars, typically with periods of less than a few days. Their scorching temperatures are considered to be totally unsuited for the development of life.

But the discovery and characterisation of exoplanets has progressed to the point where terrestrial planets are being discovered in reasonable numbers, some of which lie at star–planet separations which place them within their star's 'habitable zone'. And there is good reason to believe that such planets exist in very large numbers.

EARLY CONCEPTS of a 'habitable zone' around a star date back to Richard Bentley (1662–1742), who *'praised God's wisdom in placing our planet at just such a distance from the sun that we are neither too cold nor too hot'*. Fast forward to today, and our current assessment of the suitability of a planet for supporting life is still largely based on our knowledge of life on Earth.

With the general consensus among biologists that carbon-based life requires water for its self-sustaining chemical reactions, the search for habitable planets has focused on identifying environments in which liquid water is stable, preferably over billions of years. The presence of liquid water in turn implies that the atmospheric pressure and ambient temperature will be in ranges that promote a rich variety of organic reactions.

As of early 2022, and taking into account these and other considerations, the `Habitable Exoplanets Catalogue` lists over 20 'potentially' habitable planets, amongst them Kepler–1649, Proxima Cen, and four planets in the 7-planet system TRAPPIST–1.

IN THE SEARCH for habitable planets, and life on other worlds, two other space missions have recently joined the astronomical arsenal: as the successor to Kepler, NASA launched its Transiting Exoplanet Survey Satellite (TESS) in April 2018. It is conducting an all-sky transit search, over an area 400 times larger than covered by Kepler, over a two-year period. It aims to survey some 200 000 solar-type stars brighter than 12 mag, hoping to discover at least 50 Earth-sized planets in the process.

And the James Webb Space Telescope, launched on 25 December 2021, targets (amongst other goals) the detailed characterisation of the atmospheres of various potentially habitable exoplanets.

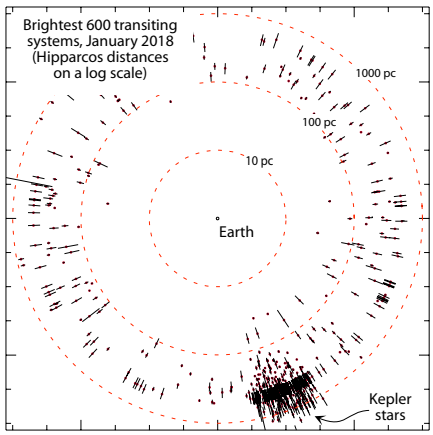
As of early 2022, TESS had discovered around 200 confirmed exoplanets, and some 5000 other candidates (`NASA Exoplanet Archive`). Its first, orbiting π Men, was announced in September 2018. In April 2019 came its first discovery of an Earth-sized planet, HD 21749 c, with an orbital period of about 8 days.

While TESS has observed only a fraction of the stars on its target list for long enough to be able to find Earth-mass planets orbiting Sun-like stars with orbital periods of about 1 Earth year, planets in *shorter period* orbits around *lower luminosity* stars can receive similar stellar irradiation as Earth.

It is this stellar flux, and the planet's atmosphere, which mainly determine the planet's surface temperature and therefore its possibility of maintaining liquid water. However, more detailed discussions of habitability also consider other factors, amongst them the star's spectral type, its metallicity and its ultraviolet flux, the effect of stellar flares, and tidal heating and tidal locking.

Meanwhile, it is a general principle of transit discoveries that the host star must normally be observed for a period long enough to see at least *three* transits of a candidate planet, allowing an orbit period to be estimated based on two of them, and confirmed by the third.

In this spirit, Kaltenegger et al. (2021) are maintaining the TESS *Habitable Zone Catalogue* which uses the first two years of observations to identify stars that have been observed for long enough to detect *three* transits of an exoplanet that receives similar irradiation as Earth. They used distances from Gaia DR2, along with broadband magnitudes and empirical relations, to determine the stellar luminosities, temperatures, radii, and masses.



Illustrating the importance of accurate distances, this shows the 600 brightest transiting systems known in 2018, pre-TESS (according to right ascension), with error bars indicating the uncertainty of their distances based on the 1 milli-arcsecond astrometry from Hipparcos.

The Kepler stars form the cluster of points at the lower right, at a few hundred pc distance. With the new Gaia data, distance errors are invisible on such a plot.

Their present catalogue contains 4239 stars within 210 pc that receive similar irradiation to Earth, of which 738 are within 30 pc. While only nine are currently known to be exoplanet hosts, there are a total of 614 stars for which TESS has observed the system long enough to be able to observe planets throughout the full temperate, habitable zone out to the equivalent of Mars' orbit.

These stars, pin-pointed in space and characterised by Gaia are, they argue, amongst the best targets for discovering habitable planets from the TESS data.

From which star system would an alien observer be able to search for life on Earth?

In the same way that an exoplanet orbit must be aligned with our line-of-sight for us to observe a transit, only a subset of favourably located stars in our region of the Galaxy could see Earth as an exoplanet, transiting slowly across the face of our Sun. Our Earth's 'transit zone' is simply a band around the Earth's ecliptic plane, projected onto the celestial sphere.

This question was considered in some detail by Heller & Pudritz (2016), who suggested that any of the 10 000 G or K dwarf stars in this zone could identify Earth as habitable (and inhabited!) using the transit method.

Wells et al. (2018) found that the maximum number of transiting solar system planets that could be observed from any particular point in the sky is three. And they estimated that there are perhaps three *known* temperate Earth-sized planets orbiting G or K stars brighter than 13 mag that are located in our Earth's transit zone.

Kaltenegger & Pepper (2020) used the combination of TESS and Gaia DR2 to identify 1004 main-sequence stars within 100 pc (amongst them 77% M-type, 12% K-type, and 6% G-type), of which more than 500 are in a position to observe a minimum 10-hr long observation of Earth's transit across the face of our Sun.

And as part of its extended mission, TESS will also search for transiting planets in the ecliptic region that could already have found life on our transiting Earth!

Breakthrough Listen, as one part of the 'Breakthrough Initiatives', is a $100 million program of astronomical observations and analysis, launched in 2015 and funded by the foundation established by Yuri and Julia Milner, describing itself as *'the most comprehensive ever undertaken in search of evidence of technological civilisations in the Universe'*.

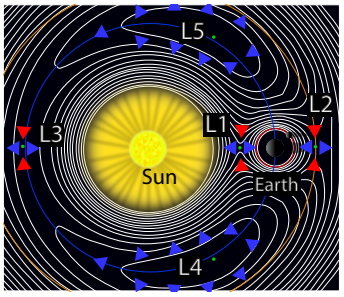
In October 2019, `Breakthrough Listen` began a collaboration with TESS scientists to look for signs of advanced extraterrestrial life. This includes expanding Listen's original target list (adding more than a thousand 'objects of interest' identified by TESS, and aided by Gaia), and refining Listen's analysis strategy, for example, by including knowledge of planet alignments to predict when transmissions are more likely to occur.

67. Where is Gaia?

GAIA orbits around the Sun–Earth Lagrange point L2, about 1.5 million km from Earth. In this essay, I will outline what L2 is, why it's special, and why a number of science missions observing deep space (including the James Webb Space Telescope) are positioned there.

I will also describe the substantial effort that is going on, within the Gaia Data Processing and Analysis Consortium, to track with high accuracy its position and its motion around this Lagrange point.

THE LAGRANGE POINTS are the location of gravitational equilibrium for small-mass objects (in our case, Gaia) under the influence of two massive bodies (in our case, the Sun and Earth). At the Lagrange points, the gravitational forces of the two large bodies and the centrifugal force balance each other, leaving small objects placed at the Lagrange points in equilibrium.



wikimedia (CC BY 3.0), Xander89

There are five such points, known as L1 to L5, all in the orbit plane of any two large orbiting bodies, for example the Sun–Earth system (also in the Earth–Moon system). L1–L3 are on the line through the centres of the two large bodies, and were discovered by Leonhard Euler in 1765.

L4 and L5 form equilateral triangles with the centres of the two large bodies, and were discovered by Joseph-Louis Lagrange, 7 years later, in 1882. Many of our solar system planets have 'trojan' asteroids gathered around their L4 and L5 points, with Jupiter hosting thousands!

Strictly, the colinear Lagrange points are in 'unstable' equilibrium, and satellites will slowly drift away from them. Detailed mathematical treatment shows that there are nevertheless families of quasi-periodic orbits around L2: 'halo' type, and 'Lissajous' type. The details need not concern us here, other than to note that small orbit adjustments are occasionally required to keep a satellite at L2 in its intended location.

VARIOUS SATELLITES have been placed at the L1 and L2 points of the Sun–Earth system, L1 being attractive when observing the Sun (e.g. ISEE–3 in the 1970s, SOHO in the 1990s), and L2 for observing deep space.

L2, of relevance for Gaia, lies on the line through the Sun and Earth, about 1.5 million km beyond the Earth (and Moon). Here, the gravitational forces of the Sun and Earth balance the centrifugal force on a body at L2, whose orbital period then matches the Earth's.

The L2 orbit offers important advantages: amongst them, it provides a stable thermal environment, stable perturbing torques, and observations free from Earth or Moon occultations. Only small and infrequent adjustments are needed to maintain the satellite's location.

The first mission to exploit L2 was NASA's WMAP (microwave background) in 2001. It was followed by NASA's WIND (2003), ESA's Herschel and Planck observatories (2009), and China's Chang'e 2 (2011). Satellites currently stationed there include Gaia (since 2013), and JWST (since 2022). Others are planned, including ESA's Euclid, PLATO, and ARIEL, as well as NASA's WFIRST.

THIS ORBIT for Gaia was optimised and selected as part of the Concept & Technology Study (ESA 2000), the choice influencing many aspects of the satellite design. Along with numerous advantages, summarised in the table, L2 was fully compatible with the baseline Ariane 5 (and actual Soyuz) launch. Its major disadvantage is for communications, complicating both ground-station visibility and telemetry rate considerations.

Parameter	Low Earth	L4/L5	L2	Geo
Thermal environment	–	*	***	–
Radiation environment	**	***	***	–
Occultations	–	*	***	*
Eclipse avoidance	–	**	***	*
Dynamic environment	–	**	***	**
Injection Δv	***	***	***	**
Maintenance Δv	–	***	***	***
Communications	**	*	**	***
Launch mass	***	**	***	*
Transfer duration	***	**	*	***
Operations	**	**	**	***
Ground facilities	***	***	***	***
Lifetime	*	***	***	*
Overall complexity/cost	–	*	***	*
Recommendation	reject	reject	baseline	reject

DELIVERED TO its L2 location, a 6 month journey after its launch in December 2013, and slowed to its final location by on-board thrusters, the satellite's position and velocity have to be monitored from the ground, so that it can be 'nudged' occasionally, again by its on-board thrusters, so as to remain more-or-less at this L2 location. This orbital adjustment is a common requirement for all satellites operating at the Lagrange points.

But Gaia has another very specific (and particularly demanding) requirement concerning our detailed knowledge of its position, and its velocity, in space.

DETERMINATION of stellar distances with Gaia (based on Earth's orbital motion around the Sun) must more strictly be derived from the *satellite's* distance from the Sun. The latter can be broken down into the Sun–Earth and Earth–satellite displacement vectors.

It turns out that the Earth's orbit is known well enough to account for the former, while the latter must be known to an accuracy of better than about 150 km. But when considering the observations of solar system objects with Gaia, e.g. the main belt asteroids, the satellite position must be known to better than about 150 m!

Similar considerations apply to knowledge of Gaia's velocity with respect to the solar system barycentre (centre of mass), necessary to account for the effect of stellar aberration due to the observer's motion combined with the velocity of light. A full (general relativistic) treatment shows that Gaia's barycentric velocity must be known, at all times, to a demanding 1.5–2.5 mm per sec.

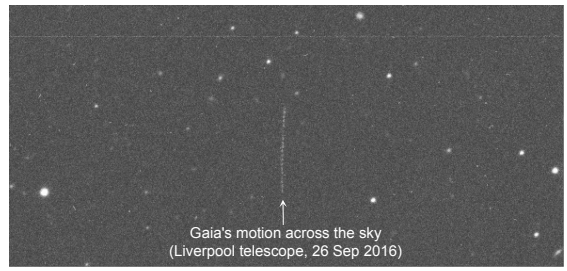
Neither of these exacting requirements, on the knowledge of Gaia's instantaneous position and velocity, can be achieved by the more standard techniques of satellite radar ranging or Doppler measurements.

The more accurate Delta Differential One-way Ranging (Delta-DOR) can contribute, and is used regularly by the Gaia operations team at ESOC (Darmstadt). In this technique, radio signals from the spacecraft are received by two widely separated deep-space optimised ground stations on Earth, and the difference in the signal arrival times is measured.

The timing differences must be adjusted to account for varying delays due to the Earth's atmosphere. This is in turn obtained by simultaneously tracking the signals from radio-emitting quasars nearby on the sky.

But the method requires two ground stations (and competition for their use is intense), and data transmission from Gaia must be suspended during execution.

THE ACCURATE RECONSTRUCTION of Gaia's orbit is therefore supplemented by a dedicated 'ground-based optical tracking' campaign (GBOT), led by Martin Altmann of the Centre for Astronomy (ZAH/ARI) of Heidelberg University. The group comprises a dozen people from Heidelberg and the Paris Observatory, the latter hosting the GBOT data centre.



Altmann (2018)

The group coordinates and reduces astrometric data obtained by tracking the satellite (at an accuracy of about 20 milli-arcsec with respect to the International Celestial Reference Frame) with a number of 2-m class optical telescopes (Altmann, 2018; Bouquillon et al., 2018).

The telescopes currently used are the 2.6-m ESO VLT Survey Telescope (VST) on Cerro Paranal in Chile, the 2-m Liverpool Telescope at the Roque de los Muchachos Observatory on La Palma, and telescopes of the Las Cumbres Observatory in Hawaii and Australia.

The task is a challenging one: Gaia is faint (around 21 mag), and relatively fast moving, with an apparent angular motion on the sky of about 0.04 arcsec per second. This has required carefully maximising the resulting positional accuracy Bouquillon et al. (2017).

In practice, the flight dynamics team at ESOC provides weekly updates of Gaia's geocentric ephemerides. The GBOT group uses these to provide sky charts of Gaia for each telescope location. The telescope teams use these to track the satellite via 10–20 sky images. These are processed by the GBOT group which returns all measurements to ESOC to create an improved Gaia orbit.

The final result is that ESOC creates two orbit files: ORB1 being a weekly update based on ESOC's own Delta-DOR tracking, and ORB2 being a more-or-less annual improved solution including both the Delta-DOR and GBOT measurements, which is then used for the subsequent Astrometric Global Iterative Solution.

THERE IS AN interesting and valuable by-product of the GBOT observations. Since the observations are made near to the ecliptic plane, and essentially at opposition (i.e. directly 'away' from the Sun, being a direction not observable by Gaia), main belt asteroids are frequent interlopers in the telescope's field of view. They are also at their brightest, which facilitates their detection, as well determination of their absolute magnitude.

Consequently, a spin-off GBOT 'asteroid discovery programme' has been active since the beginning of 2015. All their discoveries are sent to the `Minor Planet Center` for collation, and thereafter for further follow-up by observers world-wide (e.g. Carry et al., 2021).

The daily tally from VST alone amounts to some 10–80 objects, around half of them new. As of January 2022, the `GBOT project` reports that a total of 43 670 asteroids have been observed, 18 532 of them new.

68. Gaia photometry

WHILE GAIA'S EMPHATIC advance is its astrometry, the parallel acquisition of high-accuracy multi-epoch photometry for all stars is an enormously powerful complement. A survey of more than a billion stars, establishing how far they are away, and how they are moving through space, is of course a major advance. But without appropriate *photometric* information, there would be little indication of what these stars actually *are*.

In making the case for an expensive space mission, scientists cannot expect to execute from space what can be done from the ground. And multi-colour photometry *can* be acquired from the ground.

But there were compelling reasons to do this onboard. One was instrumental: significant chromatic displacements, due to asymmetric wavefront aberrations, can only be corrected if the star's colour is known.

From a scientific perspective, in a Universe in which variable stars abound, an unbiased astrometric survey must be based on stars selected according to their magnitude at each measurement epoch. Of similar importance, for statistical considerations, is that object selection should also be based on detection and thresholding at the same angular resolution as the astrometric measurements themselves.

The case for onboard photometry is compounded for transient sources such as supernovae, as well as for solar system objects moving rapidly across the sky.

MUCH EFFORT was devoted to such photometric optimisation. The two-colour photometry of Hipparcos could only hint as to each object's nature. Better data could distinguish spectral type and luminosity class, and stellar metallicity and interstellar extinction, all contributing to a more informed survey of our Galaxy.

Implementation wasn't trivial: using filter photometry, how many filters could be accommodated, and how could they be organised in the focal plane? Would every star have to be measured at each epoch in every filter? How should the fixed transit time at the focal-plane be divided up between high-accuracy (unfiltered) astrometry, and the diagnostic multi-colour photometry?

THE SCIENTIFIC BREADTH of Gaia is enormous, and multi-colour diagnostic photometry had to be optimised accordingly. With more than a billion parallaxes, Gaia is covering all phases of stellar evolution across the full range of stellar masses and populations, including pre-main-sequence stars, chemically peculiar stars, and a wide range of variable stars observed some 70 times during the nominal five-year operations phase.

Multi-colour photometry provides an important diagnostic for eclipsing binaries. And for the many tens of thousands of solar system objects, multi-colour photometry provides a decisive tool for taxonomy.

BOTH BEFORE AND AFTER mission selection by ESA in 2000, various groups devoted to specific aspects, worked together under the coordinating structure of the Gaia Science Team. Of relevance here was the Photometry Working Group, jointly led by Erik Høg (Copenhagen) and Carme Jordi (Barcelona). Many others contributed their expertise, their time, and their enthusiasm.

Amongst these were Jens Knude (Copenhagen), bringing his lifelong experience in photometry, especially concerning the intermediate-band Strömgren *uvby* system, particularly well-suited to early-type stars.

Vytas Straizys, one-time director of the Vilnius Observatory in Lithuania, and a number of his colleagues, brought their experience of the Stromvil system, optimised for robust classification across all spectral types, even in the presence of interstellar reddening.

Many meetings, and many internal reports, followed over several years. In an early design, considered by Matra Marconi Space at the end of the first industrial study in mid-1998, photometry in four broad bands was obtained from the main astrometric telescopes, along with seven medium-width spectral bands in a third, smaller telescope, with a collecting aperture of $0.75 \times 0.70\,\mathrm{m}^2$.

The next major design iteration, described in depth in the Concept and Technology Study Report in 2000, increased this to 11 intermediate bands. In 2006, this increased again to a total of 19 bands, the details described in a 35-author paper by Jordi et al. (2006).

BUT THINGS CHANGED DRAMATICALLY while the paper was being finalised, with a novel design remaining largely confidential because it was developing during the competitive bid for the industrial contract. Matra Marconi Space (who were eventually assigned the industrial contract), proposed a major overhaul of the design for reasons of operational simplicity, mass and cost.

They proposed to drop the dedicated photometric telescope, and to replace the photometric filters with two low-dispersion prisms. With the detailed involvement and extensive simulations led by Erik Høg, Carme Jordi, Anthony Brown, and Lennart Lindegren, a careful optimisation of the prism properties eventually resulted in the Gaia photometry that we have today.

The resulting focal plane assembly is common to both telescopes, and has five main functions: (a) object detection in the sky mapper; (b) astrometry in the astrometric field; (c) low-resolution spectrophotometry using the blue and red photometers, BP and RP; (d) spectroscopy using the radial-velocity spectrometer; and (e) supporting metrology, comprising both wave-front sensing and basic angle monitoring.

The primary astrometric instrument defines the unfiltered, white-light photometric G band, extending over the wavelength range 330–1050 nm. This wide-band photometric data has the highest signal-to-noise, and is therefore most relevant for, e.g., variability studies.

MULTI-COLOUR PHOTOMETRY is finally achieved by two fused-silica prisms, dispersing light entering the fields of view (Gaia Collaboration et al., 2016). The bandpasses are defined by the optical coatings deposited on the prisms, together with the telescope transmission and detector quantum efficiency.

The prisms, located in the common path of the two telescopes, are mounted on a CCD cold radiator, directly in front of the focal plane. Each are equipped with a strip of seven CCDs, which cover the full astrometric field of view in the across-scan direction. The photometers thus sample the same transits as the astrometric instrument.

The prisms disperse object images in the along-scan direction, spreading them over ∼45 pixels (at 15 mag), the along-scan window size of 60 pixels allowing for sky background subtraction. The spectral dispersion, which matched the earlier photometric-filter design of Jordi et al. (2006), results from the intrinsic dispersion of fused silica, and varies from 3–27 nm per pixel over the range 330–680 nm for BP, and from 7–15 nm per pixel over the range 640–1050 nm for RP.

For the great majority of objects, the BP and RP spectra are binned on-chip in the across-scan direction, over 12 pixels, to form one-dimensional, along-scan spectra. Unbinned, single-pixel-resolution windows (of size 60×12 pixels^2) are only used for stars brighter than $G = 11.5$ mag.

THE DISPERSED SPECTRA overlap in crowded regions. Compared with the astrometric crowding limit of around a million objects per square degree, BP/RP photometry is typically limited to about 750 000 objects per square degree. Above that, window overlap and truncation occurs and, because the onboard detection prioritises bright stars, faint-star completeness is impacted.

Methods extending this to higher density areas of particular importance, in particular Baade's Window in the Galactic centre, and the globular cluster ω Cen, have been developed (Gaia Collaboration et al., 2016).

ON GROUND, the photometric pipeline processes the data from the sky mappers, the astrometric field, and the BP and RP photometers. The source fluxes in the G band (from the astrometric field images) are turned into calibrated epoch photometry in the G band.

Integrated fluxes from the BP and RP spectra provide calibrated epoch photometry in $G_{\rm BP}$ and $G_{\rm RP}$, while the spectra themselves are wavelength and flux calibrated.

The photometric and spectroscopic data are calibrated by means of standard stars for which high-quality, ground-based spectrophotometry has been collected (Pancino et al., 2012). Extended monitoring campaigns furthermore establish the photometric stability of the standard stars (Altavilla et al., 2015).

Today, EDR3 contains 1.806 billion sources with G-band photometry, and 1.540 billion with BP and RP photometry (Busso et al., 2021; Riello et al., 2021). The median uncertainty in the G-band is 0.2 mmag at $G = 10-14$, 0.8 mmag at $G = 17$, and 2.6 mmag at $G = 19$.

DOWNSTREAM in the data processing chain, the variable star processing takes all the epoch photometry, and provides a classification of the variable type and a characterisation of the light curve (Eyer et al., 2014).

The subsequent task of 'astrophysical parameter inference' then analyses the combination of astrometry, photometry, and spectroscopy to derive discrete source classifications (for example, distinguishing single stars, white dwarfs, binaries, quasars, and galaxies), and to infer the astrophysical properties of each source.

Astrophysical parameters derived from the BP and RP data are the effective temperature, surface gravity, metallicity, and extinction (Bailer-Jones et al., 2013). From the RVS spectra of the brightest stars, α-element enhancements and individual elemental abundances can be derived in addition (Recio-Blanco et al., 2016).

The accuracy of the parameter estimates depends on the star's magnitude, and is further constrained by the degeneracy between effective temperature and extinction. As an example, for FGKM stars at $G = 15$ mag with low extinction, effective temperatures can be estimated to 75–250 K, extinction to 0.06–0.15 mag, surface gravities to 0.2–0.5 dex, and metallicity [Fe/H] to 0.1–0.3 dex.

69. HD 140283: as old as Methuselah?

UNTIL THE RELEASE of the Hipparcos Catalogue in 1997, a quarter of a century ago, there existed an unsettling paradox regarding the age of the Universe.

From its expansion, its age had been estimated at around 11 Gyr. But some stars within it had ages, derived from their luminosities, and based on evolutionary models, of around 15 Gyr. Clearly something was very wrong for science to be telling us that some of the objects in the Universe were older than the Universe itself.

FROM THE new Hipparcos distances of various nearby Cepheids, Feast & Catchpole (1997) argued that the Large Magellanic Cloud Cepheids were 10% further than previously estimated, and thus brighter.

They concluded that the overall distance scale had to be revised upwards by the same amount, with the implication that globular clusters were more distant than previously thought, that their luminosities were therefore larger, and that their 'turn-off point' (an important distance indicator for much older objects) implied younger ages than previously thought.

Their analysis pushed the age of the Universe up a little, to around 12 Gyr, and brought the oldest stellar ages down to about 11 Gyr. With this consistency better established, astronomers could breath more easily that two foundations of their science—cosmology and stellar evolution—were not, after all, incompatible.

It's not often that scientists get to resolve such a paradox, to wipe a couple of billion years off the face of a cosmic timepiece, or to add a billion years, give or take, to the age of the Universe. And I recall an excited Michael Feast being rushed by taxi from a meeting of the Royal Astronomical Society in London to an interview for BBC radio's *Science in Action* to explain these results to a wider public.

TODAY, THIS AGE TENSION has largely eased, and stellar ages determined from high-accuracy astrometric distances and spectroscopy, combined with state-of-the-art computer-based evolutionary models, generally cap their ages to less than the 13.7 billion years or so that is believed to have elapsed since the Big Bang.

LET ME RECALL here that there are two rather direct measurements, using somewhat distinct approaches, that are considered to provide the most definitive estimates of the age of the Universe today.

One, intimately tied to the 'early Universe' estimates of the Hubble constant (essay #44), is based on precision measurements of the microwave background radiation. This observable relic of the early Universe, most recently and most accurately measured by the Planck satellite, indicates an age of 13.787 ± 0.020 billion years, that is, with a formal uncertainty of a mere 20 million years (Aghanim et al., 2020).

This is estimated in the context of the Lambda–CDM model, where the Universe is assumed to contain normal (baryonic) matter, cold dark matter, radiation (both photons and neutrinos), and a cosmological constant.

The other estimate of its age is based on observations of the 'local' distance scale and expansion rate, which suggest a slightly larger value of the Hubble constant and, correspondingly, a slightly younger age for the Universe. This is both the case for distance measurements based on the younger (Population I) Cepheids (Riess et al., 2018), as well as on the older (Population II) globular clusters (Freedman et al., 2019).

THE ESTIMATE of 13.787 ± 0.020 billion years is considered to be consistent with any of the *lower limits* on its age dictated by the oldest objects within it.

For example, one such constraint comes from the measured temperatures of the coolest white dwarfs. After exhausting their nuclear fuel, white dwarfs simply cool down, albeit very gradually, as they age. The temperature of the coolest white dwarfs, and detailed models of their cooling, over cosmological timescales, must provide a lower limit on the actual age of the Universe.

Another constraint is given by the dimmest 'turnoff point' of main sequence stars in clusters. Low-mass stars spend longer on the main sequence (during their hydrogen-fusion stage) than higher mass stars, such that the lowest-mass stars that have evolved away from the main sequence set another, independent, minimum value for the age of the Universe.

BUT A SMALL number of individual stars, among the oldest known to date, continue to raise their heads above this impregnable 'age parapet'. Some of these, at several thousand light-years distance, are members of our Galaxy's distant and ancient central bulge population, with estimated ages of around 13.2 Gyr.

HD 164922 is amongst the oldest, but a particular oddity, being a metal-rich main-sequence star at a distance of just 22 pc, but with an estimated age of around 13.4 billion years. It is also one of the most ancient planet-hosting stars known in the Milky Way.

A handful of others, members of our Galaxy's halo population, but which happen to be passing through our solar neighbourhood, are more problematic. Some are sufficiently close, within a few hundred parsecs, that they are particularly well measured.

One of the oldest, at 300 pc, is BD +17° 3248, an ultra-metal-poor Population II star. Based on its thorium and uranium abundances, its cosmochronological age is estimated at 13.8 Gyr, but with an uncertainty of 0.4 Gyr, large enough to still be interpreted as younger than the age of the Universe (Cowan et al., 2002).

OF GREAT interest in this context is HD 140283, an extremely metal-deficient and high-velocity subgiant in the solar neighbourhood. It occupies a location in the Hertzsprung–Russell diagram where absolute magnitude is most sensitive to stellar age.

Currently, one of the latest and best estimate of its age, at 13.7 ± 0.7 Gyr, places it amongst the very oldest known stars in our Galaxy, and in *potential* conflict with the accepted age of the Universe (Creevey et al., 2015). Accordingly, the star has become informally known as Methuselah (the idiom *'As old as Methuselah'*, meaning extremely old, is based on the biblical character, the grandfather of Noah, who died at the age of 969!).

Today, in the era of very large surveys covering millions of stars, placing them in a detailed evolutionary context depends on the accurate knowledge of their fundamental physical parameters (notably effective temperature, surface gravity, metallicity, and radius). This, in turn, relies on stellar models which are tested and refined against a sample of 'benchmark stars', determined independently. Interest in HD 140283 has been compounded by its selection as one of the 34 FGK-type benchmark stars selected as the 'pillars of calibration' for Gaia (Jofré et al., 2014; Karovicova et al., 2020).

BEFORE PROCEEDING, let me emphasise the important difference in science, and so crucial in astrometry, between the terms 'accuracy' and 'precision'. An example is perhaps sufficient. If my height were quoted as 2.35 ± 0.01 m (or 7 ft 8.5 ± 0.5 inches in our quaint imperial system), it is clearly claimed to be a very *precise* measurement, but it is evidently not at all *accurate*.

The true *precision* of any scientific measurement may be very difficult if not impossible to quantify with rigour. Any estimates of the precision can be affected by both random and (unknown) systematic errors, and always have to be viewed with due caution.

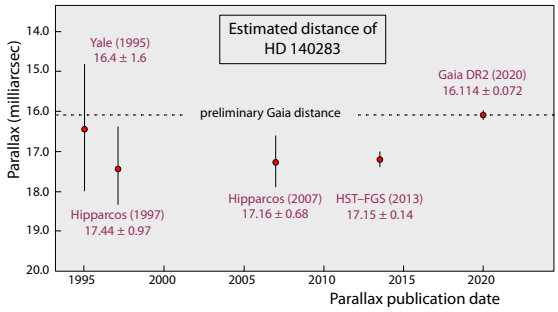
Thus if the age of Methuselah (13.7 ± 0.7 Gyr) is actually 13.0 Gyr (-1σ), we could relax. If it's 13.7 Gyr, it might be surprising, but perhaps still plausible. But if it's 14.4 Gyr ($+1\sigma$), well, Houston, we have a problem!

INTEREST IN HD 140283, from the point of view of its metal deficiency and therefore its age, dates back at least to the work of Burbidge & Burbidge (1956). By 2000, evolutionary models which included observed enhancements in the α-elements, provided 'strong evidence' in favour of an age older than 14 Gyr (VandenBerg, 2000).

A decade later, Bond et al. (2013) used a parallax from the Hubble Space Telescope Fine Guidance System, of 17.15 ± 0.14 milli-arcsec, *'five times more precise than that from Hipparcos'* (17.16 ± 0.68). They included the effects of He diffusion and enhanced O abundance, to infer an age of 14.46 ± 0.31 Gyr, where this specified error includes *only* the parallax uncertainty.

Further high-quality spectra, used to derive the surface abundances of O, Fe, Mg, Si, and Ca, resulted in a Universe-busting age of 14.27 ± 0.38 Gyr, again with the error including only the parallax uncertainty (VandenBerg et al., 2014). More recent estimates appear less in conflict (Joyce & Chaboyer, 2018; Tang & Joyce, 2021.

As of April 2022, I have not seen an age estimate for Methuselah derived using the Gaia DR2 parallax. And I will not embarrass myself by attempting to derive one.



TODAY, most astronomers would probably place their bets on the microwave background providing the most secure estimate of the age of the Universe, with any apparent conflict with stellar ages pointing to errors in the measured properties of the star, or to inadequacies in the theoretical models used to infer their ages.

But a secure conflict could point to important omissions in the theory of stellar evolution, or even to errors in our understanding of cosmology.

Future Gaia data releases, with improved parallaxes, will be an important contribution to this debate.

70. The Local Bubble

OUR SUN IS THOUGHT to lie just inside the boundaries of a warm, low-density, partially-ionised interstellar cloud ($T \sim 7000\,\mathrm{K}$, $n \sim 0.1\,\mathrm{cm}^{-3}$). The solar heliosphere results from the expanding high-velocity solar wind within this cloud, and extends from the Sun to about 100–150 astronomical units in the direction of its motion through space. The termination shock, marking the transition to the wider interstellar medium, was crossed by the spacecraft Voyager 1 in December 2004.

This and other nearby interstellar clouds are themselves located in a low-density region of the interstellar medium, referred to as the local cavity. This is partially filled with hot ($\sim 10^6$ K), low-density coronal gas, about 100 parsec in size, and referred to as the Local Bubble (Cox & Reynolds, 1987).

Detectable in soft X-rays, the Local Bubble was originally attributed to stellar winds from OB stars in the nearby Sco–Cen association, or to one or more nearby supernova explosions within the last 10 Myr.

THE SIZE AND SHAPE of the tenuous Local Bubble, and its precise origin, have been much debated, and many different observations have contributed to our knowledge of it today. Its dimensions can be determined directly by measuring the extent of the hot emitting bubble gas, or more indirectly by tracing the absence of absorption by neutral interstellar gas.

But the detailed morphology of the Local Bubble rests on our knowledge of the distances towards the respective line-of-sight stellar targets towards which the level of absorption is measured. Before Hipparcos, reasonably accurate distances from ground-based parallax measurements extended out to only about 20 pc.

Hipparcos distances out to 200–300 pc allowed the Local Bubble absorption characteristics to be defined with much greater certainty (e.g. Welsh et al., 1998). Sfeir et al. (1999) followed by mapping over 1000 lines-of-sight to reveal 'interstellar tunnels' of different widths which connect the Local Bubble to surrounding cavities. Their work supported the model of Cox & Smith (1974) in which expanding supernova-driven bubbles interact and merge to form large-scale interstellar cavities.

VARIOUS ATTEMPTS have been made to trace the origin of the supernovae that might have been responsible for the Local Bubble, with attention focusing on nearby OB associations as sites of recent star formation.

While the nearest of these, the Scorpius–Centaurus OB association, is currently about 130 pc from the Sun, Maíz-Apellániz (2001) used the Hipparcos data to extrapolate backwards in time both the Sun's position, and those of the various association members, to show that the Sco–Cen association, and especially the Lower Centaurus Crux subgroup, was much closer to the Sun 5 million years ago, making it a likely source of the few supernovae needed to produce the Local Bubble.

Other Hipparcos-based studies also concluded that the Local Bubble could have been excavated by some 20 supernova explosions around 10–20 Myr ago, some of them perhaps as close as 40 pc (Benítez et al., 2002; Berghöfer & Breitschwerdt, 2002; Fuchs et al., 2006).

NEARBY SUPERNOVA explosions are also of great interest in understanding their consequences for the geological record on Earth, their possible effects on life, and the existential threats to life on Earth in the unlikely event of a nearby supernova explosion occurring today.

Knie et al. (2004) reported direct evidence for such an event in the form of an enhancement of the ^{60}Fe concentration in a deep-sea ferromanganese crust, at far above natural terrestrial levels. Produced in Type II supernovae, with a half-life of several Myr (and detectable at ratios as low as $^{60}\mathrm{Fe/Fe} \sim 10^{-16}$), iron 60 was therefore taken as a proxy for supernova debris on Earth.

Specifically, they reported an event occurring about 2.8 Myr ago, at concentrations consistent with supernova ejecta at a distance of a few tens of parsec, and thereby supporting the hypothesis of a supernova explosion within the Sco–Cen association.

This epoch, they argued, coincides with both the onset and duration of an enhanced cosmic-ray flux, and an African climate shift towards more arid conditions, attributed to the onset of a northern-hemisphere glacial event, perhaps provoking or contributing to the Pliocene–Pleistocene boundary marine extinction.

TWENTY YEARS LATER, at least before the Gaia results became available, our picture of the Local Bubble has remained largely unchanged. It is clear that the Sun lies within a cavity of low-density, high-temperature plasma surrounded by a shell of cold, neutral gas and dust. But the precise shape and extent of this shell, the cause and timescale of its formation, and its relationship to nearby star formation remained uncertain, because of inadequate knowledge of the properties of the local interstellar medium.

The picture has sharpened significantly with Gaia. From radically new spatial and dynamical constraints, derived directly from the Gaia EDR3 distances and space motions, Zucker et al. (2022) have made a revised analysis of the three-dimensional positions, shapes, and motions of dense gas and young stars within 200 pc of the Sun, enabling some striking conclusions to be made about the nature of the Local Bubble.

THEY FIRST SELECTED stars from young clusters out to 300–400 pc with a maximum age of 20 Myr, including all young clusters associated with star-forming regions currently near the the surface of the Local Bubble (including the Lupus, Ophiuchus, Chamaeleon, Corona Australis, and Taurus Molecular Clouds), along with older stellar populations in the Sco–Cen association. Their sample totalled more than 1200 stars, with an average of nearly 40 stars per cluster.

They then performed a dynamical 'traceback' of each cluster's motion through the Galaxy, along with that of the Sun, using a numerical model of the Milky Way's gravitational potential (consisting of a bulge, disk, and dark matter halo), from 20 Myr ago to the present day.

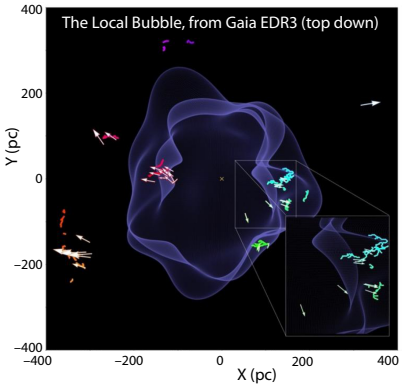
Their final step was to model the expansion of the Local Bubble, using hydrodynamic simulations to study the dynamical evolution of super-bubbles driven by clustered supernovae in a uniform interstellar medium.

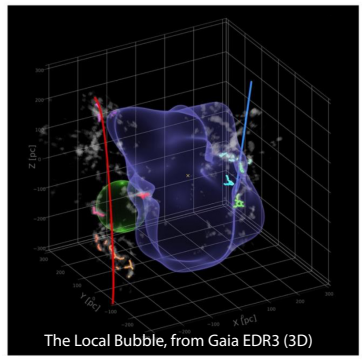
Based on the momentum injection required by supernovae to sweep up the total shell mass ($1.4 \times 10^6 M_\odot$), given its present expansion velocity (around 6–7 km/s), they estimate that around 10–20 supernovae were required to form the Local Bubble. This agrees well with earlier results, from their stellar membership and an adopted Initial Mass Function, that the Upper Centaurus Lupus and Lower Centaurus Crux clusters have produced some 14–20 supernovae over their lifetimes.

THEIR RESULTING model of the Local Bubble is shown as a top-down projection of the solar neighbourhood in the first figure, and a 3D view in the second, with the Sun shown as a yellow cross at the centre. It illustrates the inner surface of neutral gas and dust, and the 3D shapes and positions of local molecular clouds delineated at a resolution of about 1 pc. Their new Gaia-era 3D model of the Local Bubble's inner surface is shown more forcefully in their `interactive model`.

Their first major finding was quite remarkable: every well-known molecular cloud within about 200 pc of the Sun lies on the *surface* of the Local Bubble. These 'surface' clouds include not just every star-forming region in the Sco–Cen association (Ophiuchus, Lupus, Pipe, Chamaeleon, and Musca), but also Corona Australis and the Taurus Molecular Cloud, the latter lying 300 pc away from Sco–Cen on the opposite side of the bubble.

The one exception, the Perseus Molecular Cloud, at a distance of 300 pc, has likely been displaced by the recently discovered Perseus–Taurus Superbubble, also facilitated by Gaia (Bialy et al., 2021), containing Taurus on its near side and Perseus on its far side.





Zucker et al. (2022)

GOING FURTHER, they found that the young stars show an outward expansion mainly perpendicular to the bubble's surface. Tracebacks of these young star motions support a scenario where the origin of the Local Bubble was a burst of star formation, followed by death of the most massive stars as supernovae, taking place near the bubble's centre beginning 14 Myr ago.

The expansion of the Local Bubble created by the supernovae swept up the ambient interstellar medium into an extended shell that has now fragmented and collapsed into the most prominent nearby molecular clouds. This, in turn, provides robust observational support for the theory of supernova-driven star formation.

`This animation` of a flight through the Local Bubble, by Kevin Jardine, is based on data from Gaia DR2.

71. More halo streams from Gaia

In an earlier essay, #15, I looked at the discovery, using data from Gaia DR2, of the stellar 'debris' stream referred to as Gaia–Enceladus.

That study, by Helmi et al. (2018), and reported in the Journal *Nature*, demonstrated that the inner halo of our Galaxy is dominated by debris from an object which, at infall, was slightly more massive than the Small Magellanic Cloud. These authors concluded 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, and that it probably represented the last significant merger that our Galaxy experienced.

Their findings were in line with current cosmological simulations of galaxy formation, which predict that the inner stellar halo should be dominated by debris from just a few of these massive progenitors.

The Galactic disk has long been recognised as having been formed by the rapid collapse of a rotating galaxy-sized gas cloud, billions of years ago. But the idea that the Galaxy halo has been built up over billions of years from infalling low-mass objects, such as dwarf galaxies of $10^7 - 10^8\,M_\odot$, only dates back to the seminal paper by (Searle & Zinn, 1978).

Their study of 177 red giants in 19 globular clusters at distances beyond about 8 kpc, suggested that halo clusters originated within transient protogalactic fragments that gradually lost gas while undergoing chemical evolution, and continued to fall into the Galaxy after the collapse of its central regions had been completed.

This model, evidenced by the wide range of globular cluster metallicities independent of Galactocentric radius, a wide age spread of halo field stars and globular clusters, and a subset of intermediate abundance globular clusters with retrograde mean motions, has been the paradigm for the halo formation for the past two decades (e.g. Bland-Hawthorn & Freeman, 2000).

It is useful to recall that these ancient signatures of tidal infall remain accessible for study because the orbital time-scales in the outer parts of the Galaxy are several billion years. As a result, the halo retains kinematic evidence of the surviving remnants of accretion, as well as a chemical 'memory' of early low-mass stars as a result of their very long evolutionary lifetimes.

Put another way, a low-mass stellar stream follows closely its progenitor's orbit, so that the locus of any one tidal stream provides a historical tracer of an ancient Galaxy orbit. Thus mapping the line-of-sight velocities and proper motions along a stream allows the mapping of the acceleration that the stream has been subject to. And with several streams on different orbits it becomes possible to build up a reliable three-dimensional map of the acceleration field throughout the region of the Galaxy traversed by such streams.

Over the past 25 years, various of these dynamically evolved 'stellar streams' have been identified, all originating from this sort of tidal stripping. Some have their origin in disrupted satellite galaxies, others in specific globular clusters.

The first to be identified was the Arcturus stream (Eggen, 1971), comprising some 50 ancient stars deficient in heavy elements. The gaseous Magellanic stream, associated with the Large and Small Magellanic Clouds, and the Sagittarius stream associated with the Sagittarius dwarf spheroidal galaxy, were discovered in subsequent years.

Revealed by their correlated angular momentum in the Hipparcos data, and corroborated by their prominent metal deficiency, the 'Helmi stream' started life as a dwarf galaxy of $10^7 - 10^8$ solar masses, captured by the Milky Way some 6–9 billion years ago. It may be responsible for some 10% of the metal-poor stars in our Galaxy's halo beyond the Sun's orbit (Helmi et al., 1999).

Others streams have been discovered since, and variously attributed to globular cluster or dwarf galaxy origins. Several have been discovered from the Sloan Digital Sky Survey, amongst them the Acheron, Cocytos, Lethe, and Styx streams (Grillmair, 2009), as well as the polar-orbiting Cetus stream (Newberg et al., 2009).

One, the LAMOST 1 stream of more than 20 000 stars, was found from the LAMOST spectroscopic survey (Vickers et al., 2016), and another, the Phoenix stream, from the Dark Energy Survey (Balbinot et al., 2016).

BY PROVIDING THE DISTANCES and space motions of more than two billon stars, from which orbital distances and motions can be calculated, Gaia is opening a new chapter in the study of stellar streams.

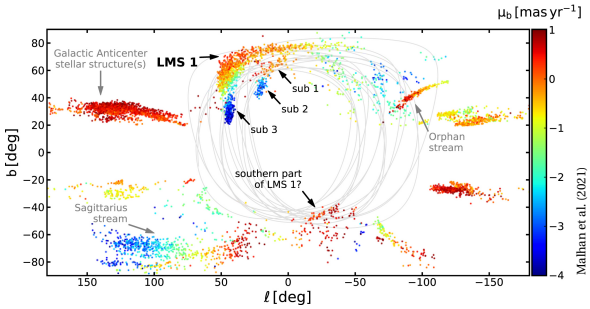
First amongst these, from Gaia DR2, was the Gaia–Enceladus stream (Helmi et al., 2018), mentioned above, and which I described in some detail in essay #15.

Others have followed. Ibata et al. (2019b) used their 'Streamfinder' algorithm, applied to the Gaia astrometry and photometry alone, to identify eight new structures at heliocentric distances between 1–10 kpc, which they named (from Norse mythology) Slidr, Sylgr, Ylgr, Fimbulthul, Svöl, Fjörm, Gjöll, and Leiptr. Spectroscopic measurements of seven of the streams have confirmed their reality, and have shown that these streams are predominantly metal-poor.

One of these, Fimbulthul, is the trailing arm of, and (gratifyingly) the same age as, the tidal stream of the massive globular cluster ω Centauri (Ibata et al., 2019a).

LMS–1/Wukong is another Gaia DR2 discovery, independently reported as 'Low-Mass Stream 1' (based on RR Lyrae and blue horizontal-branch stars) by Yuan et al. (2020), and as Wukong, using the combination of Gaia and the H3 Spectroscopic Survey, by Naidu et al. (2020).

This image, from the study of LMS–1/Wukong by Malhan et al. (2021b), shows the distribution of proper motions in Galactic latitude (μ_b, the scale ranging from 0 to just 4 milli-arcsec per year!), illustrating the rich substructure in the various streams, along with their derived orbit for LMS–1, shown as the grey curve.



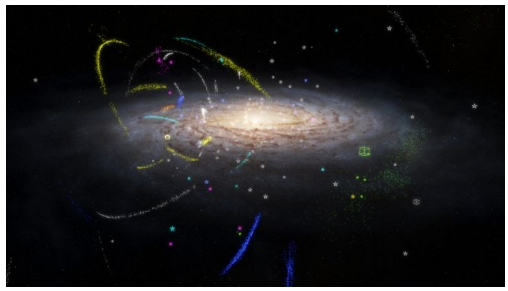
Meanwhile, the retrograde Arjuna/Sequoia/I'itoi group appears to result from three independent mergers (Myeong et al., 2019; Naidu et al., 2020).

Indeed, from their study of 5684 giants within 50 kpc from the Galactic centre, Naidu et al. (2020) attributed at least 95% of their sample stars to one of their listed structures, suggesting a halo built entirely from accreted dwarfs and the associated 'heating' of the disk.

Another stellar stream, named Nyx, comprises some 200 stars in the solar vicinity (Necib et al., 2020a; 2020b). These authors attribute Nyx to a massive dwarf galaxy dragged into the disk plane before being completely disrupted, although its extragalactic origin has been questioned (Re Fiorentin et al., 2021).

WITH THE RELEASE of Gaia EDR3, Malhan et al. (2022) assigned 170 globular clusters, 41 streams, and 46 satellite galaxies to six distinct groups, including the previously known mergers Sagittarius, Cetus, Gaia–Enceladus, LMS–1/Wukong, Arjuna/Sequoia/I'itoi, and a new merger that they named Pontus. An image from their animation is shown here.

The three most-metal-poor (C-19 with [Fe/H] = –3.4 dex; Sylgr with [Fe/H] = –2.9 dex, and Phoenix with [Fe/H] = –2.7 dex) are associated with LMS-1/Wukong, making it the most-metal-poor merger known to date.



BY MID-2022, several other studies of these halo streams have been published, and more will become possible with the future Gaia data releases. But let me now briefly consider some of the wider studies that these results of stellar streams are starting to enable.

Firstly, Malhan et al. (2021a) have shown that the morphology and dynamics of accreted globular cluster streams are sensitive to the central dark matter density profile and mass of their parent satellites. Specifically, globular cluster that accrete within 'cuspy' cold dark matter sub-haloes produce streams that are physically wider and dynamically hotter than streams that accrete inside cored sub-haloes.

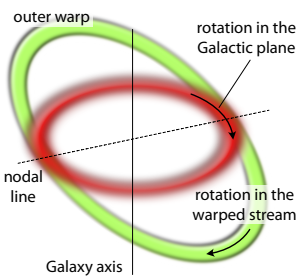
A second spin-off is that these very low-metallicity streams can probe the contribution of r-process and s-process enrichment early in their formation history. First results by Gull et al. (2021) suggest that the progenitors of some of these streams experienced one or more r-process events (such as neutron star mergers) early on, in advance of their accretion by the Milky Way.

Tutukov et al. (2021) examined the possible role of star formation and collisions of gas-rich galaxies, leading to the emergence of low-surface brightness galaxies. Stellar streams, from the decay of satellite galaxies, may thus contain the remaining dense star clusters and include stars, exoplanets, and interstellar comets.

Finally, Peñarrubia (2021) considered their role in the formation of binary stars with extremely large separations, whose origin remains poorly understood (essay #37). He found that ultra-wide binaries can arise via chance entrapment of unrelated stars in tidal streams of disrupting clusters, suggesting that these streams may be the birthplace of hundreds of ultra-wide binaries.

72. The warp of our Galaxy

A WARPED GALAXY is one with an outer disk component which is twisted upward at one side, and down at the other. Many galaxies display such a warped structure, including our own, with at least half of all spirals warped both in gas (H I) and in stars.



Various explanations have been proposed, including infall of intergalactic material, or tidal forcing through a close encounter with a companion galaxy (in our case, perhaps the Magellanic Clouds or the Sagittarius dwarf).

But the existence of many *isolated* warped galaxies may signify some general misalignment of the symmetry axes of the disk and its dark matter halo. In such cases, gas in the outer disk would settle into the symmetry plane of the halo, rather than of the disk. And, importantly, such a tilted disk, embedded in a massive halo, would precess.

PRE-HIPPARCOS STUDIES of our Galaxy's warp mainly focused on its spatial structure (using bright young objects such as Cepheids and OB stars), because its kinematic signature is most evident in the velocities *tangent* to the line-of-sight when viewed from in the plane.

A warping motion, based on the proper motions of 2000 young stars out to 0.5–3 kpc, had already been detected by Miyamoto et al. (1993). They used the 3d Ogorodnikov–Milne formulation to detect shear and rotation around two orthogonal axes in the plane, in addition to the classical (Oort) components. They concluded that young stars are streaming around the Galactic centre in a tilted sheet with a velocity of 225 km s^{-1}, and that the sheet is also rotating around the nodal line of the warp with an angular velocity of 4 km s^{-1} per kpc.

But in his dedicated review, Binney (1992) concluded that the detailed dynamical nature of warps, and their driving mechanism, remained uncertain; a conclusion largely unchanged 25 years later by Poggio et al. (2017).

THE HIPPARCOS DATA nevertheless provided proper motions with systematic and zonal errors significantly smaller than the expected warp signature.

Amongst the Hipparcos results, Smart et al. (1998) focused on 2422 distant OB stars (beyond 500 parsec) towards the Galactic anti-centre. Their choice of young stars assumes that they trace the motions of the gaseous component from which they were born, and thus more likely to trace the warp. They found that the disk is flat to approximately the solar radius, then turns up to the north in the direction of Cygnus at $l \simeq 90°$, and south in the direction of Vela at $l \simeq 270°$.

But while the spatial distribution of stars was consistent with studies based on neutral hydrogen, the velocity distribution was of opposite sign to that expected. Thus the stellar kinematics did not follow the signature of a long-lived warp, whether precessing or not.

Drimmel et al. (2000) confirmed these results using an enlarged Hipparcos sample of 4538 OB stars (also beyond 500 parsec) covering all Galactic longitudes, and using distances derived from both trigonometric and photometric estimates. They concluded that either the warp evolves on a time-scale of the Galactic rotation period or shorter (possibly originating from an impulsive interaction with the Sagittarius dwarf galaxy), or that other large but unidentified systematic motions (such as vertical oscillations) are present within the disk.

A PARTIAL EXPLANATION for these results was put forward by Ideta et al. (2000). They found that kinematic warps in oblate halos 'wind up', and disappear, within a few dynamical times. In prolate halos, in contrast, warps persist with continued alignment of the line of nodes, due to the fact that the precession rate of the outer disk increases when the precession of the outer disk recedes from that of the inner disk, and vice versa.

Meanwhile, cosmological CDM (cold dark matter) simulations suggest not only that dark matter halos surrounding individual galaxies are highly triaxial, but also that the fraction of prolate and oblate halos is roughly equal (Dubinski & Carlberg, 1991).

BEFORE TURNING to the new insights that have come from Gaia, I should open a parenthesis. Further complicating our Galaxy's disk structure are the discoveries of the Monoceros Ring (Newberg et al., 2002), and the Triangulum–Andromeda streams (Majewski et al., 2004). These are ring-like stellar density enhancements of the disk, at around twice the solar radius. Provisionally attributed to a Sagittarius-like impact, simulations suggest that the resulting interactions between the disk and halo can lead to a tightly wound spiral pattern of vertical density oscillations (e.g. Binney et al., 1998).

Other similar overdense regions, like Canis Major, are being further characterised by Gaia, and continue to complicate the picture (Carballo-Bello et al., 2021).

WITH GAIA DR1, Poggio et al. (2017) selected 758 OB stars, also measured by Hipparcos, with distances between 0.5–3 kpc. They found that the proper motions of nearby stars are consistent with a kinematic warp, while stars beyond 1 kpc are not. They concluded that the systematic vertical motions observed in the disk cannot be explained by a stable long-lived warp, and that it is either a transient feature, or that additional forces are causing systematic vertical motions that are masking the expected warp signal.

Schönrich & Dehnen (2018) also used the early Gaia–TGAS distances and proper motions to estimate the vertical and azimuthal velocities in the Galactic centre and anti-centre directions. The mean vertical motions show a linear increase with distance, consistent with the known Galactic warp. But they also reveal a previously unknown wave-like pattern, with an amplitude of $1\,\mathrm{km\,s^{-1}}$ and a wavelength of 2.5 kpc, in both directions. They attributed this to a bending wave, most likely related to the Monoceros and Tri–And overdensities.

NUMEROUS STUDIES followed with the release of Gaia DR2 in 2018. Poggio et al. (2018) extended these kinematic studies out to a distance of 7 kpc from the Sun. Combining Gaia DR2 and 2-Micron All Sky Survey photometry, they identified nearly 600 00 upper main sequence stars and more than 12 million giant stars.

The spatial distribution of the former clearly shows segments of the nearest spiral arms. The large-scale kinematics of both populations show a clear signature of the Galaxy warp, apparent as a gradient of $5–6\,\mathrm{km\,s^{-1}}$ in the vertical velocities from 8–14 kpc in Galactic radius. They argued that the presence of the warp signal in both samples, which have different typical ages, suggests that the warp is a gravitationally induced phenomenon.

Romero-Gómez et al. (2019) used DR2 to examine the vertical motions of two populations up to Galactocentric distances of 16 kpc: a young bright sample mainly comprising OB stars, and an older one of red giant branch stars. They confirmed the age dependency of the warp, both in position and kinematics, its height being around 0.2 kpc for the OB sample and 1 kpc for the older sample at a Galactocentric distance of 14 kpc, with an onset radius of the warp at 12–13 kpc for the former, and 10–11 kpc for the latter. But they also confirmed a high degree of complexity in both position and velocity, including a prominent wave-like pattern of a bending mode, different in the two samples.

A study based on 2431 Cepheids from DR2 reveals asymmetric warp-like large-scale vertical motions with amplitudes of $10–20\,\mathrm{km\,s^{-1}}$ (Skowron et al., 2019). Similar features are found in a sample of 250 000 OB stars, where the flat inner disk begins to warp at about 9 kpc from the Galactic centre (Li et al., 2019), and in a study of nearly 140 000 red clump stars from LAMOST DR4 and Gaia DR2 (Wang et al., 2020; Li et al., 2020).

IN LATER models of the global trends in vertical velocity, Cheng et al. (2020) derived a starting radius for the flare at about 8.9 ± 0.1 kpc, and a precession rate of $13.6\pm0.2\,\mathrm{km\,s^{-1}}$ per kpc in the direction of Galactic rotation.

From their 12 million giant stars in DR2, Poggio et al. (2020) found a similar precession of $10.86\pm0.03\,\mathrm{km\,s^{-1}}$ per kpc, i.e. about one-third the angular rotation velocity at the Sun's Galactocentric distance of 8 kpc.

The direction and magnitude of the warp's precession, and its consistency across stellar age groups, appears to favour our Galaxy's warp being the result of a recent or ongoing (gravitationally induced) encounter with a satellite galaxy, rather than being the dynamical relic of the ancient assembly history of the Galaxy.

But here again, the model assumptions, and these conclusions, are still contested (Chrobáková & López-Corredoira, 2021).

CURRENT WARP MODELS attempt to separate out the smaller ripples in vertical and radial velocity which are superposed on the main trend.

The present picture, interpreted in a cosmological setting, is of a galaxy disk which continuously experiences gravitational torques and perturbations from a variety of sources, which can cause the disk to wobble, to flare and to warp, in the process revealing important information on the formation history of galaxies, and on the mass distribution of their halos.

Greatly augmented by Gaia's results, our Galaxy evidently presents a unique case study due to our detailed knowledge of its stellar distribution and kinematics.

Nonetheless, a clear and consistent picture of the dynamical nature of our Galaxy's warp, and its driving mechanism, remains elusive. Hopefully, future Gaia data releases will clarify our understanding.

Meanwhile, this nice animation of our Galaxy's warp, created by Stefan Payne-Wardenaar, MPIA Heidelberg, is based on the model by Chen et al. (2019b).

73. White dwarf pollution and exoplanets

A FASCINATING CONNECTION between the observed spectra of white dwarfs (the terminal evolutionary stage for 90–95% of the stellar population), and the formation and survival of exoplanetary systems, has been forensically assembled over the past two decades.

In his history of the field of heavy element pollution of white dwarf photospheres, Zuckerman (2015) suggests that the first observational evidence for the existence of exoplanet systems actually came more than a century ago from one such 'polluted' star, van Maanen 2 (van Maanen, 1917; 1919), although Dutch–American astronomer Adriaan van Maanen did not know that he had observed a white dwarf, nor of its significance!

Today, vMa 2 is considered to be the prototype of the white dwarf DZ spectral class, with high-resolution spectra showing that its atmosphere is deficient in hydrogen, and yet at the same time rich in the so-called 'refractory' elements, with lines of Fe, Ca, and Mg.

OTHER WHITE dwarfs with photospheric metals were duly discovered, with as many as 16 terrestrial-like heavy elements in the case of GD 362 (Zuckerman et al., 2007). It was soon appreciated that their intense gravity should cause any heavy elements to sink rapidly (Schatzman, 1945), with elements heavier than helium pulled into their interiors on time scales far shorter than their cooling ages (Fontaine & Michaud, 1979).

A replenishing source for these 'polluting' metals is therefore required. Initially, accretion from the interstellar medium was favoured (e.g. Farihi et al., 2010), although further observations (e.g. Zuckerman & Becklin, 1987), and theoretical work (e.g. Jura, 2003) shifted the accretion paradigm to rocky asteroidal debris.

Studies have since suggested that heavy elements in isolated white dwarfs with $T_{\rm eff} \sim 5000 - 20\,000$ K is evidence for the presence of a wide-orbit planetary system, generally comprising a rocky debris belt and at least one planet (e.g. Mustill et al., 2014). The planet's gravitational field perturbs the asteroid orbits, so that rocky asteroidal material is occasionally accreted onto the white dwarf (e.g. Jura & Young, 2014; Chen et al., 2019a).

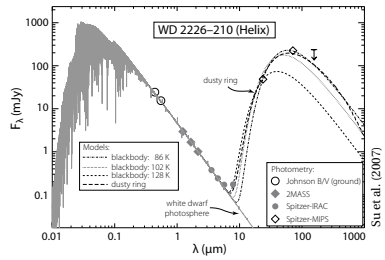
BEFORE CONTINUING to describe how Gaia is advancing the studies of white dwarf photospheric pollution, let me summarise a few points by way of background (see also Essay #29 on 'white dwarf surveys').

The first is that, while white dwarfs are extremely common in the solar neighbourhood, their very low luminosities means that any survey completeness falls rapidly with increasing distance, even within 20–50 pc. Pre-Gaia surveys listed around 20–30 000 known objects, with around 200 within 200 pc, while some 260 000 candidates have been identified from Gaia DR2.

Whether planetary companions to white dwarfs exist, having survived the red giant branch and asymptotic giant branch phases, will depend on things like the initial orbit separation and the stellar mass-loss rate. Many 'survival studies' have been made. For example, orbits with initial radii above about 0.7 au will remain larger than the stellar radius at all evolutionary stages, and white dwarfs could therefore potentially host surviving planets with orbital periods above about 2.4 years.

MEANWHILE, DUST DISKS around white dwarfs were first discovered in the early 1990s (Graham et al., 1990). Their presence is inferred from their large mid-infrared spectral excess. Major advances in their discovery and characterisation has been made from space, notably with Spitzer, and more recently WISE, and some 100 or more dust-disk systems are known today.

WD 2226–210, at the centre of the Helix planetary nebula, has been suggested as a prototype. Observed with Spitzer, the dust ring extends 35–150 au from the central white dwarf (interior to the helix structure), has a total mass of about one tenth Earth's, and is inferred to be replenished through collisional fragmentation of planetesimals (Dong et al., 2010).



Su et al. (2007)

THE MODEL favoured today attributes the heavy element pollution in white dwarf atmospheres to the accretion of rocky planetesimals which have been scattered into the tidal disruption radius of the white dwarf, where they are torn apart into dusty debris that subsequently spreads out to form a circumstellar disk. Some of this material can be perturbed by a planetary body, to be intermittently accreted by the white dwarf.

Indeed Farihi et al. (2012) eloquently described metal-enriched white dwarfs as representing astrophysical 'traps' for exoplanet debris, *'…acting as detectors that can yield the bulk composition of planetary building blocks that orbit intermediate-mass stars'*.

Drawing on models of accretion and sinking rates, they estimated that the 'high-rate' accretion time scale is less than 1000 years, such that a search of 10 000 DA stars would be needed to detect one in a high state lasting 100 yr. They argued that if Gaia provides some 10 000 metal-polluted candidates, then it may be possible to identify a single white dwarf accreting in a high state.

CURRENT OBSERVATIONS and models go further. Accreted photospheric compositions typically resemble that of bulk Earth, with at least 85% by mass in O, Mg, Si, and Fe.

Detailed results suggest that differentiation, leading to Fe-rich cores and Al-rich crusts, is common amongst Earth-mass exoplanets. It may follow that some white dwarfs might reflect purely crustal-like material, in which Si and O are the dominant pollutants.

Jura et al. (2014) even suggest that an indirect search for exoplanetary plate tectonics is possible because of a unique spectral signature in an externally polluted white dwarf signalled by the abundances of Ca, Sr and Ba.

THE DEBRIS DISKS themselves can be recognised as a reprocessed infrared flux in excess of what is expected from an isolated white dwarf. About 1–3 per cent display this infrared dust-emission signature.

But a small subset of these already rare systems also display line emission, typically strongest in the Ca II 860 nm triplet, due to the presence of a gaseous disk component. These Doppler-broadened emission features are found to be largely consistent with models for gas in Keplerian orbits in a flat disk.

Where information on the spatial distribution of the gas is available, it appears to be co-located with the dust. These findings in turn imply that some mechanism must keep generating the gas, which would otherwise re-condense on time scales of months.

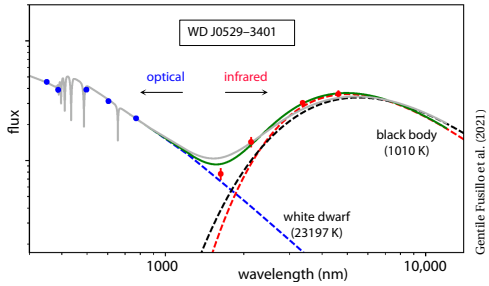
But the exact process underlying the generation of such circumstellar gas remains uncertain, and different scenarios have been proposed, including dust sublimation at the inner edge of the debris disk followed by subsequent radial spreading, and collisional cascades grinding the debris into gas.

ALTHOUGH THESE gaseous systems are therefore important for understanding the formation and evolution of planetary debris disks, their rarity is a barrier to further advances. Estimates suggest that only about 4% of white dwarfs with dusty debris disks, or less than 0.1 per cent of all white dwarfs, display emission features from a gaseous component (Manser et al., 2020).

In the first statistical study using the Gaia data, Manser et al. (2020) selected 7705 single white dwarfs from the Sloan Digital Sky Survey, and with Gaia magnitudes brighter than 19 mag. They identified five gaseous disk hosts, all of which had been previously discovered.

RESTRICTING THE 260 000 white dwarfs identified in the Gaia DR2 release by Gentile Fusillo et al. (2019) to those brighter than $G = 18.5$ mag, using the Gaia distances to constrain the models of the spectral energy distribution, and with infrared data from large area surveys (including 2MASS, UKIRT, VHS, and WISE), Gentile Fusillo et al. (2021) identified a further six white dwarfs with gaseous debris disks, bringing the total number known to just 21.

Amongst their new Gaia discoveries, WD J0846+5703 has an exceptionally strong infrared excess which contrasts with the standard model of a geometrically-thin, optically-thick dusty debris disk; WD J2133+2428 is the hottest gaseous debris disk host known, with an effective temperature of 30 000 K; and WD J0529–3401 (shown here) features a record number of 51 emission lines from five elements (H, Mg, Ca, Fe and O).



Every Ca II triplet emitting system that has both multi-epoch observations and asymmetric emission profiles show morphological variability, and those that have been studied in detail have shown that the variations are periodic in nature, and well described by the precession of a fixed intensity profile in the disk.

The period of this precession can span the range of years to decades, and this range is supported by theoretical studies that show the effects of general relativistic precession and pressure forces on the gaseous disk.

THE STUDY and understanding of heavy element pollution of white dwarf photospheres is a fascinating subject still in its infancy. Gaia will undoubtedly be contributing to its advances over the coming years.

74. Open clusters with Gaia

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

More than a thousand are known within our Galaxy, and many more are thought to exist. A few, such as the Hyades (at about 45 pc; see essay #20), the Pleiades (at about 130 pc; see essay #13), 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 gravitational attraction when 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.

They may survive as recognisable clumps for a few hundred million years, with the most massive surviving for a few billion. 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.

WITH ITS SURVEY of all stars down to around 21 mag, its accurate star positions and space motions out to 1000 pc or more, and its unprecedented multi-epoch multi-colour photometry, Gaia is revolutionising their study. Quality positions allow cluster membership to be refined, space motions convey details of their dynamics and dispersion, and detailed photometry further contributes to classifying membership and chemistry.

With several hundred papers dealing with results from Gaia even by mid-2022, this can only serve as an incomplete introduction. A more detailed review of Gaia's contribution to date is given by Cantat-Gaudin (2022).

USING Data Release 2, Cantat-Gaudin et al. (2018) examined the extensive pre-Gaia open cluster catalogue compilations of Dias et al. (2002) and Kharchenko et al. (2013). They found that, from more than 3000 identified clusters, they could confirm only around 1200.

The explanation for this is of great importance for understanding star formation in our Galaxy (Kos et al., 2018): many previously suggested clusters are simply 'asterisms', stellar over-densities arising from strong extinction patterns mainly in the direction of the Galactic bulge. And excluding them results in a cluster age function in better agreement with theoretical models of their formation and dissolution (Anders et al., 2021).

As expected, Cantat-Gaudin et al. (2018) found that the youngest clusters are concentrated near the Galactic plane and trace the Galaxy's spiral arms, while older objects are more uniformly distributed, typically further from the plane, and at larger Galactocentric distances.

AMONGST OTHER statistical studies, Soubiran et al. (2018) used DR2 astrometry, including radial velocities, to study the positions and full space velocities for 861 clusters. They found that those with ages younger than 100 Myr have low vertical velocities and are, on average, within 100 pc of the Galactic plane. Clusters older than 1 Gyr span distances to the Galactic plane of up to 1 kpc, and with a vertical velocity dispersion of 14 km s^{-1}, are typical of the thin disk population.

Tarricq et al. (2021) determined Galactic orbits for 1382 clusters, showing that those younger than 30 Myr follow rather circular orbits, fully consistent with current models of star formation in which stars and clusters are formed on circular orbits near the Galactic mid-plane.

THE FACT THAT star clusters host coeval stars of different masses makes it straightforward in principle to estimate their ages from a colour–magnitude diagram combined with theoretical stellar evolution models. An analysis of the colour–magnitude diagrams of 269 clusters by Bossini et al. (2019) concluded that the uncertainty on the absolute age of clusters is now generally

dominated by uncertainties on stellar evolution models themselves, which will provide an important foundation for future theoretical modelling.

A number of other papers are now using the Gaia data to yield simultaneous estimates of cluster properties such as distance, extinction, total mass, age, and metallicity (e.g. Kounkel et al., 2020; Perren et al., 2020).

THE PHYSICAL PROCESS driving the formation and persistence of spiral arms, in our Galaxy and beyond, remains a matter of debate – whether they are global and stationary, or local features of a more transient nature. Here, Gaia is also providing new insights.

Castro-Ginard et al. (2021) used EDR3 data to examine the birthplaces of young open clusters in the Perseus, local, Sagittarius, and Scutum spiral arms, to trace the evolution of the different spiral structures. Finding different pattern speeds for the different arms, they argued that their results question classical density waves as the main drivers for Galaxy's spiral structure, being in better agreement with models favouring transient features.

The detailed structure of the various spiral arms, inferred from the spatial distribution of young clusters, is likely to provide other constraints on spiral arm models. Amongst these are studies of the Perseus arm (Cantat-Gaudin et al., 2019a); the local arm (Xu et al., 2021); the Cepheus spur (Pantaleoni González et al., 2021); and the Sagittarius spur (Kuhn et al., 2021).

MANY PAPERS have taken up the challenge of identifying *new* open clusters in DR2 and EDR3, frequently using the accurate proper motions to identify groups of stars that are kinematically coherent, even if they may be spatially sparse. These searches are making use of a variety of advanced clustering algorithms.

Meanwhile, Koposov et al. (2017) discovered two massive disk clusters, which they named Gaia 1 and Gaia 2. At least the former appears to be a massive and luminous open cluster, rather than a low-mass globular cluster (Koch et al., 2018).

Amongst other reported discoveries, Cantat-Gaudin et al. (2018) claimed 60 new open clusters; Castro-Ginard et al. (2018) identified 31; Cantat-Gaudin et al. (2019a) found 41 in the Perseus arm; Castro-Ginard et al. (2019) discovered 53 towards the Galactic anti-centre; Sim et al. (2019) found 207; He et al. (2021) found 74 clusters with a nearest-neighbour approach; Hunt & Reffert (2021) reported 41; while in all-sky supercomputer-based searches, Castro-Ginard et al. (2020) discovered 582 clusters in DR2, and Castro-Ginard et al. (2022) found a further 664 in EDR3.

But there is likely no clear answer as to how many new clusters will be found with the Gaia data, for at the sparsest levels even pairs of comoving stars may point to long-disrupted clusters (e.g. Kamdar et al., 2019).

THE DYNAMICAL EVOLUTION of open clusters can be witnessed as stars escaping from the cluster, preferentially through their Lagrange points L1 and L2, and leading to the formation of tidal tails.

These are seen most prominently in the Hyades cluster (my essay #20). But they are also now evident in Praesepe (Röser & Schilbach, 2019); Ruprecht 147 (Yeh et al., 2019); M67 (Carrera et al., 2019); Coma Ber (Tang et al., 2019); Czernik 3 (Sharma et al., 2020); Blanco 1 (Zhang et al., 2020); IC 4756 (Ye et al., 2021); and NGC 752 (Bhattacharya et al., 2021). In the Gaia-discovered 'colliding' clusters IC 4665 and Collinder 350, each is perhaps driving the other's tidal disruption (Piatti & Malhan, 2022).

Other evidence of their dynamical evolution are 'runaway' stars, in which high proper motions pinpoint stars which have been hurled out of their natal cluster via a binary–supernova interaction, or through the disruption of single–binary or binary–binary systems.

THE DETAILED colour–magnitude diagrams of open (as well as globular) clusters also allow the identification of stars in unusual stages of stellar evolution. Gaia results are starting to appear on blue straggler stars, of uncertain origin but manifested as stars both brighter and bluer than the cluster's main-sequence turnoff.

And many new results are emerging on white dwarf in clusters, as well as on particular variability types such as young Delta Scuti stars and Cepheid variables.

FURTHER COMPLICATING an already involved picture are the groupings of young, luminous O- and B-type stars, the OB associations. Models in which OB associations are formed from the expansion of previously compact clusters are no longer favoured. Indeed, as summarised in my essay #18 on 'The Origin of OB Associations', the Gaia results are more consistent with hierarchical 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 the distinction between clusters and associations therefore becomes somewhat arbitrary, I will not attempt to touch on the extensive Gaia literature on associations here. And other spatial features are being reported from the Gaia data, amongst them structures referred to as strings or filaments (Kounkel & Covey, 2019); rings (Cantat-Gaudin et al., 2019b); snakes (Tian, 2020; Wang et al., 2022a); and pearls (Coronado et al., 2022).

AS CONCLUDED by Cantat-Gaudin (2022) in his recent review: *'The Gaia data have unlocked a deluge of new results related to many astronomical topics and transformed our ability to study star clusters and stellar structures in the Milky Way.'*

75. The local mass density

OUR SOLAR SYSTEM lies very close to (well, just a few light-years from!) the mid-plane of our Galaxy's flattened disk, and a colossal 30 000 light-years from its centre. Like all the other stars in the solar neighbourhood, the Sun (and our planetary system with it) rotates around the centre of the Galaxy, in an approximately circular orbit, once every 250 million years.

Looking out at the disk of our Galaxy, the Milky Way, stellar motions are of course not at all discernible to the naked eye. But astronomy has known for decades that, in addition to these largely circular motions, stars are 'bouncing up and down' about the Galaxy mid-plane, reaching distances of several tens of light-years above or below it, before being 'hauled' back down by the gravitational force exerted by the matter concentrated in the disk itself. The stars then fly on through the mid-plane, before continuing on their indefinite oscillatory journey.

The collective gravity of the Galactic plane results in an oscillatory motion somewhat analogous to a Focault's pendulum controlled by Earth's gravity, albeit with a somewhat longer period… of about 60–90 million years!

What is the *total* amount of matter in the solar neighbourhood determining these motions? The answer must include both visible material (mainly stars and gas, with the contribution of 'cold' gas presently being the most uncertain), along with any dark matter – mass models of the Galaxy based only on visible star counts can only be used to infer *lower* limits.

The result can be expressed in mass per unit volume, $M_\odot\,\mathrm{pc}^{-3}$ (or, for dark matter hunters, in $\mathrm{GeV\,cm}^{-3}$), or as the column density out to a given z distance from the plane, expressed in $M_\odot\,\mathrm{pc}^{-2}$. Our present understanding is that the projected mass of disk matter corresponds to roughly $70\,\mathrm{g\,m}^{-2}$, about the density of typical A4 paper, spread out across our Galaxy's mid-plane!

MEASUREMENT of the *dynamical* effects of the local mass density, from the vertical motions of stars, is potentially more robust. Analysis of the density and velocity distribution of a tracer sample of stars can provide estimates of the local density of *all* disk-like matter.

This is often referred to as the K_z problem in Galactic dynamics, where K_z is defined as the component of the Galaxy's gravitational acceleration towards the Galactic plane in the solar neighbourhood.

In a plane-stratified approximation out to a few kpc, K_z is assumed to increase monotonically with z. Studies then use some suitable 'tracer' population to estimate these forces – stars whose number density and velocities can be determined as a function of distance from the plane. Matter produces both the restoring potential (mathematically described by Poisson's equation), and is at the same time influenced by it (mathematically described by the Boltzmann or vertical Jeans' equation).

HIPPARCOS brought a number of observational advances to this problem: primarily an improved accuracy on distances and space velocities, but also an increased size of various tracer populations (such as A stars and K giants). It also provided an improved estimate of the local stellar luminosity function, such that the contribution of *visible* disk matter is better known.

Amongst the first of these studies, Crézé et al. (1998) used luminous A stars within 125 pc of the Sun – stars bright enough to be seen at meaningful distances, but not so young as to be influenced by velocity inhomogeneities as a result of recent star formation.

Crézé et al. derived a local 'dynamical' density of $0.076 \pm 0.015\,M_\odot\,\mathrm{pc}^{-3}$, and a revised assessment of the visible matter density of $0.085\,M_\odot\,\mathrm{pc}^{-3}$. These dynamical and visible estimates being broadly compatible, they argued that (with a dark massive halo still required to explain the Galaxy's rotation curve) the halo must be, to a first approximation, largely spherical.

Along with other Hipparcos-based studies, using different star samples and methodologies (e.g. Bienaymé, 1999; Korchagin et al., 2003; Soubiran et al., 2003; Holmberg & Flynn, 2004; Bienaymé et al., 2006), another important consensus emerged: that dark matter is distributed largely in the form of the halo, with little if any concentrated in the disk.

The pre-Gaia status is reviewed by Read (2014).

WITH ITS huge stellar samples and accurate space motions, it is not surprising that Gaia is opening a new chapter in this sort of dynamical study.

One of the first to tackle this problem with Gaia was a study by Widmark (2019) using DR2 (a preliminary study along similar lines based on DR1 was made by Widmark & Monari (2019)). His model for the total matter density assumed that it is symmetrical, smooth, and monotonically decreasing with distance from the mid-plane. His (Bayesian) model accounts for the position and velocity of each individual star, in a joint fit of the vertical velocity distribution and stellar number density distribution.

He did this for eight separate data samples, with different limits in absolute magnitude, each containing about 25 000 stars. In all cases, he inferred a density distribution strongly peaked within 60 pc of the Galactic plane. Assuming a baryonic model and a dark matter halo of constant density, this corresponds to an excess surface density of $5-9 M_\odot$ pc^{-2}. He concluded that there is a surplus of matter close to the Galactic plane, perhaps due to an underestimated presence of cold gas.

The model also provides an estimate of the Sun's position and vertical velocity with respect to the Galactic plane, of $Z_\odot = 4.8 \pm 2.3$ pc and $W_\odot = 7.2 \pm 0.2$ km s^{-1}.

A SIMILAR BUT MORE extensive study using DR2 was made by Widmark et al. (2021a). They repeated the analysis for 120 stellar samples in 40 spatially separate sub-regions of the solar neighbourhood. And by excluding areas of known open clusters, they aimed to quantify the sorts of spatially dependent systematic effects that have complicated this type of measurement in the past.

Their results reveal an unexpected but clear trend for all 40 spatially separated sub-regions, implying a total matter density distribution that is highly concentrated towards the Galactic mid-plane (< 60 pc), but decaying rapidly with height, and with a dependence on Galactic radius consistent with a disk scale length of a few kpc.

They suggested, in particular, that the very low matter density inferred above 300 pc is inconsistent with the observed scale height of the stellar disk, and inferred a time-varying phase-space structure that is large enough to affect all stellar samples in the same way.

GAIA IS ALSO providing new possibilities for determining the total matter density of the disk. Widmark et al. (2020) have proposed using 'stellar streams' passing through or close to the Galactic plane. They argue that the vertical component of energy for (dynamically cold) stream stars is approximately constant, such that the vertical positions and vertical velocities of their stars are related via the disk density.

This does not require the disk to be in dynamical equilibrium, and furthermore makes it possible to measure the surface density at large distances from the Sun.

SUCH METHODS that avoid assumptions of dynamical equilibrium are gaining importance as other Gaia data clearly indicate such dis-equilibrium (e.g. Salomon et al., 2020). And here we must make a short excursion.

A fascinating discovery using Gaia data has been the identification of what is called the Galaxy's 'phase-space spiral', first reported by Antoja et al. (2018) for stars that lie within 100 pc from the Sun, and subsequently confirmed to hold over a more extended volume of the Milky Way disk (e.g. Bland-Hawthorn et al., 2019). Faint spirals appear in the number density of stars projected onto the $z-v_z$ plane, which are even more prominent when characterised by their radial and azimuthal velocities.

While suggesting that our Galaxy's disk is locally out of dynamical equilibrium, the mechanism driving these phase-space spirals remains uncertain. One suggestion is that they can be (at least 'briefly') generated by a passing satellite, perhaps by a recent encounter with the Sagittarius dwarf galaxy in the case of our own.

Another is a long-lasting phenomenon which, according to N-body simulations, seems to originate spontaneously early in the life of even isolated Milky Way-type galaxies (Raha et al., 1991; Khoperskov et al., 2019): an initially axi-symmetric disk develops a prominent central bar in 2–3 rotations which, after 0.5–1 Gyr, becomes vertically unstable and buckles, breaking the symmetry with respect to the equatorial disk midplane.

First efforts in using such time-varying structures to measure the vertical gravitational potential of the Galactic disk yield a local halo dark matter density of $0.0085 \pm 0.0039 M_\odot$ pc^{-3}, and an upper limit on the surface density of any thin dark disk with a scale height < 50 pc of about $5 M_\odot$ pc^{-2} (Widmark et al., 2021b). They argue that these time-varying dynamical structures '*... can also be regarded as assets containing useful information*'.

IN A BROADER CONTEXT, the Sun's distance from the disk mid-plane, and its vertical velocity and oscillation period, probably have some connection to our own solar system's habitability, for example in relation to the frequency of stellar fly-bys (and therefore cometary impacts), or the incidence of ultraviolet or X-ray radiation.

And the local dark matter density is an important quantity for dark matter detection experiments. Other estimates come from our Galaxy's rotation curve (reviewed by Sofue, 2020), and from the analyses of halo stars, also most recently from Gaia (Wegg et al., 2019).

In their review, de Salas & Widmark (2021) conclude that most local analyses coincide within a range:

$\rho_{\mathrm{DM},\odot} = 0.4 - 0.6$ GeV cm^{-3} $= 0.011 - 0.016 M_\odot$ pc^{-3}

while more global studies give a slightly lower range:

$\rho_{\mathrm{DM},\odot} = 0.3 - 0.5$ GeV cm^{-3} $= 0.008 - 0.013 M_\odot$ pc^{-3}

(where $1 M_\odot$ pc$^{-3} = 37.5$ GeV cm^{-3}).

Future Gaia data releases will certainly allow many more advances in these important areas.

76. Gaia Data Release 3

TODAY, 13 JUNE 2022, marks the release of Gaia DR3, a very important milestone in the content and availability of the Gaia mission results.

In this summary, I will recall the context, and list some of the key statistics of this data release, otherwise simply referring to the relevant ESA Gaia www pages where `full details of DR3`, and much more supplementary information, is available.

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 specified scanning 'law', thus providing a fairly uniform scanning of the celestial sphere, as well as a progressively more robust 'separation' of the astrometric parameters (position, parallax, and proper motion) of each star. In simple terms, as more measurements are accumulated over time, the accuracy of these various quantities improves.

With its outstanding scientific results already clearly demonstrated, ESA's advisory committees approved the extension of mission operations beyond its original target of 5 years. If the operational lifetime is limited only by its cold gas propulsion system (responsible for driving the scanning law), Gaia can be expected to operate for around 10 years in total, to around 2023–24.

IN MY EARLIER essay #10 (8 March 2021) I gave a summary of the contents of the three previous data releases: DR1 of 14 September 2016, resulting from the first 14 months of satellite observations (Jul 2014–Sep 2015); DR2 of 25 April 2018, from the first 22 months of satellite observations (Jul 2014–May 2016); and EDR3 of 3 December 2020, from the first 34 months of satellite observations (Jul 2014–May 2017).

Data Release 3 combines, for the same stretch of time and the same set of observations as EDR3 (Early Data Release 3), the same primary data, i.e. with no new astrometry nor new photometric calibrations.

The importance of Data Release 3 are the many new data products, including non-single stars and solar system objects, inferred from these observations.

SO, PRECISELY as for EDR3, DR3 contains the following astrometric and photometric information:

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

six-parameter astrometric 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.

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.

THIS information (again, already available in EDR3) is summarised in the following table:

Gaia DR3	
Observations:	
– time period	Jul 2014–May 2017
– observations duration	34 months
– reference epoch	J2016.0
– catalogue release date	13 June 2022
Astrometry:	
– total number (3–21 mag)	1,811,709,771
– 5-parameter solutions	585,416,709
– 6-parameter solutions	882,328,109
– 2-parameter solutions	343,964,953
Photometry:	
– mean G magnitude	1,806,254,432
– mean G_{BP} photometry	1,542,033,472
– mean G_{RP} photometry	1,554,997,939
Radial velocities (4–13 mag)	7,209,831

WHAT IS spectacular about the new DR3 content is, for example, the radial velocity information, and the classification of variability, spectral types, non-single objects, quasars, and solar system object properties, as summarised in numerical terms below:

New results in Gaia DR3	
Sources with radial velocities	33 812 183
Sources with mean G_{RVS}-band magnitudes	32 232 187
Sources with rotational velocities	3 524 677
Mean BP/RP spectra	219 197 643
Mean RVS spectra	999 645
Variable-source analysis	10 509 536
Variability types (from machine learning)	24
Classified variables	9 976 881
Cepheids	15 021
compact companions	6 306
eclipsing binaries	2 184 477
long-period variables	1 720 588
microlensing events	363
planetary transits	214
RR Lyrae stars	271 779
short-timescale variables	471 679
solar-like rotational variables	474 026
upper-main-sequence oscillators	54 476
active galactic nuclei	872 228
Variable with radial-velocity time series	1 898
Sources with object classifications	1 590 760 469
Stars with emission-line classifications	57 511
Astrophysical parameters (BP/RP spectra)	470 759 263
Astrophysical parameters (unresolved binary)	348 711 151
Spectral types	217 982 837
Evolutionary parameters (mass and age)	128 611 111
Hot stars with spectroscopic parameters	2 382 015
Ultra-cool stars	94 158
Cool stars with activity index	1 349 499
H-alpha emission measurements	235 384 119
Astrophysical parameters from RVS spectra	5 591 594
Chemical abundances from RVS spectra	2 513 593
Diffuse interstellar band in RVS spectrum	472 584
Non-single (astrometric, eclipsing, etc.)	813 687
orbital astrometric solutions	169 227
orbital spectroscopic solutions	186 905
eclipsing binaries	87 073
QSO candidates	6 649 162
redshifts	6 375 063
host galaxy detected	64 498
host surface brightness profiles	15 867
Galaxy candidates	4 842 342
redshifts	1 367 153
surface brightness profiles	914 837
Solar system objects	158 152
epoch astrometry (CCD transits)	23 336 467
orbits	154 787
BP/RP reflectance spectra	60 518
planetary satellites	31
All-sky Galactic extinction (HEALPix levels)	6, 7, 8, and 9

THE NUMBERS are colossal, and some further details are given at the above-referenced ESA www site.

To give just a flavour, the data release includes object classification for 1.59 billion sources, i.e. the majority of the objects observed. And a wealth of astrophysical data has been distilled from the BP/RP spectra, notably astrophysical parameters ($T_{\rm eff}$, $\log g$, [M/H], and reddening) for 470 million objects, Hα emission for 235 million stars, spectral types for 217 million stars, evolutionary parameters (mass and age) for 128 million, activity indices for 1.3 million cool stars, emission-line classification for 57 000 stars, and spectroscopic parameters for 2.3 million hot stars and 94 000 ultra-cool stars.

The 10 million variables, along with their associated 'epoch photometry', have been classified (by 'supervised machine learning') into 24 classes, amongst which are 363 microlensing events and 214 planetary transits.

Amongst the solar-system results for 158 000 sources (including 31 planetary satellites) are orbital solutions and individual epoch observations for 154 000 objects, along with mean BP/RP reflectance spectra for more than 60 000 objects.

There are 6.6 million quasar candidates, with redshift estimates for most. Amongst these, the host galaxies are detected for 60 000, and 15 000 of these even have estimated surface brightness profiles.

Other highlights are all-sky Galactic extinction maps at four different spatial resolutions (HEALPix levels 6, 7, 8, and 9), and the Gaia Andromeda Photometric Survey (GAPS), consisting of the photometric time series for all 1.2 million sources located within a 5.5-degree radius field centred on the Andromeda galaxy.

MORE THAN 30 papers in the journal *Astronomy & Astrophysics* are accompanying and detailing these Gaia DR3 data. Also listed are 10 'performance verification papers', which give an overview of the science potential of Gaia DR3, and an introduction to a few of the science topics that can be addressed with it.

Topics include mapping the Galaxy's asymmetric disk; pulsations in main-sequence OBAF stars; reflectance spectra of solar system bodies; the extragalactic content; chemical cartography of the Milky Way; a 'golden sample' of astrophysical parameters; and mapping the diffuse interstellar bands at 862 nm.

ALL THIS PROVIDES a stunning preview of the science that lies ahead, as researchers around the world start to delve deep into the Gaia treasure trove.

From my retirement armchair, I express again my huge admiration of the Gaia Data Processing and Analysis Consortium and its members, some 450 scientists across Europe, who are working together, and often under very great schedule pressures, to deliver this remarkable 21st century view of our Galaxy!

77. The Galactic escape velocity

WHAT IS THE TOTAL MASS of our Galaxy? How far out does our Galaxy halo extend? The distribution of stellar velocities, and in particular the 'escape' velocity from the solar neighbourhood, holds a number of clues.

Indeed, the Galaxy's escape velocity has provided an important constraint on Galaxy models since the studies of J.C. Kapteyn and his doctoral student Jan Hendrik Oort nearly a century ago. But as perceptively pointed out by Oort (1928): *'It is very unlikely that the galactic system as a whole should contain a considerable number of stars whose velocities exceed the velocity of escape'*.[1]

MOST DYNAMICAL measurements are rather insensitive to the mass distribution in the halo's outer parts, since this exerts no net force as long as the halo is spherical or ellipsoidal. Nonetheless, local measurements can provide a constraint on the halo extent, since the escape speed from the Galaxy should exceed the largest speed of any star in the solar neighbourhood.

There are at least two caveats: (a) stars may be gravitationally bound to the Local Group of galaxies, and just passing through our Galaxy; (b) dynamical interactions involving binary stars in clusters can lead to ejected 'runaway stars' at above the escape velocity.

I WILL SKIP MOST of the mathematic details (see, e.g., Binney & Tremaine, 1987; Equation 2–192) and start by simply quoting some illustrative results for a simplified model in which the Galaxy is spherical, with a circular speed Θ_c constant out to a maximum radius r_*.

A number of nearby high-velocity stars have velocities, with respect to an inertial frame, of $\sim$500 km s^{-1}, and estimates between 1980–90 put the escape speed at around $v_{\rm esc} = 400 - 640$ km s^{-1}. Assuming, for example, $v_{\rm esc} = 500$ km s^{-1} (with $\Theta_c = 220$ km s^{-1} and $R_0 = 8.5$ kpc), yields an outer halo limit of $r_* = 4.9R_0 \simeq 41$ kpc, and a total Galaxy mass of $4.6 \times 10^{11}\,M_\odot$.

[1] I can't resist noting that Jan Oort's `Wikipedia entry` includes a photograph of him, on horseback, reconnoitring the Chilean mountains for a location for the ESO observatory! And that, in my close involvement with Leiden University, I also had the pleasure of several conversations with him in his later years.

A STATISTICAL technique to infer the escape speed from the high-velocity tail of a uniform star sample was given by Leonard & Tremaine (1990). Their estimates were, however, strongly correlated with the *shape* of the high-velocity tail and, in turn, very sensitive to the errors on distances and proper motions.

Notwithstanding these limitations, their estimates from radial velocities were in the range 450–650 km s^{-1}. But they also argued that it is not possible to estimate the mass of the Galaxy using the local escape speed without detailed knowledge of how mass is distributed beyond to the solar circle. Even today, at least pre-Gaia, the mass of our Galaxy was still considered unknown to within a factor of four (Bland-Hawthorn & Gerhard, 2016).

USING SPACE MOTIONS from Hipparcos, and radial velocities from Coravel, the same technique was applied by Meillon et al. (1997) to a set of 5307 F–M stars, including all the subdwarfs and high-velocity stars known at the time. Estimates from the total space velocities were in the slightly smaller range $v_{\rm esc} = 440 - 490$ km s^{-1}, probably due to the absence from the Hipparcos catalogue of the (faintest) highest velocity stars in the reference compilation of Carney et al. (1994).

Meillon (1999) found only 10 Hipparcos stars with $V_{\rm tot} > 350$ km s^{-1}. They were, however, able to use the intersection of the two samples (770 stars) to re-calibrate the absolute magnitudes and hence derive improved distances of about 20 non-Hipparcos stars. Together they found 98 stars with $V_{\rm radial} \geq 250$ km s^{-1}, 33 with $V_{\rm tangential} \geq 300$ km s^{-1}, and 24 with $V_{\rm total} \geq 350$ km s^{-1}. Four had $V_{\rm total} \geq 400$ km s^{-1}, and the fastest had $V_{\rm total} = 458$ km s^{-1}. They concluded that the escape velocity is unlikely to exceed 530 km s^{-1}.

Along with estimates from the RAVE radial velocity survey of $v_{\rm esc} \simeq 544^{+64}_{-46}$ km s^{-1} by Smith et al. (2007), and of 533^{+54}_{-41} km s^{-1} by Piffl et al. (2014), this largely represented the state of knowledge pre-Gaia.

Importantly, working the other way around, detailed state-of-the-art models of our Galaxy's structure and mass distribution yielded consistent escape speeds, e.g. 550 km s^{-1} according to Kafle et al. (2014).

WITH THE ONGOING availability of the Gaia results, we are therefore in a position to ask: Do the Gaia results provide an improved estimate of the escape speed from the Milky Way? Given the complex structure of our Galaxy, are there observable changes in the estimated escape speed as a function of Galactocentric radius? And do these estimates provide new constraints on our Galaxy's mass or mass distribution?

Of course, I would not be posing these questions if the answers were not respectively: yes, yes, and yes!

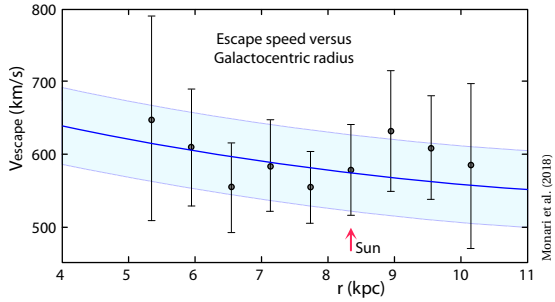
BEFORE LOOKING at the latest Gaia results, it is useful to clarify what is meant by 'the escape speed'. In principle, it could formally be defined as the velocity necessary to bring an ejected star to infinity. In practice, galaxy haloes (including our own) are not isolated systems, so that a limiting distance needs to be specified. A common definition adopts the virial radius, defined as 200 times the critical density, r_{200}, in turn containing the mass M_{200} (see, e.g., Deason et al., 2019, Section 3.2).

Another point to emphasise is that the properties of galaxy haloes are today interpreted in the context of extensive numerical simulations. For example, Auriga is a suite of high-resolution Milky Way-mass haloes, spanning a mass range $M_{200} = 1-2\times10^{12}\,M_\odot$, and employing cosmological parameters consistent with the Planck Collaboration's (2014) data release (Grand et al., 2017). It has been successful in reproducing a number of observational properties of both central disks and stellar haloes, including the rotation curves, stellar masses and star formation rates of disks, and the kinematics and number density profiles of stellar haloes.

USING THE method of Leonard & Tremaine from 1990, Monari et al. (2018) used the Gaia DR2 data to estimate the escape speed over a range of Galactocentric radii from 5–10.5 kpc, based on an assumed distance of the Sun from the centre of the Galaxy of $R_0 = 8.34$ kpc, a circular velocity at the Sun of $\Theta_c = 240\,\mathrm{km\,s^{-1}}$, and specific values for our 'peculiar' solar motion with respect to the local standard of rest.

They used the velocity distribution of 2850 counter-rotating halo stars within 5 kpc, with distance errors below 10%, and with known line-of-sight velocities; the choice of counter-rotating stars is to ensure that the kinematic tail is representative of the stellar halo. They showed that the escape velocity decreases with Galactocentric radius, as expected in any reasonable Galactic potential, with a value at the Sun of $580\pm63\,\mathrm{km\,s^{-1}}$.

Current models of our Galaxy's structure typically comprise a bulge, a disk, and a dark matter halo. Interpreted in the context of the Navarro–Frenk–White dark matter halo profile, and using the mass–concentration relation of ΛCDM cosmology, gave the model parameters $M_{200} = 1.55^{+0.64}_{-0.51}\times10^{12}\,M_\odot$ and $c_{200} = 7.93^{+0.33}_{-0.27}$ (along with the uncertainty bands shown in the figure).



FURTHER ANALYSIS using DR2 was made by Deason et al. (2019). They used the inferred assembly history and phase-space distribution of halo stars to constrain the high-velocity tail of the stellar halo.

They found it to be strongly dependent on the velocity anisotropy and number density profile of the halo stars, with highly eccentric orbits having more extended high-velocity tails. Indeed, halo stars in the solar vicinity are known to have a strongly radial velocity anisotropy, considered attributable to the early accretion of a massive dwarf satellite galaxy of around $10^9\,M_\odot$.

Accordingly, and from their sample of 2300 (also counter-rotating) stars within 3 kpc, of which 240 have $v_{\rm tot} > 300\,\mathrm{km\,s^{-1}}$, they determined a local ($R_0 = 8.3$ kpc) escape speed of $528^{+24}_{-25}\,\mathrm{km\,s^{-1}}$, which they then also used to estimate the total Milky Way Galaxy mass, $M_{200} = 1.00^{+0.31}_{-0.24}\times10^{12}\,M_\odot$, and dark halo concentration parameter, $c_{200} = 10.9^{+4.4}_{-3.3}$.

A LARGER sample of 10^7 halo stars, selected on the basis of their distances, photometry and proper motions, was similarly used by Koppelman & Helmi (2021). Their much larger sample provides a more *precise* estimate of the escape velocity, of $497\pm8\,\mathrm{km\,s^{-1}}$, which they estimate is biased low by 10%, yielding a true escape velocity most likely closer to $550\,\mathrm{km\,s^{-1}}$. Then assuming $\Theta_c = 232.8\,\mathrm{km\,s^{-1}}$, and correcting for their estimated bias, they derived $M_{200} = 1.11^{+0.08}_{-0.07}\times10^{12}\,M_\odot$, and $c_{200} = 11.8\pm0.3$. We can conclude, at least, that estimates of the Galaxy's escape velocity seems to be converging on a Milky Way mass of about $10^{12}\,M_\odot$.

DO THESE latest results have any implications for ongoing dark matter detection experiments, where the exclusion limits are affected by assumptions on the local dark matter escape velocity, the local dark matter density, and the circular velocity of the Sun?

Wu et al. (2019) found that the astrophysical uncertainties are, again, dominated by the uncertainty in the escape velocity at dark matter masses below 6 GeV, and can lead to a variation of 6 orders of magnitude in the exclusion limits at 4 GeV. They nonetheless found that the best-fit value for the escape velocity from Monari et al. (2018) leads to only minor changes on, e.g., the Gran Sasso XENON1T exclusion limits (Aprile et al., 2017).

78. Gaia's first exoplanets

WITH THE RECENT Gaia Data Release 3 (DR3, 13 June 2022), an abundance of new science is already emerging. While the first two *transiting* planets have recently been reported, based on Gaia photometry (Panahi et al., 2022), I want to look here at the first of Gaia's exoplanet discoveries based on its *astrometry*.

In my essay #19 (10 May 2021) I looked at the overall status of exoplanet discoveries, how Gaia's accurate astrometry can hope to discover many more, and also at the rather troubled history of this (in principle powerful, yet extremely challenging) approach.

To set this in context, let me recall that there are various discovery methods that astronomers have developed in the past three decades – including use of radial velocities, photometric transits, gravitational microlensing, and direct imaging. As of mid-2022, out of more than 5000 exoplanets now known, the transit method has discovered nearly 3500, the majority of these being with NASA's Kepler satellite, operated between 2009–18.

WHILE ASTROMETRY, most notably from space with Hipparcos and Hubble's Fine Guidance System, has confirmed a few exoplanets discovered by other methods, and determined improved orbits for a handful, `NASA's exoplanet archive` still only lists just one exoplanet system *discovered* by astrometry (Jean Schneider's Extrasolar Planets Encyclopaedia lists a further 14).

This object, DENIS–P J082303.1–491201, was discovered via its 246-day orbital motion using VLT–FORS2 imaging by Sahlmann et al. (2013). The host is a low-mass (0.07 solar mass) 'ultra-cool' L dwarf at a distance of just 20 pc, while the 'planet' itself weighs in at a hefty 28 Jupiter masses. Whether it is best considered a planetary system, or the low-mass tail of a stellar binary, is perhaps open for debate.

But with very accurate astrometry, exoplanet discoveries should flow in abundance. But it must be accurate! Even space astrometry with Hipparcos, at the level of 1 milli-arcsec, was demonstrably insufficient to reveal the stellar reflex motion of planets orbiting main sequence stars like our Sun, even within 10–20 pc.

USING THE BEST estimates of exoplanet occurrences, and detailed simulations of the Gaia instrument and its sky scanning, some 20 000 long-period Jupiter-mass planets could be discoverable out to 500 pc for the nominal 5-yr mission (including some 1000–1500 around M dwarfs out to 100 pc), rising to 50–70 000 for a 10-yr mission (Perryman et al., 2014a). In a stroke, astrometry could become the most powerful discovery method in astronomers' current arsenal (assuming that the instrument can be appropriately calibrated).

GAIA DATA RELEASE 3 is in two parts: Early Data Release 3 (EDR3, 3 December 2020), and the full Data Release 3 (DR3, 13 June 2022). As described in my essay #10, both use only observations between July 2014 and May 2017, i.e. a period of 34 months (compared with 14 months for DR1, and 22 months for DR2).

I have summarised the huge amount of new information in Data Release 3 in essay #77. Here I will concentrate on the processing of non-single stars and, in particular, the very first results on exoplanet astrometry.

BINARY AND MULTIPLE stars are ubiquitous throughout our Galaxy. Their census and characterisation is important for deriving certain physical properties, and in understanding the processes of star formation and evolution. Companions with masses below the threshold for nuclear fusion comprise the classes of brown dwarfs and, below about 20 Jupiter masses, planets.

Over the past century, binary systems have been discovered through a variety of methods, and they can be loosely classified into visual binaries (where both components are 'visible'), astrometric (where the companion's presence is revealed by the photocentric motion), spectroscopic (revealed as orbital variations of spectral lines), or eclipsing (where the light from one component is eclipsed by the orbital motion of the other).

With its combination of unprecedented astrometric and photometric accuracy, Gaia is in a position both to discover, and to parametrise, most of these types, simultaneously and uniformly, in substantial numbers.

BEHIND THE SCENES, Coordination Unit 4 of the Gaia Data Processing and Analysis Consortium, today led by Frédéric Arenou, is (as one of the nine data processing coordination units) dedicated to the treatment of three 'special' celestial source types observed by Gaia: non-single stars, solar system objects, and galaxies.

The details are, of course, non-trivial, and the 57-page (400+ author) paper published two weeks ago by Gaia Collaboration et al. (2023a) gives some insight into the complexities. Further details of the astrometric orbit determination in Gaia DR3 are given by Holl et al. (2023).

In the case of non-single stars, the multi-epoch astrometric observations (typically some 40 per object distributed over the 34-month time-span of DR3) are examined, and the solutions organised into four classes: those with measured orbits (whether astrometric, spectroscopic or eclipsing); those whose astrometric motion is (currently) better described using a quadratic or cubic rather than a linear proper motion; those that suggest long-period solutions of spectroscopic binaries; and those that display astrometric (positional) variations due to the variability of one component.

Skipping over the details, the resulting numbers are quite staggering: Gaia DR3 contains about 800 000 solutions with either orbital elements or trend parameters (whether for astrometric, spectroscopic and eclipsing binaries, or some combination thereof).

Of these, some 130 000 are full orbit solutions, while more than 300 000 show non-linear astrometric motion, i.e. indicative of partial orbit coverage. All will be substantially improved in future data releases as more data is included, and as the calibration models improve. And many of the present non-linear solutions will progress to full orbit solutions as the temporal baseline improves.

THE SCIENTIFIC INTEREST of these binary and multiple systems is immense: for example, they can be used to examine multiplicity across the entire Hertzsprung–Russell diagram, to clarify the statistics of white dwarf or brown dwarf companions, to quantify their fundamental properties such as mass and radius, to enable studies of, e.g., orbital period versus eccentricity, and to compare the bountiful results with detailed models of star formation and stellar evolution.

To hint at the improvements that Gaia is achieving, there are a factor of around 40 more astrometric orbit solutions in Gaia DR3 as compared to those in the (heterogeneous) *Sixth Catalog of Orbits of Visual Binary Stars* compiled by Hartkopf et al. (2001). And, reassuringly, almost all the Hipparcos targets with an acceleration (7-or 9-parameter) solution also show some sort of significant proper motion 'anomaly' in Gaia DR3.

I will look here at what these early results are starting to convey about the presence of sub-stellar mass companions, brown dwarfs and exoplanets, in Gaia's survey.

THE CATEGORY OF sub-stellar companions comprises brown dwarfs and planets, the division between the two being a complex topic but broadly and simplistically held to occur at about 20 Jupiter masses. Both have been the target of long-term radial velocity searches over the past 20–30 years. Today, with Gaia DR3, astrometry at last reaches the sensitivity to detect sub-stellar companions in very large numbers. And it enables, for the first time, determination of their 3-d orbital architectures, and true (rather than projected) companion masses.

Various tests and selection criteria by Gaia Collaboration et al. (2023a) currently yield 1843 brown dwarf and 72 exoplanet *candidates*. Of these, only 10 brown dwarfs and 9 exoplanets were previously known from radial velocity surveys, all with a reasonable agreement between (for example) their inferred orbital periods.

I will ignore here the tricky problem of fully validating these candidates (essentially, ruling out a binary star with similar mass components), and I will defer a discussion of the brown dwarf results to a future essay.

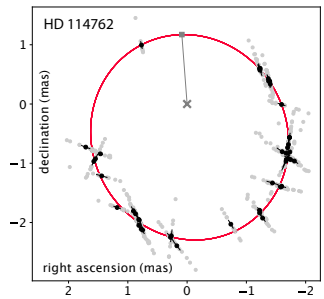
OF THEIR CANDIDATES, Gaia Collaboration et al. (2023a) list just two with *validated* orbits implying the presence of new planetary companions:

- HIP 66074 with $P = 297 \pm 2.8$ d, $e = 0.46 \pm 0.17$, and $a_0 = 0.21 \pm 0.03$ mas, of 7.3 ± 1.1 Jupiter masses; and
- HIP 28193 with $P = 827 \pm 50$ d, $e = 0.07 \pm 0.10$, and $a_0 = 0.25 \pm 0.02$ mas, of 5.3 ± 0.6 Jupiter masses.

Their sub-milli-arcsec semi-major axes are especially noteworthy, with their smallest candidate having $a_1 = 0.14 \pm 0.04$ milli-arcsec (or 140 micro-arcsec!).

Amongst their other candidates are NASA's only current astrometric listing DENIS–P J082303.1–491201, along with earlier radial velocity discoveries including GJ 876, HD 114762, HD 162020, and HD 164604. Others add to the very rare class of giant planets orbiting white dwarfs, including WD 0141–675 which, at 9.8 pc, is one of the metal-enriched systems suggestive of the capture of planetary debris (see my essay #73).

The detailed treatment by Holl et al. (2023) includes a number of impressive and compelling examples of the fitted orbits, including HD 114762 (7.15 mag, with $P = 83.74 \pm 0.12$ day, $e = 0.32 \pm 0.04$, and parallax 25.36 ± 0.04 mas) shown here.



Holl et al. (2022)

I'D LIKE to add my own congratulations to my long-term colleagues, Frédéric Arenou and Berry Holl, and their co-authors, on these spectacular results which hint at the wonderful future for exoplanet science being augmented by Gaia.

79. More insights into non-single stars

In my previous essay (#78, 27 June) I looked at just two of the papers on results from Gaia DR3 (released on 13 June 2022), and focused on the impact that Gaia is now having on the astrometric discovery of exoplanets.

These early results are emerging from the collective work of just one of the three sub-groups of 'Coordination Unit 4'. As I noted there, CU4 handles the downstream treatment of sources which demand a very different astrometric analysis compared to that of otherwise 'normal' stars in our Galaxy and beyond, specifically for non-single stars, solar system objects, and galaxies.

By focusing on those first results on exoplanets, I was picking out just a part of the work of this CU4 subgroup, and its comprehensive treatment of non-single stars. Here, I'll look at some other results.

First, let me provide the context to this essay, by repeating what I noted in essay #78. Over the past century, binary (and multiple) star systems have been discovered through a variety of methods, and they can be loosely classified as *visual* binaries (where both components are 'visible'), *astrometric* (where the companion's presence is revealed by the photocentric motion), *spectroscopic* (revealed as orbital variations of spectral lines), or *eclipsing* (where the light from one component is eclipsed by the orbital motion of the other).

With its combination of unprecedented astrometric and photometric accuracy, Gaia is in a position both to discover, and to characterise, most of these types, simultaneously and uniformly, and in substantial numbers.

Indeed, in their overall treatment of non-single stars, many other interesting results are starting to emerge. In this essay, I'll pick out a few of these described in the recent (DR3) paper on the non-single star results by Gaia Collaboration et al. (2023a): specifically considering stellar masses, brown dwarf companions, compact companions, and ellipsoidal and other variations.

Most of these topics really merit much more space, and I'll undoubtedly return to some of them in more detail in the future. Meanwhile, this is a taster!

Stellar masses: Given their fundamental role in models of stellar structure and evolution, it may seem surprising that accurate masses are known for only around a hundred stars, a situation that has advanced only modestly since the review by Andersen (1991).

This is because a star's mass can only be directly determined from the orbital motion of a binary system. Total system masses, $M_1 + M_2$, can be estimated when the period (P) and linear semi-major axis (a) of an orbital binary are known. Accurate total ($M_1 + M_2$) masses are therefore known for many systems, although the cubic dependency on a implies that errors are very sensitive both to parallax error, and to errors on a and P.

Individual stellar masses, in contrast, can only be measured under special circumstances, specifically for double-lined spectroscopic binaries (SB2 systems), and only by combining the spectroscopic orbit with that from astrometry. This is because astrometry describes motion on the *plane* of the sky (with the 7 classical orbit elements being $P, a'', e, i, \omega, \Omega, T_0$). Spectroscopy describes motion *along* the line-of-sight (with the conventional spectroscopic elements for a double-lined SB2 spectroscopic binary being $P, \gamma, K_1, K_2, e, \omega, T_0$).

Without delving into the precise meaning of these quantities, it turns out that four are in common between the two solutions (P, e, ω, T_0), such that the combination of observations can yield individual component masses. In the case of single-lined (SB1) systems, the mass of the primary has to be assumed (generally inferred from the spectral type). In the case of double-lined spectroscopic binaries (SB2) the mass ratio (M_1/M_2) is determined, and individual masses result from knowledge of the orbit inclination, i, provided by astrometry.

From the Gaia data alone, combining the available astrometric and SB2 orbits yields about 70 component masses (Gaia Collaboration et al., 2023a, Table 3). Many more will become available when the Gaia astrometry is combined with external catalogues of spectroscopic binaries (e.g. the 8105 SB2 systems from APOGEE DR17). Something similar was done for seven SB2 binaries in the Hipparcos catalogue (Arenou et al., 2000).

BROWN DWARF COMPANIONS: Due to the large 'reflex' motion induced in their host stars, early radial velocity surveys were expected to discover brown dwarf companions to solar-type stars with relative ease, and in reasonable numbers. Their masses and orbit sizes would provide important constraints on theories of formation and evolution of brown dwarfs (and planets).

But a feature of early radial velocity exoplanet surveys was the paucity of close-in ($a < 3-4$ au) sub-stellar objects with masses of $10-80 M_J$, known as the *brown dwarf desert* (Marcy & Butler, 2000). Subsequent radial velocity surveys of several thousand stars have confirmed this pattern, finding brown dwarf companions out to $P \sim 10$ yr, but with few in the mass and separation range characterising the 'desert'. Pre-Gaia, only some 100 brown dwarfs were known around F, G, or K stars.

Meanwhile, detailed models constructed before the release of DR3 predicted that Gaia should discover some 30–50 000 brown dwarfs from the astrometric (orbital) perturbations of their host stars (Holl et al., 2022).

Already, from Gaia DR3, Gaia Collaboration et al. (2023a) report 1843 brown dwarf *candidates*, of which only 10 were previously known from radial velocity surveys. Under various assumptions, these first results indicate, for example, that the frequency of brown dwarfs around M dwarfs with $P < 1000$ days is about 0.3%.

COMPACT COMPANIONS: A host star's orbital motion, 'tugged' upon by an orbiting body, does not depend on the orbiting body being visible. This opens the possibility of detecting any sort of body orbiting a star, whether it is visible (another star), of low luminosity (a brown dwarf, white dwarf, or a planet), or any other form of more 'exotic' compact object, such as a neutron star or even a black hole. Indeed, the astrometric orbits of non-single stars in Gaia DR3 are giving the first hints about the existence of all of these various classes of object.

Identification of a compact low-luminosity companion in general depends both on the *mass ratio* of the two components, but also on their *luminosity* ratio, together characterised by what Gaia Collaboration et al. (2023a) describe as their 'astrometric mass ratio function'.

Their histogram of companion masses for more than 700 compact object candidates (their Figure 36) suggests that most of these companions are white dwarfs, with a clear narrow peak at around 0.6 solar masses, as for field white dwarfs. They estimate that Gaia should increase the number of known binary systems with white dwarf components by at least a factor 10.

More exotic still, SB1 sources with a large mass function could point to stars with a neutron star or even black hole companion. It is too early to draw clear conclusions, although they hint that HIP 15429 (amongst others) may be a B5 Ib star with an orbiting black hole, although it may instead be an Algol-type or triple system.

JUST AS NASA'S Kepler satellite revealed many new astrophysical phenomena by virtue of its accurate high-cadence photometry of a very large number of stars, so Gaia is poised to capitalise on its simultaneous combination of high-accuracy astrometry, multi-epoch photometry and, for the brightest stars, radial velocities. I will give four examples of such capabilities.

First, and perhaps most easily appreciated, it is not always straightforward to distinguish a single-lined spectroscopic binary (SB1) from a long-period variable, i.e. to distinguish orbit-induced radial velocity variations from (single star) radial pulsations. Gaia offers new prospects to disentangle the two effects, and examples are given in Gaia Collaboration et al. (2023a, Table 6).

A second example are the 'ellipsoidal variables', close binary stars whose components are physically distorted by their mutual gravitation, but with orbit inclinations too small to create eclipses (such stars are not exactly ellipsoidal in shape but, rather, fill their Roche equipotential surfaces). Resulting light curves, due to the changing projected areas and surface brightnesses of the distorted stars, typically show periodic variability, albeit with amplitudes of only a few per cent at most.

Such ellipsoidal variabilities have also become of interest in exoplanet studies, and in particular based on Kepler photometry, where they were first seen in the bright system HAT–P–7 b, in which a transiting 'hot Jupiter' (of $1.7 M_J$) orbits an F6 star with $a = 4 R_\star$.

Again, Gaia Collaboration et al. (2023a, Table 6) give several examples (including V1540 Cyg) where the preliminary classification as a long-period variable is reclassified as an ellipsoidal variable by noting that the variability period is exactly one half that of its astrometric orbit. Sources for which the two periods are in good agreement generally point to true pulsating stars.

My third example is a class of sources which show a significant (and arbitrary) phase lag between the two periods. This may be revealing binary systems in which the light variation is due to star spots on the primary star, modulated in a spin-orbit synchronised binary system.

Finally, EL CVn systems are short-period eclipsing binaries consisting of an A/F primary and a low-mass pre-He white dwarf secondary. As a rare (brief) stage of binary evolution, they result from mass transfer from the pre-HeWD progenitor to the brighter primary star.

For the known 9.5 mag system HD 23692, the combined Gaia epoch photometry and radial velocity curves (their Figure 28) show that the deeper eclipse actually corresponds to the secondary component (a characteristic of such systems). And the secondary eclipse in G_{BP} is much deeper than that in G_{RP}, providing further confirmation of the higher temperature of the secondary.

GAIA'S NON-SINGLE STARS really are providing a spectacular haul of new astronomical insights!

80. Neutron stars and pulsars

NEUTRON STARS are the collapsed cores of massive stars ($8-40M_\odot$), and they result from a supernova explosion. With radii of only some 10 km, but weighing in at about 1.4 solar masses, their densities, of around 10^{17} kg m^{-3}, correspond to those of atomic nuclei.

Pulsars are the rapidly spinning, highly-magnetised manifestations of neutron stars. They emit narrow beams of radio emission parallel to their magnetic dipole axis, seen as pulses at the object's spin frequency due to a misalignment of the magnetic and spin axes.

There are two broad classes: 'normal' pulsars are isolated objects with spin periods around 1 s. 'Millisecond pulsars' are old (~Gyr) neutron stars, spun-up during mass and angular momentum transfer from a binary companion; most known still have (non-accreting) binary companions, either white dwarfs or neutron stars.

Known pulsars (a tiny fraction of those thought to exist) number around 2600, with more than 80 msec pulsars in the Galaxy disk, and some 130 in Galactic globular clusters. The fastest spin-period is the 716 Hz (1.4 msec) eclipsing binary pulsar PSR J1748–2446ad in Terzan 5. The slowest, PSR J2144–3933, has a period of 8.51 s.

DISTANCES are important for understanding their physics (e.g. Jennings et al., 2018). They provide constraints on the nuclear equation of state (from their radii), and on energy transport in neutron star winds. Space motions point to their birth sites, while transverse velocities constrain the physics of supernova core collapse, as well as the evolution of close binary systems.

Precision astrometry is also crucial in determining accurate pulsar spin-down rates (via their luminosities) as well as their (relativistic) binary orbital decay rates (by properly accounting for the kinematic Shklovskii effect, and that of Galactic acceleration).

However, distances to neutron stars are difficult to measure. For most radio pulsars, they can be estimated indirectly through the observed pulse dispersion measure (the line-of-sight integral of the electron density), although this in turn relies on models of the Galactic electron density distribution.

Distances can occasionally be inferred from their association with star clusters, from spectroscopic parallaxes of binary companions, or from annual variations in pulse arrival times (requiring a timing accuracy of better than a micro-sec for distances of 1 kpc). Parallaxes can also be determined from astrometric imaging observations with VLBI at sub-milliarcsecond accuracy.

Incidentally, other than the Crab pulsar, examined in the context of Gaia DR2 by Fraser & Boubert (2019), and which I will discuss next week (essay #81) focusing on supernovae remnants, no pulsar is intrinsically bright enough at optical wavelengths to be visible with Gaia.

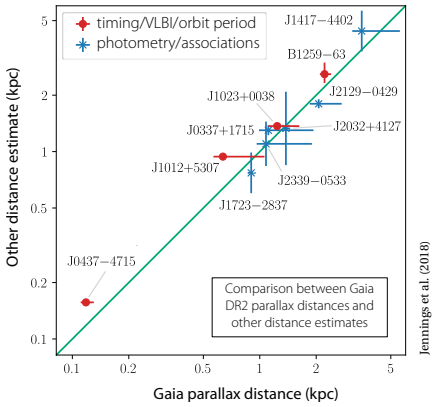
GAIA IS, however, contributing in the specific case of pulsar companions in binary systems, where the brighter orbital companion serves as a proxy, circumventing the problems otherwise arising from their small sizes and general lack of detectable optical emission.

The more than one billion astrometric solutions of Data Release 2 (2018) already included the binary companions to a number of known pulsars, including white dwarf companions to millisecond pulsars and the non-degenerate components of 'black widow' and 'redback' systems (msec pulsars with an ablating companion).

Jennings et al. (2018) worked with the 2424 (non-globular cluster) objects from the ATNF catalogue of known pulsars, selected the 188 in binary systems, and further restricted it to sources that could be securely cross-matched to the Gaia DR2 catalogue on the basis of their combined position and proper motion uncertainties. They found, finally, 22 counterparts in DR2, of which 12 have statistically significant parallax measures.

Amongst them are specific systems used in tests of alternative theories of gravity (PSR J0348+0432 and PSR J0337+1715), in theories of accretion (PSR J1023+0038) and of particle acceleration (PSR B1259–63).

Notwithstanding the fact that some pre-Gaia distance estimates are more precise (if not always more accurate) than those from Gaia DR2, they found a generally reasonable agreement between the previous estimates of parallaxes and proper motions with those from Gaia.



Taken together, the Gaia DR2 results already constitute a small but important improvement in the pulsar distance scale, with the subset of millisecond pulsars with Gaia distances helping to improve the sensitivity of pulsar timing arrays to nano-Hertz gravitational waves.

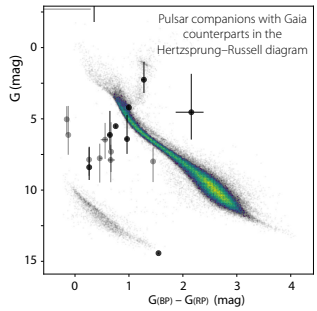
Their results also point to other areas that will be the subject of further detailed study. For example, the distances predicted by both favoured dispersion measure models (NE2001 by Cordes & Lazio 2002, and YMW16 by Yao et al. 2017) appear to be underestimated compared with those from Gaia, confirming earlier suspicions.

Their results on individual objects also identify areas in which several pre-Gaia binary and orbit models merit some revision and refinement. Thus for the nearby system J0437–4715 (actually one of the closest known neutron stars), the Gaia distance is only 77% of that derived from orbital derivative models, and is provisionally attributed to unmodeled binary reflex motion, which should be resolved with future Gaia data releases.

For J1417–4402, the Gaia distance lends support to the inference that its companion fills its Roche lobe, and disfavours the dispersion-measure distance (of 1.6 kpc). For J1816+4510, the Gaia proper motion disagrees with that from pulsar timing, requiring further investigation. For J2032+4127, the wide-orbit, long-period (46 yr) binary may have affected the Gaia astrometric fit, especially for the proper motion term. And for J2215+5135, the Gaia proper motion (1.9 ± 0.8 mas yr^{-1}) is much smaller than the pre-Gaia estimate (189 ± 23 mas yr^{-1}), the latter now considered to be in error.

While the majority of massive stars have a stellar companion, most pulsars appear to be isolated. This implies that most massive binaries break apart due to strong natal 'kicks' received in supernova explosions. But since the monitoring of newly discovered pulsars has rarely been carried out for long enough to conclusively rule out multiplicity, the observed binary fraction can still be subject to strong selection effects. A study also based on Gaia DR2 to improve estimates of the binary fraction has been made by Antoniadis (2021).

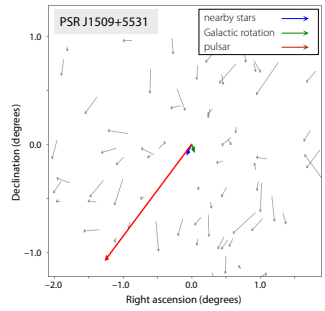
Searching for companions to the 1534 pulsars with positions known to better than 0.5 arcsec, he found 22 matches to known pulsars, and eight new possible companions to young pulsars.



Examining their photometry and kinematics, he confirmed that the observed multiplicity fraction is small, even though the number of binaries in the DR2 sample below the sensitivity of Gaia and radio timing could still be somewhat higher. For example, he constrained the binary fraction of young pulsars to be below about 5%, and set an upper limit on the Galactic neutron star merger rate of 7×10^{-4} per year.

His results on individual objects, including new candidates, again point to areas in which pre-Gaia binary and orbit models merit further revision and refinement.

As already noted, the accurate measurement of pulsar velocities provides important clues concerning supernova physics and the subsequent evolution of binary systems. In this spirit, Yang et al. (2021) also used Gaia DR2 to selected a sample of 24 young (< 3 Myr) pulsars with accurate parallaxes, using them to measure the velocity of their local standard of rest and their local velocity dispersion (by selecting 20 or more Gaia stars within 20–50 pc of each selected source).



The median difference between their calculated local standard of rest and the local Galactic rotation model is 7.6 km s^{-1}, small compared to the typical velocity dispersion of 27.5 km s^{-1}. But for pulsars away from the Galactic plane, the differences grow significantly, up to 40 km s^{-1}.

More importantly, the fact that the stellar velocity dispersion in the vicinity of low-velocity pulsars is often comparable to their transverse velocities, confirms that the intrinsic velocities of neutron star progenitors must be considered when estimating their natal kicks, and their subsequent evolution.

With results to date exploiting only astrometry based on Gaia DR2, and the contributions of Gaia's distances and proper motions already having an impact on statistical studies and on the improved understanding of individual objects, much more in this area can be expected from future Gaia data releases.

81. Supernova remnants

In my previous essay (#80) I looked at Gaia's ongoing contributions to the study of neutron stars, and in particular to the study of pulsars (their rapidly spinning, highly-magnetised manifestations) and, amongst them, to the occurrence of pulsars in orbital binary systems.

As I described there, pulsars are the dense, compact, collapsed cores of massive stars ($8-40M_{\odot}$), which result from a supernova explosion, with the most rapidly spinning 'millisecond pulsars' resulting from mass and angular momentum transfer from a binary companion.

Since a companion star can be 'flung out' during the impulsive burst of the supernova explosion, another important route to the study of neutron stars and pulsars is by searching for high-velocity stars receding from the sites of known supernova explosions. Enter astrometry!

A supernova remnant is the prominent structure resulting from a supernova explosion. The explosion expels much of the remaining stellar material at velocities up to 10% of the speed of light. At these highly supersonic speeds, a strong shock wave forms ahead of the ejecta, heating the ambient interstellar medium to millions of degrees Kelvin. The shock slows as it sweeps up material, but it can expand over centuries, and tens of parsecs, before it slows to below the local sound speed.

Of the 295 supernova remnants catalogued by Green, prominent examples include the Crab Nebula (discovered in 1731, and associated with a supernova recorded by Chinese astronomers in 1054); Tycho (SN 1572, named after Tycho Brahe who recorded its original explosion); Kepler (SN 1604, named after Johannes Kepler); Cas A (from around 1667, in the constellation Cassiopeia, and one of the brightest radio sources); the 40 000-yr old Spaghetti Nebula (discovered in 1952); SN 1987A (a young supernova in the Large Magellanic Cloud observed in February 1987); and G1.9+0.3 (the youngest known in our Galaxy, discovered in 2008).

Of their remnant pulsars, only the Crab pulsar (16.5 mag) is bright enough at optical wavelengths to be observable directly by Gaia. Only a handful of others have known optical emission, the next brightest being PSR B0540–69 (23 mag), and the Vela pulsar (23.6 mag).

Providing a theoretical framework for search programmes, Renzo et al. (2019) estimated that 20–30% of massive binary systems merge prior to the supernova event and, of the remainder, some 80–90% become unbound. Only a minority of those ejected are 'runaways' (loosely defined as velocities above 30 km s^{-1}), with the majority having much lower velocities.

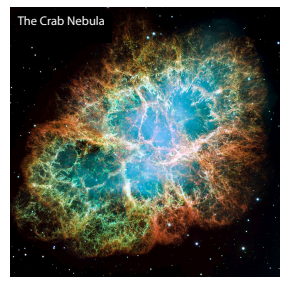
Various searches have been made for these surviving binary companions of massive stars which exploded as supernovae within our Galaxy, focused on finding escaping stars 'pointing back' to nearby supernova remnants.

Much more comprehensive searches are now possible using the Gaia astrometry. Using Gaia Data Release 2, Fraser & Boubert (2019) made an extensive search for surviving binary companions to three of the best-studied supernova remnants in our Galaxy.

Consistent with previous work, they found no plausible binary companion to either the Crab or Cas A supernovae. Examining nearly 80 000 Gaia sources within a 0.5 degree radius, they also ruled out any luminous ($L > L_{\odot}$) companion to the Vela supernova, while noting a faint kinematic companion more likely to be a background star. They concluded that none of these supernovae had a stellar companion at the time of explosion.

The distance to the Crab pulsar and its supernova remnant is still only poorly known, although it is widely assumed to be around 1.93 kpc (Trimble, 1973).

The significantly larger Gaia DR2-based parallax distance, $3.37^{+4.04}_{-0.97}$ kpc, derived by Fraser & Boubert (2019), would have implications for models of its underlying pulsar physics.



Hubble Space Telescope mosaic

They also showed that Gaia's multi-epoch photometry already hints at a measurable secular decrease of the pulsar's luminosity. Being attributed to synchrotron radiation losses, this will also provide new constraints on models of pulsar physics.

HIGH-VELOCITY 'runaway stars' are not only generated in binary supernova events, but also by ejection following the gravitational interaction of stars in dense stellar systems, such as globular clusters or stellar associations. A supernova origin would be invoked if the runaway star's trajectory could be traced back to the position of the supernova remnant (or to its common origin with the remaining neutron star), and further supported (for example) if supernova debris can be identified in the runaway star's atmosphere via high-resolution spectroscopy (e.g. Przybilla et al., 2008).

Using the astrometry of Gaia DR2, Neuhäuser et al. (2020) examined the 68 pulsars with known transverse velocities in the Scorpius–Centaurus–Lupus association. Amongst other results, they connected the runaway star ξ Oph to the radio pulsar PSR B1706–16, tracing both objects back to a common origin 1.78 ± 0.21 Myr ago. They also argued that the moderate distance of the implied supernova, 107 ± 4 pc, makes it likely that this event contributed to the ^{60}Fe deposit that has been detected in the Earth's crust and ocean sediments, and attributed to such a recent nearby supernova (e.g. Knie et al., 1999).

PREVIOUS SEARCHES for this type of runaway star have generally been restricted to luminous O and B type stars because of their young ages, and their detectability out to large distances. The comprehensive Gaia data now makes this type of large-scale search feasible for runaway stars of *all* spectral types. Spectroscopy of any candidates can hope to verify their young ages by means of the standard 'lithium test'.

Lux et al. (2021) made such a search for 12 nearby supernova remnants, including Vela, Puppis A, and the Cygnus Loop. For each, they selected all Gaia stars brighter than 17 mag within a search radius defined by the remnant's age and distance. They then traced back the projected trajectories of each, using their proper motions and the age of the supernova remnant, to obtain their coordinates at the time of the supernova explosion.

Of their 12 selected supernova remnants, they excluded runaway stars brighter than 17 mag in three, confirmed the previously known candidate HD 37424 in Simeis 147 (G180.0–01.7, the 'Spaghetti Nebula') and, from high-resolution spectra of 39 stars, found one or more candidates in each of the remaining eight. Three of these supernovae could possibly have ejected two runaway stars. Further confirmation of these possible candidates will require higher resolution spectroscopy.

A SIMILAR SEARCH of 10 supernova remnants (with many of the targets in common with the study by Lux et al., 2021), but now based on Gaia EDR3, was made by Kochanek (2021). He confirmed the one known example of an unbound binary, HD 37424 in G180.0–01.7, but found no other examples.

The fraction of stars that are in binary or triple systems at the time of a supernova explosion, and the fraction of these companions that survive the explosion, provide crucial constraints for evolution models and predictions for gravitational wave source populations. Kochanek (2021) concluded (for example) that 72% of supernovae producing neutron stars are not binaries at the time of explosion, 14% produce bound binaries, and 12.5% produce unbound binaries.

Incidentally, one of his target supernova remnants, G039.7–02.0 (W50) contains the exotic system SS 433: an eclipsing X-ray binary, with the primary being a stellar-mass black hole, and its relativistic precession resulting in both blue and red Doppler shifts at visible wavelengths. As another example of the present Gaia astrometry being at odds with the pre-Gaia consensus, the distance to SS 433 implied by the Gaia EDR3 parallax (6.1–13.9 kpc) disagrees with estimates from kinematic models of its jets, viz. around 5.5 ± 0.2 kpc (Blundell & Bowler, 2004) or 4.5 ± 0.2 kpc (Marshall et al., 2013).

MY FINAL EXAMPLE of the application of Gaia data to the understanding of supernova remnants relates to our present understanding of the formation of stellar mass black holes. While observations of interacting black hole binaries, as well as the LIGO gravitational wave results, provide clear evidence that stellar-mass black holes exist, there appears to be no immediate prospect of directly investigating their formation. Meanwhile, modern theoretical models suggest that some 10–30% of stellar core collapses fail to result in a supernova, and instead become black holes.

This led Kochanek et al. (2008) to propose a search for failed supernovae by looking for massive stars which suddenly 'vanish'. In contrast to transient searches looking for the appearance of new sources such as supernovae, the idea is to do the opposite – to search for the disappearance of massive stars. With the systematic observation of nearby galaxies in order to monitor some million supergiants, such a disappearance should occur yearly, since these massive stars must end their lives through core collapse within about a million years.

Indeed, from monitoring with the Large Binocular Telescope and others, at least one such candidate has been identified, N6946–BH1 (Basinger et al., 2021). The $300\,000\,L_\odot$ red supergiant progenitor underwent an outburst in 2009, and has since disappeared!

An alternative to searching for these individual progenitors is to use the local stellar population to infer the likely progenitor mass. Kochanek (2022) used the Gaia EDR3 parallaxes of luminous stars in the vicinity of the Vela pulsar to characterise a representative stellar population, starting with 37 000 stars within 250 pc of the pulsar. His analysis concluded that the Vela pulsar progenitor was of low mass, in the range $8.1 - 10.3 M_\odot$.

82. Gaia's galaxy survey

THE PRIMARY OBJECTIVE of the Gaia space astrometry mission is, of course, to characterise the stellar content of our own Milky Way Galaxy.

But passing across the telescope's fields of view as it scans the sky are other objects which can be measured at the same time. Amongst them are more than 150 000 minor bodies of the solar system, and more than 500 000 quasars. For both of these types of object, their (typically) star-like images makes their detection no more complex than (and indeed operationally on-board indistinguishable from) the detection of stars.

Early in the technological study phase of Gaia (in the late 1990s), it also became clear that distant compact galaxies could also be detected, albeit with some revision of the on-board detection and CCD sampling strategies. And it became progressively more evident that there was a strong science case for their inclusion.

To set the scene, I should recall that, even at 20–21 mag, most of the focal-plane CCD pixels are 'empty' of target stars. Indeed, it is a basic principle of the Gaia measurements that all of the CCD pixels are *not* read out and sent to the ground: the penalty for doing so, in terms of CCD read-out noise, and the data volume to be transmitted to the ground, would be prohibitive.

Instead, stars and star-like objects are detected on-board, as they enter the field, by the sky mappers. Above a given signal-to-noise threshold, information on the target's brightness, location, and transverse motion are transmitted to the subsequent focal plane detectors (astrometry, photometry, and radial velocity spectrometer), where only the small region related to the source is read out, and only these 'windows' are transmitted to ground.

THE OPTIMISATION of a detection and sampling strategy for galaxies, i.e. extended objects characterised by faint surface brightness variations, was not at all straightforward. It was the subject of numerous detailed technical reports and iterations, led by Erik Høg, Mattia Vaccari, Frédéric Arenou and Lennart Lindegren. The status at the time of the 2000 Concept & Technology Report was summarised by Vaccari (2002).

Essentially the problem was to optimise three aspects: (i) the angular areas over which the average surface brightness and local sky background values are computed; (ii) the detection area and signal-to-noise limit such that only useful data are transmitted to the ground (and without being swamped by less interesting regions of the Milky Way or the zodiacal light); (iii) a specific sampling scheme for galaxies, providing a trade-off between angular resolution, CCD read-out noise and telemetry volume. As for stars, a larger sample size would yield a smaller error on the mean surface brightness, along with a lower read-out noise and reduced telemetry rate, but also a lower angular resolution.

For example, a typical galaxy of $I = 17$ would be detected about 60% of the time (or some 50 times on average during a 5-year mission) with a signal-to-noise exceeding 4, using a detection area of 2×2 arcsec^2.

A statistical model of galaxy number density, size and surface brightness distribution, was also developed to characterise the 'typical' faint galaxy to be targetted. Results suggested that there could be about three million galaxies brighter than this limit at more than $\pm 15^\circ$ from the Galactic plane (i.e. where galaxy detection should not be compromised by the high density of stars).

THE ULTIMATE GOAL could never be a survey of *all* galaxies, but rather a high-reliability multi-colour catalogue extending to low Galactic latitudes, and with high-angular resolution (~0.18 arcsec) imaging of all sufficiently high-surface brightness galaxies.

The result would represent a homogeneous and well-defined sample for two main purposes: (i) for a statistical analysis of the photometric structure of the central regions of tens of thousands of well-resolved galaxies at an unprecedented angular resolution, and (ii) for the study of the large-scale structure of the local Universe, aiming to further clarify important features such as large filaments, 'walls', and the Supergalactic Plane.

The resulting scientific case occupied two pages of the 100+ pages dedicated to the overall scientific case for Gaia in 2000 (ESA–SCI(2000)4).

DETAILS OF THE final on-board detection and sampling algorithms are given in the paper describing the overall mission (Gaia Collaboration et al., 2016), and in the DR1 documentation (de Bruijne et al., 2017).

With the adopted detection algorithm, unresolved early-type elliptical galaxies and galaxy bulges are preferentially detected, even with radii of several arcsec, while late-type spiral galaxies, even those with weak bulges, will remain mostly undetected. In any case, the sampling function for resolved galaxies is complicated, since the onboard detection depends on the contrast between any point-like, central bulge and any extended structure, all convolved along the scanning direction.

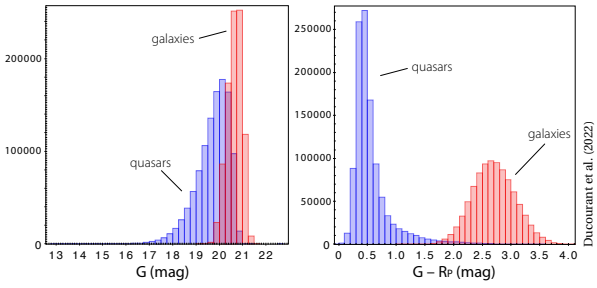
The on-ground processing is carried out within Coordination Unit 4 of the Gaia Data Processing and Analysis Consortium. As one of the nine coordination units, this is dedicated to the treatment of three 'special' celestial source types observed by Gaia: non-single stars, solar system objects, and 'extended objects'.

This latter task, of interest here, treats all sources considered to be extended, aiming to classify the sources and to quantify their morphology. It also analyses all sources separately classified as quasars to search for underlying host galaxies, or other features that could disqualify their use in the reference-frame alignment.

THE RESULTS forming part of Gaia DR3 (released on 13 June) are described by Ducourant et al. (2023). Their adopted surface brightness profile fitting 'pipeline' works by simulating a large set of two-dimensional image profiles, then iteratively searching for the one that best reproduces the one-dimensional observations.

As they point out, although their future list of quasar candidates will be based on the results of Gaia's own classification algorithms, such results were not available for the DR3 processing, which used instead their own compilation of quasar candidates from the existing literature. The sky coverage of each of the merged source catalogues being somewhat heterogeneous, their resulting input list is similarly heterogeneous across the sky.

Their starting list of extended sources similarly contained some 1.7 million galaxies with an entry in Gaia DR3. It is reasonably homogeneous across the sky, except in the Galactic plane which is mostly empty. Their distributions in magnitude and colour are shown here.

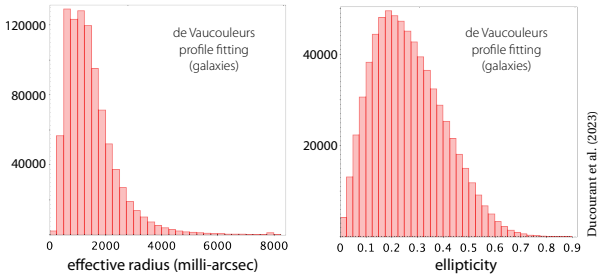


FOR THE CLASS of known quasars, their light profile model allows the source structure to be decomposed into two components: the central quasar and a possible surrounding host galaxy. The central quasar is expected to be point-like, its apparent angular extent being essentially due to the line spread function of the Gaia instrument. It is modelled by a circular exponential profile with a fixed scale length (39.4 mas, corresponding to the instrument's line spread function).

Given that the form of the host galaxy is known to be spiral for distant quasars, but could be bulge-dominated for more nearby quasars, their adopted (Sérsic) profile model has the flexibility to represent either.

For 1 103 691 previously known quasars which were analysed in DR3, the great majority (1 031 607) were classified as point-like, either based on their low integrated flux in the astrometric field data, or on the result of the fitting process. The pipeline processing detected a host galaxy around 64 498 of these, some 6%, and they published the surface brightness profiles of the underlying host galaxy for a subset of 15 867 quasars with robust solutions. The distribution of the Sérsic index describing the light profile of the host galaxies peaks at 0.8, with a mean value of 1.9, indicating that their host galaxies are consistent with disk-like morphologies.

The pipeline also analysed 940 887 galaxies, assuming two different model profiles (a 'de Vaucouleurs' profile $\propto r^{1/4}$, and a Sérsic profile $\propto r^{1/n}$) and provided robust solutions for 914 837. In contrast to the underlying quasar hosts, the distribution of the Sérsic indices confirms that Gaia indeed detects mostly elliptical galaxies, and that very few disks are detected.



WHILE ANY SCIENTIFIC interpretation is a subject for the future, we can only be impressed by the promise of the present results, both in terms of pure numbers, and in the robustness of the profile fitting.

Looking to the future, and in particular to Data Release 4, the number of transits per source will double compared to the data interval used for Gaia DR3, with a corresponding improvement in angular coverage. As a consequence, it is expected that the starting source lists will be at least twice as large as the present ones.

The input list will also improve, complemented by the classification treatment supplied by the relevant DPAC classification unit.

83. The Andromeda photometric survey

THE ANDROMEDA GALAXY, M31, is a large barred spiral galaxy, visible to the naked eye under dark conditions. With a diameter of 65–70 kpc, and at a distance of around 780 kpc, it is the nearest large galaxy to our own (and of similar mass), and indeed the largest member of the Local Group of galaxies.

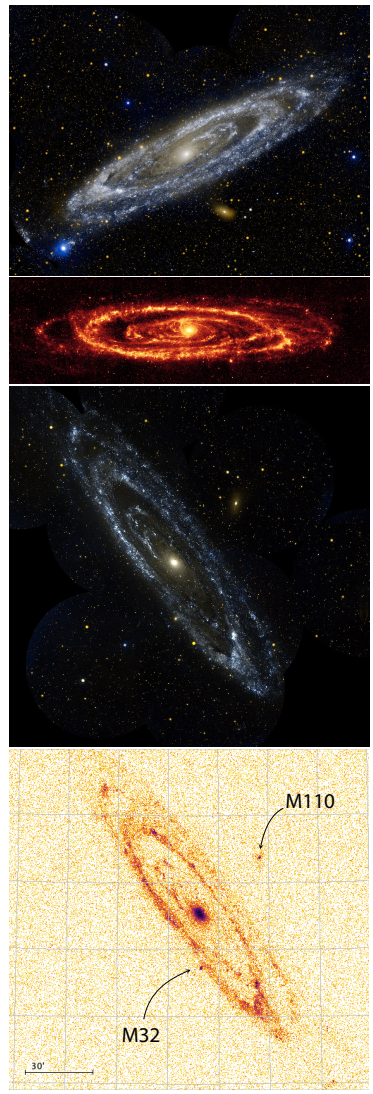
The formation and subsequent evolution of M31 has been pieced together from various lines of evidence, including its star formation history (Davidge et al., 2012). Current understanding is that it was formed roughly 10 billion years ago from the collision and subsequent merger of smaller protogalaxies, resulting in a high rate of star formation, and most of the galaxy's halo and extended disk. A subsequent close passage with the Triangulum Galaxy (M33) 2–4 billion years ago resulted in further bursts of star formation across its disk, and also in disturbances in the outer disk of M33.

Other interactions with satellite galaxies, perhaps M32 or M110, resulted in its metal-rich giant stellar stream (Ibata et al., 2001). And a merger 100 Myr ago may have created the counter-rotating gas disk in its centre, as well as a relatively young stellar component.

SPECTROSCOPIC STUDIES have provided detailed measurements of the rotational velocity as a function of radial distance from its centre. Its spiral arms, punctuated by a series of H II regions, are tightly wound, although more widely spaced than in our own galaxy.

Of its other major components, Andromeda hosts a dense compact star cluster at its centre, containing a black hole of $1-2\times10^8 M_\odot$. Its extended halo appears to be roughly comparable to our own, with stars being generally metal-poor, and increasingly so with distance.

VARIOUS TECHNIQUES have been used to estimate its distance, including the use of Cepheid variables, an eclipsing binary, and stars at the tip of the red giant branch. With a reasonable agreement amongst them, they yield a distance of 778 ± 34 kpc. Corresponding to a parallax of ~1.3 micro-arcsec, this is well beyond reach of Gaia's individual star measurements.



Andromeda seen with (top to bottom): NASA's GALEX (optical); NASA's Spitzer at 24 μm (JPL–Caltech/K. Gordon, Univ. Arizona); GALEX (ultraviolet). Bottom: star counts in the centre of the 5°.5-radius field, on a log scale, from Gaia DR3 (Evans et al., 2022).

THE RELEASE of progressively better Gaia catalogues at roughly 2-year intervals is driven by the increasing quantity of satellite data with time, along with numerous iterative aspects of the data treatment. It means that the scientific community does not have to wait for the final data products. And it also means that user experiences, and any identified shortcomings of an instrumental, calibration, or data treatment nature, can be fed back to ensure appropriate improvements in future releases.

In the first data release (DR1, 2016), G-band 'epoch photometry' (i.e. the time-series photometry) was made available for 3194 variable stars. In the second data release (DR2, 2018), this number increased to half a million variable stars. With the third data release (DR3, June 2022) about 10 million variables are being made available, along with their time series (Eyer et al., 2023; Rimoldini et al., 2023). The present plans of the Gaia Data Processing and Analysis Consortium are to issue the epoch photometry for all stars with the fourth data release, DR4, around 2025.

A specific initiative to encourage user experience and feedback on these plans for epoch photometry is the Gaia Andromeda Photometric Survey, GAPS, made available as part of the DR3 data release (Evans et al., 2023). It provides epoch photometry for all sources, variable and constant, from a limited region of the sky.

Amongst the fields considered for this specific initiative were the Ursa Minor dwarf spheroidal galaxy, the intermediate-age open cluster NGC 2516, the Kepler exoplanet field, and the PLATO long-duration fields.

The final choice for this specific data release was largely determined by the satellite scanning law: Andromeda is in a region where the number of scans varies significantly (the 10th and 90th percentiles ranging from 10 to 57 observations). It also encompasses regions of very different star densities, allowing the effects of crowding to be assessed. And with a radius of 5.5 degrees centred on M31, the chosen field includes many Milky Way stars, resulting in a Hertzsprung–Russell diagram where all sequences are well populated.

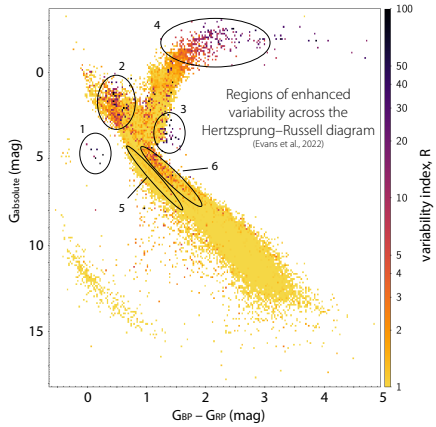
THE RESULTING data set contains epoch photometry in G, B_P and R_P for 1 257 319 sources, of which the majority have between 30–45 observations. Due to the particular geometry of the scanning law over the 34-month data collection interval of DR3, some 15% have more than 50 observations. In contrast, the central region of M31 has much fewer observations due to crowding (due to the presence of multiple sources in the CCD window for the G-band observations, and to the larger overlapping windows in B_P and R_P). Improvements in both aspects are expected in future data releases.

In terms of star colour (Evans et al., 2023, Figure 9), the field average away from M31 is $B_P - R_P \sim 1.4$, whereas in its spiral arms the average colour is ~1.0.

THE SIMULTANEOUS OBSERVATION in three passbands provides an important means for detecting variability. Evans et al. (2023) describe their use of the variability index, R, which characterises the residuals, compared to the mean, in the three passbands (G, B_P and R_P) by means of a *principal component analysis*.

Their Hertzsprung–Russell diagram colour-coded with this R variability index provides a convenient way to visualise the location of all types of variable star. Although in the diagram shown here (their Figure 19), the restriction to stars with better than a 10% parallax error means that they are unlikely to be part of M31 itself (note also that the diagram shows mean values in a given area, rather than individual points).

Many of these variables, but not all, have also been identified as part of the variability processing for Gaia DR3 (Eyer et al., 2023; Rimoldini et al., 2023).



MEANWHILE, Evans et al. (2023) have classified the most prominent variability regions, as designated in the figure, as follows:

1. The seven highly variable blue objects below the main sequence are cataclysmic variables.
2. The classical instability strip on the main sequence is formed by δ Scuti stars (p-mode pulsators), along with γ Doradus stars (g-mode pulsators) at lower luminosity.
3. The clump with strong variability at the edge of the giant branch are RS Can Ven-type variables.
4. The very red giants are long-period variables; they include Mira variables, semi-regular variables, and OSARGs (OGLE small amplitude red giants).
5. The faint ridge in the middle of main sequence seems to have been unexpected. A similar feature was present in the diagram of Gaia Collaboration et al. (2019, Fig. 8).
6. The ridge at the top of the main sequence comprises different variable types, many linked with binaries.

MORE IN-DEPTH analyses of the Gaia variables, and further consideration of many other aspects of the Andromeda Galaxy as seen by Gaia, remain substantial tasks for the future.

84. Gaia's microlensing events

I INTRODUCED THE topic of gravitational microlensing in my earlier essay #11. Briefly, in general relativity, matter distorts spacetime, and the path of electromagnetic radiation is deflected as a result. If there is almost strict alignment between two stars along our line-of-sight, light rays from the more distant background object (the source) are bent by the gravity of the foreground object (the lens) to create images on Earth which are distorted (and possibly multiple), and which may be highly focused and hence significantly amplified.

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.

It is this *photometric* manifestation of microlensing which has been responsible for the many thousands of events that have been discovered to date. The *astrometric* effect, due to the associated tiny motion of the photocentre, is undetectable from the ground, but should (as I described in essay #11) be observable by Gaia.

Microlensing relies on the chance (and very rare) alignment of a background source, an intervening lens, and the observer. 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. This phenomenon has led to the discovery of more than 350 exoplanets to date, almost all from the ground-based networks OGLE, MOA, and KMT.

THE EXPECTED DISCOVERY of many more photometric microlensing events has long been part of the scientific case for Gaia. And their prompt discovery and possible follow-up has long been part of the planning for the photometric 'pipeline processing'. This demands careful calibration, comprehensive variability processing, and the rapid issuing of 'science alerts' for unusual and important 'burst' phenomena – including supernova and microlensing events (Hodgkin et al., 2021).

I described the goals and organisation of Gaia's science alerts activity in essay #36. At that time (September 2021), nearly 50 microlensing events had been discovered as part of the alerts pipeline (some of which were also detected by OGLE), most in the Galactic plane (and in particular in its bulge) where source densities (and hence lensing cross-sections) are highest.

As of mid-2022, the science alerts task has reported more than 400 microlensing candidates, including unusual events such as Gaia16aye, a rotating binary lens (Wyrzykowski et al., 2020); Gaia18cbf, a very long duration event spanning 491 days (Kruszyńska et al., 2022); and Gaia19bld, a particularly bright event, $I = 9.05$ mag at its peak (Rybicki et al., 2022; Bachelet et al., 2022.

MICROLENSING EVENTS are, today, being discovered from the ground, not only by the dedicated search teams such as OGLE, MOA and KMT, but also serendipitously by other large-scale photometric surveys such as the All-Sky Automated Survey for Supernovae (ASAS), and the Zwicky Transient Facility (ZTF).

Gaia provides only relatively sparse and irregular sampling of its targets: thus, in the 34-month period covered by DR3, the average number of observations per source is about 40, with only some 20 in the bulge, but reaching some 100 at intermediate ecliptic latitudes.

But the Gaia observations also offer several important advantages: a high photometric accuracy; a wide range of magnitudes; its all-sky coverage; and its simultaneous 3-colour sampling (G, $G_{\rm BP}$ and $G_{\rm RP}$).

In addition, the detailed astrometric and spectroscopic measurements which will be available with Gaia DR4, will be important ingredients for more complete microlensing solutions.

AS PART OF the Gaia Data Release 3 package, covering 34 months of data acquired between 2014–2017, the study and classification of the two billion objects with photometric time-series has been an important activity. The 'variability processing unit', CU7, has for example generated 1 840 947 142 associated light curves.

Within this, the systematic search for photometric microlensing events is described by Wyrzykowski et al. (2023). They worked with two samples: sample A based on an automated procedure and selection criteria tuned to the Gaia light curves; and sample B comprising specific events selected from the pre-existing literature.

For the former, the identification of rare microlensing events requires very efficient search techniques. In common with techniques developed for ground-based searches, their approach is to first identify all possibly variable stars exhibiting a single episode of brightening, and then study the narrowed-down sample in more detail with more computationally-expensive tests.

The first part ('Extractor'), used the G, G_{BP}, and G_{RP} data for a series of simple tests, in particular searching for a significant number of points brighter than the median; a brightening episode in all three pass-bands; and constancy of the light curve outside of the event. This resulted in 98 million sources for more detailed investigation. The second part, 'microlensing specific object studies', performed a microlensing model fit to the data.

Sample B consisted of 1527 suitable events already known from OGLE–IV. Two more were included from the ASAS–SN survey, and a further 7 had already been identified by the alerts system during the DR3 time interval.

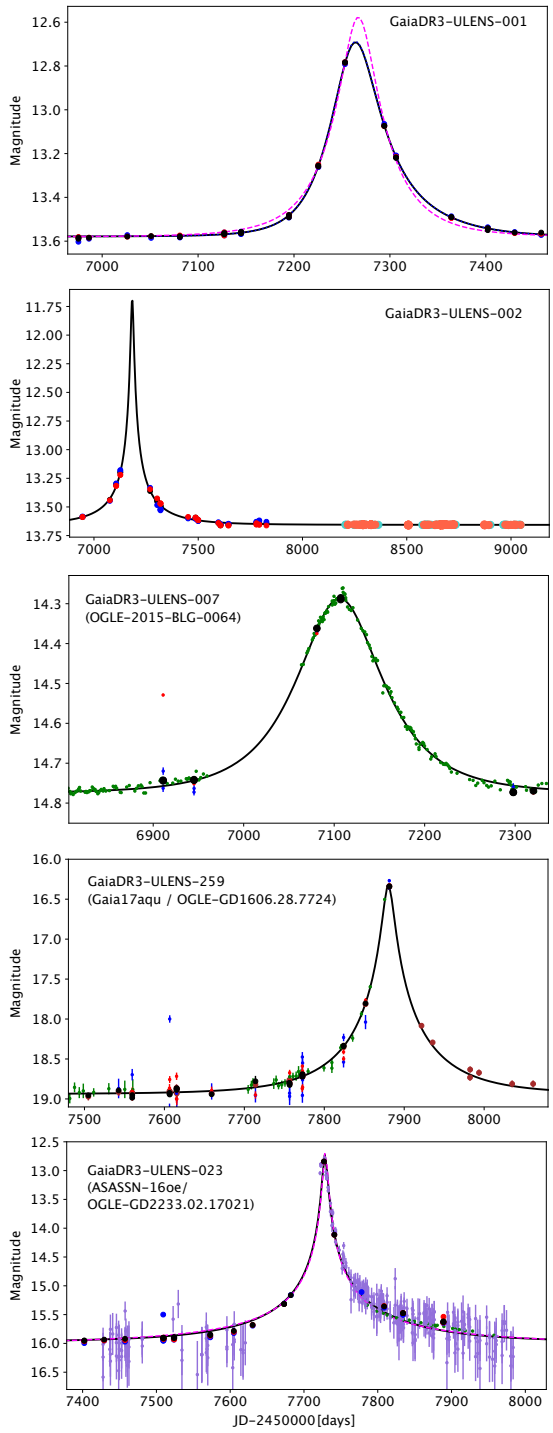
THEIR FINAL COMPILATION comprises 163 events from sample A, and a further 233 events from sample B, with 33 in both, leading to a total of 363 Gaia microlensing events (designated GaiaDR3-ULENS-NNN). 90 were previously unknown, while 273 were independently identified by other surveys, notably ASAS and ZTF.

The majority of their events (228, or 63%) are located towards the Galactic bulge, with most concentrated at latitudes $|b| < 7°$ (their Figure 2). The event density roughly follows the stellar density, and is concentrated towards the Galactic plane. The remainder are mostly concentrated to Galactic longitudes $|l| < 50°$. Gaia EDR3 parallaxes gave source distances for 346 events.

Their estimated completeness ranges between 30% in the bulge to some 80% in the disk, and it is fairly complete for events longer than about 60 days. They consider that less than 1% are false events.

AMONGST THEIR discoveries: #001 is one of the events discovered in the DR3 data, but missed by previous surveys. #002 was previously classified as a long-period variable by the ATLAS survey. #007 was discovered independently by OGLE. #259 (Gaia17aqu) is one of the seven events flagged by the Gaia science alerts system. And #023 is a long-duration event discovered by the ASAS–SN survey in 2016.

Some splendid animations illustrating the lensing phenomena, with some real Gaia events, are given at the University of Warsaw `microlensing group's page`.



Example Gaia microlensing events (Wyrzykowski et al., 2022; Figures 17–21). If the events were independently discovered (by OGLE, ASAS–SN, or Gaia Science Alerts) their alternative names are also given, and the plots include those data. For clarity, the key to the data points, and their model residuals, are not shown.

85. Radial velocities: what wavelength?

GAIA'S THIRD data release, DR3, was issued on 13 June 2022. Alongside more than 1.8 billion stars with their (preliminary) astrometric solutions, it included more than 33 million sources with radial velocities.

To put this in context, when the Hipparcos catalogue of 120 000 stars was released in 1997, radial velocities were known for only some 20 000. High-accuracy radial velocity surveys undertaken by Coravel, amongst others (today often focused on the search for exoplanets) have since measured several thousand more.

Subsequently, and on a larger scale, and in the northern hemisphere between 2005–2010, the Sloan Digital Sky Survey's SEGUE extension obtained spectra of 240 000 stars, with typical radial velocity accuracies of 10 km s^{-1}, with SEGUE–2 observing a further 120 000.

Complementing SEGUE in the southern hemisphere, RAVE (RAdial Velocity Experiment) was a multi-fiber spectroscopic survey using the 1.2-metre UK Schmidt Telescope of the Australian Astronomical Observatory and covering 20 000 square degrees, some one half of the sky. Conducted between 2004–2013, and partly motivated by Gaia, RAVE acquired some 574 000 spectra for around 483 000 individual stars.

IN MY ESSAY #8, I summarised the efforts that went into acquiring radial velocities for Hipparcos, and the challenges of obtaining them from the ground.

And I mentioned the main motivation for acquiring radial velocities for as many stars as possible in the first place: viz., to provide the third component of the star's space velocity (given that astrometry can only measure the projected motion of the star across the sky, termed its 'proper motion'). Without it, kinematic and dynamical insights of individual objects – and indeed entire populations – are accordingly greatly restricted.

In preparing for Gaia, it was evident that *multiple* radial velocity measures can also yield information on orbits and orbital companions (whether stars or planets), which themselves can distort knowledge of the star's space motion. But even a single radial velocity measure would provide much valuable information.

WHILE IT WAS desirable to obtain the highest accurate for all target stars, in practice faint and bright regimes can be loosely distinguished. Faint targets will mostly be distant stars, which will be of interest as tracers of Galactic dynamics. In these cases, the uncertainty in the tangential space motion will be dominated by the parallax error, and a radial velocity accuracy of around 5 km s^{-1} would be adequate for statistical purposes.

Brighter stars, < 15 mag, will be of more individual interest. Here the radial velocity will be valuable for the determination of perspective acceleration (essay #34), while multiple observations would provide a binarity indicator, also averting errors in the star's space motion.

IN THE FOLLOWING, I will look back to the period leading up to the selection of Gaia in 2000, and recall why the decision was made to acquire radial velocity observations on-board the satellite itself, and the considerations which influenced the choice of spectral range.

THE PARALLEL acquisition of non-astrometric data, both photometric and radial velocity, in the framework of a deep astrometry mission was considered by Perryman (1995). The possibility of acquiring the necessary spectroscopic data (for radial velocities), through dedicated ground-based multi-fibre spectroscopic facilities, was also considered (Bastian, 1995). Such programmes were considered as feasible down to magnitudes of 14–15 mag (and indeed helped to catalyse the southern hemisphere RAVE survey), but would nevertheless be logistically complex and very expensive.

There was also a clear scientific return in acquiring *several* radial velocity measurements per star, and doing so not only well spaced in time but, preferably, simultaneously with the astrometric measurements. Given the large number of targets which would be observed, it was not considered practical to acquire such radial-velocities from ground-based facilities.

Superior accuracy due to the absence of atmospheric scintillation, and an improved control of systematics, also argued for their acquisition from space.

HAVING MADE the decision to acquire as many radial velocities as possible with Gaia itself, the choice of the spectral region was studied in considerable detail as part of the Gaia Concept and Technology Study (ESA–SCI(2000)4), and in particular in a series of papers by Ulisse Munari from the University of Padua (Munari, 1999; Munari & Castelli, 2000; Munari et al., 2001. Let me summarise the broad logic driving the eventual choice.

Most Gaia stars are intrinsically red, and made even redder by interstellar absorption. Thus, a red spectral region was to be preferred. To maximise the radial velocity signal even for metal-poor stars, strong, saturated lines are desirable. Broad lines also allow the radial velocity to be accurately derived from moderate-resolution spectra: only a sampling of the lines is needed to derive the radial velocity, while oversampling leads to only marginal improvements in accuracy.

Beyond the Hα line at 656.3 nm, three strong spectral features are present in late-type stars: the potassium (K I) doublet near 768 nm, the sodium (Na I) doublet at 819.4 nm, and the calcium (Ca II) triplet near 860 nm.

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

The Na I doublet is stronger than the K I doublet, but again has a complex luminosity dependence, for example being weak in giants. Its two lines are also rather close in wavelength (0.16 nm), so that high spectral resolution is required. They are similarly affected by TiO bands. Also, both the K I and the Na I region have no lines in earlier (B and A) stars, so that no radial velocity information can be derived from them in such cases.

THE CA II LINES are strong throughout most of the Hertzsprung–Russell diagram, with $W \geq 0.3$ nm for all dwarfs from F8 to M8, and $W \simeq 0.6$ nm in supergiants. The lines appear around B8, and dominate in M stars, with a strong luminosity effect. Being non-resonant, there are no problems with contamination by interstellar components. They are strong enough that useful radial velocities can be derived by cross-correlation even in low signal-to-noise spectra. The Ca II lines were duly considered an optimum choice (Munari, 1999).

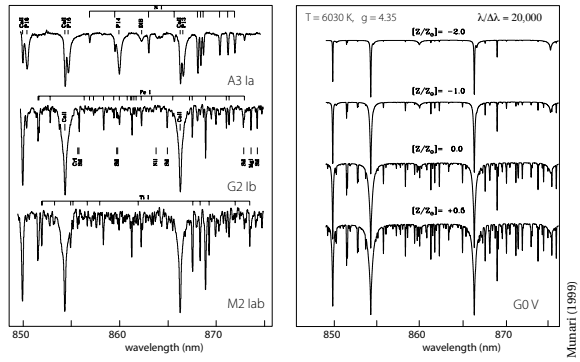
A further advantage of the Ca II infrared triplet region is that, in early-type stars (B, A and early F), where the Ca II lines disappear, the Paschen series of hydrogen appears instead. These lines are also strong, also allow for radial velocity determination, and are also important for classification (Frémat et al., 1996). Finally, the 860 nm spectral region is free from telluric absorption lines, so that follow-up (or complementary) observations from ground can be undertaken in a homogeneous way.

THIS SPECTRAL REGION also contains numerous other features of intrinsic astrophysical interest.

First, the region is not affected by strong molecular bands, but at the same time several metallic lines are present (N I 872.89 nm, Si I 874.26 nm, Mg I 873.60 nm, Ti I 873.47 nm and He I 873.34 nm), which allow for a detailed abundance analysis and quantitative spectral classification for the brighter stars, based on the ratios of equivalent widths.

Second, stars with peculiar spectra are also easily detected in this spectral range. For example, the infrared Ca II spectrum of the proto-typical mass-loosing star P Cyg shows a characteristic line-profile prominent in the Paschen and He I lines, and is easily detected even in low signal-to-noise spectra. Classical Be stars are also easily found, as well as classical T Tau stars (which have strong Ca II emission). Active late-type stars generally display emission cores in the Ca II triplet.

Third, while there is no resonant *interstellar* line in the Ca II triplet region, a medium-intensity diffuse interstellar band (DIB) is present at 862 nm. The correlation of its equivalent width with absorption is rather tight, suggesting that it could be used as one of the indicators to build a detailed reddening map, especially for high values of interstellar extinction (Munari, 1999).



Example spectra in the Ca II triplet region (0.025 nm per pixel). Left: the effect of temperature from A to M stars; right: the effect of metal abundance in G stars (Munari, 1999).

IN FUTURE ESSAYS, I will look more at the implementation of the radial velocity spectrometer on-board Gaia, at the associated data processing, and at the huge range of scientific results that Gaia's radial velocities are enabling today.

While a number of other scientists picked up the huge challenge of optimising the Gaia radial velocity instrument, and addressing the complexities of the data analysis, I wanted to revisit here the early foundations on which the implementation was based. Although Ulisse Munari moved on to other activities, I wanted to credit the key role he played in suggesting and quantifying this important direction for Gaia.

86. Radial velocities: their acquisition

THE RADIAL VELOCITY spectrometer, or RVS, obtains spectra of all stars brighter than about $G_{\rm RVS} \sim 16$ (the magnitude in the spectroscopic bandpass). These are used to determine the radial velocity of each sufficiently bright star through its measured Doppler shift.

Together, a star's proper motion (its angular motion projected on the plane of the sky) and its radial velocity (along the line-of-sight) combine to provide the star's full three-dimensional space velocity. This is required for kinematical and dynamical studies of individual stars, as well as of larger stellar populations.

For nearby, fast-moving stars which show 'perspective acceleration' (essay #34), radial velocities are also required for a rigorous astrometric treatment. And multi-epoch radial velocity measurements provide a powerful indicator of binary or multiple star systems.

For progressively brighter stars, various other important quantities can be derived from the spectra (Recio-Blanco et al., 2016), including coarse stellar parameters (for $G_{\rm RVS} \lesssim 14.5$ mag), astrophysical information such as interstellar reddening, α-element abundances and rotational velocities (for $G_{\rm RVS} \lesssim 12.5$ mag), and even individual abundances for elements such as Fe, Ca, Mg, Ti, and Si (for $G_{\rm RVS} \lesssim 11$ mag).

THE SPECTROSCOPIC INSTRUMENT is highly integrated with the astrometric instrument (Gaia Collaboration et al., 2016; Cropper et al., 2018). It uses the same telescopes, a dedicated section of the same focal plane, and the same sky-mapper and astrometric field (AF1) combination for object detection and confirmation. The selection of the fainter objects for observation with RVS is, however, based on an on-board flux estimate derived from the $R_{\rm P}$ photometer, which is collected just before the object enters the spectroscopic instrument field.

The radial velocity spectrometer is an integral-field spectrograph, in which the spectral dispersion of all objects in the combined field of view is achieved through an optical module (with unit magnification) in the common path of the two telescopes, actually between the final telescope mirror (M6) and the focal plane.

This module contains a blazed-transmission grating plate, four fused-silica prismatic lenses (two with flat surfaces and two with spherical surfaces), and a multi-layer interference bandpass-filter plate to restrict the wavelength range to 845–872 nm.

As I detailed in essay #85, this spectral range was carefully selected to cover the important Ca II infrared triplet, which is suitable for radial-velocity determination over a wide range of metallicity, signal-to-noise ratio, temperature, and luminosity class (in particular for abundant FGK stars), and which also provides a robust metallicity indicator and stellar parameterisation.

This narrow wavelength range also covers the hydrogen Paschen series, from which radial velocities can be derived for early-type stars, as well as a prominent diffuse interstellar band (DIB), at 862 nm, which appears to be an excellent tracer of interstellar reddening.

THE DISPERSED LIGHT from the spectrometer illuminates a dedicated area of the focal plane containing 12 CCDs, arranged in three strips of four CCD rows. As a result, objects observed by the spectrometer have 43% fewer focal plane transits than in the astrometric and photometric fields, typically some 40 over 5 years.

The grating plate, with $R = \lambda/\Delta\lambda \approx 11\,700$ (giving a dispersion of 0.0245 nm per pixel) disperses images in the along-scan direction, spread over ~1100 pixels. The along-scan window size is 1296 pixels, to allow for background subtraction, and window-placement 'errors'.

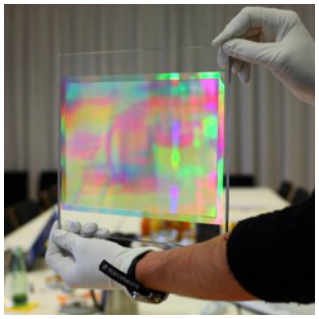
In common with the devices used throughout the focal plane, the CCDs are back-illuminated, with an image area of 4500 lines along-scan and 1966 columns across-scan. Each pixel is 10 μm × 30 μm, corresponding to 58.9 × 176.8 milli-arcsec on the sky. As with the astrometric and photometric fields, the CCDs are operated in time-delayed integration (TDI) mode: photoelectrons are integrated over the CCD as the images cross the focal plane, perfectly synchronised with the spacecraft spin.

For the radial velocity spectrometer (as well as the $R_{\rm P}$ photometers) the CCDs are red-enhanced, including an anti-reflection coating centred on 750 nm.

FOR THE MAJORITY of objects, the spectra are binned on-chip in the across-scan direction over 10 pixels, to form one-dimensional along-scan spectra. Single-pixel-resolution windows (of size 1296×10 pixels^2) are only retained for stars brighter than $G_{\rm RVS} = 7$ mag.

The object-handling capability of the radial velocity spectrometer is limited to about 35 000 objects per square degree. In areas exceeding this stellar density, only the brightest objects are allocated a detector window. As for the photometers, the data quality is progressively compromised in dense areas by contamination and blending from nearby sources.

THE PROCESSING of the RVS data is carried out, on the ground, within an extensive and dedicated processing 'pipeline', described for Gaia DR2 by Sartoretti et al. (2018), and for Gaia DR3 by Katz et al. (2023).



RVS grating demonstrator model (Airbus Defence and Space)

The pipeline takes care of the basic spectroscopic calibrations, including the wavelength scale, geometric calibration of the focal plane, the treatment of straylight, and the effects of CCD charge-transfer-inefficiency.

The noise-dominated faint object transit spectra are subsequently 'stacked' to derive mission-average radial velocities through cross-correlation techniques. For the brightest subset, epoch spectra and epoch radial velocities are preserved. Iterations between calibrations and source parameters are performed entirely within the RVS pipeline.

FOR THE VAST MAJORITY of (faint) stars, the individual spectra are too noisy to derive transit-level radial velocities. As a result, a single, end-of-mission composite spectrum will be constructed by co-adding the spectra collected during all of the RVS CCD crossings obtained throughout the mission lifetime.

A single, mission-averaged radial velocity is then extracted from this composite spectrum by cross-correlation with a synthetic template spectrum. The cross-correlation method finds the best match of the observed spectrum to a set of predefined synthetic spectra (e.g. with different atmospheric parameters) and subsequently assigns the astrophysical parameters of the best-fit template to the observed target.

For the few million brightest targets, single field-of-view transit spectra will be retained to derive associated epoch radial velocities. For this subset, the radial velocities of the components of (double-lined) spectroscopic binaries is also being estimated.

Radial velocity spectrometer results in Data Release 3	
Sources with radial velocities	33 812 183
Sources with mean $G_{\rm RVS}$-band magnitudes	32 232 187
Sources with rotational velocities	3 524 677
Mean RVS spectra	999 645
Variables with radial-velocity time series	1 898
Astrophysical parameters from RVS spectra	5 591 594
Chemical abundances from RVS spectra	2 513 593
Diffuse interstellar band in RVS spectra	472 584

THIS BRIEF SUMMARY cannot do justice to the considerable effort, by scientists and engineers, in designing, optimising, constructing, and processing the data from the RVS system (Cropper et al., 2018).

Today, August 2022, Gaia is 8 years into its data collection phase, having been extended beyond its initially foreseen operational lifetime of 5 years in view of its spectacular performances. Hopefully it will remain operational for a total of around 10 years, limited only by its (cold-gas) attitude control system.

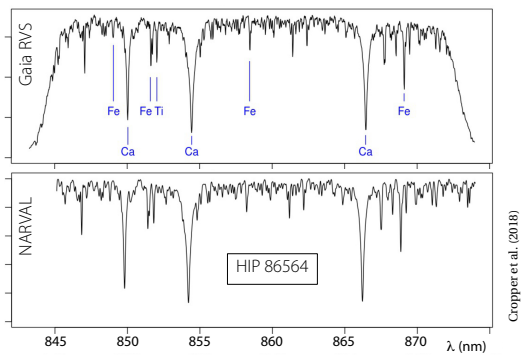
Meanwhile, results for only the first 34 months of mission data (Jul 2014–May 2017) have been processed and released by the Gaia Data Processing and Analysis Consortium, as Gaia Data Release 3 on 13 June 2022.

Along with 1.8 billion sources with astrometric data, the table above summarises the current radial velocity results. My essay #76 gives a summary of the other DR3 results, while the ESA Gaia www pages provide more complete and other supplementary information.

THE FIGURE SHOWS the first public RVS spectrum, the $V = 6.7$ mag K5 star HIP 86564 from a single 4.4 sec exposure (top), and with the NARVAL spectrograph at the Observatoire du Pic du Midi (convolved to the same spectral resolution, bottom). Along with the prominent Ca triplet, lines of Fe and Ti are also visible.

Already dwarfing previous spectral surveys, and with Data Release 4 expected to contain more than 100 million radial velocities, an exceptional resource for Galactic science awaits the world's astronomers!

I will return in my next essay, #87, to look more at the scientific results from RVS contained in Data Release 3.



87. Radial velocities: results from DR3

GAIA'S radial velocity spectrometer (RVS) is obtaining spectra of all stars brighter than about $G_{\rm RVS} \sim 16$ (the magnitude in the spectroscopic bandpass). These will primarily be used to determine the radial velocity of each sufficiently bright star through its measured Doppler shift. But for millions of the brightest stars, the spectra also yield astrophysical information such as interstellar reddening, α-element abundances and rotational velocities, and even individual abundances for elements such as Fe, Ca, Mg, Ti, and Si.

In earlier essays, I have described why these radial velocity measurements were undertaken as part of the Gaia mission (#8), the choice of the wavelength region for the on-board radial velocity spectrometer (#85), and how the spectra are actually acquired (#86).

While Gaia is, today (Aug 2022), 8 years into a possible 10-year mission, results for only the first 34 months of data (Jul 2014–May 2017) have been processed and released by the Gaia Data Processing and Analysis Consortium, as Gaia Data Release 3 on 13 June 2022.

Along with 1.8 billion sources with astrometric data, DR3 provides radial velocities for more than 33 million down to $G_{\rm RVS} \sim 14$ mag. When DR4 is released in 2025, based on 66 months of data, results for the 100 million sources down to $G_{\rm RVS} \sim 16$ mag should be available.

ALTHOUGH I WILL NOT detail the complexities of the pipeline processing needed to create these results, it is worth stressing that many effects must be included and carefully calibrated (Katz et al., 2023).

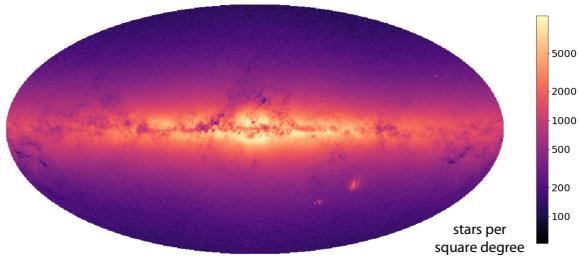
Amongst them are the effects of straylight, calibration of the line-spread function (both along- and across-scan), the effects of CCD bias, saturation, and dark current, the flagging of cosmic rays, the handling of crowded regions and overlapping samples. All this is followed by the detection of emission lines and spectral breaks, eventually leading to the extraction of astrophysical information including the radial velocity and rotational broadening velocity.

But here I will focus exclusively on some of the first scientific results that have accompanied the DR3 radial velocities, recently reported by Katz et al. (2023).

Radial velocity spectrometer results in Data Release 3

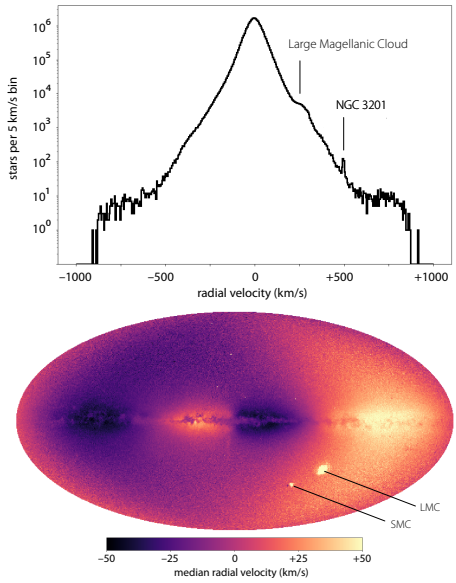
Limiting magnitude of RVS results in DR3	$G_{\rm RVS} \sim 14$
Temperature interval, $G_{\rm RVS} < 12$	3100–14 500 K
$G_{\rm RVS} > 12$	3100–6750 K
Sources with radial velocities	33 812 183
Sources with mean $G_{\rm RVS}$-band magnitudes	32 232 187
Sources with rotational velocities	3 524 677
Mean RVS spectra	999 645
Variables with radial-velocity time series	1 898
Astrophysical parameters from RVS spectra	5 591 594
Chemical abundances from RVS spectra	2 513 593
Diffuse interstellar band in RVS spectra	472 584

I WILL ILLUSTRATE these remarkable results through just some of the diagrams included in their paper. The first, below, simply shows the sky distribution, in Galactic coordinates (centred on the Galactic centre), of the 33 million stars having a combined radial velocity in GDR3. The sampling of the map is approximately 0.2 square degrees.



THE TOP FIGURE on the next page shows the histogram of the combined radial velocities. The bulk of the sources belong to the Milky Way disk, and are centred around ~0 km s^{-1}.

Two specific groups of objects stand out in the right wing of the distribution: the Large Magellanic Cloud appears as a bump around +262 km s^{-1}, while the high-velocity globular cluster NGC 3201 is visible as a spike around +494 km s^{-1}.

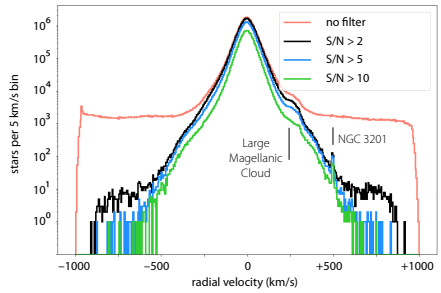


Beneath that, the map of the median radial velocities as a function of Galactic longitude and latitude reveals the rotation of the Galactic disk (projected along the lines-of-sight), manifested by the alternation of bright areas (with positive median velocities) and dark areas (with negative median velocities).

Several objects whose radial velocities differ from those of their close environment (and which are sufficiently populated to affect the median value), are visible by their local contrast. In addition to the Large and Small Magellanic Clouds appearing as bright spots in the lower right region of the sky distribution, the Sagittarius dwarf galaxy is visible as a faint quasi-vertical stripe below the Galactic centre. Several globular clusters and other compact objects appear as small dots in the image, including the globular clusters 47 Tuc and Omega Cen.

THE FIGURE at the top of the next column embraces the distribution of high-velocity stars, those beyond the $\pm 250-500\,\mathrm{km\,s^{-1}}$ wings of the measured radial velocity distribution. Samples are colour-coded, and show all 37.5 million sources successfully processed by the pipeline (red), the 33.8 million sources that passed the validation filters, including a cut at S/N = 2 (black), the 25.3 million with S/N ≥ 5 (blue) and the 13.3 million with S/N ≥ 10 (green).

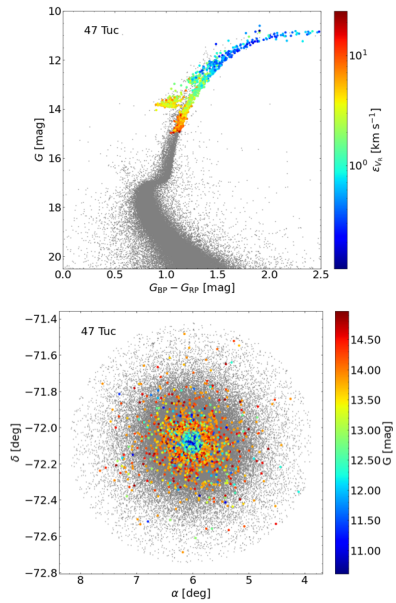
The flat extended wings of the red curve are likely to be dominated by spurious velocities (stars with velocities above $750\,\mathrm{km\,s^{-1}}$, notably the hypervelocity stars, being extremely rare in the Galaxy), while the higher S/N cuts remove most of the spurious solutions. Evidence of the higher velocities of the Large Magellanic Cloud stars, and especially the high velocity globular cluster NGC 3201, are clearly seen after applying the filters.



AMONGST PARTICULAR objects discussed by Katz et al., including the open clusters NGC 2516 and Mamajek 4, and the LMC, I conclude with two figures illustrating the radial velocities in the globular cluster 47 Tuc.

In the colour–magnitude diagram (top) all stars with DR3 photometry are shown as grey dots. Stars with a DR3 radial velocity are colour-coded by their formal uncertainties. In the sky distribution in equatorial coordinates (bottom), stars are colour-coded by their G magnitude. The limit of the spectroscopic processing is visible as a relatively sharp cut around $G \sim 15$ mag.

Radial velocities are measured throughout the cluster, including the densest regions of the core. They cover the upper part of the red giant branch, the asymptotic giant branch, and the horizontal branch. The formal uncertainties improve with magnitude, and are of order a few $\mathrm{km\,s^{-1}}$ at the level of the horizontal branch and of a few hundreds $\mathrm{m\,s^{-1}}$ at the tip of the red giant branch.



LOOKING TO THE 100 million radial velocities expected in DR4, 'deep learning' methods are being developed to best identify spurious velocities, minimise the contamination, and maximise the completeness.

88. Pinpointing exoplanets

I HAVE ALREADY COVERED several aspects of exoplanet research being enabled by Gaia, amongst them the expected number of astrometric discoveries (essay #19), the importance of star distances in determining the *radii* of transiting planets (essay #21) and, notably, Gaia's first astrometric and photometric discoveries (essay #78).

AT THE END OF JULY, I had the pleasure of attending the 2022 Sagan Summer Workshop, on the campus of the California Institute of Technology, Pasadena. This annual event honours the memory of the late great Carl Sagan. In the context of these troubled times, I thought I'd mention that Sagan was born of a Ukrainian refugee (as was my wife). And, as an example of his prescience, and his oratory, do take a moment to listen to him testifying to US Congress on the increasingly alarming consequences of climate change… back in 1985!

The topic of this year's workshop was 'exoplanet science in the Gaia era', with some 100 in-person and 800 remote early-stage researchers. The scientific organising committee was chaired by Gaia exoplanet guru Alessandro Sozzetti (Torino) and senior scientist and educator at the American Museum of Natural History (New York), Jackie Faherty. Do take another moment to listen to Jackie enthusing about Gaia back in 2019!

Contributions from the Gaia side included Anthony Brown's 'Introduction to Gaia', Orlagh Creevey's 'Revolution in stellar astrophysics', and Alessandro Sozzetti's 'Gaia's exoplanet survey', detailing the big expectations for Gaia's astrometric discoveries. Orlagh Creevey (OCA, Nice) leads the coordination unit responsible for the machine-based source parameterisation, and I listened in delight and awe to the resulting statistics (essay #76).

OTHER EXAMPLES of Gaia's contributions to exoplanet science discussed included use in calibrating stellar ages and Galactic kinematics, and ideas for future experiments, including Japan's JASMINE, ESA's Gaia-NIR, and the Roman Space Telescope (formerly WFIRST).

But I will select here just three areas where Gaia is helping to vet, and to pinpoint, these other worlds.

LOCATED ON THE Caltech campus is NExScI, the NASA Exoplanet Science Institute. Of its various tasks, NExScI has developed and maintains several public astronomy data archives and software tools, amongst them the superb NASA Exoplanet Archive, a data service that collates and cross-correlates astronomical data and information on exoplanets and their host stars.

Jessie Christiansen, the Archive's project scientist, gave a fascinating talk which included how Gaia is helping to validate exoplanet candidates being identified by the space transit missions Kepler and TESS.

SOME BACKGROUND. In analysing the enormous amount of data on time-series photometry returned by these transit search missions, it is a huge and challenging task to confirm that a transit *candidate* is truly a transiting planet. Some numbers from NASA's Exoplanet Archive hint at this challenge: as of today, 2711 confirmed planets have been discovered by Kepler, operated between 2009–18. But there remain another 2056 candidates yet to be confirmed (or refuted).

The current statistics from TESS, launched in 2018, are even more dramatic: the archive lists 233 confirmed planets… but with 5808 candidates yet to be confirmed!

Why is it that a regular transit-type signature cannot be taken as unambiguous evidence for a transiting *planet*? As one example, while the fractional brightness changes in a binary-star eclipse are normally much larger, occasionally such an eclipse is grazing, or light from the star is diluted or 'blended' with light from a background or companion eclipsing binary nearby on the plane of the sky. Such 'false-positives' plague both ground and space searches.

In consequence, before upgrading a periodic transit-like signal from a candidate to a confirmed planet, other possibilities must first be excluded. False signals include stellar binaries with grazing eclipses; background eclipsing binaries; background star–planet systems; eclipsing binaries in a hierarchical stellar triple system; eclipsing binaries presenting only secondary eclipses; and a planet orbiting a physical stellar companion.

While the most robust confirmation can be obtained from radial velocity measurements, the majority of the Kepler host stars are faint, making such measurements impractical if not impossible. Accordingly, many clever techniques have been developed to assist confirmation.

The key point underpinning Gaia's role here is that Kepler's pixel size is large: 4 arcsec × 4 arcsec. Within this sky area, binary stars, and even unassociated background stars and eclipsing binaries, can be superposed on the candidate. And if the stellar multiplicity is not understood, inferred planet parameters will be wrong!

Because Gaia identifies stellar binaries down to 0.7 arcsec separation, it can assist with candidate rejection. Gaia quantifies the astrometric 'noise' (the 'RUWE', in Gaia-speak) for each of the 1.46 billion sources with a full astrometric solution in Data Release 3, allowing many false candidates to be swiftly excluded.

The problem is still more acute for TESS, where the pixel size is a hefty 21 arcsec × 21 arcsec, and the assistance of Gaia is proving correspondingly more powerful.

MY SECOND FOCUS is on how Gaia is contributing to clarifying the orbital projection of radial velocity exoplanet candidates. Of more than 5000 exoplanets known today, around 1000 have been discovered by radial velocity measures of the host star. The crucial issue here is that radial velocities are sensitive only to the component of the star's motion *along* the line-of-sight. And it follows that the planet mass can only be estimated to within the uncertainty of $\sin i$, where i is the (unknown) inclination of the orbit to the line-of-sight.

Accordingly, radial velocity discoveries are accompanied by their quoted 'minimum' mass, which may well be planet-like. But the true mass, which can only be provided by Gaia's two-dimensional measurements on the plane of the sky, may be much larger.

Here I will give just two examples of Gaia's contributions from a rapidly expanding literature. The first is HD 92987 b, with a minimum mass of 17 Jupiter masses, and a semi-major axis of 9.6 au, and considered as a promising candidate for direct imaging (Rickman et al., 2019). But the $2087 \pm 19\,\mathrm{m\,s^{-1}}$ astrometric signal reveals that the orbiting companion is not close to its minimum mass, but instead a $0.25 M_\odot$ star viewed at a near-polar orbital inclination of 175° (Venner et al., 2021).

My second example is the historically interesting HD 114762 b, with a minimum mass of $11 M_\mathrm{J}$ (Jupiter masses). It was classified as a 'probable brown dwarf' by Latham et al. (1989), although it has been considered as one of the first possible exoplanet discoveries. Examination of the excess astrometric noise from Gaia vindicates the discoverers' conservative claim, and yields an inclination of about $i = 6°$, and hence a companion mass of around $100 M_\mathrm{J}$, placing it firmly in the brown dwarf (rather than planet) regime (Kiefer, 2019).

MY FINAL EXAMPLE, the contribution of Gaia to the discovery of directly 'imaged' exoplanets, is again the subject of a rapidly expanding literature.

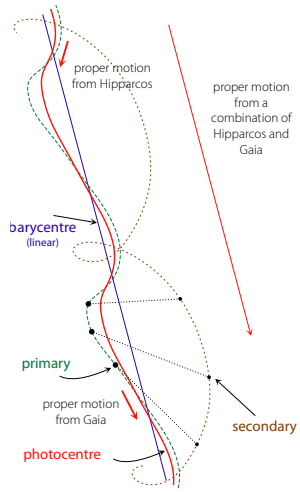
Exoplanet imaging represents a considerable technical challenge, both from ground and from space, because of the close (angular) proximity of planets to their host stars, and the very small ratio of the planet-to-stellar brightness. But a number of important objectives nevertheless motivate its pursuit.

These include gaining more direct confirmation of the existence of exoplanets; discovering long-period planets in wide orbits ($a \gtrsim 20$ au) that cannot easily be probed by other techniques; identifying and characterising orbital motion; studying formation mechanisms and planet–disk interactions in young protoplanetary disks in which planet formation is ongoing; as a precursor to more extensive spectroscopic investigations; and as a first step towards a far-future goal of resolved spatial imaging of an exoplanet surface.

Today, the number of 'imaged' systems is around 60. Amongst them, Jason Wang's spectacular time-lapsed movies illustrate planet orbital motion in β Pictoris, Fomalhaut, 51 Eridanus, and the 4-planet system HR 8799.

WITH ORBIT PERIODS of 20–30 years or more, space astrometry can exploit the near 30-year time interval between the Hipparcos and Gaia proper motions, because each samples the star's reflex motion at very different epochs in a long-period planet orbit.

These 'accelerating systems' can then be used to determine masses for previously imaged planets, or as 'dynamical beacons' to pinpoint where a new planet must lie.



Thus Currie et al. (2020) imaged the companion to the Sun-like star HD 33632 which, at a projected separation of 20 au, induces a 10.5σ Hipparcos–Gaia astrometric acceleration, and gives a dynamical mass of $46 \pm 8 M_\mathrm{J}$.

Others have also been reported: by Damasso et al. (2020) for π Mens; Franson et al. (2022) who obtained a dynamical mass of $61 \pm 4 M_\mathrm{J}$ for the companion to HD 984; Kuzuhara et al. (2022) who found a $28 M_\mathrm{J}$ companion to HIP 21152 in the Hyades cluster; Bonavita et al. (2022) who found 10 companions out of 25 'accelerating' stars; and Kervella et al. (2022) who carried out an even larger statistical survey.

With more imaging observations, and future Gaia releases, many more imaging discoveries lie in the future.

89. A revolution in stellar astrophysics

In essay #76, I gave a summary of the contents of Data Release 3 (DR3), issued by the Gaia Data Processing and Analysis Consortium (DPAC) on 13 June. While DR3 covers just the first 34 months of mission data, it already contains 1.5 billion 5-parameter astrometric solutions.

But DR3 also includes a wealth of 'extracted' astrophysical data, which I summarise at the end of this essay. This is motivated by the fact that the chemical and physical characterisation of stellar spectra is essential for understanding the nature and evolution of stars and Galactic stellar populations. Observations from the ground over the past century have provided 'only' a very heterogeneous collection of chemical abundances for about two million stars in total, with fragmentary sky coverage.

The wealth of data now available from Gaia follows from the pre-launch mission optimisation, state-of-the-art space instrumentation, and rigorous processing and machine-learning classification techniques on ground; one person taking a mere 1 second to inspect every Gaia source, would require 200 years, working 8 hours a day!

The title of this essay is taken from the presentation by Orlagh Creevey at the Sagan Summer Workshop in July. Orlagh leads Coordination Unit 8 of the Gaia DPAC which undertakes the source classification and 'parameterisation'. This involves taking all relevant calibrated data, and extracting key astrophysical quantities based on automated classification techniques.

The 9 coordination units which today form the Gaia DPAC developed from various 'working groups' set up in the late 1990s… the preparatory and operational work has extended over more than 20 years! For most of this time, Coryn Bailer-Jones (MPIA, Heidelberg) led the CU8 work, with Orlagh Creevey (OCA, Nice) taking over its responsibility in June 2018. There are more than 70 contributors, including from Nice, Heidelberg, Bordeaux, La Coruña, Bruxelles, Athens, Catania, Madrid and Torino. The computations are executed in CNES, Toulouse.

Here I will provide the 'big picture' of the various parameterisation activities, aiming to cover more of the scientific results of this work in the future.

The 'astrophysical parameters' included in Gaia DR3 provide the intrinsic properties of each stellar object (such as effective temperature, age, and chemical composition), as well as other inferred properties such as redshifts of distant sources and object classification.

CU8 does this by taking the mean (calibrated) BP and RP spectra, and the mean (calibrated) RVS spectra, along with astrometry (distances and proper motions) and photometry (G, $G_{\rm BP}$, $G_{\rm RP}$), and subjects all these to the 'Astrophysical parameters inference system' (Apsis).

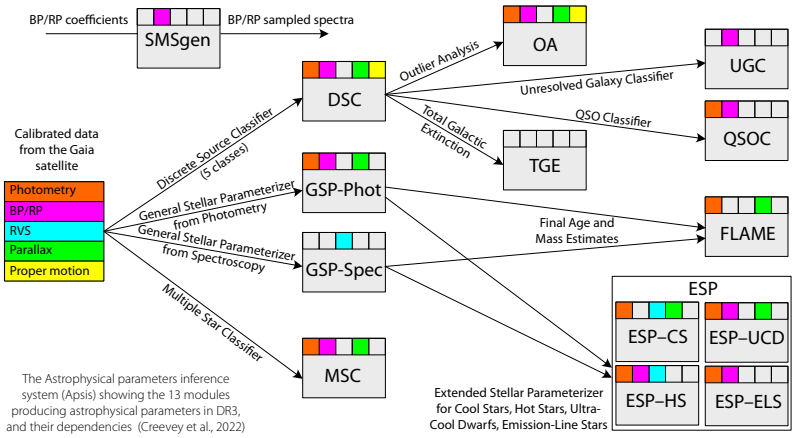
Within this Apsis pipeline, described pre-launch by Bailer-Jones et al. (2013), 13 modules use different combinations of input data and/or models to produce astrophysical parameters for stars and sub-stellar objects, as well as galaxies and quasars.

The overall workflow is shown in the attached figure, which shows each of the 13 Apsis modules, colour-coded to show which of the input data sets are used to generate the output of each module. The details of these various modules, their outputs, and their validation, are described in a series of papers accompanying Data Release 3.

Details of the huge complexity underpinning Apsis are given in the three principal Apsis papers: Creevey et al. (2023b) gives an overview of the Apsis methods and content, along with the detailed data products, and the associated 10 tables of the Gaia archive; Fouesneau et al. (2023) focuses on the stellar content (including parameter determination); and Delchambre et al. (2023) focuses on the non-stellar content.

Three other Gaia release papers are work module specific: Andrae et al. (2023) details the derivation of specific parameters from the photometry (GSP–Phot); Recio-Blanco et al. (2023) details the derivation of specific parameters from the RVS spectra alone (GSP–Spec); and Lanzafame et al. (2023) details the derivation of stellar chromospheric activity and mass accretion from the Ca II infrared triplet in the RVS spectra.

In the following, I will give a concise overview of these various modules to give the overall picture.



The Astrophysical parameters inference system (Apsis) showing the 13 modules producing astrophysical parameters in DR3, and their dependencies (Creevey et al., 2022)

OF THE TWO 'General Stellar Parameterizer' modules, GSP–Spec uses projection and optimisation methods to best match the mean RVS spectra with a large grid of *theoretical* spectra computed using MARCS models, with a range of atmospheric parameters ($T_{\rm eff}$, $\log(g)$, metallicity [M/H], [α/Fe]) and chemical abundances ([X/Fe]), spanning the full parameter space of Galactic stellar populations (Recio-Blanco et al., 2023).

GSP–Phot estimates $T_{\rm eff}$, $\log(g)$, [M/H], absolute magnitude $M_{\rm G}$, radius R, distance, line-of-sight extinctions (A_0, $A_{\rm G}$, $A_{\rm BP}$ and $A_{\rm RP}$), as well as the reddening $E(G_{\rm BP}–G_{\rm RP})$ by forward-modelling the BP/RP spectra, apparent G magnitude, and parallax using a Markov Chain Monte Carlo (MCMC) method.

OF THE VARIOUS 'classifiers', the Discrete Source Classifier (DSC) classifies sources probabilistically into five empirical classes (star, white dwarf, physical binary, quasar, galaxy).

The Unresolved Galaxy Classifier (UGC) estimates the redshift of unresolved galaxies. It does this by applying a supervised machine-learning model to their sampled BP/RP spectra.

The QSO Classifier (QSOC) aims to determine the redshift of sources that are classified as quasars by the DSC module. It is based on the cross-correlation between a rest-frame quasar template and an observed BP/RP spectrum, evaluated at a range of trial redshifts. The module predicts redshifts in the range $0.0826 < z < 6.1295$, along with the corresponding uncertainty.

The Multiple Star Classifier (MSC) infers stellar parameters by assuming that the BP/RP spectrum is a composite spectrum of an unresolved coeval binary system, and that the two components have a flux ratio in the BP/RP spectrum between 1–5.

The Total Galactic Extinction module (TGE) uses a subset of giants with extinction estimates provided by GSP–Phot as extinction tracers, to construct all-sky maps at various resolutions of the total Galactic foreground extinction.

IN THE OTHER PROCESSES, FLAME (Final Age and Mass Estimates) takes the output spectroscopic parameters from GSP–Phot and GSP–Spec, along with astrometry and photometry, to derive the evolutionary parameters: radius, luminosity, mass, and age.

ESP–CS computes a chromospheric activity index from the analysis of the RVS Ca II infrared triplet.

ESP–HS processes the BP and RP spectra, along with the RVS spectra when available, to provide a spectral type for all stars, and stellar parameters for $T_{\rm eff} > 7500$ K.

ESP–UCD is dedicated to the analysis of ultra-cool dwarfs, and it produces $T_{\rm eff}$ for these stars.

Finally, ESP–ELS analyses emission-line stars, and provides class probabilities and labels, along with a measurement of the Hα equivalent width.

THE NUMBERS derived by Apsis for Data Release 3 are quite staggering: source classification and probabilities for 1.6 billion objects; interstellar medium characterisation and distances for up to 470 million sources, including a 2-d total Galactic extinction map; and 6 million redshifts of quasar candidates and 1.4 million redshifts of galaxy candidates. The astrophysical parameters include many stellar spectroscopic and evolutionary parameters for up to 470 million sources.

These comprise $T_{\rm eff}$, $\log(g)$, and [M/H] (470 million using BP/RP; 6 million using RVS); radius (470 million), mass (140 million), age (120 million), chemical abundances (5 million), diffuse interstellar band analysis (0.5 million), activity indices (2 million), Hα equivalent widths (200 million), and further classification of spectral types (220 million) and emission-line stars (50 000).

THESE CATALOGUES, based entirely on Gaia data, provide the most extensive homogeneous database of astrophysical parameters to date. As Chas Beichman, Director of the NASA Exoplanet Science Institute (NExScI) at Caltech, put it recently, *'… [The next generation] will never know the struggle to determine a distance or a spectral type thanks to you and the Gaia team.'*

90. Astrophysics of our Galaxy

ALMOST 20 YEARS ago, the Geneva–Copenhagen spectroscopic survey established a major advance in characterising the nearby stellar population: it was a magnitude-limited (~ 8.4 mag), and kinematically unbiased sample of F and G stars within about 40 pc, yielding ages, metallicities, and kinematic properties of around 14 000 such stars in the solar neighbourhood (Nordström et al., 2004).

Among many subsequent applications, they revisited the age–metallicity, age–velocity, and metallicity–velocity relations of the solar neighbourhood. It was, they said, a data set '*... unlikely to be superseded in essential respects until the results of the Gaia mission and/or the RAVE project become available*'. That time has come.

IN MY PREVIOUS essay I summarised how astrophysical parameters are being estimated from the totality of Gaia data, resulting in (for example) spectroscopic and evolutionary parameters for up to 470 million stars.

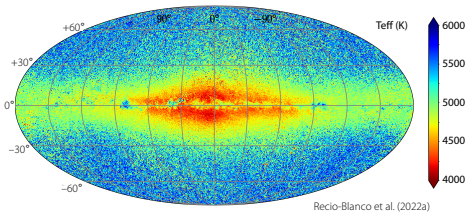
Since the publication of this Data Release 3 in June 2022, several papers have published some early results, illustrating how these data are providing a totally new insight in understanding the structure, formation, and evolution of huge numbers of stars in our Galaxy. I will present here just a few of these, based on two specific data samples, to give a flavour of what Gaia is providing.

THE FIRST SAMPLE was constructed and discussed by Gaia Collaboration et al. (2023b). It makes use exclusively of the stellar parameters and chemical abundances derived from the Gaia RVS spectra by the GSP–Spec module, described briefly in essay #89.

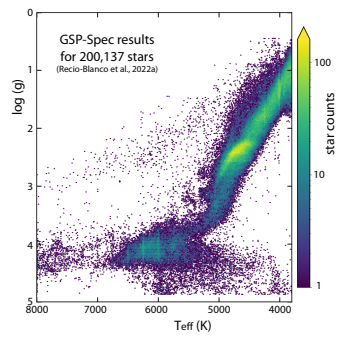
The second is the 'Golden Sample' of Creevey et al. (2023a). This utilises all relevant Gaia data, i.e. mean (calibrated) BP and RP spectra, and the mean (calibrated) RVS spectra, along with astrometry (distances and proper motions) and photometry (G, $G_{\rm BP}$, $G_{\rm RP}$).

The latter comprises six carefully defined subsets of around 3 million young massive disk (OBA) stars, 3 million FGKM stars, 20 000 ultra-cool dwarfs, and smaller numbers of 15 740 carbon stars, 5863 solar analogues, and a sample of spectrophotometric standard stars.

THE FIRST THREE IMAGES are examples of the stellar characterisation provided by the radial velocity spectrometer data, the first two from Recio-Blanco et al. (2023), the other from Gaia Collaboration et al. (2023b).



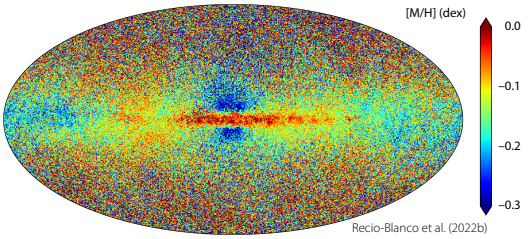
The first shows the distribution, in Galactic coordinates, of effective temperatures, $T_{\rm eff}$, throughout the Galaxy, for their 5.6 million stars. It shows the giant star population dominating the Galactic disk and bulge regions. In-plane interstellar extinction leaves its imprint on the underlying populations: in higher extinction regions, cool giant stars at large distances become too faint in the RVS wavelength domain, and the median of the temperature distribution becomes hotter. Nearby fainter dwarf stars in the foreground dominate the regions above and below the Galactic plane, increasing the median $T_{\rm eff}$ values.



The analogue of the Hertzsprung–Russell diagram showing surface gravity, $\log(g)$, versus $T_{\rm eff}$ (the Kiel diagram), is shown here, colour coded according to star density. The parameters' precision is evident from the well-defined evolutionary sequences. For example, the clear 'red clump' presents a metallicity dependence that is independent from that of the red giant branch, as expected. Additionally, younger, more massive stars populate the hotter metal-rich sequence with $\log(g) \lesssim 3$. These stars are located in the Milky Way spiral arms.

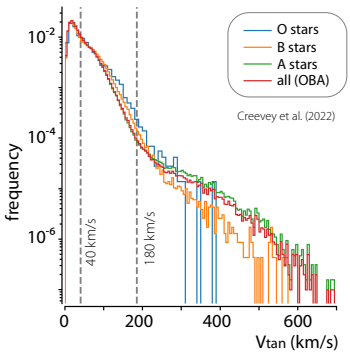
My third example from the RVS data shows Galactic stars in the GSP–Spec database, colour-coded according to the median stellar metallicity, [M/H]. In the Galactic plane, the higher metallicity of the thin disk is evident.

In the central regions of the Galaxy, there is a more metal-poor mix of bulge and thick disk populations while, far from the plane, there are both nearby high-metallicity thin disk stars, and more distant metal-poor halo stars.



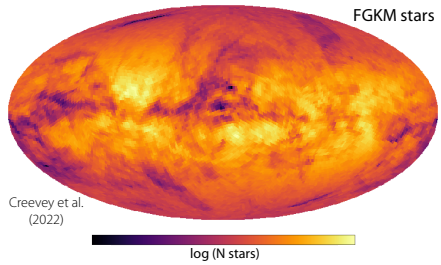
The next four images are examples taken from the 'Golden Sample' of Creevey et al. (2023a). Again, I include them as representative illustrations of the huge wealth of information contained in these catalogues.

Firstly, their O-, B- and A-type (OBA) star sample yielded 3 023 388 sources (compared with 34 055 from a pre-Gaia SIMBAD query). It was constructed as being the best targets to study the structure and dynamics of star-forming regions, as well as of the Galactic spiral arms. This is because, being of intermediate to large mass, they evolve rapidly, and hence typically do not migrate far from their birth association or cluster.



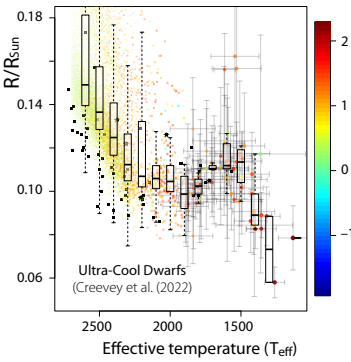
The figure shows their transverse velocities constructed from their distances and proper motions, with thin disk stars 'defined' as having tangential velocities $v_{\rm tan} < 40\,{\rm km\,s^{-1}}$, thick disk stars as $40 < v_{\rm tan} < 180\,{\rm km\,s^{-1}}$, and halo stars having $v_{\rm tan} > 180\,{\rm km\,s^{-1}}$. The vertical dashed lines indicate these transverse velocity limits, and these correspond well to the inflections in the histograms for this hot star sample.

Stars of F, G, K, and M spectral type form the majority of the stars of our own Galaxy. They hold a record of how our Galaxy was formed, and how it has evolved. Their sample (constructed using constraints on $T_{\rm eff}$, log(g), [M/H], radius, mass, age, evolutionary stage, and spectral type), contains 6.3 million stars, which is shown here as a projection on the Galactic plane.



Ultracool dwarfs (UCDs), defined as sources with spectral types M7 or later, are objects at the faint end of the main sequence. The class includes the coolest hydrogen-burning stars, as well as brown dwarfs. Despite a fundamental differences in internal structure across this stellar/sub-stellar regime, the atmospheric properties overlap at the boundary, and it has been difficult to distinguish between the two based on photometric or spectroscopic properties alone.

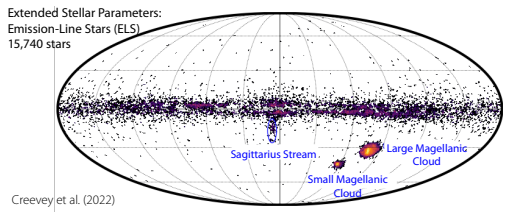
For their high-quality sample of 21 068 sources, they complemented $T_{\rm eff}$ with radii inferred from the Stefan–Boltzmann law (making use of infrared photometry). This figure, showing radii as a function of $T_{\rm eff}$, suggests a clear minimum in the mass–radius relation relation at $T_{\rm eff} \sim 2000$ K.



My final example from their work is of carbon stars. These are a class of Asymptotic Giant Branch (AGB) stars having C-enriched atmospheres, showing strong C_2 and CN molecular bands either due to mass transfer in binary systems, or to the pollution by nuclear He fusion products from the inner to the outer layers.

Because they belong to a late stage of stellar evolution where mass-loss occurs, carbon stars are important contributors to the interstellar medium, and provide good reference cases to study the physical processes affecting the end of the life of low-mass stars.

Their sample of 15 740 carbon stars is shown here in Galactic coordinates. It reveals a strong concentration towards the Galactic plane, with others in the Magellanic Clouds, and the Sagittarius stream.



91. Cerium and the Galaxy infall history

In this essay, I will bring together two of the remarkable new avenues of research into the formation and evolution of our Galaxy that are being opened up by Gaia. And I will discuss the consistency between them – a consistency that only Nature could conjure up!

I will need to explain what cerium is, why it is relevant to astronomy, how it is measured by Gaia, and what its occurrence tells us about the formation of our Galaxy.

Cerium, atomic number 58, despite its classification as one of the rare-Earth elements, is not at all rare in the Earth's crust; at 66 ppm, it lies (just) behind copper, and is far more abundant than, say, lead or tin. Its two most commonly occurring isotopes are ^{140}Ce, produced in both the s- and r-processes (slow- and rapid-neutron capture) of stellar nucleosynthesis, and ^{142}Ce, produced only in the r-process (e.g. Lawler et al., 2009).

While the formation sites of the r-process elements are still somewhat uncertain (and may include the cataclysmic merging of neutron stars, or of neutron star–black hole binaries), those of the s-process elements are better (if still imperfectly) understood. Of interest to us here, the 'main' s-process elements such as Ce and Nd are produced in low- and intermediate-mass asymptotic giant branch (AGB) stars, making use of neutrons produced by the ^{13}C$(\alpha,\mathrm{n})^{16}$O reaction.

Meanwhile, massive stars ($8-10\,M_\odot$) are mostly responsible for the 'weak' s-process, producing Sr, Y, and Zr. Here, the neutrons are mainly provided by the ^{22}Ne$(\alpha,\mathrm{n})^{25}$Mg reaction in convective He-burning cores and C-burning shells. Low-metallicity, low-mass AGB stars can produce the elements of the third peak (such as Pb) through the 'strong' s-process.

As a result, studying the Ce content of stars across the Galaxy provides a probe of these different production sites. Various studies over the past decade have typically found flat trends of [Ce/Fe] versus [M/H], perhaps with a small decreasing trend for large [Fe/H] values.

Others have suggested a decreasing cerium abundances with age, at least up to about 8 Gyr, both for open clusters and field stars, and an increasing abundance with age for stars older than 8 Gyr.

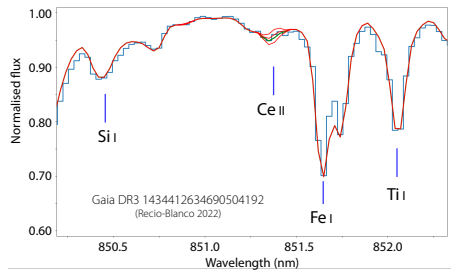
The recent Gaia Third Data Release (DR3, June 2022) contains a homogeneous analysis of millions of high-quality stellar spectra, obtained by the radial velocity spectrometer on Gaia, reduced and calibrated on the ground within the Gaia Data Processing and Analysis Consortium (DPAC) by Coordination Unit 6, and analysed by the GSP–Spec module of Coordination Unit 8, responsible for the source classification and 'parameterisation' (Recio-Blanco et al., 2023).

As I have described in essay #89, this leads to abundance estimates for up to 13 individual elements in the atmospheres of several million stars, and the possibility for unprecedented chemical mapping of our Galaxy.

And among the abundances derived in GSP–Spec are three of the heavy-elements produced by these neutron-capture processes: Ce, Zr and Nd. In common with all Gaia abundances, these are estimated by comparing the mean observed spectra with an extensive grid of synthetic spectra covering a wide range of stellar atmospheric parameters, and with varying abundances.

A first analysis of the Gaia Ce data has been made by Contursi et al. (2023). Among the 5.5 million stars parameterised by GSP–Spec, 103 948 have a derived cerium abundance, more than doubling the numbers previously available. Selection of a high-quality subset according to the flags provided, their resulting sample comprises about 30 000 Ce abundances distributed in the disk and halo components.

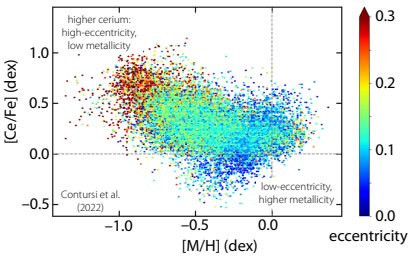
The example shown here, centred on the triplet of Ce II lines around 851.375 nm, has a derived Ce abundance compared to iron, $[\mathrm{Ce/Fe}] = 0.26^{+0.12}_{-0.08}$ dex.



EQUIPPED WITH this sample of almost 30 000 stars with known cerium abundances, the next step is to use the surface gravities determined by the Gaia classification and parameterisation process to establish their nature, and the distances and proper motions provided by Gaia to examine their dependencies on location or kinematics within the Galaxy.

From the former, Contursi et al. (2023) first established that the sample is mainly composed of stars with $\log(g) < 3.5$, that is both red giant branch and asymptotic giant branch stars.

From the astrometry, they found that the closest, being more Ce-poor, more metal-rich, and concentrated within ±500 pc from the Galactic plane, probably belong to the thin disk population. Stars with higher Ce abundances are at the same time generally metal-poor, and preferentially located at larger distances from the Sun, and at larger distances from the Galactic plane.



These hints are confirmed by their kinematic behaviour: this diagram of Ce abundance with respect to metallicity, colour-coded with the eccentricity of their Galactic orbits, shows that stars with higher metallicity and low Ce abundance are on more circular orbits typical of thin disk stars, while those with higher Ce abundances are on more eccentric orbits, and so unlikely to belong to the thin disk population.

MORE DETAILED examination of their orbital angular momentum also reveals a small number of lower angular momentum halo stars. And, interestingly, eleven of these stars are associated with previously identified halo substructures, notably the Helmi (2), Thamnos (2), and Enceladus (7) streams. The data was also used to map out the [Ce/Fe] gradient as a function of Galactic radius, and the vertical gradient as a function of the distance from the Galactic plane.

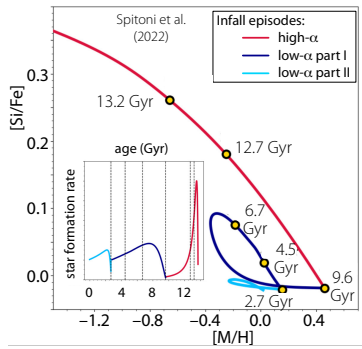
The fascinating thing about these distributions of cerium throughout our Galaxy is that they provide direct support to the latest models describing the chemical evolution of our Galaxy. Let me explain.

I MUST STEP BACK to mention the recent 'two-infall' model of Spitoni et al. (2021), designed to reproduce the Apache Point Observatory Galactic Evolution Experiment project (APOGEE DR16) results of [X/Fe] versus [M/H] abundance ratios (where X stands for various α-elements in the solar neighbourhood). That work concluded that high- and low-α sequence stars were formed by two independent episodes of gas infall.

Things got more complicated with Gaia DR3, when Gaia Collaboration et al. (2023b) identified the presence of young metal-poor stars with low [α/Fe] values. Spitoni et al. (2023) then showed that this low-α population itself can be modelled by two sequential infall episodes, resulting in a total of three discrete infall events.

IN MORE DETAIL, Spitoni et al. concluded that their data could be well reproduced by a 'three-infall' Galactic chemical evolution model. The first event, the high-α infall, had an accretion timescale of 0.103 Gyr. The second, termed low-α part I, had an accretion timescale of 4.110 Gyr, and with a delay between the first and second of 4.085 Gyr. The third, low-α part II, had an accretion timescale of 0.150 Gyr, and with a delay between the second and third of 11 Gyr, started some 2.7 Gyr ago.

And, importantly, the model is consistent with the latest cerium results from Gaia which, in the process, independently confirms the quality and reliability of the Gaia DR3 astrophysical parameter determination!



AT THE END of this summary of one of many aspects of 'Galactic archaeology' that is being revolutionised by the Gaia data, I'd like to make a few observations.

The first is to recall the difficult days when one of the many tasks I was juggling as Project Scientist was to convince senior ESA management of the merits of including the radial velocity spectrometer (the counter-argument being that ESA does not fly experiments that can be conducted on the ground). I feel ever more vindicated!

The second is to emphasise the huge effort invested by Gaia scientists in optimising all aspects of the satellite payload, including the radial velocity instrument, two decades ago. This was matched by the remarkable expertise of the industrial prime contractor, Airbus Defence & Space, in its design and construction, and by the teams carrying out the data processing on the ground today (and in particular here, those in CU6 and CU8).

Finally, I continue to be awed by the staggering complexity of Nature, in assembling a vast and unimaginably puzzling system that is so hospitable for life. And, indeed, in humanity's success in comprehending so much of what Nature has created… especially as aided by the huge numbers of accurate star positions from Gaia!

92. Diffuse interstellar bands

INTERSTELLAR 'EXTINCTION' occurs because of the absorption and scattering of electromagnetic radiation by dust and gas in the interstellar medium lying between source and observer (for Gaia, in space, we can ignore the additional effects of Earth's atmosphere!).

For stars lying near to the Galactic plane, and within a few kpc of the Sun, extinction in the visible amounts to about 1.8 mag per kiloparsec. Shorter wavelengths being more strongly attenuated, extinction causes objects to appear redder than their intrinsic spectrum would suggest, an effect referred to as interstellar 'reddening'.

While the effect of extinction is a reasonably smooth function of wavelength, superimposed on its general shape are absorption features that can have a variety of origins, and which can give clues as to the chemical composition of the interstellar gas and dust grains. Prominent absorption features include a pronounced bump at 217.5 nm (in the ultraviolet), a water ice feature at 3.1 μm, and silicate features at 10 and 18 μm.

For Gaia, the effects of extinction and reddening can be estimated for *each source* as part of the astrophysical parameter determination, using the RP and BP spectra, and the radial velocity spectra (see essay #89). This is allowing the construction of a detailed all-sky map of interstellar extinction, which I will describe elsewhere.

DIFFUSE INTERSTELLAR BANDS, DIBs, are additional absorption features seen in stellar spectra, first reported just over a century ago, but whose detailed origins still remain something of a mystery. Their wavelengths do not correspond with any known (ion or molecular) spectral lines, and so the material which is responsible for the absorption cannot easily be identified.

Their name further conveys the fact that DIBs are conspicuously broader than interstellar atomic lines, presumably due to unresolved rotational structure.

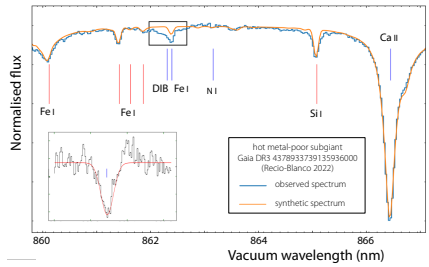
In his review, Herbig (1995) summarised the status at the time by noting that of the 127 confirmed DIBs in the region between 0.4–1.3 μm, only two had been (tentatively) identified with a specific 'carrier'. Around 600 bands have now been discovered, at ultraviolet, visible and infrared wavelengths (Fan et al., 2019).

Because DIB strengths increase roughly in proportion to colour excess, they were originally suspected as being produced by interstellar grains. But current evidence favours unknown species of free polyatomic molecules, such as polycyclic aromatic hydrocarbons and other large C-bearing molecules, either neutral or ionised. Today, only one carrier is securely identified: ionised buckminsterfullerene ($C60^+$), with five absorption bands in the near-infrared (Linnartz et al., 2020).

One important observational result, established pre-Gaia, is that the strengths of most DIBs are not strongly correlated with each other. This means that there must be many carriers, rather than one carrier responsible for all DIBs. Also significant is that the strength of DIBs is broadly correlated with the interstellar extinction, even though they are unlikely to be caused by dust grains.

THE CHOICE OF WAVELENGTH range for Gaia's radial velocity spectrometer was driven by a number of considerations, which I summarised in my essay #85. Amongst these, and within the adopted wavelength range (845–872 nm), was a known medium-intensity diffuse interstellar band at 862 nm, first identified in 1975. The correlation of its equivalent width with absorption was known to be rather tight, suggesting that it could be used as one of the indicators to build a detailed reddening map, especially for high values of interstellar extinction (Munari, 1999).

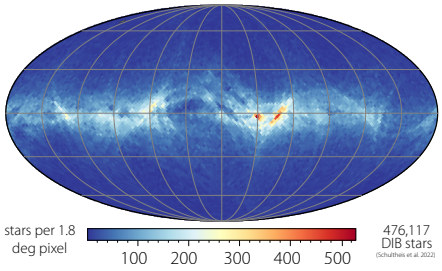
The figure is an example RVS spectrum of one Gaia star from DR3 in the wavelength region of the DIB feature. The difference between the observed spectrum (blue line), and the best-fit synthetic spectrum (orange line) brings out the DIB feature as shown in the inset.



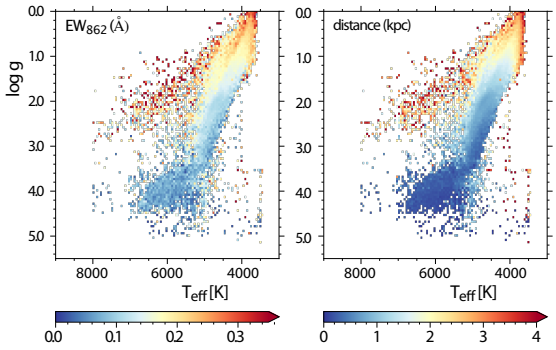
THE 862 NM diffuse interstellar band is characterised as part of the GSP–Spec module, which I described briefly in essay #89. In addition to the equivalent widths and their uncertainties, the tabulations in DR3 list their characteristic central wavelength, line width, and various quality flags (Recio-Blanco et al., 2023).

The first major study of the 862 nm DIB in Gaia DR3 was reported by Gaia Collaboration et al. (2023c); they also review the pre-Gaia work on this specific feature (notably with RAVE), which I will not discuss further.

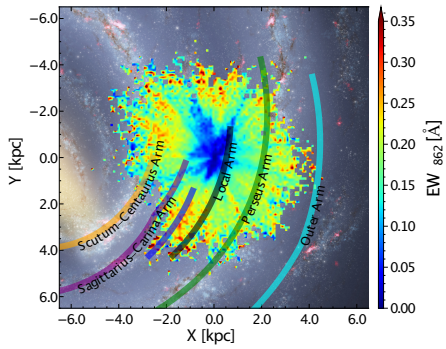
Their sample contains 476 117 stars with valid DIB measurements, with their high-quality sample comprising 141 103 stars. The former is, by an order of magnitude, the largest homogeneous full-sky sample obtained so far. The distribution, given here in Galactic coordinates, shows that stars with a significant DIB feature are concentrated towards the Galactic plane, as expected, but with significant sub-structure.



IN THESE TWO 'Kiel diagrams', which are analogues of the Hertzsprung–Russell diagram showing the star's surface gravity versus effective temperature, i.e. $\log(g)$ vs $T_{\rm eff}$, the first is colour-coded as a function of the absorption feature's equivalent width, and the second according to stellar distances. The very similar trends in both diagrams is striking, and clearly indicates that stars with larger distances have larger equivalent widths.



By also showing that massive stars at 2–4 kpc distance are located in the most nearby spiral arms (notably the Sagittarius/Carina, Local and Perseus arms), they were also able to demonstrate that the 862 nm line provides an excellent tracer of spiral arm structures.



THE EQUIVALENT WIDTHS are also generally well-correlated with estimates of interstellar reddening (with E(BP–RP) derived from the Gaia BP/RP spectra by the GSP–Phot module), with both showing larger values in the Galactic plane.

But they also drew attention to some notable differences, in particular that the scale height of the 862 nm carrier is smaller than that of the dust, and that it is concentrated within the inner kpc from the Sun.

Another striking difference between the 862 nm DIB carrier and dust distributions is that the former features are present in the Local Bubble around the Sun, even though this region is known to contain almost no dust.

To first order, they found that the spatial distribution of the 862 nm DIB follows a simple slab model. They derived its local density and scale height, which can be used to predict its expected equivalent width towards any star up to about 3 kpc from the Sun. Specifically, they determined a local density of 0.019 ± 0.004 nm/kpc, and a scale height of $98.6^{+11.1}_{-8.5}$ pc. The latter, being smaller than the dust scale height, indicates that the DIBs are more concentrated towards the Galactic plane.

FINALLY, taking advantage of the full sky coverage, they determined its rest-wavelength in the Galactic anticentre direction, yielding $\lambda_0 = 862.086 \pm 0.0019$ nm, the most precise determination to date.

Based on this rest-wavelength, the Galactic rotation curve within 1–2 kpc from the Sun shows a remarkably good agreement between the 862 nm diffuse interstellar band velocities and the CO gas velocities, reinforcing the suggestion that the associated carrier(s) could be related to gaseous macromolecules.

STUDIES OF THE correlation between different diffuse interstellar bands are important for gaining insight into their origin. Although the strong line at 862 nm was the only confirmed DIB in Gaia's RVS spectral range, Munari et al. (2008) considered a weaker one at 864.8 nm to be a second promising candidate.

From 8458 DR3 spectra, Zhao et al. (2022) confirmed its existence, and showed that both equivalent widths are well correlated, albeit with the 862 nm line being better correlated with the E(BP–RP) extinction measure.

93. The mass of the Milky Way

WHAT IS THE MASS of our Milky Way galaxy? Why is it important to know? Why is it so difficult to measure? And what is Gaia contributing to our knowledge?

IN THEIR REVIEW of the structural and kinematic properties of our Galaxy, and the state of knowledge pre-Gaia, Bland-Hawthorn & Gerhard (2016) placed its mass at somewhere between $0.5-3\times10^{12}\,M_\odot$.

Despite being one of its most fundamental parameters, and despite decades of intense effort, today's best mass estimates still span a significant range.

Why is it important to know? Accurate determination of the mass profile of the Milky Way has implications for our understanding of the dynamical history of the Local Group, including both its past evolution and any future interactions. And, perhaps even more fundamentally, the mass of a galaxy (and its mass distribution) are intrinsically linked to the formation and growth of structure in the Universe.

Consequently, a better understanding of the mass and mass distribution of our own Galaxy provides a better picture of where it sits in a cosmological context, in particular whether it is typical or atypical and, hence, how much of what we learn about the Milky Way can be applied to other galaxies.

WHY IS IT so difficult to measure? The main problem is that, while the masses of the principal baryonic components (the supermassive black hole at its centre, the central bulge, and the flattened disk) are reasonably well determined, its other major component, the halo, is dominated by dark matter. And it is our inability to see dark matter directly that gives rise to the greatest uncertainty in estimating its mass.

The problems do not end here. We can only hope to measure the mass of the dark matter which has had time to 'virialise' in the Galaxy, the so-called 'virial mass' defined within some adopted 'virial radius'. How far out to measure is usually expressed as the mass within a region in which the average density exceeds some multiple of the mean density of the Universe.

It is the distribution of dark matter which controls the motion of distant 'tracer' objects. These include halo stars and globular clusters as powerful probes of the inner halo, while dwarf spheroidal galaxies offer better coverage further out. In their compilation of total mass estimates between 1999–2014, Bland-Hawthorn & Gerhard (2016, Table 8) also quote estimates from the kinematics of halo streams, from the kinematics of hypervelocity stars, and from modelling of the escape velocity.

But all of these estimates remain sensitive to assumptions such as which of its satellites are bound, the shape and extent of the dark matter halo, and the velocity anisotropy of the halo and of its satellite system.

And this takes us to one of the key problems faced by all mass estimates based on kinematics: while we need to know the total space velocity of each tracer, we seldom have knowledge of all three components of motion; typically, we only have line-of-sight velocities.

Since the Sun is relatively close to the Galactic centre, most line-of-sight velocities probe mainly the component of motion in the Galactocentric radial direction, providing little information about their tangential motions. As a result, estimated masses pre-Gaia depended strongly on the assumptions used for the tangential motions: the so-called mass–anisotropy degeneracy.

ACCORDING TO Bland-Hawthorn & Gerhard (2016), the analysis of halo star kinematics typically results in relatively low values, $M_{200}\lesssim10^{12}\,M_\odot$ (where M_{200} denotes the mass within 200 kpc). Here, the main uncertainties are the lack of stellar tangential velocities from proper motions, along with the need to extrapolate from spatially limited samples to the scale of the virial radius.

Satellite galaxies reach to larger radii, with uncertainties coming from small numbers and, again, a lack of proper motions. Estimates based on satellite and globular cluster kinematics typically result in higher values, $M_{200}=1-2\times10^{12}\,M_\odot$ if the Leo I dwarf galaxy (with its large line-of-sight velocity) is assumed to be bound to the Milky Way; or $M_{200}\lesssim10^{12}\,M_\odot$ if it is considered as unbound (Bland-Hawthorn & Gerhard, 2016).

ONE ESTIMATE of our Galaxy's total mass that doesn't depend on the choice or details of a tracer population is the so-called 'timing mass'.

In the original formulation by Kahn & Woltjer (1959), the Milky Way and M31 proto-galaxies are assumed to have had a small separation at the time of the Big Bang, subsequently moving apart as they participated in the Hubble flow. If there is sufficient mass, their expansion is reversed. Given an estimate of the age of the Universe, together with their present separation and approach velocity, the equations of motion can be solved to give the mass of the Galaxy, along with that of the Local Group.

Timing mass estimates in subsequent work (e.g. Li & White, 2008) have been lowered due to improved relative space motions, and more accurate estimates of the solar reflex motion. With current mass estimates for M31 being comparable to that of the Milky Way, the timing mass provides an upper limit to the Galaxy's virial mass of $\lesssim 1.6 \times 10^{12} M_\odot$ (van der Marel et al., 2012).

ACCORDING TO Bland-Hawthorn & Gerhard (2016), the halo star kinematic studies provided (at least pre-Gaia) the largest and most reliable data sets. They gave an average of $M_{200} = 1.1 \pm 0.3 \times 10^{12} M_\odot$, or equivalently $M_{\rm vir} = 1.3 \pm 0.3 \times 10^{12} M_\odot$, consistent with the upper limit from the timing mass.

This compares with a *baryonic* mass of $M_{\rm b} = 6.3 \pm 0.5 \times 10^{10} M_\odot$ (stars and cold gas), or $8.8 \pm 1.2 \times 10^{10} M_\odot$ when including a contribution from the hot corona. This, in turn, leads to a baryonic mass fraction (out to some chosen virial radius) of $6 \pm 1\%$, well short of the 'universal' value of about 16%.

GIVEN THESE complexities of the determination of our Galaxy's mass, it should not come as a surprise that the advent of the Gaia data has led to a flurry of some 20–30 papers to date, targetting a re-analysis of some of these methods, and in particular using the more complete space motions of halo stars, globular clusters (see essay #30 for Gaia's contributions), and dwarf spheroidal galaxies (see essay #31). In this fast-changing landscape, I will provide a current snapshot.

EMPLOYING DATA FROM DR2, Callingham et al. (2019) inferred the Galaxy's total mass by comparing model satellites in the EAGLE cosmological hydrodynamics simulations with the dynamics of 10 of the 'classical' Milky Way satellites with six-dimensional phase-space measurements (i.e. positions and velocities), including updated proper motions from Gaia (including the LMC, the SMC, Draco, and Ursa Minor).

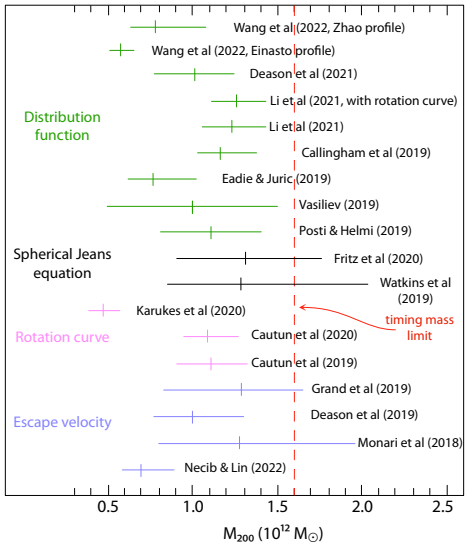
In this 2-d space, the orbital properties of satellite galaxies vary according to the host halo mass, which can then be inferred from the likelihood that the observed satellite population is drawn from this distribution function. They inferred a mass of $1.17^{+0.21}_{-0.15} \times 10^{12} M_\odot$.

Fritz et al. (2020) used a somewhat similar approach based on Gaia DR2 proper motions of 45 satellite galaxies to find a mass of $1.43^{+0.35}_{-0.32} \times 10^{12} M_\odot$ within 273 kpc, extrapolating to a virial mass of $1.51^{+0.45}_{-0.40} \times 10^{12} M_\odot$.

Watkins et al. (2019) used the Gaia DR2 proper motions of 34 halo globular clusters within 21.1 kpc. They determined the mass of the Milky Way inside the outermost globular cluster as $0.21^{+0.04}_{-0.03} \times 10^{12} M_\odot$, leading to an implied virial mass of $1.28^{+0.97}_{-0.48} \times 10^{12} M_\odot$.

VARIOUS MASS determinations have followed with Gaia EDR3 in 2020, using various related methods and assumptions, including modelling halo tracers using spherical, power-law distribution functions.

As one example, Wang et al. (2022b) used Gaia EDR3 proper motions for about 150 Milky Way globular clusters, combined with constraints from the rotation curve (over distances of 5–25 kpc). They derived a total mass in the range $0.536^{+0.081}_{-0.068} \times 10^{12} M_\odot$ to $0.784^{+0.308}_{-0.197} \times 10^{12} M_\odot$, depending on the dark matter profile (Zhao or Einasto), values at the lower end of current estimates.



THESE, and some of the other results based on Gaia DR2 and EDR3, are shown here (adapted from Wang et al., 2022), colour-coded by methodology.

What can we conclude from this present state of affairs? Perhaps first, that while the formal errors are often much smaller than they were pre-Gaia, and while a value of $\sim 10^{12} M_\odot$ appears favoured, current estimates are still affected by the choice of tracers and model assumptions. Second, the Gaia results are nonetheless providing much new insight into its physical complexity. Third, that the Galaxy mass, dominated by its dark matter halo, remains relatively poorly characterised… our imperfect understanding of this dark matter halo, and its substructure, epitomised by the 'too-big-to-fail' problem (Boylan-Kolchin et al., 2011), still persists!

94. The mass of the Local Group

In my previous essay, I looked at the how the Gaia astrometry, and in particular the kinematics of globular cluster and local group dwarf spheroidal galaxies, are providing an improvement in our estimates of the mass of our Galaxy. I will continue here on the related question of the total mass of the Local Group.

The 'Local Group', a term introduced by Edwin Hubble in 1936, is an 'overdensity' of around 80 known galaxies, extending across some 3 Mpc, and dominated by the Milky Way and Andromeda (M31). Both are spiral galaxies with masses of order $10^{12} M_{\odot}$, and both are accompanied by their own systems of smaller satellite galaxies, including many 'ultra-faint' dwarf spheroidals.

In the case of our Galaxy, these include the Large and Small Magellanic Clouds, along with Sagittarius, Draco, Carina, Sextans, Fornax, and many others. Tied to Andromeda are M32, M110, NGC 147, and many others. A more complete list of known members, and some illustrative graphics, are included at this wiki page.

Over the past 20 years, much new insight has been gained about the evolution and merger history of our own Galaxy, specifically as it 'devoured' smaller nearby galaxies over its multi-billion year history.

Gaia is helping to clarify the picture, revealing various debris streams, including the Gaia–Enceladus and Milky Way merger some 10 Gyr ago which contributed to the formation of our Galaxy's thick disk (Helmi et al., 2018), and the 'phase-space spiral' attributed to a recent crossing of our Galaxy's disk by the Sagittarius dwarf galaxy 300–900 Myr ago (Antoja et al., 2018).

In the case of Andromeda, a dominant merger some 2 Gyr ago has been proposed to explain its compact and metal-rich satellite M32 as the stripped core of the disrupted satellite, as well its rotating inner stellar halo, and its associated giant stellar stream (D'Souza & Bell, 2018).

In any successful cosmological model, all of these, and many other properties of the Local Group, must also be consistent with the increasingly detailed predictions of large-scale numerical models of the formation and evolution of structure in the Universe.

As I outlined in essay #93, the mass of the Local Group can be estimated, along with that of our own Galaxy, by modelling the kinematics of various 'tracer' objects (notably halo stars, satellite galaxies and globular clusters) around its most prominent members.

Underpinning these methods are various assumptions tied to our presently favoured cosmological models, for example, that the dark matter is clustered around the major galaxies, and not distributed more uniformly throughout the Local Group.

Indeed one of the early triumphs of the ΛCDM (Lambda cold dark matter) cosmological model in the 1990s was its prediction that massive galaxies like the Milky Way should be surrounded by many dark matter dominated satellite halos, which were only discovered in the predicted numbers over the past 20 years (essay #31).

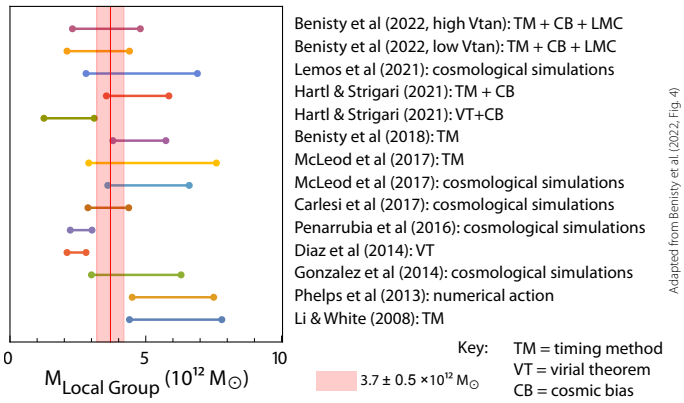
Other 'tensions' with respect to ΛCDM have been noted over the past few years as the accuracies of relevant observations improve. A detailed compilation is provided by Perivolaropoulos & Skara (2022).

Amongst these, which I mention as perhaps one of the easiest to appreciate, is the 'too-big-to-fail' problem (Boylan-Kolchin et al., 2011), an inconsistency between the predicted mass of dark matter sub-haloes in ΛCDM theory, and the observed central mass of the brightest satellite galaxies in the Local Group.

For these and other reasons, mass estimates of the Local Group (amongst various other observables) continue to provide a powerful test of cosmological models.

Two of the most developed methods for determining the mass of the Local Group are via cosmological simulations, and via the so-called 'timing method'.

The former infers the Galaxy's total mass by comparing model satellites in (for example) the EAGLE cosmological hydrodynamics simulations with the dynamics of a number of the 'classical' Milky Way satellites with six-dimensional phase-space measurements (i.e. positions and velocities), most recently including updated proper motions from Gaia (including the Large and Small Magellanic Clouds, Draco, and Ursa Minor).



The timing method, which I introduced in essay #93, is based on the hypothesis that the Milky Way and M31 proto-galaxies had a small separation at the time of the Big Bang, subsequently moving apart as they participated in the Hubble flow. Given an estimate of the age of the Universe, together with their present separation and approach velocity, the equations of motion can be solved to give the mass of the Galaxy, along with that of the Local Group.

The timing method has been progressively refined since its original formulation by Kahn & Woltjer (1959), who assumed, for example, that the Milky Way and M31 are on an exactly radial orbit. Einasto & Lynden-Bell (1982), later showed that the problem retained an analytic solution even if M31 has some tangential motion.

Since then, numerous details have been refined, including the tidal effects of galaxies beyond the Local Group; the introduction of a cosmological constant, which contributes an additional expansion term, and increases the Local Group mass by some 10%; and the effects of hierarchical growth of both galaxies by comparisons with cosmological simulations.

OF SPECIFIC INTEREST here are the latest refinements of the orbit of M31, and indeed M33, as a result of the improving observational accuracy of their tangential motions, most recently using Gaia DR2 (van der Marel et al., 2019). They determined the proper motions of a number of massive stars in each galaxy, relative to surrounding background quasars, to derive a centre-of-mass proper motion of $(65 \pm 18, -57 \pm 15)$ μarcsec yr^{-1} for M31 (1084 stars), and $(31 \pm 19, -29 \pm 16)$ μarcsec yr^{-1} for M33 (1518 stars), highlighting *'the future potential of Gaia for proper motion studies beyond the Milky Way'*.

Similar values for M31 were determined from Gaia EDR3 by Salomon et al. (2021), who noted that their values were consistent with (but more accurate than) earlier Hubble Space Telescope measurements that predicted a future merger between Andromeda and our Milky Way Galaxy (Sohn et al., 2012).

IN A SUBSEQUENT development, Benisty et al. (2022) adjusted the approach velocity between the Milky Way and M31 for the effects of our Galaxy's motion towards the Large Magellanic Cloud.

They gave two estimates of the total mass of the Local Group: $3.4^{+1.4}_{-1.1} \times 10^{12} M_\odot$ when using a higher tangential velocity for M31 of 80 km s^{-1}, and $3.1^{+1.3}_{-1.0} \times 10^{12} M_\odot$ when using a lower value of 60 km s^{-1}. This can be compared with a *known* mass of $3.7^{+0.5}_{-0.5} \times 10^{12} M_\odot$, based on the likely mass of the four most substantial members (the Milky Way, M31, M33 and the LMC), with the minor members contributing about $0.7 \times 10^{12} M_\odot$.

I MUST MAKE another aside. The Large Magellanic Cloud is the most massive satellite galaxy of the Milky Way, at more than a tenth of our Galaxy's mass. Just past its closest approach of about 50 kpc, and passing the Milky Way at more than 300 km s^{-1}, this 'fly-by' is affecting our Galaxy in various ways… including dislodging the Milky Way disk from the Galactic centre of mass!

Petersen & Peñarrubia (2021) used Gaia DR2 data to show that the Milky Way disk is moving with respect to stellar tracers in the outer halo, at about 32 km s^{-1}, and in a direction that points to an earlier location on the trajectory of the Large Magellanic Cloud.

The timing method, I recall, rests on modeling the relative orbit of our Galaxy and M31. And this will be affected by any perturbation of our Galaxy's equilibrium, including this recent pericentric passage of the LMC, along with any by-products of its merger history.

Accounting for this perturbation of the Milky Way disk lowers the inferred Local Group mass by 10–20% compared to a static halo. Specifically, Chamberlain et al. (2023) gave an updated total mass of between $4.0^{+0.5}_{-0.3} \times 10^{12} M_\odot$ or $4.5^{+0.8}_{-0.6} \times 10^{12} M_\odot$ depending on their adopted kinematic measurements of M31.

SLOWLY BUT SURELY, and in numerous ways, Gaia is contributing to our understanding of the Local Group, its total mass, and its implications for cosmology.

95. Our Galaxy's tumbling disk

I WILL LOOK AT two large-scale dynamical phenomena which are, today, believed to affect the bulk motion of our Galaxy's disk with respect to its dark matter halo.

The first, not yet observed, but tied to their primordial origin, would be a remarkable confirmation of one of the predictions of the cosmological ΛCDM structure formation paradigm. The other, related to the orbit of our neighbouring Large Magellanic Cloud, results in a huge disk perturbation which has recently been measured using the astrometric data in Gaia Data Release 2.

IN THE DECADE or so preceding the launch of Gaia, the simple view in which our Galaxy's disk is fixed in orientation within a spherical dark matter halo was being challenged by a number of studies and simulations.

It seemed increasingly inevitable that the disk–halo orientation is likely to vary with time, due to effects including the infall of misaligned gas (Binney & May, 1986; Roškar et al., 2010, and the tidal effects of the Large Magellanic Cloud (Weinberg & Blitz, 2006). These tidal effects presumably extend to ancient infall satellites and their associated debris (e.g. Dodge et al., 2023).

Meanwhile, simulations also showed that, in galaxies more generally, the inner disk and outer halo often decouple, with average misalignments of $30-40°$.

OTHER COMPLICATIONS were also becoming evident. Amongst these, simulations of galaxy formation were predicting that most galaxy halos 'tumble', in inertial space, with a typical rotation rate of around 2 radians per Hubble time (Bailin & Steinmetz, 2004).

Furthermore, analytical arguments (e.g. Nelson & Tremaine, 1995), as well as simulations (Dubinski & Kuijken, 1995), suggested that dynamical friction in the inner galaxy regions should couple the inner disk to the halo, at least within the region dominated by baryons rather than dark matter (Bailin & Steinmetz, 2004).

It follows that, if the angular momentum vectors of the disk and halo remain coupled, the plane defined by the disk stars should rotate at some 30 μarcsec yr^{-1} with respect to the quasar reference frame.

In 2014, with colleagues David Spergel and Lennart Lindegren, I looked at the detectability, with Gaia, of a time-varying orientation of our Galaxy's disk with respect to its dark matter halo (Perryman et al., 2014b). Our simulations suggested that any residual rotation of the Gaia reference frame with respect to an inertial reference system (approximated to by observable quasars) should be at the level of some $0.2-0.5\,\mu$arcsec yr^{-1}.

Although we must await future Gaia results for confirmation, this accuracy of the reference frame should, in due course, allow for the detection of such a cosmological 'tumbling' motion of our own Galaxy's disk.

It has since been shown, from cosmological hydrodynamical simulations, that the same effect may be detectable in other Local Group galaxies. For example, Earp et al. (2019) found average tilting rates for their sample of $3.8\pm2.3°$ Gyr^{-1}, compared to a detection limit of some $0.27°$ Gyr^{-1}.

IS THERE any evidence that our Galaxy's disk is tilted with respect to a (non-spherical) dark matter halo? One such example, also advanced by Gaia, is provided by two significant stellar 'overdensities' in the outer halo: the Virgo stellar stream, or Virgo Overdensity, discovered in 2006 (Duffau et al., 2006), and the Hercules–Aquila Cloud, discovered in 2007 (Belokurov et al., 2007).

Han et al. (2022) found that the debris from the Gaia–Enceladus merger, some 8–10 Gyr ago, and discovered in the Gaia data (essay #15), can be associated with these overdensities only if the halo potential is tilted, while orbit integration within a *spherical* halo would result in the rapid phase mixing of any initial asymmetries. In their tilted model, the Hercules–Aquila Cloud and the Virgo Overdensity arise as 'apocentric pile-ups' of the merger debris, which then persist in space over many Gyr.

They concluded that a coherent picture is now emerging, in which the Galaxy's stellar halo out to 30 kpc is dominated by debris from this ancient radial merger, which is itself globally asymmetric, and demarcated by the two apocentres of the Virgo Overdensity (above the disk), and Hercules–Aquila Cloud (below the disk).

Further support for this picture, also coming from the Gaia DR2 data, is the fact that these features occupy opposing octants of the Galaxy, overlap in orbit properties, and exhibit metallicity distributions indistinguishable from that of Gaia–Enceladus.

More specifically, Iorio & Belokurov (2019) used the DR2 data to construct slices parallel to the Milky Way disk, showing that within 30 kpc from the Galactic centre the halo is triaxial, with the longest axis misaligned by $\sim 70^\circ$ with respect to the Galactic x-axis. The procedure again exposed the large diffuse overdensities, identified with the Hercules–Aquila Cloud and the Virgo overdensity, and aligned with the semi-major axis of the halo.

Furthermore, they constructed the kinematics of the inner halo by mapping the proper motions of high-luminosity RR Lyrae stars, demonstrating that the inner halo is dominated by stars on highly eccentric orbits, with pericentres around 1 kpc and apocentres at 15–25 kpc from the Galactic centre (Simion et al., 2019).

FOR THE REMAINDER of this essay, I will explore the other effect on our Galaxy's disk that I mentioned at the start: the consequences of the existence of the Large Magellanic Cloud. First, we need some background.

High-precision proper motions of stars in the Large and Small Magellanic Clouds made with the Hubble Space Telescope around 2006, suggested transverse velocities of ~200–300 km s^{-1}, and led to orbit solutions suggesting that the Clouds are either on only their first infall about the Milky Way, or are on an eccentric long-period orbit of more than 6 Gyr (Kallivayalil et al., 2013).

Gómez et al. (2015) demonstrated that previous estimates of the Large Magellanic Cloud's orbital period and apocentric distance derived assuming a *fixed* Milky Way are significantly shortened in models where the Milky Way moves freely in response to the gravitational attraction of the LMC, and that the probability of models favouring a first infall is accordingly reduced. Including this interaction, they found that the Milky Way centre of mass within the inner 50 kpc can be significantly displaced, by as much as 30 kpc in position and 75 km s^{-1} in velocity, over time scales of only 0.3–0.5 Gyr.

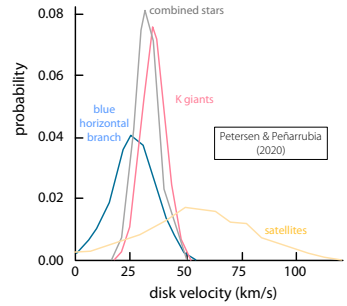
Garavito-Camargo et al. (2019) used numerical simulations to explore the expected consequences as these satellite galaxies orbit within the dark matter halos of their hosts. For example, they found that the Large Magellanic Cloud should generate a pronounced trailing 'wake', in both the stellar and dark matter halos, leading to overdensities and distinct kinematic patterns that should be observable with Gaia.

Further numerical simulations by Petersen & Peñarrubia (2020) confirmed that the Milky Way disk may be moving at 40 km s^{-1} relative to the barycentre prior to the Large Magellanic Cloud infall, dislodging the Milky Way disk from the Galaxy's centre of mass in the process.

WITH THIS BACKGROUND it should come as no surprise that, as the most massive satellite galaxy of the Milky Way, and just past its closest approach of about 50 kpc, and moving at more than 300 km s^{-1}, the LMC is likely to affect our Galaxy in a number of ways.

Petersen & Peñarrubia (2021) have quantified these effects using three classes of distant tracer objects: 543 K giants and 292 blue horizontal branch stars, selected from and with radial velocities determined by SDSS–SEGUE, and with distances and proper motions from Gaia DR2, along with 33 satellite galaxies with suitably determined orbits, again constrained by Gaia.

They found that the Milky Way disk is moving with respect to these various stellar tracers in the outer halo (at distances of 40–120 kpc) at a velocity of 32 ± 4 km s^{-1}, in the Galactic coordinate direction $l = 56 \pm 9^\circ$, $b = -34 \pm 10^\circ$. Interestingly, this direction points not towards the present-day location of the LMC, but rather to an earlier location on its orbital trajectory. This is a consequence of the fact that the distant tracers have long dynamical times, and are thus slow to react to the LMC's gravitational pull. The result is that the Milky Way disk moves in response to the LMC while the outer halo does not; in other words the distant halo appears to move relative to the Milky Way disk.



The significant reflex motion, and the location of the apex, is consistent with a massive Large Magellanic Cloud falling in for the first time. Its magnitude is also consistent with numerical models where the infall mass is greater than about $10^{11} M_\odot$. Being significantly larger than the mass enclosed within its luminous radius, $\sim 1.7 \times 10^{10} M_\odot$, suggests that the infalling LMC was surrounded by an extended dark matter halo.

THEIR RESULTS indicate that dynamical models of our Galaxy cannot neglect the perturbations induced by the LMC infall, nor can observations of the stellar halo be treated in a reference frame that doesn't correct for the disk's reflex motion.

And they pose a number of questions: what was the structure of the stellar and dark matter halos prior to its infall? What is the trajectory of the LMC since the start of its infall, and how did the dynamics of dark matter alter its path? Does the LMC lose its dark matter halo to tides, and where does the stripped dark matter end up in our Galaxy? And what are the implications for the dynamics of the Local Group more widely, including the orbit of our nearest massive neighbour, Andromeda?

Gaia DR3 will surely assist with some of the answers!

96. Is the Earth flat?

Is the Earth round, or is it flat? It is a fascinating question. Asking it shows the power of enquiring minds to probe the accepted wisdom. Answering it has opened up a huge range of insights into the natural world.

As a *scientific* question, the solution to which must fit consistently within a vast range of natural phenomena, it was answered more than two millennia ago, by the Greeks of the Hellenestic period, some 300 BCE.

Aristotle drew on arguments such as southern constellations rising higher above the horizon the further south one travels, while the shadow of the Earth on the Moon during a lunar eclipse is circular. Cleomedes, and later Eratosthenes, determined Earth's circumference by measuring the length of the Sun's shadow at one location in Egypt (Alexandra) at noon on the summer solstice when the Sun was directly overhead at Syene.

Amongst various methods for determining the Sun's distance, estimates based on the Transit of Venus gave successive improvements in the 1600s, 1700s, and 1800s.

'Direct' demonstration of Earth's spherical form followed from Ferdinand Magellan's circumnavigation of the globe between 1519–1522, while the Transglobe Expedition of 1979–1982 was the first to make a surface circumnavigation via the poles.

Any remaining scientific doubt was firmly excluded by the advent of Copernicanism in the mid-1500s. This placed the Sun at the centre of the solar system, with the Earth and other planets in orbit around it. Newton's laws of gravity gave both a mathematical foundation to this hypothesis, as well as enormous predictive power which has been verified countless times since.

In practice, the Copernican 'revolution' took some decades to gain full acceptance, accompanied by a reluctance to accept the idea that the Earth is really moving through space. But from that time, until at least the mid-19th century, the idea that the Earth was 'round' rather than flat was largely unquestioned.

Since then, the scientific method of iterating observation, hypothesis, verification and prediction, has enabled scientific advances at an enormous pace.

The scientific method has led to innumerable advances across physics, chemistry, geology, biology, medicine, and beyond. In theoretical terms it has led to a detailed, deeply interconnected understanding of fields such as electromagnetism, optics, relativity, quantum mechanics and particle physics.

In practical terms this has, in turn, given us railways, cars, planes, televisions, mobile phones, the internet, satellite navigation, MRI… and so on. These practical advances are embraced by the vast majority, even if their roots are not necessarily well understood.

Let me continue to use the term 'round' although we now know that the Earth, largely because of its rotation, is a slightly flattened 'ellipsoid'. We can explain its form, and its structure, as arising from the complex process of planetary formation, all shaped by the force of gravity.

Along with these ideas come a vast range of experimental results, and a truly remarkable self-consistent picture, of the Earth's formation, its structure and its geological history, embracing seismology, radioactivity, plate tectonics and a host of other phenomena.

A huge range of natural phenomena is explained by today's framework of physics and mathematics: we can comprehend and predict seasons on Earth, along with the 'wobbling' effects of precession and nutation. We can explain the motions of the planets, of planetary moons, and of other solar system bodies. We can predict eclipses, planetary transits, stellar occultations, and the path of artificial satellites like the Space Station across the night sky. It gives us Earth observation and meteorological satellites, and a multitude of geostationary, telecommunications, and global navigation satellites.

And a whole host of even more 'subtle' natural phenomena are explained (and predicted) by our current understanding of physics: amongst them, stellar and Galactic aberration, relativistic precession of the orbit of Mercury, and lunar libration, small changes in perspective and rotation which bring slightly different regions of the Moon's surface into view at different times.

The list of successes is indeed impressive. And our modern high-tech society rests firmly upon them.

WHY, THEN, DO SOME people, in the face of such towering and overwhelming evidence to the contrary, and in the absence of any other self-consistent theory, cling vigorously to the idea that the Earth is flat?

According to those who have studied this more deeply, a fraction of those that profess belief in a 'Flat Earth' probably do so to be provocative. But let me place these people to one side, and focus on those who sincerely believe that the Earth is flat. Could they be right?

THE RISE IN THE BELIEF in a flat Earth seems to have taken hold in the mid-1800s. It appears to have developed alongside the major and rapid advances that were occurring across all of the sciences at the time.

Inevitably, perhaps, some people were uncomfortable with these advances, concerned that the natural world was being interpreted in a way that was becoming ever more deeply disconnected with the sorts of phenomena that could be experienced, explained, and understood by any one individual. Should one really accept explanations that could not, immediately and categorically, be confirmed by a rational mind?

Into this space stepped Samuel Rowbotham. In what was a primitive experiment even then, he found no noticeable curvature on a 10-km long straight drainage ditch of the Bedford Levels in East Anglia. He became convinced of the flatness of the Earth, started to lecture about it, and published a book on the subject in 1865.

The movement was continued by others, amongst them Samuel Shenton, and it remained active into the early part of the 20th century, when it went into decline. But it was revived in 1956 as The Flat Earth Society.

ONE DIFFICULTY in grasping a clear pecture of 'the' flat-Earth hypothesis is that different proponents often hold somewhat different viewpoints. But as I understand it, the central tenets are that the North Pole is at the centre of the flat Earth's disk, and its circumference is marked by an ice wall which is guarded (some say by NASA) to prevent anyone going beyond.

In this model, gravity is due to the flat Earth accelerating upwards, continuously, at $9.8\,\mathrm{m\,s^{-2}}$. The Sun is a 'small' object not far above the Earth's surface (as are the Moon, planets and stars), and its motion gives rise to the seasons. I am not clear what stars are in this model, or how anything of our Galaxy's structure is interpreted.

But if we accept that the Earth is flat, are the Sun and Moon and other planets also flat? This question was put to the then president of the Flat Earth Society, Daniel Shenton, in a Guardian podcast with David Adam and Brian Cox in 2010. *'I'm not sure about them,* he said. *'It seems to me from the evidence that I've seen that they're probably spheres, but it's definitely not proven.'*

His view on space was that *'As far as the evidence of the space age, anything like that can be easily faked.'*

YOUGOV is an internet market research and data analytics firm. In 2019, they polled views on which science-based conspiracy theories the British believe in.

You can see, from this link, the significant fraction of people who consider that the threat of climate change is over-exaggerated; that vaccines have harmful effects which are not being fully disclosed to the public; that the moon landing was staged; and that the Earth is flat.

Results of a similar poll in the US in 2018, found that only 84% said with certainty that the Earth is a globe, a fraction even lower among young people at a mere 64%.

THERE IS AN important message for educators and scientists, and one which has been eloquently expounded by science communicator Sabine Hossenfelder. The belief in a flat Earth has, perhaps, less to do with the shape of the Earth, and more to do with the way today's big science is conducted and communicated. The relevant institutions are too large and complex for any individual to easily comprehend.

A consequence can be that without a tolerable understanding of 'the scientific method', by which conjectures *can* be refuted, but only with solid evidence, it is not unreasonable that some will choose to mistrust anything that cannot be easily verified. While this may not matter too much in the question of whether the Earth is round or flat, the consequences will be more serious in the case of, say, uninformed views on vaccines (such as MMR and Covid), or on climate change.

In the same Guardian podcast, Robert Winston argued that in hoping to tackle climate change, wider society must be 'on side', implying that scientists somehow need to communicate even better with society. And in the spirit of the flat-Earth proponents, he even emphasised the point that we should not *necessarily* trust Governments, for they may well have other (in particular, short-term) agendas, predicated on their own survival.

Brian Cox concluded that: *'Science is all about checking, but at the same time one has to accept the settled view of experts… As an idea, to be a sceptic, I like the idea of a Flat Earth Society… It's just that, in this case, they've picked the wrong thing to be sceptical about.'*

WHAT HAS ALL THIS to do with Gaia? Simply that this advanced space mission continues to add enormous weight to our global understanding of Nature, underpinned by the Earth's rotation, a massive Sun some 160 million km from Earth, and stars at vast distances.

As just one example, we can now predict, months in advance, and to within a minute or so, hundreds of star occultations by minor solar system bodies (essay #24).

Advocates of a flat Earth might enjoy the challenge of trying to predict *just one* such stellar occultation, say, a few months in advance and with a timing accuracy of better than a minute. Do we have any takers?

97. Life on other worlds?

I WILL DEVIATE from discussing Gaia directly to review the philosophical debates that have raged since antiquity on whether 'populated' worlds exist beyond our own. I draw largely on the extensive literature chronicled by Michael Crowe (1986). It provides an interesting historical background to the search for exoplanets, and life on other worlds, which *are* being assisted by Gaia.

THE ANCIENTS of both Greece and Rome were already deeply divided. Arguing for the 'plurality' of life beyond Earth was Epicurus (341–270 BCE), who wrote *'…there are infinite worlds both like and unlike this world of ours'*, adding that *'we must believe that in all worlds there are living creatures and plants and other things we see in this world'*.

Opposition to this pluralist position dates back to at least the time of Plato and Aristotle. Plato (428–348 BCE) was confident that *'there is and ever will be one only-begotten and created heaven'*.

Post-Copernicus, Galileo Galilei, René Descartes, Pierre Gassendi and Isaac Newton viewed such speculation as being at the limit of legitimate science.

BUT BETWEEN 1500–1750, plurality gained the support of many prominent figures. As summarised by Crowe: *'Presented with exceptional appeal by Fontenelle, given legitimacy in scientific circles by Christiaan Huygens and Isaac Newton, reconciled to religion by Richard Bentley and William Derham, …[it was] integrated into philosophical systems by George Berkeley and Gottfried Wilhelm Leibniz, and taught in textbooks by Christian Wolff'*. Many leading astronomers and philosophers of the 18–19th centuries contributed to the debate.

During the second half of the 18th century, plurality was embraced by a number of influential astronomers. Crowe argues that the writings of pioneers such as Thomas Wright, Immanuel Kant, and Johann Lambert, were permeated by pluralism and the extraterrestrial life debate. Kant's various texts suggest that the 'starry heavens' that so filled him with awe were *'a densely denizened domain wherein millions of inhabited planets orbited suns clustered in endless hierarchies of systems'*.

THE RESEARCH of the period's foremost astronomer, William Herschel, was deeply influenced by ideas of extraterrestrial life. Indeed, his genius notwithstanding, Herschel considered that the Moon had *'the great probability, not to say absolute certainty, of being inhabited'*, his lunar observations evidencing 'vegetation', 'turnpike roads', and 'circuses'.

He referred to the 'inhabitants' of Mars, Saturn, and Uranus, and of their satellites, and to the fact that every star *'is probably of as much consequence to a system of planets, satellites, and comets, as our own Sun…'*, while his observations of sun spots suggested that even the Sun *'…is most probably also inhabited…by beings whose organs are adapted to the peculiar circumstances of that vast globe'* (Herschel, 1795).

Many of his renowned contemporaries, including François Arago, Jérôme Lalande, and Johann Schröter, were also involved in the quest. Johann Bode advocated extraterrestrials with fervour and influence, considering that essentially every celestial body – sun, star, planet, and satellite – was populated with rational beings.

Pierre Simon Laplace's 'nebular hypothesis' of 1796 explained the solar system as resulting from the rotation and contraction of a primitive solar 'fluid', implying that planets are an expected result of stellar evolution, and thus providing support for the pluralist contention that most stars were encircled by planets

And from the cooling times of the Earth and planets, the Comte de Buffon even tabulated, in 1775, the times at which life began and would end on each.

PHYSICOTHEOLOGIANS embraced extraterrestrials as evidence of God's omnipotence, while the religious authorities saw plurality as the product of divine power and generosity.

Thomas Paine's *Age of Reason* of 1793 essentially rejected Christianity because of pluralism. James Mitchell's lecture *On the Plurality of Worlds* to the London Mathematical Society in 1813 argued for life on all planets and satellites. Thomas Dick (1837) actually estimated the population of the solar system bodies (including the rings of Saturn!), as 21 894 974 404 480.

In his *Complete System of Astronomy* of 1797–1808 Samuel Vince, Plumian Professor of Astronomy at Cambridge, considered that novae were *'the destruction of that system at the time appointed by the Deity for the probation of its inhabitants'*.

Other eminent early 19th century astronomers were pluralists and even proponents of lunar life. Historical research suggests that Carl Friedrich Gauss proposed a giant figure in Siberia to signal to the Moon or Mars, while Johann von Littrow suggesting a kerosene-lit canal in the Sahara for similar purposes.

In a letter to the astronomer and journal editor Franz Gruithuisen, Bremen astronomer Wilhelm Olbers stated that *'I hold it to be very probable that the Moon is inhabited by living, even rational creatures, and that something not wholly dissimilar to our vegetation occurs on the Moon'*.

Support for extraterrestrials received a setback with a lecture by the German astronomer, physicist and mathematician Friedrich Bessel in 1834. By appealing to the sharpness with which stars are occulted by the Moon's limb, he argued that it lacks any significant atmosphere, and that it therefore has neither air nor water.

Hegel raised philosophical objections to plurality, and in 1853 William Whewell published an essay claiming that many of the pluralist arguments were scientifically defective and religiously dangerous.

This led to renewed debate involving, among others, physicist David Brewster, Oxford Savilian professor Baden Powell, astronomers Lord Rosse and Camille Flammarion, populariser Richard Proctor, mathematician Augustus De Morgan, and biologist Charles Darwin.

The advent of spectroscopy, pioneered by William Huggins, Anders Ångstrom, Henry Draper, Jules Janssen, Joseph Norman Lockyer, and Angelo Secchi, led to more quantitative stellar investigations, but also provided further substance for the pluralist debate; in fact *'to a greater extent than has been previously noted, [they] saw themselves as involved in that debate'*.

In their first major spectroscopic study of nearly fifty stars, Huggins & Miller (1864) identified numerous spectral lines coinciding with the terrestrial elements. And they concluded that *'There is therefore a probability that these stars are... like our Sun, surrounded by planets.'*

And they went on: *'It is remarkable that the elements most widely diffused through the host of stars are some of those most closely connected with the constitution of the living organisms of our globe... On the whole we believe that the foregoing [observations] contribute something towards an experimental basis on which a conclusion... may rest, viz. that at least the brighter stars are, like our sun, upholding and energising centres of systems of worlds adapted to the abode of living beings'*.

Subsequent debates surrounded the suggested observational evidence for life in our solar system. Amongst these were varying features in the lunar crater Linné (Schmidt 1866; as summarised by Moore, 1977); as well as consequences of the lunar figure for habitability (Hansen 1856). There were various reports of Martian or Venusian 'signals', involving Nikola Tesla, Lord Kelvin, and Guglielmo Marconi, as well as claims of lunar vegetation by William Henry Pickering (1902).

Even a mere 80 years ago, the then Astronomer Royal Harold Spencer Jones (1940, p207) considered it *'almost certain that there is some form of vegetation on Mars'*.

A century ago saw the protracted debate (1877–1910), and substantial literature, on the question of canals on Mars. Schiaparelli, Lowell, and Flammarion were the main advocates, with Antoniadi, Campbell, and Maunder leading the anti-canal and anti-life arguments.

Sixty pages in Crowe (1986) detail the controversy, its now-favoured explanation in terms of the Maunder–Cerulli theory of unconscious tendencies to see discrete details in lines or circles, and other assessments of its context and scientific consequences (Shklovskii & Sagan, 1966; Sagan & Fox, 1975; Wells, 1979).

Today, the existence of (admittedly more primitive) life elsewhere in the solar system is still under serious investigation, and the debate remains unresolved.

In the search for life on the growing number of known exoplanets, an important consideration is the optimum spectral range (optical, infrared, or both), and which spectral lines offer the best prospects of indicating the presence of life. The arguments are largely founded on carbon-based life as presently known.

The search for extra-terrestrial *intelligence*, SETI, is motivated by the belief that intelligent life is likely to emerge under conditions similar to those on Earth.

Loosely connected but not implicit in such searches are various unproven and somewhat inconsistent postulates, amongst them the 'anthropic principle', which suggests that no assertion can be made about the probability of intelligent life based on a sample set of one); the 'mediocrity principle' (which, given the existence of life on Earth, asserts that life typically exists on Earth-like planets throughout the Universe); and the 'fine-tuning hypothesis' (which asserts that the natural conditions for intelligent life are implausibly rare).

Attempts to quantify the probability that intelligent life exists elsewhere in the Galaxy include consideration of the 'Drake equation', and of the 'Fermi paradox', viz. *'If other advanced civilisations exist, where are they?'*. Resolution of these questions may still lie far in the future.

*****

This essay is based on my text in The Exoplanet Handbook, Cambridge University Press (2nd edition, 2018).

98. Boyajian's Star(s)

NASA's KEPLER mission targeted the discovery of exoplanets through the transit method, and was operated between 2009–2018. It observed some 530 000 stars, and detected more than 2500 planets.

Amongst its many discoveries, remarkable for their sheer number but especially for the richness of the physical phenomena uncovered, are numerous 'hot Jupiters' and other planets orbiting very close to their host stars, often with periods of only some 3–4 days, or less!

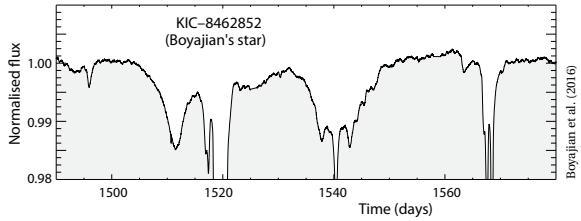
Having been 'delivered' there, by some mix of tidal migration, scattering, and resonance effects, they will eventually spiral in towards their host stars as the angular momentum of their orbital motion is transferred to the host star rotation through tidal coupling.

The end result is some combination of tidal disruption, disintegration, and planet engulfment. The disintegration of such highly irradiated planets releases large amounts of gas and dust into an exospheric tail.

THE FIRST SUCH (presumed) disintegrating planet was around the 16 mag star Kepler–1520, which exhibits transit-like features every 15.7 h, varying in depth between 0.2–1.2%. It was promptly explained as a disintegrating rocky planet with a trailing cloud of dust of varying optical depth, created and constantly replenished by thermal surface erosion (Rappaport et al., 2012).

Amongst a few other such candidates was the curious KIC–8462852, discovered through the contribution of the amateur 'Planet Hunters' collaboration on the basis of its unusual light-curve (Rappaport et al., 2012), and subsequently often known as Boyajian's star. The discovery paper reported pronounced dimming by up to 20%, lasting between 5–80 d, and with an irregular cadence and unusual profile. It has been subsequently monitored intensively at both radio and optical wavelengths.

Considerable speculation continues to accompany the unusual lightcurve. Interpretations range from occulting clouds of exocomets, a disrupted exomoon, and effects due an orbiting binary star. It has even been attributed to a 'swarm of megastructures', and hence as an outstanding SETI target (Wright et al., 2016).



WHILE THE ATTRIBUTION of the curious and complex lightcurve to some sort of alien-civilisation structure is not considered mainstream, the existence of such new and difficult-to-explain variability phenomena has certainly inspired the continued search for possible signatures of alien life.

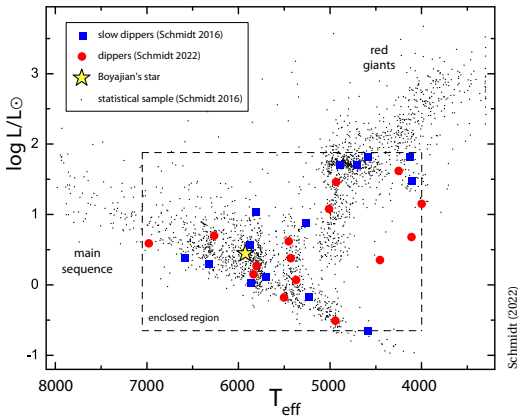
Meanwhile, I will focus here on the search for, and Gaia's contribution to, additional examples of the sorts of lightcurves exhibited by Boyajian's star, characterised by infrequent episodes of small brightness dips and a long-term decline in brightness between the dips.

EVEN 4–5 YEARS after its discovery, no fully satisfactory mechanism has been established for its behaviour. Explanations are made more challenging because there was only this one known example.

To search for other stars showing similar 'dipping' behaviour, Schmidt (2019) used the extensive photometric data from the Northern Sky Variability Survey (NSVS) and the All Sky Automated Survey for Supernovae (ASAS–SN).

He reported 21 stars identified as possible dippers, with 15 showing similarity to Boyajian's star (he classified these as 'slow dippers', with narrow dips of less than a few days, and fewer than three 'dips' per year), and the other 6 likely to be more extreme examples of the same phenomenon ('rapid dippers' with more than eight 'dips' per year). With this search covering some 15 times more stars than Kepler, his 15 'slow dippers' are consistent with finding just the one amongst the Kepler sample.

A further extension to the search, over a larger region of the sky, yielded an additional 15 new 'slow dipper' candidates (Schmidt, 2022).



USING DISTANCES from Gaia Early Data Release 3 to establish their luminosities (and slightly modifying the preliminary conclusions in his first paper), Schmidt (2022) showed that the dipper candidates occupy a restricted region of the Hertzsprung–Russell diagram, a region which embraces both main sequence stars of around 1 solar mass, and stars in the red giant region near the evolutionary track for stars of around 2 solar mass. He considered those near the main sequence as likely to be analogues of Boyajian's star.

The various groupings are shown in the above figure, along with a statistical sample of stars representative of the selected dippers selected from Gaia DR2. His analysis suggests that the dipper stars are clustered significantly more in the Hertzsprung–Russell diagram than a random sample, therefore concluding that his searches have isolated a physically meaningful group of stars.

THINGS GET a little more interesting, and undoubtedly a little more contentious, when it comes to examining the distribution of these stars on the sky. Here, Schmidt (2022) drew attention to an apparent clump of 12 dippers between $254 - 303°$ in right ascension (plus Boyajian's star, incidentally at a Gaia DR3 distance of 444 ± 2 pc, or 2.2545 ± 0.0099 mas), while the density of stars elsewhere is much lower. He rules out the appearance of the clump as originating from the higher star density near the Galactic plane.

Using the Gaia EDR3 parallaxes, the next step was to calculate the three-dimensional coordinates of the dipper candidates, using a frame in which the origin is at the Sun, the x axis is in the direction of the centre of the clump, y is towards the west, and z towards the north.

Four of the 12 clump stars are more than 1000 pc from the Sun, and clearly separated from the others. The remaining 8 stars in the clump, as well as Boyajian's star, form a structure that is about 50% longer in the x direction than in the y or z directions. The parallax uncertainties are all smaller than 0.8%, much too small to explain the apparent elongation of the clump.

Schmidt continues: 'Since no fully satisfactory explanation for the behaviour of Boyajian's star… has been found, it is premature to try to explain the existence of the clump. However, the possibility that extraterrestrial civilisations might have developed interstellar travel and expanded beyond their original planetary systems has been widely discussed in connection with the search for extraterrestrial intelligence. If this is actually possible, it could lead to a clump of stars with inhabited planets over an extended region of space… With these considerations in mind, I suggest that the dippers in the clump and other stars in the same region would be appropriate targets for SETI searches.'

THE SEARCH FOR unusual photometric lightcurves is undoubtedly still in its infancy. And the application above illustrates the power of Gaia in establishing the location of identified targets in the Hertzsprung–Russell diagram, and indeed their location in space.

Such surveys will be advanced dramatically by the Vera C. Rubin Observatory (previously LSST), where full survey operations are to begin in October 2024. But stimulated by the discovery of complex variables like Boyajian's star, renewed emphasis is also being placed on more robust algorithmic searches.

For example, Chakraborty et al. (2020) report a detection algorithm which makes minimal assumptions about the shape of a periodic signal. Applied to 240 000 TESS sources from the first-year southern sky survey, they found 377 previously unreported periodic signals, amongst which 26 are ultra-short periods, 313 are likely eclipsing binaries, 28 appear planet-like, and 10 are miscellaneous signals. Interestingly, none display the unusual behaviour displayed by Boyajian's star.

In their 'Vanishing and Appearing Sources during a Century of Observations' project, Villarroel et al. (2020) are using existing survey data to search for exceptional astrophysical transients, perhaps extending to evidence of technologically advanced civilisations.

THERE IS at least one other case where multiple unexplained sources appear to be clustered on the sky. Villarroel et al. (2021) identified 9 (apparent) transient sources in a region about ten minutes of arc across on a plate taken in April 1950 as part of the Palomar Observatory Sky Survey. All are absent on both previous and later photographic images, and absent in modern surveys with CCD detectors which go several magnitudes deeper. Were the plates subjected to an unknown type of contamination? Did they result from solar reflections from objects near geosynchronous orbits? Or what?

IN MY NEXT essay, I will look at one other specific search for alien civilisations that is ongoing, and also aided by the Gaia results: the search for Dyson spheres.

99. Searching for Dyson spheres

In my previous essay, I looked at Gaia's contribution to the search for additional examples of the curious lightcurve, discovered from Kepler photometry, of Boyajian's star. Some of the interest in this star arose because of the early inference that its apparent variability could be a manifestation of one specific type of 'technosignature' originating from an alien civilisation.

In my essay after this, I will assemble some top-level ideas about the origin of life on Earth, the search for life on other worlds, and what our current failure to detect alien life forms, or signals or other signatures of alien civilisations, might be informing us about the 'Fermi Paradox', viz., if other civilisations exist, where are they?

Here I want to look at one specific search for other civilisations which is being assisted by Gaia: the search for so-called 'Dyson spheres'. But let me be clear that I am not attempting to review the ever-growing literature on the Search for Extraterrestrial Intelligence (SETI), or on the topic of Messaging Extraterrestrial Intelligence (METI): in what will hopefully become an annual review, and providing an excellent entry into the relevant literature, 'SETI in 2020' (Wright, 2022), and 'SETI in 2021' (Huston & Wright, 2022b) cover 75 relevant papers published in 2020, and 98 published in 2021, respectively.

Gaia is not directly related to most of these ideas, searches, and philosophical considerations. But the search for Dyson spheres is already making use of the Gaia data in a very fundamental way.

In some of the earliest discussions of SETI, Soviet astronomer Nikolai Kardashev developed the idea of the 'Kardashev scale' as a method of measuring a civilisation's level of technological advancement, based on the amount of energy it is able to use (Kardashev 1964). The classification rests on the premise that any civilisation's energy needs grow continuously with time.

On this Kardashev scale, a Type I civilisation is one that can harness *all* the energy that reaches its home planet from its parent star. For Earth, this value is around 2×10^{17} W, around 10 000 times higher than that presently consumed on Earth, some 2×10^{13} W.

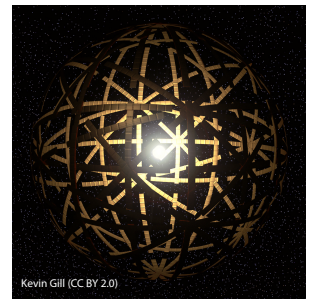
A Type II civilisation is one that has developed to the point that it requires, and can actually consume, most of the total energy emitted by its parent star. And a Type III civilisation requires, and is able to capture, 'all' the energy emitted by its host galaxy.

Assuming of a growth rate of 1% per year, Kardashev estimated that it would take humanity 3200 years to reach Type II, and 5800 years to reach Type III. The aim of such a classification was to guide the search for extraterrestrial civilisations, on the assumption (for example) that a fraction of the energy used by each type is assigned to communicate with other civilisations, for which a very substantial energy source is required.

Without going into further details of the whys and hows, I will focus here on what is often considered as one possible manifestation of a Kardashev Type II civilisation, viz. the presence of a 'Dyson sphere' around a planet's host star.

A Dyson sphere is a (hypothetical) megastructure that encompasses a star, and which captures a significant fraction of its emitted energy. The concept attempts to explain how a space-faring civilisation would meet its energy requirements once this exceed what can be generated from the home planet's resources alone. Because only a tiny fraction of a star's energy emissions naturally reaches the surface of any orbiting planet, building structures encircling a star would enable a civilisation to harvest substantially more energy.

The idea was popularised by physicist Freeman Dyson (1923–2020) in his paper 'Search for Artificial Stellar Sources of Infrared Radiation' (Dyson, 1960). He speculated that such structures would be the logical consequence of the escalating energy needs of a technological civilisation, and would be a necessity for its long-term survival.



A FAIRLY EXTENSIVE literature has grown up around this topic, with various suggested hypotheses as to their detailed structural form, including complete or partial shells, rings, and nets (e.g. Wright, 2020).

The relevant point for efforts to detect such structures is that they would alter the light emitted from the star system through absorption and re-radiation, typically resulting in increased *infrared* radiation in the system's emission spectrum. And if the fraction of starlight so affected was significant, the resulting spectral anomalies could be detected over interstellar distances.

Notwithstanding the immense theoretical and practical difficulties involved in their construction, various searches for Dyson spheres, through identification of a star's modified spectrum, have been carried out. Dyson himself argued that the waste heat of the 'energy metabolism' of this extraterrestrial technology would likely be detectable at 10 μm, and that a search of the sky at mid-infrared wavelengths would be worthwhile.

PRACTICAL SEARCHES began with the launch of IRAS in 1983, creating the first mid-infrared sky survey with plausible sensitivity. Progressive (unsuccessful) searches were then made by Slysh (1985), Timofeev et al. (2000), Jugaku & Nishimura (2004), and Carrigan (2009).

Subsequently, Arnold (2005) suggested using Kepler to identify transiting megastructures; Teodorani (2014) described the use of Spitzer as a more sensitive probe for infrared excesses; Wright et al. (2014) proposed using the WISE survey to search for Dyson spheres, and Chen & Garrett (2021) used the LOFAR Two-metre Sky Survey (LoTSS) to search for more extreme Type III civilisations.

The discovery of Boyajian's star in 2015 (essay #98) raised early (but subsequently retracted) speculation that the first Dyson sphere may have been discovered (Wright et al., 2016; Wright & Sigurdsson, 2016).

IN ALL this work, a continuing difficulty has been to distinguish anomalous Galactic stars from the far more ubiquitous extragalactic infrared (AGN) sources.

And this is where Gaia is contributing: instead of searching for sources with an infrared *excess*, underluminous stars (due to the 'missing' luminosity intercepted by the Dyson sphere) can be identified as a result of their discordant trigonometric and spectroscopic parallaxes (Zackrisson et al., 2015; Wright et al., 2016).

A first application of this approach, using Gaia DR1, has been reported by Zackrisson et al. (2018). As these authors note, with later Gaia data releases, both trigonometric and spectroscopic distance estimates can be obtained from the Gaia data, but in this early analysis, they combined the parallax distances provided by the Tycho–Gaia Astrometric Solution (TGAS) with spectrophotometric distances from Data Release 5 of the ground-based radial velocity survey RAVE.

Candidate Dyson spheres can then also be tested through a comparison of predicted and observed fluxes at mid- and far-infrared wavelengths from space telescopes including IRAS, WISE and Herschel.

Of the 230 000 objects in common between Gaia DR1 and RAVE DR5, they focused on the main-sequence FGK dwarfs, for which current stellar models are expected to produce reliable results. They were left with just 6 outliers, of which TYC 6111–1162–1 was considered to be the most appropriate for further investigation.

For this object, they finally adopted the spectrophotometric parallax listed by RAVE as 4.51 ± 0.8 mas, corresponding to a distance 222^{+50}_{-35} pc, while Gaia DR1 gives a trigonometric parallax 8.72 ± 0.53 mas, implying a distance almost a factor two smaller.

Interpreted as a partial Dyson sphere, it would correspond to a fractional star coverage of 0.77, and a dimming of the host star by around 1.6 mag at optical and near-infrared wavelengths.

Any excitement about the possible discovery of a Dyson sphere was, however, short-lived. A footnote to the article reports that Gaia DR2, released while the paper was under review, confirms that the Gaia DR1 parallax was indeed somewhat preliminary, replacing it with a significantly smaller DR2 parallax of 5.75 ± 0.17 mas. And the astrometric solution may have been further affected by an unseen binary companion, which will become clearer from later data releases.

DESPITE ANOTHER FAILED SEARCH, the power of the method has nonetheless been validated. And future results based on Gaia DR3 should represent another major step forward: in numbers of objects, improved astrometric solutions, identification of binary companions, and by providing both trigonometric and spectroscopic distances from the same global data set.

Other complications remain: one effect that could give rise to spurious Dyson-sphere candidates is grey dust, i.e. line-of-sight material that obscures starlight without causing significant spectral reddening.

SEARCHES FOR ALIEN civilisations are clearly in their infancy, and further ideas along these lines have recently been detailed (e.g. Haqq-Misra et al., 2022; Huston & Wright, 2022a). Yet more exotic approaches to the detection of advanced alien civilisations are being examined, amongst them the proposal to detect signatures of 'star-lifting' as a means of extending the main-sequence lifetime of a host star (Scoggins & Kipping, 2023).

Meanwhile the importance of searching for *both* bio- and techno-signatures, especially in the context of understanding the longevity of alien civilisations, has been emphasised by (Wright et al., 2022). I will return to the issue of what non-detections of either might be telling us about our place in the Universe in my next essay.

100. The Fermi Paradox

I TOUCHED ON the Fermi paradox in my essay #55. It refers to the question famously posed by the eminent nuclear physicist Enrico Fermi in 1950: *'If other advanced civilisations exist, where are they?'*. Alternatively formulated as *'If alien civilisations existed, they would be here'*, or as the 'Great Silence' problem, it is a deceptively simple question that presents a challenge for theories assuming a naturalistic origin of life and intelligence.

Following my last two essays on the nature of 'Boyajian-type stars', and the search for Dyson spheres, I wanted to look at it in a slightly wider context... even though it is not immediately linked to new Gaia results.

SEARCHES FOR LIFE, and more so *intelligent* life, are faced with the question: 'what is life?' Interestingly, despite abundant empirical data, there is no generally accepted definition. In biology, it is considered a characteristic of organisms that display all, or most of, a set of phenomena including metabolism, growth, adaptation, reproduction, regulation, and response to stimuli.

An alternative came through NASA's efforts to create a working definition for its astrobiology programmes. This defined life as: 'a self-sustaining chemical system capable of Darwinian evolution'. Nonetheless, difficulties of definition aside, intuitive concepts of life and intelligence are probably sufficiently developed to allow relevant research to advance.

PROGRESS IN UNDERSTANDING the *origin* of life on Earth is a big and complex story – involving chemistry, biology, and geophysics (and many hypotheses) – and based on the fact that life on Earth is dependent on the specialised chemistry of carbon and water. Indeed, complex organic molecules are found widely in the solar system and beyond, and it is these that are considered to have provided the starting materials. Candidates for first life include hydrothermal vents and extremophiles.

A few relevant timescales are: 13.7 Gyr for the origin of the Universe, 4.5 Gyr for the origin of our solar system, 3.5 Gyr for the first life forms, and 300 000 for modern humans. And we should note that many other stars are much older, while many others are much younger.

FIVE MAJOR STAGES of the origin of life have been suggested: starting from inorganic molecules and the synthesis of the simplest organic building blocks, continuing through self-organisation into more complex molecules, leading to the first reaction pathways, the spontaneous aggregation into macrostructures and, eventually, a system capable of being shifted from chemical equilibrium, maintaining homeostasis (regulation) and undergoing evolution.

But the processes are challenging: Howland (OUP, 2000) suggested that believing that a living organism would inevitably arise through self-assembly is *'...similar to a tornado descending on a junkyard and producing, by self-assembly, a jet airliner'*.

IT IS WORTH stressing that there is, today, no clear evidence for life beyond Earth. Elsewhere in the solar system, the present status is broadly as follows. In the absence of liquid water, neither the Moon nor Mercury are considered to have ever been habitable. Venus and Mars are both considered uninhabitable today, although both may have been habitable earlier in their history.

Amongst the various icy moons of Jupiter (Europa, Ganymede and Callisto), Saturn (Titan and Enceladus), and Neptune (Triton), liquid water oceans may exist beneath surface ice layers, so providing possible habitats for life. And there are various plans for space missions to probe some of them for possible biosignatures.

IN THE ONGOING search for life on exoplanets, spectroscopy is considered a key tool. Which spectral features are best targeted tends to rest on the idea that life is C/H_2O based. For example, O_2 may be a possible but not necessary by-product, with O_2 and O_3 absorption considered as promising biosignatures. With CH_4 resulting from anaerobic decomposition of organic matter, the (high disequilibrium) existence of both CH_4 and O_2 may be strong evidence for the presence of life.

Wider considerations of this type motivated the ESA Darwin and NASA TPF missions to target *infrared* observations for key spectral lines, such as H_2O at 6–8 μm, CH_4 at 7.7 μm, O_3 at 9.6 μm, and CO_2 at 15 μm.

THE DRAKE EQUATION was formulated by Frank Drake (1961) as an aid in trying to estimate the number of technical civilisations in the Galaxy today, at or beyond our technology level. It takes the form:

$$N = R_{\rm SF} \times f_{\rm p} \times n_{\rm e} \times f_{\rm i} \times F_{\rm i} \times F_{\rm c} \times L$$

where $R_{\rm SF}$ is the star-formation rate averaged over the Galaxy lifetime; $f_{\rm p}$ is the fraction of stars with planetary systems; $n_{\rm e}$ is mean number of planets in such systems suitable for life; $f_{\rm i}$ is the fraction of such planets on which life originates; $F_{\rm i}$ is the fraction of such planets on which 'intelligence' develops; $F_{\rm c}$ the fraction of such planets developing a communicative phase; and L is the mean lifetime of such technical civilisations.

Our ability to attach plausible probabilities to each of these terms decreases (strongly) down this chain. Estimates for N accordingly vary widely, from 10^6 (Sagan 1966), to much less than 1 (Barrow & Tipler, 1988).

THE SEARCH for intelligent life, SETI, is meanwhile motivated by the belief that intelligent life is likely to emerge under conditions similar to Earth's.

Amongst past or present searches are those undertaken at radio wavelengths, including an all-sky survey at 1–10 GHz funded by US Congress until 1993, and later continued as Project Phoenix (using the Parkes 64-m, Green Bank 140-ft, and the Arecibo 300-m); the Allen Telescope Array (42 antennae operational since 2007); and Yuri Milner's (2016, 100M$) 'Breakthrough Listen' initiative using the Parkes and Green Bank telescopes.

Optical searches have been made based on intercepting targeted interstellar laser communications, arguing that pulses of 1 TW over many pico-seconds might be detectable across our Galaxy.

Others are being made for anomalous celestial objects such as Oumuamua (discovered in 2017, #25), and Boyajian's star (in 2015, #98), and even signatures of alien engineering (e.g. Dyson spheres, #99). And I referred to a couple of the most renowned 'false alarms' in essay #55, including the episode during the discovery of pulsars in 1967, and the 'Wow' signal of 1977.

WHAT MAKES Earth habitable in the first place? The answer appears to be a remarkable combination of many fortuitous circumstances. Amongst them, our Sun appears to be a particularly suitable host star (not too massive, not too light), and one without superflares.

Our Earth is at the 'perfect' distances from the Sun, resulting in liquid surface water, and it placing us not only in the 'habitable zone', but also in the 'continuously habitable zone' over billions of years. Earth is also of a 'perfect' mass – less massive planets retain no atmospheres, while much more massive can becomes gaseous – while massive enough for plate tectonics, which reprocesses CO_2 to $CaCO_3$.

It has an atmosphere which resulted from outgassing and nebular gas capture, and possesses liquid water resulting from cometary impacts early in the Solar System formation, the resulting water oceans are being not too shallow, and not too deep. Relatively few major asteroids impacts have occurred in the past billion years.

Meanwhile our Moon (originating from an early giant impact) appears to stabilise Earth's spin axis, assisting climate stability. And Jupiter, in is outer orbit, acts as an attractor for potentially impacting asteroid. There have been no nearby/recent supernovae (which could wipe out life), while the metal content of our solar system, including supplies of rare-earth elements, is also auspicious. The list of favourable 'coincidences' goes on.

AND SO TO the Fermi paradox: if alien civilisations exist, as is presumably quite likely, where are they?

Existence probabilities are largely founded on arguments of scale, which run broadly, and in order of increasing uncertainty: (a) that the Sun is relatively young, and there are billions of stars in the Galaxy significantly older; (b) that some are likely to have Earth-like planets which, if Earth is typical, may develop intelligent life; (c) some of these are likely to develop interstellar travel; (d) under plausible assumptions, including a civilisation's average lifetime, they are likely to have explored or colonised the Galaxy on time scales of 1–100 Myr.

The paradox then arises as the contradiction between probability estimates of the existence of such civilisations, and the absence of evidence either for intelligent life elsewhere in the Galaxy (via detected 'signals'), or actual visits (e.g., of alien space probes).

ONE INTERPRETATION of the Fermi paradox is that alien civilisations simply do not exist. In this scenario, Earth is unique (within the Galaxy, or Universe) in harbouring intelligence. This may be because, even though life appeared early in Earth's geological history, it may nonetheless be a highly improbable occurrence.

Contrasting with the uniqueness view, many other explanations for the absence of 'visits' have been put forward. They can be grouped as physical (some fundamental difficulty makes space travel impossible); sociological (extraterrestrials have not visited Earth, either by choice, or due to the limited longevity of technological civilisations); temporal (advanced civilisations have arisen only recently, with insufficient time to reach us); or visits have occurred, but are not presently observed.

Today, resolution of the paradox remains unclear. It may be that intelligent life on Earth is unique in the Galaxy. Or it may be that the failure of searches to detect advanced civilisations is subject to so many interpretations as to be, at present, without significance.

Or it may signal bleak prospects for Earth in terms of impending natural or human-driven catastrophes.

101. The nearest black hole

GAIA IS ENABLING the discovery of two important classes of stellar-mass black holes in our Galaxy: *isolated* black holes, and those in binary systems which lack significant mass transfer. Amongst the latter, recently reported, is the nearest known black hole, at just 480 pc. I will start with a 'big picture' for background.

A BLACK HOLE is a region of spacetime where gravity is so strong that nothing, neither particles nor electromagnetic radiation (including light), can escape. Since the work of Karl Schwarzschild in 1916, black holes were known to be consistent with the theory of general relativity. But they were viewed more as a mathematical curiosity, at least until the discovery of neutron stars in 1967 stimulated interest in gravitationally collapsed compact objects as a possible astrophysical reality.

The first observational evidence for a black hole was Cygnus X–1, a high-mass X-ray binary discovered in 1964, and characterised (from the 1970s) as comprising a blue supergiant orbiting a $20M_\odot$ black hole at about 0.2 au. A stellar wind from the supergiant provides material which forms a hot accretion disk around the black hole, generating the observed X-rays. And the compact object is considered too small to be anything other than a black hole. Gaia DR3, incidentally, pinpoints Cygnus X–1 at 2250 ± 80 pc (0.4439 ± 0.0149 mas).

IN THE 50 years since, stellar-mass black holes have been inferred to exist in some 20 of the 300 or so (low-mass and high-mass) X-ray binaries now known in our Galaxy. With masses spanning $5-20M_\odot$ or more, they are thought to have formed by gravitational stellar collapse, often accompanied by a gamma-ray burst. The Hipparcos Catalogue, by the way, included just eight of the then-known high-mass X-ray binary systems.

Despite this small number of suspected stellar-mass black holes, there are held to be many millions throughout our Galaxy. But if they exist in wider-separation binaries, there will be little or no accreting matter to signal their presence through high-energy radiation. And those that are isolated, perhaps the majority, could only be discovered through gravitational lensing.

I SHOULD MENTION, for completeness, although they are not part of this story, that there are also 'supermassive' black holes, weighing in at *millions* of solar masses, which are believed to exist in the centres of most galaxies, including our own. They are believed to grow through the accretion of other stars or, over cosmological epochs, by mergers with other black holes; in February 2016, the LIGO–Virgo collaboration announced the first direct detection of gravitational waves, representing the first observation of a black hole merger.

RETURNING TO stellar-mass black holes, the mass distribution of the X-ray binary candidates is tantalising: it peaks around $7-8M_\odot$, with few if any below $5M_\odot$. This suggests that a 'mass gap' exists between the highest measured masses of neutron stars in binary radio pulsars, of around $2.1-2.3M_\odot$, and the lowest-mass black holes (Sahu et al., 2022).

The mass distribution falls off above $10M_\odot$, and few (X-ray binary) black hole masses are known above about $15M_\odot$. Indeed, the only known exception in our Galaxy is the latest mass determination, from VLBA as well as Gaia EDR3 astrometry, for the black hole in Cygnus X–1, of $21.1 \pm 2.2M_\odot$ (Miller-Jones et al., 2021).

More information on black hole masses is therefore of great interest. And there are two 'classes' of occurrences for which Gaia is contributing.

THE FIRST is through ongoing microlensing searches. This approach is important because of the evidence suggesting that a substantial fraction of stellar-mass black holes are single. And in consequence, being less affected by binary interactions, isolated black holes may provide a more direct probe of their formation.

Isolated black holes are, as I mentioned above, extremely difficult to detect directly, with microlensing offering the only current discovery method, and with no convincing candidates known pre-Gaia.

But 'normal' microlensing alone is insufficient: the well-known degeneracy between lens mass and lens–source relative velocity can only be resolved if accurate *astrometry* accompanies the photometry (essay #11).

In a 100-author paper by members of the OGLE, MOA, PLANET, μFUN, MiNDSTEp, and RoboNet microlensing teams, Sahu et al. (2022) observed the 2011 microlensing event MOA–2011–BLG–191 (also known as OGLE–2011–BLG–0462) with the Hubble Space Telescope at eight epochs over six years.

They measured an astrometric deflection of the background star's apparent position, which together with the parallactic signature of the Earth's motion on the microlensing light curve, led to a lens mass of $7.1 \pm 1.3 M_\odot$, and a distance of 1.58 ± 0.18 kpc. Gaia EDR3 astrometry enabled the measurement of the astrometric deflection by providing the proper motions of the background stars in the HST images during the 6-yr of observations to properly characterise the reference frame.

Finding that the lens itself emits no detectable light, and with a lens mass higher than that of a white dwarf or neutron star, they confirmed its black hole nature. Their mass measurement is, they emphasise, the first for an isolated stellar-mass black hole using any technique.

Their astrometry also provides an absolute proper motion for the black hole. Being offset from the mean motion of Galactic disk stars at similar distances by a transverse space velocity of 45 km s^{-1}, furthermore suggests that it received a 'natal kick' during the supernova explosion that led to its formation.

THE SECOND class for which Gaia is contributing, and somewhat more decisively, is in the case of non-accreting (or weakly accreting) black hole binaries. Some have recently been discovered in longer-period spectroscopic binaries, occupying the neutron star/black hole 'gap' with masses of $3 - 4.5 M_\odot$.

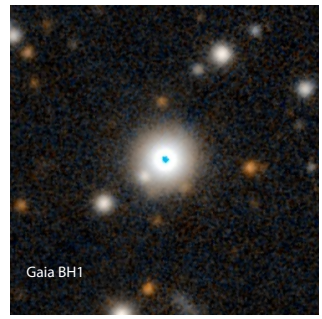
In parallel, the substantial nearby binary star population being discovered and characterised through Gaia astrometry has demonstrated the possibility of detecting many more wide binary systems containing quiescent black holes, with masses for some hundred or more in prospect (e.g. Chawla et al., 2022; Janssens et al., 2022).

Indeed, Gaia DR3 already includes orbital solutions for more than 100 000 single-lined spectroscopic binaries, more than a factor 10 in sample size compared with all previous studies, and for which accretor masses can probably be derived (El-Badry & Rix, 2022).

AMONGST THIS potential future haul, and using the binary catalogue included as part of Gaia Data Release 3, two papers have independently reported the discovery of the first such binary system, the $G = 13.8$ mag Gaia DR3 4373465352415301632, comprising a main-sequence sun-like star and a massive *non-interacting* black hole candidate (Chakrabarti et al., 2023; El-Badry et al., 2023).

Selected for more careful study on the basis of its high mass ratio, and its location close to the main sequence, follow-up spectroscopy and radial velocity monitoring of this system validated and refined the Gaia orbit solution.

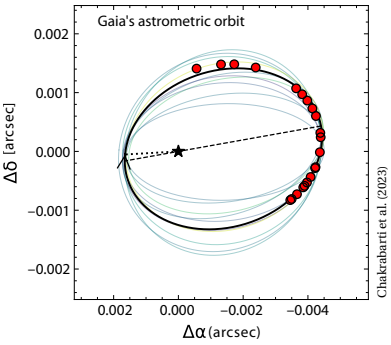
The spectral energy distribution of the visible star is well described by a single star model (a slowly-rotating G dwarf with $T_{\rm eff} = 5900$ K, $\log(g) = 4.5$, $M_1 = 0.9 M_\odot$). Importantly, no contribution from another luminous source is required to fit the observed photometry. Joint modeling of the radial velocities and astrometry constrains the companion (black hole) mass to $M_2 = 9.8 \pm 0.2 M_\odot$.



El-Badry et al. (2022)

Other noteworthy properties are that the main-sequence star is at most only moderately metal-poor, [Fe/H] = –0.30. And it is on a Galactic orbit similar to thin disk stars, suggesting that it formed in the Milky Way disk with at most a weak natal kick.

Both papers conclude that this binary system harbours a massive black hole on an eccentric ($e = 0.45 \pm 0.02$), long-period (185.4 ± 0.1 d) orbit, much longer than the typical black hole X-ray binaries discovered to date. And at a distance of only 480 pc, it is the nearest known black hole by a factor of 3. Its discovery hints at the existence of a sizeable population of 'dormant' black holes in binary star systems.



Chakrabarti et al. (2023)

El-Badry et al. (2023) argue that explaining the system's formation with standard binary evolutionary models is challenging. Specifically, it is not clear how the luminous star would survive a common envelope event under standard assumptions, and difficult for it to end up in a wide orbit afterwards. They conclude that formation models involving a triple system, or dynamical assembly in an open cluster, may be more plausible.

THE DISCOVERY OF black holes around luminous stars from their measured orbital accelerations now provides a new avenue for understanding the different possible formation channels of black holes in our Galaxy.

With estimates suggesting that there are perhaps a million similar systems in the Milky Way, future Gaia data releases may discover upwards of a hundred.

102. The heart of the Milky Way

A RECENT PAPER based on Gaia Data Release 3, entitled 'The Poor Old Heart of the Milky Way' (Rix et al., 2022), caught my attention as providing a fascinating and substantial advance in understanding the earliest phases of our Galaxy's formation.

OUR GALAXY comprises a flattened disk with a number of spiral arms, an extended massive halo, and a spherical bulge with an associated bar-like extension. The present ΛCDM (cold-dark matter) cosmology can explain an overall picture. But many details are still unclear: Are spiral arms transient or long-lasting? Why is the disk warped? What is the origin of its central bar?

Our knowledge of the spherical halo has advanced enormously over the past two decades (through a combination of astrometric, photometric, and spectroscopic surveys): for example, we know that tidal debris from the disrupted Sagittarius dwarf satellite dominates from 20–50 kpc, with stars mostly on highly inclined orbits. Between 5–25 kpc the Gaia–Sausage–Enceladus satellite debris dominates, with stars on highly eccentric orbits. And a growing number of distinct but less-prominent accreted components are also being identified.

MEANWHILE, details of the inner Galaxy have proven much more elusive, complicated by extinction and the large distances of the stellar populations.

Amongst recent advances, Gaia data have been used to clarify the properties of inner globular clusters as tracers of a very old Galaxy component, with 'Kraken' (Kruijssen et al., 2020) and 'Koala' (Forbes, 2020) having been identified as possible examples of early accretion.

Belokurov & Kravtsov (2022) used Gaia EDR3 data to reconstruct stellar kinematics in the inner Galaxy down to metallicities of −1.5 in [M/H], and identified an *in situ* proto-Galactic structure 'Aurora'. The constituent stars shows little net rotation, and may have preceded the Milky Way's ancient disk.

Conroy et al. (2022) used kinematics and ages from EDR3 and the Hectochelle H3 survey to identify proto-Galactic stars down to −2.5 in [M/H] with ages $\gtrsim$ 13 Gyr.

THE EARLIEST PHASES of our Galaxy's star-formation and enrichment history are reflected in the orbits and abundances of these old metal-poor stars. But to be more quantitative about how our Galaxy came into existence, we need to phrase our objectives more precisely. Here, I will borrow extensively from the introduction given by Rix et al. (2022).

In the context of the hierarchical formation of massive disk galaxies, the oldest and most metal-poor stars are expected to be a mix that either formed within one of the main overdensities which coalesced early on to form the proto-Galaxy (termed *in situ* formation), or formed early on in distinct satellite galaxies that eventually merged with the main body (accreted formation).

If the *in situ* stars formed in a more massive potential well than the accreted stars, the difference in origin should also be reflected in their abundance patterns, such as [α/Fe] vs [Fe/H]. However, at early epochs, this distinction may be blurred, with high-resolution simulations of the formation of Milky Way-like galaxies showing that fragments of comparable mass may rapidly coalesce early on (say, $z > 4$) in a sequence of major mergers.

ONE CHOICE IN terminology could then be to label only the most massive fragment as the *in situ* component, and to label all others as accreted – even though, together, the latter may constitute the majority of the total mass. But the same simulations show that, later on, it may not be possible to differentiate what was *in situ* and what was accreted at very early epochs.

Alternatively, one could label collectively all the major pieces that coalesced very early on, say at $z \geq 5$, as the proto-Galaxy, and then differentiate subsequent additions at slightly later epochs (say, $z < 3$) either as *in situ* (if the material was brought in as gas), or as accreted (if the stars formed in a distinct potential well that then subsequently merged).

Partly as a result of their findings, Rix et al. (2022) adopted the second terminology, referring to the oldest parts of the Milky Way as the proto-Galaxy, without trying to single out one of the contributing pieces as *in situ*.

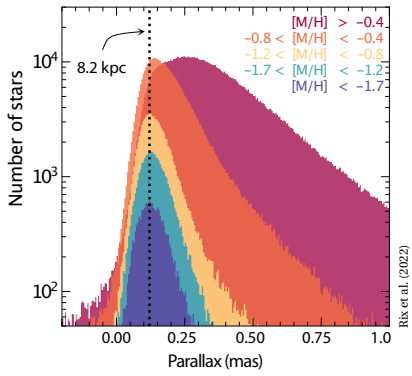
WITHIN THIS FRAMEWORK, Rix et al. (2022) set out to derive a more comprehensive picture of the dynamical properties of metal-poor stars in the inner Galaxy (within 5 kpc), drawing on *'... the immense wealth of new information that Gaia DR3 now affords'*.

They selected giant stars within 30° of the Galactic centre, derived [M/H] estimates from the BP/RP spectra, and combined them with Gaia's RVS velocities to obtain Galactic orbits for 1.5 million stars in the inner Galaxy.

Their primary goal was to see whether there is an extensive ancient and metal-poor stellar population at the 'heart' of the Milky Way. And to learn whether the bulk of these stars are a tightly bound proto-Galactic population, rather than stars accreted from more distant satellites (whose location in the inner Galaxy merely reflects the pericentre phase of orbits that take them out to the 'classical' halo at at distances beyond the solar radius).

THEY WERE ABLE to identify metal-poor samples of stars towards the inner Galaxy a factor 100 larger in size than previously available: specifically, more than 4000 stars with [M/H] < -2, some 20 000 with [M/H] < -1.5, and 70 000 with [M/H] < -1.

Their samples reveal a large metal-poor population ([M/H] < -1.5) in the inner few kpc of the Galaxy. Various lines of argument imply that this population is centrally concentrated: for example, the parallax distribution (shown here) implies that most stars are within 5 kpc of the Galactic centre. And their reconstructed orbits show that most have apocenters within 5 kpc.



These orbits also imply that most of this population could not have been identified in many of the previous surveys that focused on Galactic radii $\lesssim 5$ kpc. At the same time, their analysis shows that a minority (but nonetheless a large number) of stars with [M/H] < -1.2 currently in the inner Galaxy are on high-eccentricity orbits that can take them to > 10 kpc. These are presumably members of the accreted halo, simply found near the pericentre of their orbits.

Using the sample members with detailed abundances from SDSS (DR17), in particular [α/Fe], [Al/Fe] and [Mn/Fe], they explored which of these stars have abundance patterns attributable to a proto-Galactic (or *in situ*) origin, as opposed to an accreted origin. They found that almost all stars that remain tightly bound within the inner Galaxy, say with apocentres within the solar radius, have abundances that suggest a proto-Galactic origin. Stars on high-eccentricity orbits with with apocentres above 10 kpc show abundances typical of accreted stars; presumably pericentric members of the disrupted Gaia–Sausage–Enceladus.

By considering, for each mono-abundance population, the mean ratio of their angular momentum to the total action, they also quantified their net Galactic rotation, and the extent to which they resemble disk-like orbits. They found that only the most metal-poor members, [M/H] < -2, show no rotation, while populations with higher [M/H] also have higher rotation.

ALL OF THIS, they conclude, fits a picture in which this metal-poor heart of the Milky Way constitutes the most ancient 'proto-Galactic' component of our Galaxy. Their findings are worth setting out in detail (Rix et al., 2022, Section 4).

Much of their low-metallicity population is confined to Galactic radii within the solar radius, i.e. tightly bound at the bottom of the Galaxy's potential well, arguing against an origin through accretion from a once-distant satellite. And the vast majority of stars with [M/H] < -1.5 are distinctly α-enhanced, arguing for rapid enrichment in a deep potential well.

The total stellar mass seen at [M/H] < -1.5 within 5 kpc corresponds to $\sim 5 \times 10^7 M_\odot$. But obvious dust obscuration implies a large correction, in turn suggesting that the total stellar mass of this metal-poor heart of the Milky Way is much higher, $M_* \gtrsim 10^8 M_\odot$. Such a high mass of strongly α-enhanced stars at such low [M/H] is most likely to occur at the centre of a rather massive halo, rather than within a low-mass satellite system.

Their findings of a correlation between rotation and increasing metallicity is expected if star formation occurs – as time goes on – in gas that has increasingly settled into a disk, as shown in recent simulations of disk galaxy formation (Gurvich et al., 2023).

Finally, much of this central component has [M/H] values well below that of the centrally concentrated, α-enhanced, thick disk, whose oldest members (at around -1 in [M/H]) have ages of around 12.5 Gyr. If much of this metal-poor component predates chemically, and therefore temporally, the oldest α-enhanced disk, it must have formed before that, corresponding to $z \gtrsim 5$. The stellar population at the heart of the Milky Way should therefore be, almost exclusively, ancient.

I WILL END by quoting Rix et al. (2022), that their analysis *'... reflects the astounding information content of the Gaia DR3 data, particularly the B_P/R_P spectra'*.

103. Stellar rotation... for 3 million stars

ALL CELESTIAL BODIES rotate... galaxies, stars, and planets amongst them. The rotation of individual stars is inherited from the angular momentum of the gas and dust out of which they formed (albeit subsequently modified by various evolutionary processes), itself ultimately attributed to quantum density fluctuations in the early rapidly expanding Universe.

Since stars are fluid objects, they are neither perfect spheres, nor do they rotate as solid bodies. Instead, their rotation produces an equatorial bulge due to centrifugal forces, and generally involves differential rotation in which the angular velocity is a function of latitude. Differential rotation is believed to have a significant role in the generation of a star's magnetic field, which then interacts with the stellar wind, a process gradually slowing the star's rotation rate over its lifetime.

As a result of these and other processes, rotation is an important quantity in understanding many fundamental stellar properties, including the origin and evolution of a star's angular momentum, the role of rotation in the creation and persistence of stellar magnetic fields, magnetic braking, and in the mixing of chemical elements deep within their interiors.

Rotation also provides the basic ingredient for age determination based on gyrochronology, which exploits the star's gradual angular momentum loss over time.

THERE ARE TWO principal approaches by which stellar rotational velocities can be determined: either from its effect on the spectrum of the star, or by timing the motion of active features on the stellar surface.

The former exploits the fact that different regions of the star are moving at systematically different line-of-sight velocities, which leads to broadening of its absorption lines. From this broadening (which is further modified by effects such as microturbulence) the star's projected rotation velocity can be determined.

The latter exploits the rotational modulation of active surface features, in particular star spots and faculae. This can result in significant photometric modulation at the rotation period, recognisable if the star is monitored by accurate photometry over long periods of time.

TO PROVIDE some indicative numbers, our Sun rotates with a period of 25.05 days at the equator ($1.997\,\mathrm{km\,s^{-1}}$), and 34.4 days at its poles, and has an oblateness $(b/a-1) \sim 8\times10^{-6}$. Extremely rapid rotating stars include Regulus (α Leo A), with a projected rotational velocity of $317\,\mathrm{km\,s^{-1}}$, a rotation period of 15.9 h (86% of the break-up velocity), and an oblateness of 0.32. Other rapidly rotating stars include Vega and Achernar.

At the other extreme, KIC–11145123, with a rotation period of 100 d, has been claimed as the 'roundest' object in Nature, with an inferred difference of just 3 km between its polar and equatorial radii (Gizon et al., 2016).

THE STUDY of stellar rotation has been transformed within the past decade. The compilation of Glebocki et al. (2000), derived primarily from spectral line-broadening, contained just 11 000 stars of all spectral types and luminosity classes, albeit restricted to a few nearby open clusters and a few thousand bright stars.

The Kepler mission (2009–18) provided a substantial advance by characterising photometric modulation due to surface star spots and faculae. Amongst its 133 000 main-sequence targets, the combination of high photometric precision and densely-sampled long-duration observations resulted in some 34 000 rotational periods in the range 0.2–70 days (McQuillan et al., 2014). A further 30 000 were similarly identified from the extended Kepler K2 mission (Reinhold & Hekker, 2020).

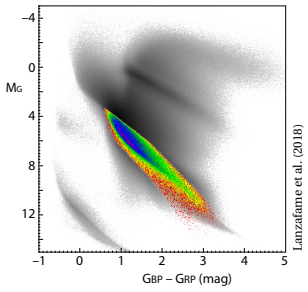
This vast bounty of rotational periods from Kepler has had a profound scientific impact, ranging from advances in the understanding of star spots and faculae, of differential rotation, of gyrochronology and the effects of magnetic braking, of the consequences for asteroseismology and, for planet-hosting stars, of the effects of tidal interactions and stellar obliquities.

Interpretation of the Kepler (and other) rotation data has also been substantially augmented by Gaia astrometry. Examples include quantifying the unexplained bimodal period distribution found for field M dwarfs (Davenport & Covey, 2018); and providing an updated view of stellar rotation as a function of mass and age in nearby open clusters (Godoy-Rivera et al., 2021).

In addition to its role in aiding interpretation of stellar rotation through its astrometric distances, Gaia is also determining new rotation periods. It is able to characterise rotation using these two distinct methods: rotational modulation (using its accurate photometry) and spectral line broadening (using its RVS spectrometer data). And, for both, doing so in even larger numbers.

Although Gaia's light curves are not as densely sampled in time as Kepler, the same principles of detection apply. Gaia's advantages include 3-colour time-series space photometry (G, G_{BP}, G_{RP}) covering the whole sky, and to much fainter magnitudes, $G \sim 21$.

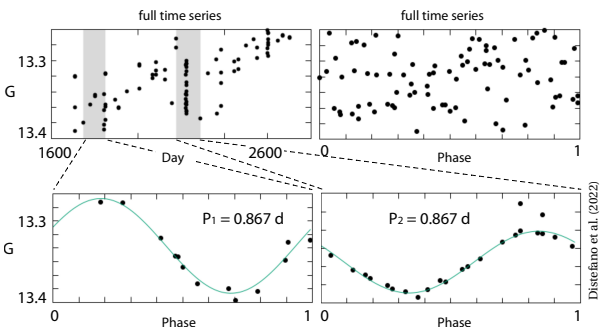
In their study of the first 22 months of mission data in Gaia DR2, Lanzafame et al. (2018) started with the 500 000 stars already classified as variable, and identified nearly 150 000 rotation periods (and modulation amplitudes) in stars where the flux modulation is induced by surface inhomogeneities (viz. star spots and faculae).



Candidate stars, which they term BY Dra variables, embrace dwarfs (with amplitudes of milli-mag), sub-giants, and T Tauri stars (with amplitudes up to several tenths of mag). They occupy the region of the M_G versus $G_{BP} - G_{RP}$ diagram shown in colour here. Again, this will open new possibilities for understanding the evolution of angular momentum and dynamo action.

An updated version of the same pipeline has been applied to solar-like variable stars in Gaia DR3 by Distefano et al. (2023). They started with 474 000 stars with variability attributed to magnetic activity. Of these, some 430 000 are newly discovered variables, and some 150 000 have rotation periods below 1 day, i.e. fast rotators poorly sampled by previous surveys.

Some 150 000 resulted in a 'well-defined' rotation period, meaning that the period was consistent over restricted data segments, even though the total time-series loses its overall coherence, as illustrated here for their object Gaia DR3 6503897945888972544.

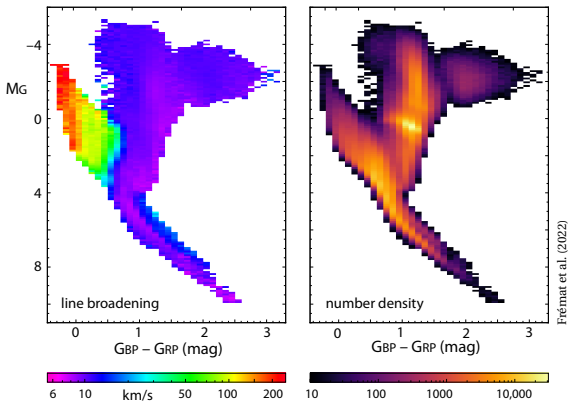


The numbers get even bigger with the results from the Radial Velocity Spectrometer (RVS) spectra, which cover the wavelength range 846–870 nm. As described in Essay #87, Gaia DR3 (June 2022), contains radial velocities for 33 million stars with effective temperatures between 3100–14 500 K, and $G_{RVS} < 12$ mag.

In brief, these radial velocities are based on a comparison of the observed RVS spectrum with a library of synthetic spectral models covering a range of temperature, gravity, metallicity and reddening (and 'broadened' to match Gaia's along-scan line-spread function, due to the contribution of its optics, the detector, and the satellite spin rate).

Various physical effects intrinsic to the star can also contribute to broadening of the spectral lines (related to quantum mechanics, particle interaction, and velocity fields on scales shorter than the photon mean free path). In most cases, their spectral impact is sufficiently well described by classical atmospheric modelling, and the spectral line shapes can usually be well replicated by keeping the effective temperature, surface gravity, metallicity, and microturbulence fixed.

Ignoring the origin of such second-order effects, the DR3 pipeline also determines an *additional* line-broadening contribution ('vbroad'), assumed due exclusively to axial rotation, to best fit the observed spectra.



Details of this additional line-broadening term are discussed by Frémat et al. (2023). The resulting numbers are spectacular: the DR3 line-broadening catalogue contains 3 524 677 stars with T_{eff} ranging from 3500–14 500 K, and $G_{RVS} < 12$. And they found reasonable agreement between the DR3 values and those in other catalogues, including GALAH, APOGEE, and LAMOST.

The DR3 line-broadening catalogue represents, by far, the largest stellar sample for the determination and interpretation of stellar rotation across the Hertzsprung–Russell diagram. Their Figure 16 (above) shows the median amplitude of the line-broadening term (left) and its corresponding number density (right) across the Gaia colour–magnitude diagram.

104. Light deflection. . . by Jupiter

IN GENERAL RELATIVITY, the presence of matter (energy density) distorts spacetime, and the path of electromagnetic radiation is deflected as a result.

Formally, the equations of light propagation are derived from the general relativistic Maxwell equations. For the solar system, they can also be derived in the limit of geometrical optics since wavelength-dependent relativistic effects are much smaller than 1 microarcsec.

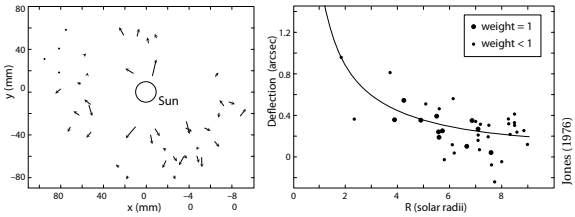
For a spherically symmetric gravitational field, the general expression for the deflection angle reduces to the 'post-Newtonian' formula $\delta\chi = \frac{2GM}{bc^2}\,\frac{(1+\gamma)}{2}\cot(\chi/2)$, where G is the gravitational constant, M is the mass of the deflecting body, b is the impact parameter, c is the velocity of light, and χ the angular separation between the deflecting body and the source.

Inclusion of the term γ, which appears in the so-called Parameterised Post-Newtonian (PPN) approximation (relevant for the slow-motion, weak-field limit), is convenient for comparing 'metric' theories of gravity and, in particular, assessing how well general relativity matches specific experimental tests. In general relativity $\gamma = 1$, while the classical derivation of light bending yields only the first part of the leading coefficient in this expression, viz. the factor 1/2. Conventionally, experimental results are expressed in terms of $\frac{1}{2}(1+\gamma)$.

For grazing incidence at the solar limb, the expression reduces to $\delta\chi \simeq \frac{1}{2}(1+\gamma)\,1.7505\,\mathrm{arcsec}$. In other words, with $\gamma = 1$, stars observed close to the solar limb are deflected by an angle of 1.7 arcsec.

LIGHT DEFLECTION has been observed, with various degrees of precision, on distance scales of 10^9 – 10^{21} m, and on mass scales from 10^{-3} – $10^{13}\,M_\odot$, the lower range from planetary deflection at 200 arcsec from Jupiter (Treuhaft & Lowe, 1991), and the upper range from the lensing of quasars (Dar, 1992).

Observational confirmation based on the 1919 solar eclipse in Brazil (Dyson et al., 1920) was of limited accuracy, while the most recent *ground-based* astrometric solar eclipse campaign, from Chinguetti (Mauritania) in 1973, yielded $\frac{1}{2}(1+\gamma) = 0.95 \pm 0.10$ (Jones, 1976).



Light deflection from the 1973 total solar eclipse at Chinguetti

THE DEVELOPMENT of radio interferometry and radio VLBI greatly improved the measurement of light deflection: a 1995 VLBI measurement of 3C 273 and 3C 279 yielded $\frac{1}{2}(1+\gamma) = 0.9996 \pm 0.0017$ (Lebach et al., 1995), and a later analysis of nearly 2 million VLBI observations of 541 radio sources from 87 VLBI sites yielded $\frac{1}{2}(1+\gamma) = 0.99992 \pm 0.00023$ (Shapiro et al., 2004). And Fomalont & Kopeikin (2003) reported VLBA–Effelsberg observations of 'light' bending of the quasar J0842+1835 at 3.7 arcmin from Jupiter.

THAT THE EFFECT of relativistic light bending had to be accounted for in the Hipparcos data analysis is evident from the magnitude of the effect. As noted above, according to general relativity, light bending at the solar limb reaches 1.75 arcsec. But given the Hipparcos measurement accuracies, the effect remains significant over almost the entire celestial sphere, amounting to about 4 milli-arcsec even at 90° from the Sun, i.e. for star light arriving orthogonal to the ecliptic.

The approach adopted for Hipparcos was to introduce the predicted light bending, at each measurement epoch, as fully deterministic as far as effects due to the Sun (and the Earth in the case of the NDAC consortium) were considered.

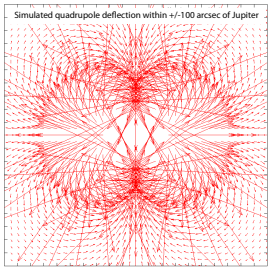
The measurements were then assessed for systematic departures from the prescriptions of general relativity, with the objective of assessing whether general relativity provides an adequate framework, or whether an alternative theory of gravity is indicated. The detailed analysis of the Hipparcos great-circle abscissae residuals gave $\frac{1}{2}(1+\gamma) = 0.9985 \pm 0.0015$ (Froeschlé et al., 1997).

ALTHOUGH THE Hipparcos accuracy was improved upon by VLBI, it is noteworthy that the former was derived from observations at large solar angles, i.e. not constrained to within a few solar radii, and with a single instrument. And while the same framework applies equally to Gaia, the geometrical model for Hipparcos (at the level of 1 milli-arcsec) required major refinements in the case of Gaia (at around 1 micro-arcsec).

The adopted model, formulated within the Barycentric Celestial Reference System (BCRS), actually comprises five successive transformations (Klioner, 2003). Indeed, parallax and proper motion are no longer effects which can be considered separately, or independently, of the chosen coordinate-dependent model.

A FURTHER complication when considering light bending, whether by the Sun, Earth, or the other planets, is that they are not spherical, but rather (due to rotation) better approximated as oblate spheroids.

Their resulting gravitational potential can be written as the expansion $U = -GM(C_0 r^{-1} + C_1 r^{-2} + C_2 r^{-3} \ldots)$, thus defining the monopole moment (C_0), the dipole moment (C_1), the quadrupole moment (C_2), and so on.



For the Sun and Jupiter, the monopole terms at grazing incidence are 1.7 arcsec and 16.27 milli-arcsec respectively. But while the quadrupole term at grazing incidence is only around 1 micro-arcsec for the Sun, it is a significant 240 micro-arcsec for Jupiter (Klioner, 2003; Ludl, 2011).

A final caveat is that the geometry of light deflection is similar to that of the parallax displacement, although differing in direction and in their dependence on χ: varying as $\cot(\chi/2)$ for light bending, and as $\pi \sin\chi$ for parallax. This allows the two distinct effects to be separated in the observation equations, although they remain highly correlated.

In summary, light deflection measurements to date have succeeded in confirming the monopole light deflection predicted by general relativity. The quadrupole light deflection term has not yet been measured.

THE POSSIBILITY of making these tests was highlighted in the ESA Gaia Phase A report (2000), with an expected precision of about 5×10^{-7} for γ, based on multiple observations of $\sim 10^7$ stars with $V < 13$ mag.

Gaia's scanning law means that, while $\chi_{\min} = 45°$ for the Sun, scans very close to grazing incidence are possible for Jupiter, albeit complicated by the effects of scattered light. Nonetheless, the observation of stars with very small impact parameters are required, because while light-bending due to the monopole moment falls off as r^{-2}, that of the dipole falls off as r^{-3}.

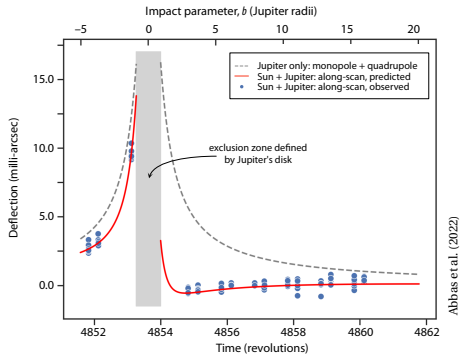
WHILE IT IS too early to have any insight into Gaia results for light-bending by the Sun, observations of stars close to Jupiter's limb have recently been reported using star positions from Gaia DR3 (Abbas et al., 2022).

But we should first go back to a time just after the launch of Gaia in 2013, when a study of the optimum precession and spin phases of the 'scanning law' was made by Klioner & Mignard (2014). They took into account the predicted configuration of three stars with $G < 15.75$ close to Jupiter's limb (within 6 Jupiter radii) which would take place between 22–26 February 2017.

These high-cadence observations were duly obtained, with the closest of the three observed at just 7 arcsec from Jupiter's limb. The closest bright star, with $G = 12.78$ mag, was observed on 25 almost consecutive transits spread over a few days.

The analysis, based on differential measurements, required the consideration of scattered light, differential aberration of several milli-arcsec along-scan, differential proper motions of order tens of micro-arcsec over 24 hours, and the non-linear differential light deflection, actually dominated by Jupiter's monopole.

The measured deflection, of around 10 milli-arcsec, represented the combined effect of Jupiter and the Sun (projected in the along-scan direction), and is mainly dominated by Jupiter's monopole. It was detected with a signal-to-noise of 50 at closest approach. It is the closest measurement made to Jupiter's limb in the optical, the highest signal-to-noise at any wavelength, and again provides excellent agreement between the Gaia observations and the predictions of general relativity.



Light deflection due to Jupiter, measured by Gaia

IT SETS THE STAGE for future efforts to disentangle an along-scan component of some 100 micro-arcsec at closest approach due to Jupiter's quadrupole moment from the much larger monopole deflection of around 10 milli-arcsec. And with Data Release 3 covering data only up to May 2017, two similar events, predicted to occur during September 2018 and April 2019, should be available in future data releases.