Scientific Background on the Nobel Prize in Physics 2019

PHYSICAL COSMOLOGY
AND
AN EXOPLANET ORBITING A SOLAR-TYPE STAR

The Nobel Committee for Physics
Physical Cosmology
and
an Exoplanet Orbiting a Solar-Type Star

“for contributions to our understanding of the evolution of the universe and Earth’s place in the cosmos”

with one half to

James Peebles

“for theoretical discoveries in physical cosmology”

and the other half jointly to

Michel Mayor and Didier Queloz

“for the discovery of an exoplanet orbiting a solar-type star”

Modern cosmology has revealed the history of the Universe and uncovered new unexpected components of matter and energy. In parallel, the Sun has been found to be far from the only star in our galaxy to host planets. The new findings show a wide diversity of planetary systems. As a result, our understanding of the Universe has changed in profound ways during the past few decades, and with that our view of our place in the Cosmos. This year’s Nobel Prize in Physics focuses on these ground-breaking discoveries.

Physical cosmology

Cosmology has developed into a science characterised by precision through evermore accurate measurements of temperature anisotropies in the Cosmic Microwave Background (CMB), along with studies of the expansion history of the Universe, as well as sky surveys providing detailed mapping of large-scale structures.

This exciting development has been possible thanks to ground-breaking discoveries in the theoretical framework of cosmology over the past half century. This year’s Nobel Laureate James Peebles has made seminal contributions in this area. Through detailed modelling, with the help of analytic as well as numerical methods, he has explored fundamental properties of our Universe and uncovered unexpected new physics. We now have a unified model capable of describing the Universe from its earliest fraction of a second up until the present and into the distant future.

Modern cosmology is based on Einstein’s theory of general relativity and assumes an early era, the Big Bang, when the Universe was extremely hot and dense. A little less than 400,000 years after the Big Bang, the temperature had decreased to about 3,000 K, enabling electrons to combine with nuclei into atoms. Because no charged particles were left that could easily interact with the photons, the Universe became transparent to light. This radiation is now visible as the CMB. Due to the cosmological redshift, its temperature is currently just 2.7 K — a factor of about 1,100 lower since the decoupling of matter and radiation. In figure 1, the source of the CMB can be seen as a screen that prevents us from easily looking back in time further than to a few hundred thousand years after the Big Bang.
One of the very first to propose that the Universe started with something like a Big Bang was the American horror writer Edgar Allan Poe in his prose poem *Eureka* [1]. As an explanation as to why the sky is dark at night, often referred to as Olber’s paradox after the German astronomer Heinrich Wilhelm Olbers, Poe suggested that the Universe had a beginning. In *Eureka*, he even proposes that it started out as a “primordial particle”, which then exploded.

The first to formulate a mathematical theory for the expanding Universe, using Einstein’s newly developed theory of general relativity, was the Russian mathematician and cosmologist Alexander Friedman [2] in 1922. He further developed his theory in 1924 [3]. These ideas were rediscovered in 1927 by the Belgian Catholic priest and astronomer Georges Lemaître [4], who later introduced the notion of a “primeval atom” [5]. He argued that the galaxies were receding from each other, and that this could be explained if the Universe expanded. In 1924, the Swedish astronomer Knut Lundmark [6] had made a similar observation, albeit with less rigor and accuracy. A more general acceptance that the Universe was in fact expanding came with the observations by the US astronomer Edwin Hubble in 1929 [7].

It is easy to derive the basic equations that describe the expansion of the Universe, the Friedman equations, even without the use of general relativity. To see this, let us for simplicity assume a homogenous universe. We pick an arbitrary point, at rest relative to matter, draw a sphere around it with radius $R$, and assume the sphere will grow as the universe expands. On the surface of the sphere, we introduce a small test mass with mass $m$. The total energy of the test mass is given by

$$E = \frac{m k^2}{2} - \frac{G M m}{R}.$$
where \( M = \frac{4\pi}{3} \rho R^3 \). A simple rearrangement gives

\[
\left( \frac{\dot{R}}{R} \right)^2 = \frac{8\pi G}{3} \rho - \frac{k c^2}{R^2},
\]

where \( k = -\frac{2\epsilon}{mc^2} \). Identifying \( H = \frac{\dot{R}}{R} \) as the Hubble constant, this becomes the first Friedman equation. By rescaling \( R \) one can set \( k = \pm 1, 0 \). To correctly interpret the meaning of \( k \), we need to appeal to general relativity, where it is identified as the spatial curvature. The value \( k = 0 \) corresponds to the critical density of a flat universe given by

\[
\rho_c = \frac{3H^2}{8\pi G}.
\]

Observations show that the total energy density of the Universe is very close to this value. Defining \( \Omega = \frac{\rho}{\rho_c} \), we have \( \Omega < 1 \) for a universe with negative curvature, \( \Omega = 1 \) for a flat universe, and \( \Omega > 1 \) for a universe with positive curvature.

There are several different components of energy in the Universe. Matter in the form of pressureless dust has an energy density that dilutes with volume, described by \( 1/R^3 \), while radiation disperses according to \( 1/R^4 \), due to the loss of energy caused by redshift. In the early Universe, radiation dominated the energy density of the Universe until a bit before recombination. Moreover, in the framework of general relativity, and to account for the possibility that the Universe could have been static, Einstein introduced an additional term in 1917 [8], corresponding to a constant energy contributing to \( \rho \), the cosmological constant, \( \Lambda \).

Multiplying the Friedman equation with \( R^2 \), to think of it as energy conservation, makes it easy to figure out what is actually happening. On the left of figure 2 we see the effective gravitational potential in the case of matter or radiation. When \( k > 0 \), the Universe reaches a maximum size and then re-contracts. If \( k < 0 \), it may keep on expanding forever.

**Figure 2:** The effective gravitational potential \( V \) as a function of the the scale factor \( R \), without a cosmological constant on the left and with a cosmological constant contributing to the energy density on the right. The red arrows show what happens in a few cases. The red dot represents Einstein’s static, and unstable, universe.
From observations, we know that the amount of ordinary baryonic matter in the form of nucleons, present in stars, clouds of gas, and the like, is no more than 5% of the total energy density at present. In addition, dark matter contributes 26% of the Universe’s critical density. It might as well be called invisible matter, as it neither emits nor absorbs light and is so far only known through its gravitational effects.

The most important component is the cosmological constant, representing a constant energy density unaffected by expansion. It is often called dark energy to account for the possibility that it might vary over time and space. That is, dark energy is not necessarily the constant introduced and thought to be related to the vacuum energy in quantum field theory. Observations show that dark energy contributes the remaining 69% of the critical density. As the other components of matter become diluted by the expansion, the dark energy will become evermore important with time (unless its energy density starts to decrease).

On the right in figure 2, a cosmological constant has been added to the potential with dramatic results. We note how the potential slopes downwards and gives rise to the accelerated expansion. The static universe of Einstein is marked as the unstable point at the maximum of the potential. That dark energy can force galaxies to accelerate away from each other may seem counterintuitive, but it is a direct consequence of the unusual properties of dark energy. The accelerated phase is indicated in the right-hand half of figure 1, where the number of galaxies thins out.

In figure 1, the hot Big Bang is the fire in the middle of the diagram, indicating that a preparatory phase, such as inflation, might exist before the hot Big Bang. Inflation is postulated as a scenario with a period of rapid acceleration, which would explain several properties of our Universe, such as its flatness.

The cosmic components we have discussed, together with the equations that determine how they interact and evolve, constitute the standard model of cosmology, sometimes called ΛCDM. The model is a triumph of physical cosmology – the systematic application of the physical laws to the evolution of the Universe. One of its most important originators is James Peebles. For his own review of the subject, see his textbook [9].

The birth of physical cosmology

In the late 1940s, Ralph Alpher, Robert Herman and George Gamow formulated a crude model of a hot Big Bang. An important motivation for their work was the need to explain the origin of the elements [16]. Evgeny Lifshitz [11] and Gamow [12] also made early attempts to understand the formation of galaxies. Gamow used the Jeans length, introduced by the British physicist James Jeans [13], which determines how large an object needs to be to collapse gravitationally. In 1948, Gamow argued that structures should not begin to form until the density of radiation was roughly equal to the density of matter, and found that this should happen at a temperature of a few thousand degrees. That same year [14], Alpher and Herman suggested that the present Universe should have a temperature of around 5 K. Not many physicists at the time thought the resulting radiation would be possible to observe. Rare exceptions were Andrei Doroshkevich and Igor Novikov [15].

The spring of 1965 was dramatic for cosmology. In a paper dated 13 May and published in The Astrophysical Journal Letters [16], Arnold Penzias and Robert Wilson described their discovery of what was to be identified as the cosmic background radiation, which was awarded the Nobel Prize in Physics 1978. The discovery was unexpected. Only through contact with a team consisting of Robert Dicke, Peebles, Peter Roll and David Wilkinson at Princeton University did Penzias and Wilson become aware of the cosmological explanation, described in a paper in the same issue, dated 7 May [17]. In his book The First Three Minutes [18], Steven Weinberg relates how Penzias and Wilson happened to learn about a young astrophysicist – Peebles – who, inspired by Dicke,
had predicted a thermal background radiation with a present temperature of about 10 K. The full story with its twists and turns is told in a book co-edited by Peebles in 2009 [19].

While Dicke's team [17] was not the first to suggest the existence of the CMB, their paper went further and discussed the reasons to expect a hot initial state of the Universe, which would explain the cosmic background radiation. The key is the connection between temperature and the density of matter, which determines how much helium is produced. What is important is the matter density when the temperature has dropped low enough to destroy the deuterium that is produced and prevent it from turning into helium. The denser the Universe, the more helium. These ideas were developed in detail by Peebles [20a,b] and followed up by other authors [21]. This approach to nucleosynthesis is quite different from the work of the preceding decades, when it was thought that heavier elements also could have been produced in the Big Bang.

As early as their landmark 1965 paper [17], based on the recently observed temperature of the Universe, the authors discussed a constraint on the amount of baryonic matter (i.e. matter consisting of nucleons that can participate in the formation of elements) in the Universe. This is one of the pillars of the Big Bang model. The authors also noted that the actual amount of matter suggested by astronomers' observations is far greater and that large amounts of exotic matter are necessary to satisfy this gap.

A key contribution is a paper by Peebles alone from the same year, 1965 [22]. He had already submitted it to The Astrophysical Journal on 8 March 1965, revised 1 June and published 15 November. The first sentence of the abstract states: "A critical factor in the formation of galaxies may be the presence of a black-body radiation content of the universe." This work, together with other contributions by the late Russian cosmologist Yakov Zeldovich [23], can be viewed as the starting point of physical cosmology, where the laws of physics are applied to the Universe at large. This is the moment when cosmology embarks on its way to become a science of precision and a tool to discover new physics.

Physical cosmology gets its freckles

The first researchers to predict anisotropies in the background radiation were Rainer Sachs and Arthur Wolfe [24]. Their idea was conceptually simple: variations in the depth of the gravitational potential affect the observed temperature of the CMB. First, an over-dense region cools the photons as they climb out of its gravitational potential, leading to a relative decrease in temperature given by

$$\frac{\delta T}{T} = \frac{\delta \Phi}{c^2}.$$

Second, time dilation at the last scattering surface where photons decouple, as we look back at an earlier and hotter Universe, contributes to an increase in temperature. This can be seen through

$$\frac{\delta t}{t} = \frac{\delta \Phi}{c^2}.$$

Because the Universe is dominated by matter at the time of recombination, the temperature follows

$$T \sim \frac{1}{a} \sim t^{-2/3},$$

and we find, adding the two contributions,

$$\frac{\delta T}{T} = \frac{\delta \Phi}{3c^2}.$$

Sachs and Wolfe did not have a theory for how the fluctuations arose, but their work inspired the development of observational techniques to find these variations, dubbed the SW effect.
The SW effect dictates the amount of anisotropies in the CMB at large scales. At small scales, the physics is more involved. Initial fluctuations in the density will lead to propagating acoustic waves in the hot plasma of coupled photons and baryons, which in turn will leave an imprint in the CMB. Andrei Sakharov [25] was one of the first to discuss the importance of acoustic waves, but only in a cold model without photons. Others who were thinking along these lines early on include Peebles and Zeldovich. Joseph Silk came up with an important general result in 1968 [26], when he realised that the amplitude of the anisotropies in the CMB are damped at small scales due to diffusion.

A breakthrough in the understanding of the acoustic waves, and the peaks they cause in the power spectrum of the CMB, came through the works of Rashid Sunyaev and Yakov Zeldovich [27], as well as of Peebles and Jer Yu [28]. Sunyaev and Zeldovich [27] explained the physics behind the acoustic peaks and their periodic nature. Peebles and Yu [28] had a different focus, using numerical methods to calculate and predict what actually could be measured. In their paper, they worked out power spectra of density fluctuations for different cosmological parameters. In particular, they presented the curve shown in figure 3, which is remarkably similar to the actual measurements by the Planck satellite in figure 4, obtained more than four decades later.

![Figure 3](image.png)

*Figure 3. Power spectrum for a flat universe according to Peebles and Yu [27], showing the acoustic peaks. The normalization is fixed to peak value unity.*

The scales characterising the acoustic peaks in the power spectrum tell us about the physics of the Universe. Together, the horizon scale and the somewhat smaller sound horizon, are of particular importance. The sound horizon tells us how far sound waves have had time to travel. It is determined by the speed of sound, which is given by $c_s = \frac{c}{\sqrt{3}}$ for a pure photon gas, and less than this for a baryon-photon fluid. The actual size of these scales, characterising the anisotropies that appear in the sky, is affected by the geometry of the Universe. Their angular size in the sky is surprisingly large. In an expanding universe, objects look smaller with distance only up to a point. If you look back far enough in time they will appear larger. What is important is neither the distance to the object when the light is received, nor the time it took for light to travel, but the
distance when the light was emitted. The sound horizon determines the position of the first acoustic peak. A scale of about 400,000 light years, viewed at distance of 40 million light years, has an angular size of a bit less than a degree. Currently, these regions are at a distance of more than 45 billion light years.

In order for these predictions to be possible, Peebles and Yu [28] had to make an assumption concerning the primordial power spectrum that initiates the sound waves. They argued that the most reasonable possibility is a spectrum without any characteristic scale. A scale invariant spectrum was also argued at the same time by Edward Harrison [29], who used quantum fluctuations of the metric in the early Universe. A couple of years later, Zeldovich [30] further explored the consequences of a scale invariant spectrum. Earlier, Peebles and Dicke [31] used a scale-invariant spectrum in the context of structure formation.

All of this was long before inflation was proposed [32-36]. According to inflation models, quantum fluctuations conjectured to occur during the inflationary era should leave their imprint by influencing when inflation ends (see [37] based on a model presented in [38]). The magnitude of these quantum fluctuations is supposed to be determined by the Hubble constant during inflation. Because the Hubble constant is expected to decrease as inflation proceeds, the spectrum is not expected to be exactly scale invariant. Measurements have confirmed that this is indeed the case.

Physical cosmology matures

Researchers have been aware of indications of unknown components of matter in the Universe for a long time. One of the first to propose such dark matter, Lundmark [39] began by studying stellar kinematics in galaxies. Based on his observations, he saw the need for considerable amounts of Dunkle Materie, the German phrase he coined for dark matter. A few years later, Fritz Zwicky [40, 41] drew the same conclusion while studying the motion of galaxies in the Coma galaxy cluster. More recently, optical galaxy rotation curves were obtained by Vera Rubin and Kent Ford [42] among others. An important contribution to the field was the observation by Jeremiah Ostriker and Peebles [43], who found that the galactic halo of our Milky Way must contain large amounts of dark matter in order for the flat galactic disk to remain stable. This finding was an inspiration for subsequent research.

Dark matter in the form of neutrinos or other weakly interacting particles help to structure the formation of matter, by allowing it to start to clump even before radiation has decoupled from baryonic matter. Hot dark matter in the form of light and fast-moving neutrinos causes structures first to form at very large scales. Unfortunately, this does not fit observations, which led physicists to explore other exotic possibilities, eventually classified as warm dark matter.

During the 1980s, a crisis developed in cosmology. Calculations based on an open universe, with a density less than the critical density, did not predict anisotropies compatible with observations. If the Universe had been open, the anisotropies would already have been discovered. Yet there were no sign of them. On the other hand, if the density of ordinary matter had been at the critical value, the galaxies we have observed could never have formed. In addition, in order for the amount of light elements to be correctly predicted by theory, the amount of ordinary matter that exists could not exceed that already found.

The ground-breaking work by Peebles on cold dark matter [44] is the first to consider non-relativistic, and thus cold, dark matter and its effect on structure formation. Through the introduction of non-relativistic cold dark matter, he was able to couple anisotropies in the CMB to large-scale structures in the Universe. In particular, in his 1982 paper [44], Peebles predicted a temperature anisotropy given by $\frac{\delta T}{T} = 5 \times 10^{-6}$, consistent with the actual measurements by the Cosmic Background Explorer (COBE) some years later. The theory was further developed by other researchers in the mid-1980s [45, 46].
In 1984, Peebles [47] took the next crucial step by reintroducing the forsaken cosmological constant – considered to be a superfluous term by most theorists for over half a century – arguing that it makes sense in the context of structure formation. Peebles was inspired by the contemporary theory of inflation and its prediction of a flat universe with a critical density. Because the matter density measured is much too small for the Universe to be flat, the cosmological constant might make up for the deficit. Other work briefly mentions a cosmological constant [48]. However, structure formation only works in combination with the cold dark matter introduced by Peebles in 1984 [47].

Another problem where the cosmological constant could come to the rescue is with the so-called age problem, where the estimated age of the oldest stars exceeded the age of the Universe. As illustrated in the right-hand side of figure 2, if the Universe is flat, and thus has to pass over the bump in the potential, it will slow down while doing so – a pale copy of Einstein’s static universe. As a consequence, the Universe would be older than if the dark energy was neglected [47-50].

All the components of the standard model of cosmology were in place by 1984, through the combination of Peebles’ two key papers [44, 47]. His breakthroughs came more than a decade before the conclusive measurements of the accelerated expansion of the Universe and five years before Weinberg’s argument based on the anthropic principle [51]. Given that there were no good reasons based on fundamental physics for why the cosmological constant should be small, Weinberg argued that all values are a priori equally likely. Using the existence of galaxies as a constraint, he concluded that the most likely value of the cosmological constant is comparable to or a bit larger than the contributions from other matter components. In the middle of the 1990s, the argument for a cosmological constant was strong [52]. In 1995, Jeremiah Ostriker and Paul Steinhardt introduced the notion of concordance cosmology [53], to summarise how well the different pieces of the puzzle fit together.

At the time, two more ground-breaking discoveries were made in observational cosmology. In 1992, the elusive anisotropies in the CMB were finally observed by COBE [54] (Nobel Prize in Physics 2006 to John Mather and George Smoot). In 1998, the accelerated expansion of the Universe was discovered using bright thermonuclear supernovae as distance indicators [55-56] (Nobel Prize in Physics 2011 to Saul Perlmutter, Brian Schmidt and Adam Riess).

Further observational evidence

Around the turn of the millennium, observational cosmology experienced remarkable breakthroughs. Ground-based as well as balloon-borne experiments, e.g., TOCO [57], BOOMERanG [58] and Maxima [59], achieved sufficient angular sensitivity to resolve the first acoustic peak in the CMB power spectrum, providing the first observational evidence for the flatness of the Universe.

While these experiments only probed small patches of the sky, the Wilkinson Microwave Anisotropy Probe (WMAP) [60] was launched in 2001 to study the CMB anisotropies over the entire sky. The nine-year-long mission revolutionised the accuracy of the measurements of the early Universe, putting the standard model of cosmology to strict tests [61]. Temperature intensity maps, as well as polarisation measurements, were used to accurately measure the fraction of baryons, dark matter and dark energy, as well as the overall geometry of the Universe. Furthermore, the data led to important bounds on the sum of the mass of the neutrino species and verified a key prediction from inflation, besides flatness, namely that the large-scale temperature fluctuations are slightly more intense than the ones at small scale.

The Planck satellite, which launched in 2009 and operated for 4.5 years [62], took observational cosmology to an even higher level of precision. Operating at nine frequencies, Planck could reach an angular resolution of just 10 arcminutes and a temperature resolution of one part in a million.
The satellite achieved unprecedented accuracy for all of the parameters of the standard model [63].

These accurate values are extracted from the power spectrum shown in figure 4. For example, the age of the Universe is now known with better than 1% accuracy to be 13.8 billion years. The density of the cosmic constituents was measured with a comparable significance, and when combined with supernova and large-scale structure observations, the margin for a potential time evolution of dark energy has been severely bounded, i.e., the observational evidence for a cosmological constant, $\Lambda$, is very strong [62]. Similarly, the statistical evidence for dark matter exceeds 100 standard deviations, a remarkable triumph for physical cosmology.

Figure 4. Anisotropies in the temperature of the CMB as measured by the Planck satellite. The acoustic peaks are clearly visible.

A summary of the physics of the acoustic peaks

All the key ingredients of modern cosmology become visible after a close look at the acoustic peaks in figure 4 – especially the first three peaks. As we have seen, the detailed structure of the peaks is determined by the physical content of the Universe. The angular size of the structures, and in particular the position of the first peak, is fixed by the geometry of the Universe. As illustrated in figure 5, the “spots” in the CMB will look larger if the Universe has positive curvature, analogous to a sphere, and smaller if the curvature is negative, like a saddle. The actual position shows that our Universe is very nearly flat with a density that is critical.
Figure 5. The angular size of spots in the CMB are determined by the geometry.

The first peak, as well as all the odd peaks, are caused by baryonic matter falling into gravitational wells. The even-numbered peaks correspond to decompressions as radiation pushes back. The more baryonic matter, the deeper the fall into the gravitational potential, and the more pronounced the first peak is relative to the second. The relative height between the first and the second peaks implies that the amount of baryonic matter is only 5% of the critical density.

The higher peaks correspond to more oscillations and probe earlier times when the radiation played a more important role. In particular, the third peak corresponds to a compression followed by a decompression, and then yet another compression of the photon-baryon fluid. Dark matter does not bounce back after the first compression because it is unaffected by radiation. It can therefore provide a gravitational well for the baryons to fall into the second time. This means that dark matter enhances the third peak. Its measured amplitude suggests that 26% of the Universe is composed of dark matter.

We can now make a simple calculation to determine the amount of dark energy. Working in units of the critical density, the first peak tells us that the Universe is flat and that the total sum needs to add up to one:

\[ \Omega_A = 1 - 0.05 - 0.26 = 0.69 \]

Hence, we find that 69% of the energy content of the Universe at present is in the form of dark energy, in agreement with direct measurements of how the Universe expands [55-56].

Outlook

In addition to its profound success in explaining the structure and evolution of the Universe, precision cosmology is also a tool to discover new physics. We still do not understand the physics of the cosmological constant. Perhaps its value is not constant, and perhaps a time-varying dark energy plays an important role in the evolution of the Universe. Peebles has already contemplated such a possibility [64]. The nature of dark matter is also not known. Favourite explanations include new particles, such as supersymmetric partners of the known ones or axions, which are hypothetical particles that could explain an important observation about the strong nuclear
forces. Until such a new particle is discovered, we cannot be sure that the current theoretical explanation of cold dark matter is the right one.

The way theory and observations now fit is astounding and the number of parameters are few. Still, there are observations that cannot be fully explained at the present time [65]. Measurements of the Hubble parameter in the late-time Universe do not quite match what is predicted from CMB physics. The explanation is currently unknown. Systematic errors in the measurements could potentially be responsible, or, perhaps new physics is still hiding somewhere out there.

Physical cosmology, with its interplay between observations and theory, is a tremendous success story that over the past half century has changed the way we view our Universe. Once, cosmology was a subject full of unfounded speculations and little data. It is now an exact mathematical science, where evermore accurate observations play a key role. The era of discovery is not over. As the measurements become more precise, new and unexpected phenomena are likely to be discovered. Physical cosmology will have more surprises in store, and Peebles is the one who has shown us the way to discover them.

References – Physical cosmology


[58] P. de Bernardis et al., *A flat Universe from high-resolution maps of the Cosmic Microwave Background radiation*, Nature **404**, 955 (2000)


An exoplanet orbiting a solar-type star

Since ancient times, humans have speculated whether there are worlds like our own, with points of views at the extremes expressed thousands of years ago [1]. In modern times, the possibility of observing planets orbiting stars other than the Sun was proposed more than 50 years ago [2], and builds on the measurement of stellar radial velocities. However, the formidable technical challenges remained a major obstacle for several decades after this idea was first proposed by Otto Struve, in 1952 [2].

Struve was unable to find compelling reasons why “hypothetical stellar planets” could not be much closer to their parent stars than is the case in the Solar System. We now know that there are no such reasons, and that our own Solar System may not be typical at all.

Several observational campaigns started in the early 1980s with the aim to observe stellar companions [3–8]. The use of words like “substellar companion” or “low-mass companion” in the titles of publications describing this new field of research reflects a certain scepticism at the time towards the search for exoplanets as a high-priority scientific objective.

The principle of measuring radial velocities by means of the Doppler effect is shown in figure 1. If the inclination angle $i$ is 0°, the plane of the orbit is parallel to the sky, “face-on”, which means that the observer on Earth sees the orbit face-on and no Doppler shift occurs. The other extreme is an “edge-on” observation ($i = 90°$), in which case the planetary mass can be determined directly from the Doppler shifts. In general, because the inclination angle is unknown, only $M_{\text{planet}} \times \sin (i)$ can be determined, setting a lower limit to the mass of the planet.

Somebody monitoring our Solar System from a distance would observe a radial velocity change of $\pm 13$ m/s of the Sun over 12 years, owing to the orbital motion of Jupiter around the Sun. This imposes severe challenges on any observational instrument, not least of which is to ensure that it is extremely stable, say $\leq 2$ m/s, over several years.

Different strategies were chosen in order to measure Doppler shifts. Gordon Walker and his group, including Bruce Campbell, at The University of British Columbia (UBC) in Vancouver, Canada, consulted the pre-eminent molecular spectroscopists in Canada (and the world) at the time: Gerhard Herzberg, the Nobel Prize in Chemistry 1971 Laureate, and his colleague Alexander Douglas [9]. They recommended the use of a cell of hydrogen fluoride (HF) gas as the source of a reference spectrum to be compared with the stellar spectra. As a reference, HF was an excellent choice, but less so from a practical point of view. The compound is toxic and highly corrosive. The UBC group conducted their search for substellar companions at the Canada-France-Hawaii Telescope (CFHT), a 3.6-m reflector. The HF absorption cell was inserted ahead of the slit of a coudé spectrograph, so that absorption lines from HF could be superimposed on a star’s light. This technique allowed radial velocity measurements with a precision of 13 m/s.

Geoffrey Marcy (University of California, Berkeley) and Paul Butler (then a PhD student at the University of Maryland) used a similar approach [10] as the UBC group, but with an absorption cell of molecular iodine (I$_2$) instead of HF. The researchers also consulted Gerhard Herzberg in this case. The spectrum of I$_2$ is routinely used by laser spectroscopists as a reference. Marcy and Butler made their observations with an echelle spectrograph at the 3-m reflector at Lick Observatory of the University of California, located at Mount Hamilton, east of San Jose.
Figure 1. The principle of measuring the radial velocity by means of the Doppler effect. The star and the orbiting planet move around their common centre of mass, causing Doppler shifts due to stellar wobble. Stellar absorption lines that arise when radiation from the interior passes the stellar atmosphere will be red- and blue-shifted depending on whether the star is moving away or towards Earth. These Doppler shifts give information about the planet’s orbital period around the star and also set a lower mass limit. (Reproduced from Las Cumbres Observatory, a worldwide network of telescopes.)
Michel Mayor at the University of Geneva and his collaborators had been studying stellar multiplicities at the Haute-Provence Observatory in the southeast of France, when they designed a new echelle spectrograph. In collaboration with André Baranne at the Marseille Observatory and colleagues from the Haute-Provence Observatory, they built the ELODIE spectrograph [11], an updated version of CORAVEL, which had been in use for more than a decade at the Haute-Provence Observatory. In order to survey more than just very bright stars, Mayor and collaborators chose a solution that did not include an absorption cell and a slit. Instead, they had an optical fibre-fed echelle spectrograph with the clear intention of avoiding the disadvantage with a cell, for which suitable objects are limited to bright stars in the vicinity of the Solar System. The intention with ELODIE was to expand the number of objects for which precision Doppler spectroscopy could be applied. Figure 2 shows the working principle of ELODIE.

![Figure 2. Schematic layout of the cross-correlation spectrograph ELODIE. The echelle spectrograph was fed by optical fibres, one from the starlight and the other from a thorium-argon hollow cathode calibration lamp. The echelle grating had a blaze angle of 76°, giving a resolving power of 42,000 by working in high diffraction orders with a relatively low groove density. After order separation, the spectra were recorded by a 1024×1024-pixel CCD camera over a wavelength range from about 390 to 680 nm. The exposure time for an individual star was 30 minutes, and data could be reduced while working online at the observatory. ELODIE had a radial velocity accuracy of 13 m/s and was designed to allow a large number of stars to be included in the observational campaign at the Haute-Provence Observatory. The absorption spectrum is simplified here; in practice, about 5,000 absorption lines were used. (Redrawn from fig. 1, ref. [11].) The situation in the beginning of 1995 did not look very promising. A decade and a half of searching the skies had turned up nothing. Only one earlier report, by A. Wolszczan and D.A. Frail, of planets orbiting a pulsar showed promise, but only because the pulsar made the planets easier to detect [12]. The millisecond radio pulsar PSR1257+12 provided a “built-in” timing system...](image-url)
that was used to conclude that at least two Earth-sized bodies were orbiting the central body. However, this technique could not be used for solar-type stars. The observation was made in the microwave region at the 305-m Arecibo Observatory radio telescope in Puerto Rico. In contrast, researchers had been using the optical region to search for exoplanets orbiting solar-type stars. The planets orbiting the pulsar may have been the result of a supernova explosion in connection with the formation of the rapidly rotating neutron star (pulsar) PSR1257+12 and therefore are not representative of solar-type planetary formation. And in fact, we now know that planet formation around pulsars is probably rare, as only a few of the more than 2,000 known pulsars have a planetary system.

Gordon Walker and collaborators, including Stephenson Yang, a co-author of their 1988 paper [3], reviewed the situation in the August 1995 issue of the journal Icarus [13]. Not only did they review the 21 bright, solar-type stars they had studied over the past 12 years, but also other searches for Jupiter-mass companions. They concluded that no planets of Jupiter-mass or larger had been detected orbiting solar-type stars. The last sentence in the abstract states, “This absence presents an interesting challenge to theories of planet formation.”

It is the irony of destiny that the breakthrough paper [14] authored by this year’s Laureates, Mayor and Didier Queloz, was received by Nature on 29 August – the same month that Walker and his colleagues published their review in Icarus. Mayor and Queloz reported their discovery at the ninth Cambridge Workshop of Cool Stars, Stellar Systems and the Sun in Florence on 6 October, and their paper was accepted for publication on 31 October and published on 23 November. Their transformational discovery forever changed our conception of humankind’s place in the Universe.

**The discovery**

The ELODIE echelle spectrograph allowed Mayor and Queloz to plan an observing programme that included 142 stars, many more than had been possible in earlier campaigns by other groups. As early as the fall of 1994, they found that the radial velocity of the star 51 Pegasi in the constellation Pegasus had a periodic variation of just about four days.

This was surprising because based on the only data point available at that time – our own Solar System – a Jupiter-mass companion ought to have a much longer period. A period of only four days would put the Jupiter-mass companion to 51 Pegasi at a distance of only 0.05 astronomical units (AU), one-hundredth of the distance between Jupiter and the Sun.

On the other hand, the short period gave Mayor and Queloz opportunities to study several full cycles. Another advantage with the very short period was that it could be checked very quickly by other radial velocity groups. In a note added in revision of the breakthrough paper [14], the Laureates thanked “a team working at the Lick Observatory, as well as by a joint team from the High Altitude Observatory and the Harvard-Smithsonian Center for Astrophysics” for having confirmed the discovery, naming Marcy, Butler, R. Noyes, T. Kennelly and T. Brown (see also [15]). Figure 3 shows the orbital motion of 51 Pegasi from the breakthrough paper [14]. The half-amplitude of the velocity variation was measured to be 59 m/s, more than a factor of four larger than the ELODIE precision of 13 m/s.

The following year, Marcy and Butler published the discoveries of two Jupiter-mass planets orbiting 70 Virginis [16] and 47 Ursae Majoris [17], respectively. A full account of the confirmation of the Jupiter-mass companion 51 Pegasi b was published shortly after [18].

The orbital period of 51 Pegasi b was determined to be 4.23 days and nearly circular, and the mass $0.47 \times M_J / \sin (i)$, where $M_J$ is the mass of Jupiter [14]. The surface temperature was estimated to be 1,300 K, compared to Jupiter’s 130 K.
The discovery of the first exoplanet orbiting a solar-type star [14] was initially met with some reservations. It was well known that stellar pulsation and star spots in combination with rotation potentially could lead to false positives. The extremely short orbital period for a Jupiter-mass planet was also difficult to reconcile with the structure of our own Solar System. However, Mayor and Queloz convincingly argued against such stellar effects in their breakthrough paper [14], and the fast verification by other groups also strengthened their case. Other researchers soon realised that 51 Pegasi b could not possibly have been formed at 0.05 AU, but rather at a much larger distance from the host star, say 5 AU, and that migration had moved it into close vicinity to the host star [19]. Migration had been theoretically predicted [20–23] as a result of the interaction of the protoplanetary disk and the planet, so observations that supported this migration were not a total surprise.

Five years after the discovery, when the first review “post–51 Pegasi” appeared [1], 34 exoplanets had been discovered orbiting Sun-like stars – and all scepticism was long since gone.

**Exoplanets – a new and vibrant field of astrophysics**

The discovery of a Jupiter-mass companion to 51 Pegasi was the starting point of a new field of astrophysics that virtually exploded after 1995 – the study of exoplanets and planet formation. Thus, in terms of the impact on the astronomical community and emergence of new programmes, the discovery by Mayor and Queloz can be compared to the discovery of the CMB by Arno Penzias and Robert Wilson in 1965 and awarded with the Nobel Prize in Physics 1978.

Whereas the radial velocity Doppler spectroscopy method dominated the first five years of exoplanet research, other methods were soon developed. When a planet transits its host star as seen from the Earth, some of the starlight is blocked, resulting in a decrease in the measured
photon flux. This is the principle for the observation of transiting planets, and the first such observations were reported in early 2000 [24, 25]. The transit method was first exploited in space by the Convection, Rotation and planetary Transits (CoRoT) satellite [26], launched by ESA in 2006, and came into its prime when NASA launched the Kepler satellite in 2009 [27]. Given the superior stability these two satellites could offer, which is important for the transit method, the photometric database of observed exoplanets expanded into the thousands during the nine-year-lifetime of Kepler.

The observation of a Jupiter-mass companion to a solar-type star, but with an extremely short orbital period, challenged the common view of how planets are formed and brought into focus earlier migration predictions [20–23]. Figure 4 shows the distribution of known exoplanets in terms of their mass, radius and orbital period, together with the Solar System planets.

![Figure 4](image)

**Figure 4.** The distribution of known exoplanets as a function of their orbital periods and mass (left panel) and radius (right panel). In addition to the radial velocity and transit methods, which have been used to discover the vast majority of exoplanets, imaging and microlensing also have been used. Most of the exoplanets discovered via the radial velocity method do not transit, and hence only their masses are known, and not their radii. The opposite holds for transiting planets. In some cases, an exoplanet can be studied by both methods, in which case both the radius and mass can be determined. Exoplanets in the upper left corners are denoted “hot Jupiters”, to the right of those are the “warm Jupiters”, and below are the “super Earths”. (Reproduced from figure 3.1 in Exoplanet Science Strategy, National Academies Press, 2018.)

The census of exoplanets shown in figure 4 is a striking demonstration of the diversity of planetary systems. It also shows that the Solar System is the exception rather than the rule, with the caveat, of course, that the exoplanet detection methods have an observational bias favouring planets close to the host star. What is clear is that planetary systems are ubiquitous, and the challenge now is to explain the diversity of planetary systems rather than how the Solar System was formed [28], and to predict which fraction of planetary systems are similar to the Solar System.

The understanding of the physical processes leading to planet formation have advanced significantly from the classic 1969 reference [29] during the past two decades thanks to the discovery of exoplanets, while at the same time the complexity of the problem has increased.
Planets are born out of the circumstellar disk of gas (primarily hydrogen and helium) and dust grains (amorphous silicate, carbon compounds and ice) swirling around a newborn star (see figure 5). Dust particles clump together by electrostatic interaction to form bigger aggregates. This formation process is critically dependent on the collision energies involved, which depend on the turbulence in the disk (which is not well known) and the radial migration towards the central star owing to interaction with the gas. When larger pebbles are formed, further growth occurs because of gravitation. This leads to the formation of planetesimals that range in size from a few hundred meters to 100 km in diameter. Collisions of planetesimals may lead to their destruction, but also to the formation of larger bodies. When the planetesimals become larger, pebble accretion becomes the dominant growth mechanism [30]. This leads to the formation of protoplanets and eventually planets. Many details in this chain of events remain poorly understood, as discussed by Morbidelli and Raymond [28].

Figure 5. Artist’s view of a young star surrounded by a protoplanetary disk composed of gas (mainly hydrogen and helium) and dust. Planets are formed in the protoplanetary disk in two steps. In the first step planetesimals are formed by dust particles sticking to each other. In the second step the largest planetesimals grow by pebble accretion to form protoplanets. (Credit ESO/Luís Calçada.)

Present and future

More recently, the NASA satellite mission Transiting Exoplanet Survey Satellite (TESS) was launched on 18 April 2018. During its two years in orbit it will survey 85% of the sky, corresponding to an area 400 times larger than that covered by the Kepler satellite. In particular, TESS will look for exoplanets orbiting stars near our Solar System. These planets can then be further characterised by means of ground-based observations.

Today more than 4,000 exoplanets in about 3,000 planetary systems have been confirmed. Transiting exoplanets are particularly suitable for detecting atmospheres, and the first fingerprint
in the form of the sodium resonance doublet at 589.3 nm had been observed already in 2001 [31].
Since then, molecules in the gas phase, such as carbon dioxide and water, also have been observed.
Atmospheres have been observed primarily for gas-giant planets, but very recently, researchers
have successfully detected water on smaller non-gaseous planets [32].

The recent observations of Earth-like planets with a planetary surface that in principle can
support liquid water, located in what is denoted “the habitable zone” of their host stars [32–35],
raise the question of whether life can exist on these planets. While presently undetected, future
satellite missions such as the CHaracterizing ExOPlanets Satellite (CHEOPS), the James Webb
Space Telescope (JWST) and Planetary Transits and Oscillations of stars (PLATO), as well as
ground-based telescopes such as the Extremely Large Telescope (ELT), are well equipped to
search for possible signs of life in the atmospheres of Earth-like planets [36], such as ozone and
methane.

As a final note, recent research has identified the potential for using exoplanet atmospheres for
studying different climate systems [37]. Just as with the physics of planet formation, the diversity
of exoplanets opens up new territory in the research of the dynamics of different kinds of
atmospheres and other aspects of climate. New observational techniques that are rapidly being
developed now will expand the parameter spaces for which theories can be tested. In the long run,
this new area of research will give us a better understanding of the terrestrial atmosphere.

References – An exoplanet orbiting a solar-type star


[2] O. Struve, Proposal for a project of high-precision stellar radial velocity work, Observatory
72, 199 (1952)


341, 961 (1989)

[5] G.W. Marcy and K.J. Benitz, A search for substellar companions to low-mass stars,

Suppl. Ser. 88, 281 (1991)


[8] A.A. Tokovinin, The frequency of low-mass companions to K and M stars in the solar

Astron. Rev. 56, 9 (2012)


