The search for worlds like our own

Journal Article

How to cite:

Fridlund, Malcolm; Eiroa, Carlos; Henning, Thomas; Herbst, Tom; Lammer, Helmut; Léger, Alain; Liseau, René; Paresce, Francesco; Penny, Alan; Quirrenbach, Andreas; Röttgering, Huub; Selsis, Franck; White, Glenn J.; Absil, Olivier; Defrère, Denis; Hanot, C.; Stam, Daphne; Schneider, Jean; Tinetti, Giovanna; Karlsson, Anders; Gondoin, Phillippe; den Hartog, Roland; D’Arcio, Luigi; Stankov, Anna-Maria; Kilter, Mikael; Erd, Christian; Beichman, Charles; Coulter, Daniel; Danchi, William; Devirian, Michael; Johnston, Kenneth J.; Lawson, Peter; Lay, Oliver P.; Lunine, Jonathan and Kaltenegger, Lisa (2010). The search for worlds like our own. Astrobiology, 10(1) pp. 5–17.

For guidance on citations see [FAQs].

© 2010 Mary Ann Liebert, Inc.
Version: Version of Record
Link(s) to article on publisher's website:
http://dx.doi.org/doi:10.1089/ast.2009.0380

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online’s data policy on reuse of materials please consult the policies page.
The Search for Worlds Like Our Own


Abstract

The direct detection of Earth-like exoplanets orbiting nearby stars and the characterization of such planets—particularly, their evolution, their atmospheres, and their ability to host life—constitute a significant problem. The quest for other worlds as abodes of life has been one of mankind’s great questions for several millennia. For instance, as stated by Epicurus ~ 300 BC: “Other worlds, with plants and other living things, some of them similar and some of them different from ours, must exist.” Demokritos from Abdera (460–370 BC), the man who invented the concept of indivisible small parts—atoms—also held the belief that other worlds exist around the stars and that some of these worlds may be inhabited by life-forms. The idea of the plurality of worlds and of life on them has since been held by scientists like Johannes Kepler and William Herschel, among many others. Here, one must also mention Giordano Bruno. Born in 1548, Bruno studied in France and came into contact with the teachings of Nicolas Copernicus. He wrote the book De l’Infinito, Universo e Mondi in 1584, in which he claimed that the Universe was infinite, that it contained an infinite amount of worlds like Earth, and that these worlds were inhabited by intelligent beings. At the time, this was extremely controversial, and eventually Bruno was arrested by the church and burned at the stake in Rome in 1600, as a heretic, for promoting this and other equally confrontational issues (though it is unclear exactly which idea was the one that ultimately brought him to his end).

In all the aforementioned cases, the opinions and results were arrived at through reasoning—not by experiment. We have only recently acquired the technological capability to observe planets orbiting stars other than
our Sun; acquisition of this capability has been a remarkable feat of our time. We show in this introduction to the Habitability Primer that mankind is at the dawning of an age when, by way of the scientific method and 21st-century technology, we will be able to answer this fascinating controversial issue that has persisted for at least 2500 years. Key Words: Terrestrial exoplanets—Habitability—Planet-detection methods—Bioastronomy. Astrobiology 10, 5–17.

1. The Quest for Other Habitable Worlds

This so-called “great question” can also be formulated as a clearly defined science goal. Since the origin of life most likely requires a stable supply of energy, planets that could host life are likely to orbit within what is known as the habitable zone (HZ), a region relatively close to the parent star. The HZ is defined such that liquid water is likely present on a particular planet in that region, given that life is thought to depend exclusively on this substance to exist. Further, bodies that host life, as we currently understand it, will likely be rocky worlds that are somewhat similar to Earth, Venus, or Mars. Or, in principle, they could resemble icy moons akin to those we find in the Jupiter system. Answers as to how large or how small such a world can be and still host life will very likely require empirical data.

Acquisition of such data, along with studies that address the prevalence of terrestrial planets, as well as their properties and ability to host life as we know it, now constitute high-priority objectives (themes) in the long-term science plan (Cosmic Vision 2015–2025) of ESA, as well as in similar plans under development at NASA.

An underlying theme within the Cosmic Vision plan is the need to place our own Solar System into the context of the rest of the Universe. To accomplish this, the following fundamental questions must be addressed: Is our Solar System a rare occurrence or even unique? Does life arise on suitable planets almost automatically? Is the spontaneous formation of life something that occurs rarely on a galactic scale?

Serious discussion as to the design and development of instruments capable of acquiring the required data to answer such looming questions has only occurred over the course of the last 25 years.

In 2007, the first space mission dedicated to the search for “rocky” or terrestrial planets was launched by the French space agency, Centre National d’Études Spatiales (CNES), with the active participation of ESA and the space agencies of Austria, Belgium, Brazil, Germany, and Spain. CNES reported the first actual terrestrial planet in the beginning of 2009. While this object is larger than Earth—it’s radius is 1.68 times that of Earth—its discovery signaled the beginning of the search for planets like our own. It is very likely that such objects will be found in the near future, either from space, from the ground, or very likely from a combination of the two.

The European Space Agency and NASA are currently working on development of space missions capable of meeting the challenges involved with the study of such terrestrial planets. The mission concepts that, to date, are most seriously under consideration are Darwin (ESA; Léger et al., 1996a, 1996b; Fridlund, 2000; Cockell et al., 2009) and the Terrestrial Planet Finder (TPF) (NASA; Beichman, 2000). Currently under study by ESA is a mission called PLATO, which follows in the footsteps of the highly successful CoRoT and Kepler missions (successfully launched in the spring of 2009) (see Fig. 1).

As a consequence of three recent ESA planning exercises (Horizon 2000, Horizon 2000+, and Cosmic Vision 2015–2025), the agency has undertaken several Darwin-related studies since 1996. While these studies have focused on system aspects and enabling technology (including the development of hardware such as breadboards and technology demonstrators, which has resulted in hundreds of articles in the technical literature and dozens of doctoral theses in Europe alone), the investigation of the scientific issues has been central to the entire endeavor. How one goes about translating the phrase “to put Earth and life on it into context” is not a trivial issue; therefore, ESA has appointed a number of science teams to address this. The largest and most recent such team is the Terrestrial Exoplanet Science Advisory Team or TE-SAT. Since 2002, the eleven members of TE-SAT and associates have worked under the following terms of reference, to

- Define mission science goals compatible with the ESA Cosmic Vision program;
- Consider and identify design concepts, instrumentation, and implementation plans that satisfy the scientific priorities and identify necessary technological developments;
- Develop quantifiable science drivers for the major components of the system, including software and operations;
- Assess the impact of the science drivers on design parameters (orbit, mission duration, optics size and layout, detectors, back-ends, etc);
- Monitor the progress of the ongoing technological studies and developments;
- Consider relationship and overlap with other planet-finding missions with emphasis on NASA’s TPF mission;
- Make recommendations on readiness for the next study phase;
- Prepare material and reports for the review process by ESA’s advisory bodies;
- Explain the science goals to the larger community and the general public.

The review-type articles reported in this journal issue are the updated results of TE-SAT’s deliberations. The group held 11 meetings between January 1, 2003, and December 31, 2006. Officially, TE-SAT has also worked in conjunction with the TPF Science Advisory Group (TPF-SAG), in that two scientists on TE-SAT have been members of TPF-SAG, and vice versa. Further, the scientist in charge of the Darwin studies and the Darwin study manager have also been part of TPF-SAG, and the scientist and manager in charge of the NASA studies have participated in all TE-SAT meetings. This collaboration was carried out under the auspices of a Letter of Agreement between ESA and NASA, which was signed in August 2002 and remained valid until December
31, 2006. This collaboration focused primarily on an early determination of scientific parameters that uniquely defined the two different mission concepts, with the intention of providing input for trade-offs between concepts.

2. Problems Related to the Direct Detection of Earth-Type Exoplanets

The direct detection of a planet the size of Earth, orbiting its parent star in the HZ, constitutes a challenging problem, given that the signal detected from such a planet would be between about \(10^{10-11}\) (visual wavelength range) and \(10^{6-7}\) (mid-IR spectral range) times fainter than the signal received from the nearby star. Selecting the appropriate spectral region with which to attempt detection is governed by this contrast problem, as is selection of a region with which the characterization of the planet and its habitability is optimized. This has recently been addressed by a number of researchers, for example, Selsis (2002) and Traub (2003).

A fundamental aspect of the problem of directly detecting a planet, whose light output is faint compared to that of a strong parental stellar flux, is the huge contrast ratio. Several techniques of detection have been suggested, most notably the use of (1) giant (100 m class) ground-based telescopes that...
operate in the visual or near-IR, (2) large (10 m class) space-deployable coronographic telescopes that operate in the visual, and (3) interferometers—either on the ground or deployed in space—that operate in the visual or mid-IR. Selection of the right technology is a complex issue but includes elements such as compatibility with the scientific case, difficulty of implementation, and time schedules and costs. ESA has selected an interferometer that operates in the mid-IR and utilizes free-flying telescopes (i.e., no connecting structures) for detailed study and possible implementation. This is commonly referred to as the Darwin mission study (Fridlund, 2000), which uses the new technology of nulling (or destructive on-axis) interferometry (Bracewell, 1978; Bracewell and MacPhie, 1979). For the Darwin study, the spectral region between 6 and 20 μm was chosen. This is a region that contains (among others) the spectral signatures of CO₂, H₂O, CH₄, and the O₃ molecules found in the terrestrial atmosphere. The presence or absence of these spectral features would indicate similarities or differences relative to the atmospheres of Solar System planets such as Venus, Earth, and Mars. The same reasoning has been applied to NASA’s Terrestrial Planet Finder Interferometer (TPF-I). These two studies have been carried out in parallel for the last 10 years and have recently reached a consensus on the more detailed design, such as configuration, sizes, and detailed technological issues. On the other hand, the parallel NASA Terrestrial Planet Finder Coronagraph study is considering the visual–near-IR wavelength region, which requires an extremely large single aperture telescope and a complex occulting mechanism that would have the capability to observe different spectral signatures. It has become clear, however, that a complete study of a terrestrial exoplanet at interstellar distances would ultimately require complete spectral coverage. A step-wise approach to the problem would require development and deployment of the technologies sequentially. A key problem then would be which system is to be developed and launched first. This would primarily be decided in conjunction with issues such as scientific completeness, cost, complexity, and readiness.

The basic concept of “nulling” (or “destructive”) interferometry is to sample the incoming wave front from the star and its planet(s) with several (>2) telescopes that individually do not resolve the system. By applying appropriate phase shifts between different telescopes in this interferometric array, destructive interference is achieved on the optical axis of the system in the combined beam. At the same time, constructive interference is realized a short distance away from the optical axis. Through the right choice of configurations and distances between telescopes, it is possible to place areas of constructive interference on regions representative of the HZ of an exoplanetary system, which would achieve the required contrast to discriminate between the planet and the host star. The first practical demonstration of nulling on the ground was undertaken in February 1998 (Hinz et al., 1998). Using the Multiple Mirror telescope on Mount Hopkins, Arizona, these authors were able to cancel out the image of a star: α Orionis. The ability of the interferometer to suppress the entire Airy pattern was demonstrated in that performance. In this case, the nulled image had a peak in intensity of 4.0% and a total integrated flux of 6.0% of the constructive image. When compared to the 10⁻⁵⁻⁶ contrast required to carry out the science goals of Darwin, this is of course only the beginning. Nevertheless, more-recent experiments in the laboratory have demonstrated contrasts of about 10⁶ in broadband conditions and at the appropriate wavelengths. The technology has, therefore, already matched the level where exoplanets could be reached. Of course, interferometry in general and nulling interferometry specifically is only one of many technological hurdles that have required study. These include issues such as phase shifting, delay compensation, formation flying of multiple spacecraft, and the manufacture and testing of special optical components, the most important being monomode fibers (used to filter away low-frequency spatial orders or “speckle noise”). Over the course of the past 9 years, ESA (and NASA as well) has initiated a large number of studies in these areas. Since the technologies involved with nulling interferometry in space touch on so many new and interesting (as well as potentially profitable) areas, the funding of these studies has been supported at levels of tens of millions of Euros. In this context, the technology of integrated optics serves as a good example. Here, the whole system of beam combination, if considered in terms of implementation with bulk optics in a “classical” sense, is large, complicated, sensitive, and difficult to align and test. If beam combination is transferred onto a silicon chip, however, with canals that pass through and carry out mixing and de-multiplexing, all the difficulties of bulk optics are overcome. While this technique is already in use on several ground-based interferometers, where it is producing impressive results, the potential is great for such a silicon chip to be used in conjunction with (or to completely replace) electromechanical switches to improve the efficiency of such devices by several orders of magnitude.

The Darwin mission, as recently described in the literature (e.g., Wallner et al., 2006; and, most recently, the community’s proposal to ESA’s Cosmic Vision Programme), is only one example of a mission scenario designed to address the issue, though it is very likely close to the preferred implementation. The nulling interferometry itself could be implemented in a wide variety of different configurations/architectures (e.g., Coulter, 2003; Karlsson et al., 2004), which would be constrained by, among other parameters, the number of telescopes and the necessary background and starlight suppression. Recent work in the area suggests that requirements on the shape of the null can be relaxed when unavoidable noise contributions introduced by instrumental errors are taken into account (Dubovitsky and Lay, 2004; Kaltenegger and Karlsson, 2004; Kaltenegger et al., 2006). Thus, configurations with fewer telescopes should be investigated as viable candidates for the exoplanetary missions, which would potentially reduce complexity and cost.

The overarching mission-specific science goal, as formulated in this study and then perpetuated into Cosmic Vision, can be specified as follows:

2.1. To detect and study Earth-type planets and characterize them as possible abodes of life

In Fridlund (2000), the high-level scientific requirements were described with regard to the following questions:

- Are we alone in the Universe?
- How unique is Earth as a planet?
- How unique is life in the Universe?
Therefore, the following would need to be achieved:

- Detect exoplanets within their HZs;
- Determine these planets’ physical and orbital characteristics (period, eccentricity, inclination, etc.);
- Determine the presence of an atmosphere and the object’s effective temperature via the spectrum of the planet;
- Determine the composition of the atmosphere. The presence of water, ozone, and oxygen for an Earth-type planet, inert gases for a Mars- or Venus-type planet, and hydrogen and methane atmospheres for a Jupiter-type planet or “primordial” Earth-like planets.

These high-level science drivers can be translated into specific observational requirements and allow for specification of mission requirements. These are detailed in Fridlund and Kaltenegger (2008), but can be summarized briefly as follows:

- The minimum number of single, Sun-type (F–K) and M-dwarf (the most common stars in the Galaxy) stars to be surveyed for terrestrial exoplanets in the HZ during primary mission should be 165–500.
- 165 under the added condition that significant amounts of dust (10 times the level in the Solar System) are present (see below)
- 500 Sun-like stars (complete sample of F, G, and K single stars out to 25 pc + a large number of M dwarfs) under the conditions of similar levels of dust to the Solar System
- Completeness of survey (probability that a planet in the HZ has not been missed for a specific star) should be 90%.
- Spectral signatures to be observed for each planet in a detected system should be CO2, H2O, CH4, and O3.

3. Exoplanets

In 1995, the search for planets outside our Solar System met with success by way of ground-based radial velocity surveys (Mayor and Queloz, 1995). It should be pointed out, however, that pioneering milestones were passed more than 50 years ago, when Reuyl and Holmberg (1943) reported a 0.008 solar mass component to the star 70 Oph, orbiting in a 17-year orbit. This study, based on 30 years of astrometric observations, is typical of early studies, of which there are a large number. Although this early work was important, since it set the scene for later work, not one such early astrometric observation has been confirmed. In 1952, Struve suggested a program by which to search for the radial velocity signature of a potential planet, as suggested by Struve, 1952), as outlined in the section above.

The period of 84 days placed the object in an orbit similar to that of Mercury around the Sun, and the velocity amplitude of about 0.6 km s⁻¹ implied a mass of the companion as low as 0.011 solar masses, or 11 times the mass of Jupiter (M_Jup). These authors suggested that the companion was probably a brown dwarf but could possibly be a giant planet. As was pointed out by Latham and his coworkers, since the inclination of the orbit to the line of sight was unknown, the mass of the companion could, however, be considerably larger than this lower limit. Today, this object is still tentatively classified as a brown dwarf, but, since there are indications that the inclination is very low (Mazeh et al., 1996), it is still not totally excluded that this is a giant planet.

The theory behind the detections, as suggested by Struve (1952), is that a potential planet would cause its parent star to undergo a reflex motion around the star-planet center of mass, as viewed from Earth. The deflection would have a radius \( r_{\text{star}} = r_{\text{pl}} \times (M_{\text{pl}}/M_{\text{star}}) \) and a period \( P \). This would lead to the perturbation of three (in principle) measurable variables:

- the radial velocity (a small periodic Doppler shift in the radial velocity of the absorption lines in the stellar spectrum, which need to be measured over a relatively long period of time—typically at least three times the planetary orbital period),
- the astrometric position,
- the arrival times of electromagnetic radiation.

The last of these was detected first. In the timing of radio signals from a number of pulsars, several Earth-sized planets have been detected (e.g., Demianski and Prosynski, 1979; Shabanova, 1995; Wolszczan, 1997). These objects are generally considered to be different from what we call planets in a “classical” sense. It is unlikely that they are objects formed around a star that later underwent a supernova explosion and managed to survive the cataclysm (though it has been suggested that they formed around another star and were later captured by the pulsar). A more favored model is that these objects formed after the explosive event in an accretion disk composed of debris material surrounding the pulsar (Bani et al., 1993; Phinney and Hansen, 1993; Podsdiakovs, 1993; Phillips and Thorsett, 1994).

The unambiguous detection of a planet beyond our own Solar System was reported by Mayor and Queloz (1995). The technique utilized was an indirect one that relied on the measurement of the reflex motion of the parent star, which was caused by an unseen (planetary) component, with respect to the common center of mass of the star-planet system (as suggested by Struve, 1952), as outlined in the section above.

The radial velocity amplitude \( A \) of a star of mass \( M_{\text{star}} \) due to a companion of mass \( M_{\text{pl}} \) with orbital period \( P \) is going to be:

\[
A = (2pG/P)^{1/3}[M_{\text{pl}} \sin (i)/(M_{\text{pl}} + M_{\text{star}})^{2/3}]/(1 - e^2)^{1/2}
\]

where \( i \) is the inclination of the orbit, \( G \) is the gravitational constant, and \( e \) is the eccentricity.

The first planet found via this method orbits the Sun-type star 51 Pegasi (Mayor and Queloz, 1995). This report was rapidly followed by a paper that announced planetary-mass bodies orbiting around the stars 70 Vir and 47 Uma (Marcy and Butler, 1996). Several groups, using this and other methods (see below), have now (October 2009) put more than 400
planets on the map of our immediate neighborhood in the Galaxy.

Most of the Sun-type stars brighter than ≈8 magnitude (about 1000 objects) are now being monitored by one or another of the 10 or so different radial velocity programs initiated after the first successes. In this category, we have an almost complete sample within 30 pc of the Sun. The extension to fainter stars is progressing faster as more dedicated and specialized instruments become available on large telescopes with time available for planet searches.

The breakthrough that allowed the first detection of bona fide planets was the ability to measure radial velocities consistently with the necessary precision—typically down to a few meters per second (the current limit is less than 1 meter per second). The precision of the method is, therefore, dependent on the accuracy with which we can measure the displaced spectrum of the star over periods from a few days to several years. This can be done by introducing a stable calibrating source (gas cell with hydrogen fluoride or iodine, which produces a stable reference spectrum) into the light path and utilizing regions of the stellar spectrum where the star itself shows a suitable complement of spectral lines. An alternative (and also very successful) method is to pipe the light from a calibration lamp (e.g., Th-Ar) into the spectrograph along the same light path as the stellar light. The most important part is the stability of the spectrograph to external influences. The spectral line requirement limits the type of star that is likely to be detected with this kind of search. Improvements of the method will likely allow, in the near future, precision of a few tenths of a meter per second. It is also necessary to compensate for the motion of Earth, including the gravitational perturbations by the other planets in the Solar System. Other techniques that utilize absolute accelerometry (e.g., Connes, 1994) have promised similar ultimate possible precision (±1 m s⁻¹). Since Earth itself causes a reflex motion of order 0.1 m s⁻¹ in the solar spectrum, this mass range is currently not detectable. The lowest-mass planets to date are the newly discovered Gliese 581c (Udry et al., 2007), which weighs in at 5.1 Earth masses [of course times the factor range is currently not detectable. The lowest-mass planets to calculate the average density, which turned out to be 5.5 g cm⁻³—radial velocity method from the ground. This allowed us to transit (see below) from which the radius of the body could be calculated. This is because the CoRoT spacecraft observed a planetary occultation, together with many others, has also been observed with amateur equipment—see Oksanen in Sky & Telescope, January 2001, page 14. In the case of the first observed transit, the planet orbits the star HD 209458 every 3.52 days at a distance of about 0.05 AU. The occultation lasts about 2.5 hours; and, from this observation, the inclination is found to be 87.1° (Charbonneau et al., 2000).]

Another indirect search method is to obtain astrometric data and, thus, track a star’s path across the sky, measuring the wobble introduced by the rotation around the common center of mass of the star-planet system. Reports of the detection of planetary companions to some nearby stars have been frequent during the last century. Barnard’s star, 70 Ophiuaci, and 61 Cyg have all at several times been the central objects of claims for detected planetary systems (e.g., Reuyl and Holmberg, 1943; Struve, 1952). None of these observations, however, have been confirmed. In contrast, by using Hipparcos data, upper masses for a number of planets have been determined. Among these, we find the first exoplanet detected—that in the system of 51 Peg (Perryman et al., 1996).

A first attempt toward an absolute mass determination for the outermost planet of Upsilon Andromedae (10.1 M_Jup +/− 4.7 M_Jup) (Mazeh et al., 1999) has also been carried out by utilizing Hipparcos data.

The Hubble Space Telescope is equipped with sophisticated Fine Guidance Sensors that can be used for detailed astrometric studies. With use of these sensors, planet(s) in close proximity to the Solar System have been discovered orbiting low-mass red dwarf stars.

The ESA’s Gaia mission promises large statistical surveys of massive planets. NASA’s Space Interferometer Mission (SIM—see below) will also carry out relative astrometry on a number of nearby stars and thus possibly detect planetary deflections.

Struve (1952) suggested that it would be possible to detect the occultation of a star by its planets. When a planet transits its primary, it will, if the geometry is favorable, induce a measurable loss of light in the stellar flux. The flux loss will be

\[ \frac{D}{F} = \left( \frac{R_p}{R_{\text{star}}} \right)^2, \]

which implies a mass of 0.63 M_Jup and a planetary radius of 1.27 R_Jup. This could be done, since the orbital radius is well known from the

Gliese 581c is found at a distance of about 6 pc, which implies a multitude of targets for missions of the Darwin/TPF type. (It must be stressed that, if Darwin was flying today, the actual properties of this planet—including its habitability—would have been determined within 20 hours).
radial velocity measurements, and the stellar radius is known to good accuracy from stellar evolution theory. The actual shape of the light curve during the occultation (Mazeh et al., 2000) then constrains the planetary radius and orbital inclination to a very high precision. The average density of the planet, as it turns out, is only half that of the major gaseous giant planets in our own Solar System, which immediately rules out the possibility of a rocky, terrestrial body, given that such a body would be significantly smaller than 1.27 R_{\text{Jup}}. The planet is thus a gas giant and larger than Jupiter for a lower mass, since proximity to its primary results in a surface temperature of 1200 K. Such temperatures would, however, only affect the outer 1% of the planet, and the large diameter speaks to the planet’s evolution in that the flux from the star retards the cooling of the planetary interior (Lunine, 2001). A giant planet formed in isolation would cool in a brief time (~10^6 years); thus it would also shrink rapidly from its original distended state. For a planet in very close proximity to a star, as is the case for HD209458b, the atmospheric temperature profile is flattened, and the rate by which heat can be transported outward from the interior is reduced. Thus the contraction will be retarded. It can also be shown in these models that the planet must have arrived at an orbital radius of ~0.05 AU within, at most, a few tens of millions of years after formation. Otherwise, it would take longer than the present age of the Universe for the external heat to diffuse inward far enough to expand the radius to the observed value. The observation of a single transit has thus shown that the so-called “hot Jupiters” either form in place or migrate inward within, at most, ~a few times 10^7 years (Burrows et al., 2000; Lunine, 2001).

Today, more than 60 confirmed exoplanets that occult their primaries have been reported. Most of these objects have been shown to possess larger diameters than their masses would indicate from modeling work and are thus “puffed up,” as is the case for HD209458b. Whether Lunine’s (2001) suggestion is viable, however, remains to be seen.

Utilizing the occultation method from space, where uninterrupted observations with high photometric accuracy are possible, was an early consideration. ESA began to develop mission concepts based on this method between 1992 and 2002 (STARS: Badiali et al., 1996. Eddington: Favata et al., 2000). These mission concepts and their associated technological development have been embodied in the CoRoT mission (see below), and NASA is currently implementing this strategy in its discovery-class mission Kepler, which launched in March 2009 (Borucki et al., 1997).

At the time of writing, the CNES/ESA/Brazil mission CoRoT (Convection, Rotation and planetary Transits; Rouan et al., 2000) has been in operation for 3 years. Having the dual objectives of detecting planetary transits and conducting studies of acoustical modes in stars, this mission was launched on December 27, 2006, into a near-perfect orbit around Earth. After a month of verification, it began scientific operations. Planned as a 3-year mission but extended to 6 years in October 2009, it has at the time of writing reported about 10 planets. Currently, CoRoT science teams are following up on hundreds of candidates. The precision in the photometry is such that AF/F of about 1–2×10^-3 is achievable, given that enough eclipses are observed to allow epoch folding (about 10 instances). This means that Earth-sized planets are, in principle, reachable. Restrictions come from the satellite’s position in low-Earth orbit. This allows uninterrupted observation for periods of a maximum of 180 days. To detect and verify an Earth-type planet (in an Earth-type or “habitable zone” orbit) around another (Sun-like) star using this method, the star must be observed continuously for a significant part of the typically 2–3 years required to view at least three occultations. The first occultation is considered to be the discovery, while the second gives the orbital period, and the third provides confirmation.

Nevertheless, CoRoT has achieved remarkable success. In February 2009, the scientists involved with the exoplanetary search and use of this spacecraft reported (at a CoRoT symposium in Paris) the discovery of the first unambiguously detected Earth-like planet. Its diameter is 1.68 times that of our own planet, and its mass is slightly less than 5 Earth masses. This means that its average density is of the same (close) order as that of the terrestrial planets Mercury, Venus, and Earth (Léger et al., 2009; Queloz et al., 2009). This planet orbits a K0V star at a distance of only about 5 stellar radii and must therefore be very hot indeed. Furthermore, one or several other planets (also in the “super-Earth” category) orbit farther out in this system. The discovery has demonstrated that there is a possibility for further discoveries in the CoRoT material, since only a small fraction of the collected data has been thoroughly analyzed and followed up.

Given the poor statistics at present, we do not know which type of star will likely have an Earth-type planet, nor do we know what the mass spectrum of lower-mass planets would be. Therefore, we need to observe very large samples of objects simultaneously to ensure that we will have a detection. Since by necessity such aggregates of stars are going to be distant (in order to fit within the field of view of the telescope), the method can—except in very rare instances—only produce statistical information. CoRoT will observe large samples of stars for several years and pick up transits of planets with masses down to about one or a few Earth mass(es). For CoRoT, the observation limits are set by the number of passages that are detected. CoRoT will be able to observe transits of Earth-sized planets in the HZ of late K- and M-type dwarf stars; and it is possible that, by using a different observational method, CoRoT will be able to search for an Earth-sized planet orbiting in the HZ around a G-type star (at about 1 AU), though it remains to be seen if this scheme can be implemented.

Missions like this can thus provide us with statistics about the number of systems containing Earth-like planets. It is not likely that it will be possible (for time reasons) to observe individually the nearby stars that will be the targets for detailed studies by interferometers. Nevertheless, it is clear that CoRoT will be able to determine the frequency of Earth-like worlds orbiting different classes of stars and, as a consequence, influence the design of missions like Darwin and TPF.

3.1. The formation of exoplanets

The presence of planets orbiting other stars is a natural consequence of the process of star formation. The process through which stars will form out of interstellar clouds is thought to be reasonably well understood—at least qualitatively. Dense molecular cores collapse under conservation of angular momentum, which leads to the emergence of a
protostar surrounded by a rotating disk and envelope (e.g., Fridlund et al., 2002). This disk contains sufficient angular momentum to prevent all material from collapsing onto the surface of the accreting stars, which grow at a slower pace through an interface in the inner part of the disk. The forming star eventually reaches the point where density and pressure allow full thermonuclear burning if the mass is above a certain value—a limit model calculations show to be about 0.08 $M_{\odot}$ (note that this corresponds to $\approx 80 M_{\text{Jup}}$). Deuterium burning in the cores can proceed at lower temperatures and consequently at lower masses, which leads to the formation of so-called brown dwarf stars. The disk drives the accretion onto the star by losing angular momentum. This poses a major problem in our understanding of the star-formation process as to exactly how excess angular momentum is shed; but interactions between the star, the disk, and magnetic fields, which cause extensive mass loss in directed outflows, are thought to be crucial to the process. At least during the formation of low- and intermediate-mass stars, an accreting disk will exist in a stable configuration for long enough—some $10^5$ to $10^6$ years—for a centrifugally supported disk containing the material for planetesimals to form.

Bodies with a mass below $\approx 12 M_{\text{Jup}}$ are too small to sustain nuclear fusion (neither hydrogen nor deuterium burning in the cores). The minimum Jeans mass (for formation via fragmentation) is $\approx 7$–20 $M_{\text{Jup}}$. This provides a useful mass boundary of 7–12 $M_{\text{Jup}}$ for a definition of what a planet is and what a star is (see Perryman, 2000, and references therein for a more detailed discussion about the brown dwarf problem as well). The picture that emerges is of material in the star-formation disk growing into planetesimals through mutual collisions, which are eventually guided by gravitational interactions into special orbits (while sweeping other regions clear). The dynamical timescale for these growth processes is between $10^5$ to $10^6$ years, while the actual formation of terrestrial planets takes $\approx 10^7$ to $\approx$ a few times $10^8$ years (Perryman, 2000). In the outer parts of these proto-solar systems, planetesimals of terrestrial size may grow larger and accrete residual gas (closer to the proto-sun the temperature is too high for these processes), which leads to planets of about Jupiter mass. Alternative models for the formation of the major planets in our Solar System that involve gravitational instabilities in the solar nebula have also been proposed (see references in Perryman, 2000). Most of the details of the described scenario remain unknown, such as the role of viscosity and the redistribution of angular momentum. To a certain extent, theoretical work has been significantly hampered by having had only one system to study. The search for exoplanets is thus also driven by the need to acquire the empirical knowledge needed to understand fully the origin and evolution of our own and other solar systems.

A significant number of the known planets also have very high eccentricities. In principle, all (26) of the planets more than 0.3 AU from their stars are in high-eccentricity orbits ($e > 0.1$). In contrast, most of the closer-in objects are in more-circular orbits, though there are a few exceptions (including two objects with $M \sin(i) > 14 M_{\text{Jup}}$ and thus above the “deuterium-burning limit”). Prior to the discovery of 51 Peg, most theories for planetary formation predicted that exoplanets would be in circular orbits similar to those characteristic of the Solar System (e.g., Boss, 1995). Theorists have had problems explaining the formation of these so-called “hot Jupiters”—the massive planets orbiting very near their parent stars—as well as the high value of the eccentricity characteristic of so many of the confirmed planets. In principle, most theories now deal with planetary migration, which involves a planet spiraling inward toward its central star due to its interaction with a gaseous disk that contains a significant amount of dust as well (Goldreich and Tremaine, 1980; Lin and Papaloizou, 1986). Tidal torques, viscous drag, or both cause the orbital decay of the protoplanet. A growing planet orbiting within the disk will generate spiral density waves in the inner and outer disk regions. The inner disk causes a positive torque, while the external disk causes a negative torque, with a finite residual tidal torque that causes the planet to lose orbital energy and migrate toward the star (Ward, 1986). A large protoplanetary core can initiate a runaway gas accretion. The gravitational torques will clear a gap inside the planet’s feeding zone. The protoplanet should migrate toward the gap on a timescale that is proportional to the viscous evolution time of the disk. Tanaka and Ida (1999) showed that in a minimum mass nebula a protoplanet with 1 martian mass at 1 AU should be able to survive migration and final assimilation into the star itself. The growth toward protoplanets is then caused by a shuts merger of the smaller planetesimals. Migration could also lead to an accumulation of planetary building blocks orbiting at close distances from the star, which would allow the building of giant planets at the small distances observed. (Bodenheimer et al., 2000).

However, not all models require post-accretion migration to explain the peculiar orbital semi-axes and eccentricities. For instance, the dynamical interaction between two or more giant planets has been invoked by Rasio and Ford (1996) and Weidenschilling and Marzari (1996) as a mechanism by which to generate instabilities that could lead to the ejection of one planet while another is left in a smaller and eccentric orbit. This could be the case for 70 Vir or 16 Cyg B. Moreover, for very close periastron distances, tidal dissipation would circularize the orbit of the planet in less than 1 billion years (which could have been the case for 51 Peg or t Boo). Armitage and Hansen (1999) suggested that, if a 5 $M_{\text{Jup}}$ planet could form sufficiently early (in the lifetime of the disk) in a disk of 0.1 solar masses (which is 1 order of magnitude larger than the minimum mass nebula), its presence could drive gravitational instabilities in an already marginally stable massive disk and cause disk fragmentation (see Fig. 1). The fragments would then rapidly accumulate into giant coplanar planets whose gravitational interaction with the largest planet would lead to systems of planets on eccentric orbits. Several numerical simulations have been produced that generate plausible planetary systems, at least on a dynamical basis. These simulations are mainly focused on the formation of either giant planets (Levison et al., 1998) or terrestrial planets (Wetherill, 1996) through the study of the dynamical and collisional evolution of a swarm of planetesimals under a wide range of initial conditions. In both cases, the authors were able to produce a large number of bodies with characteristics similar to those presented by our own Solar System planets as well as by exoplanets with small semimajor axes, large eccentricities, or both. Levison et al. (1998) were able to obtain planetary bodies with masses ranging between $4M_{\text{Jup}}$ down to Neptune-sized planets. Also, the eccentricity in several stable systems
has been pumped up after considerable gravitational scattering between interacting planets, which produces eccentricity as large as that of 16 Cyg B or 70 Vir.

The formation of rocky inner planets (inside the HZs of stars) was investigated by Wetherill (1996). Wetherill’s results show that the distribution of the semimajor axis of the planet is centered around 1 AU and is quite insensitive to the mass of the star (for variation between 0.5 and 1.5 solar masses). At the same time, it heavily depends on the location of the inner edge of the solar nebula (where high temperatures preclude the condensation of solid particles from which planetesimals are formed) and on the distance of the orbit of a Jupiter-like planet. If no giant planet is present, Earth-sized planets could extend out to 4 AU.

3.2. The search for terrestrial exoplanets

Today, it is known that planets exist outside our Solar System. These systems, to date, however, do not look like our own system. The current search from the ground is now aimed at detecting a system that would be more like our own. Specifically, we would like to detect a system with one or more Jupiter-mass planet(s) in circular orbit(s) outside 3 AU distance from the primary and without any large planets orbiting farther in toward the HZ. Nevertheless, the recent success with CoRoT, as well as the number of low-mass planets detected through radial velocity measurements (see above), seem to indicate a class of objects with masses about 5 times that of Earth, which orbit very near their primaries. A few cases of objects in the same class orbiting what appear to be red dwarfs have also been found to orbit very far away from their primaries (these last discoveries were made via the detection of gravitational lensing magnification of the stellar light—see, e.g., Beaulieu et al., 2006).

Even today, two possibilities must be considered with regard to terrestrial planets of the type and location found in the Solar System:

- Such terrestrial planets are common throughout the universe or at least in our neighborhood.
- Such terrestrial planets are extremely rare or even nonexistent elsewhere.

Given these two cases, it is clear that a search for terrestrial exoplanets in any sample of stars must be performed in such a large sample that even a negative result is meaningful. The primary scientific case for the next objective of exoplanetology can be defined thus:

To search for and study Earth-like planets in the HZs around Sun-type stars in a large enough sample and provide a statistically significant answer to whether Earth, the one habitable planet we know, is a unique occurrence.

In this context, then, what is a solar-type star? To be specific, it is a G2 main sequence star of roughly 4.5 Gyr of age; and, should investigators wish to search several hundred objects, the search radius would have to be extended to several hundred parsecs or more. To have a statistically significant sample of single stars within, say, a 25 pc radius, it would be necessary to include all F, G, K, and some M dwarfs that are currently known. Further, even within their specific types, these stars are going to be vastly different, depending on differences in age, metallicity, activity, conditions for habitability, and other physical properties. Also, the definition and existence of the so-called HZ is going to be a function of the stellar properties.

To study any newly discovered terrestrial planet, it will be necessary to analyze the planetary light in detail. The first step will obviously be to detect the planet more than once. Apart from confirming the existence of the object, this would allow a determination of its orbit and, under certain assumptions about the albedo, its mass. An analysis of the planetary light would also allow for determination as to whether the planet has an atmosphere and, if it does, what the composition of the atmosphere is. All of this is with restrictions and to a greater or lesser level is going to be model dependent.

The habitability of any discovered body will, of course, depend on these parameters, but the most exciting aspect is that, in principle, it will be possible to determine whether life exists on terrestrial exoplanets if they are close enough. This will depend, however, on the existence of so-called biomarkers.* Life—at least as we understand it—modifies its surroundings. All life on Earth exists, more or less, out of equilibrium with the rest of our planet. The most obvious cases of biomarkers are the existence of oxygen in the atmosphere today and the existence of methane in the earliest phase of the history of life. It should be noted that such biomarkers involve chemical equilibrium, and if, for example, all life was somehow removed from Earth, essentially all oxygen would be removed from our atmosphere in the relatively short time of 4 million years. It would be converted into carbonates, silicates and ferro-oxides (rust), plus water.

3.3. Comparative planetology and astrobiology—two new sciences based on traditional ideas

What is meant by “putting Earth into context”? In studies carried out to date, the notion of “putting Earth into context” has meant taking an approach whereby a number of issues that remain unanswered with regard to the world we inhabit are investigated. This approach has resulted in a focus on comparative planetology and astrobiology, two new sciences that have received a great deal of attention over the past decade. Comparative planetology has been defined in one of two ways: (1) the comparison between different bodies in our Solar System or (2) the comparison between bodies that have their abdomes in different star systems. Here, we take the latter approach, given that we do not know how well a comparison between Earth and, for example, Mars or Venus could be applied to our current, past, or future conditions. On the other hand, the study of a large number of systems similar to ours that are well distributed in a grid of physical parameters, such as, for example, stellar masses, ages, and abundances of heavy elements, could provide the framework by which to place our world into context.

Astrobiology, defined as the science of life external to Earth, is the second step in the search for our own identity; and, given it is in its infancy as a discipline, its first empirical results are recent. Astrobiology seeks to understand the

---

*The term biomarker is used here to mean detectable atmospheric species or set of species whose presence at significant abundance strongly suggests a biological origin.
occurrence of life on this planet and the possible evolutionary pathways beyond the present, and has two fundamental branches:

(1) The issue of life can be addressed within our Solar System, for example, on Mars or perhaps in the liquid water beneath the icy crust of Jupiter’s moon Europa. This branch has to do with the emergence of life on Earth, since it is possible that life arose multiple times on several worlds only to be snuffed out except possibly on Earth. It is possible as well that life on one world was transferred (through, e.g., meteoric impacts) to others in the Solar System. Questions such as the latter await in situ investigations in detail on other planets in the Solar System—something which is currently the goal of the expansion of the Mars exploration program in the context of several space agencies (e.g., ESA’s ExoMars program).

(2) Astrobiology is concerned with life outside the Solar System.

To an extent, this is a field more relevant to the core question: “Are we alone in the Universe?” Given that our Solar System formed all at once and under the same initial conditions, it can be argued that the complete Solar System—from the Sun to the Oort Cloud—represents one instance in the formation of life. Of course, if we could find life—fossil or otherwise—in, for example, the liquid water reservoirs of Europa that was based on other fundamental parameters than terrestrial life, the situation would be somewhat different. Nevertheless, in advance of the discovery of such “alien” life-forms, we must attempt to detect signs of life at interstellar distances if the fundamental question is to be addressed.

It can thus be seen that the two “new” sciences are deeply interconnected.

In summary, questions that need to be answered are as follows:

- Do other solar systems exist?
- Do such systems possess planets like our own Earth, that is, terrestrial planets? Do these planets orbit around stars that have enough luminosity and duration to allow for stable conditions for life?
- Are the planetary orbits within the continuously habitable zone?
- Do these planets have physical conditions such that life, in principle, is possible?
- Are there signs of life that can be detected (“biomarkers”)?

So far, only the first of these questions has been answered. We know of more than 400 planets orbiting in more than 270 star systems. The second question, however, is in the process of being answered. On December 27, 2006, the European CoRoT mission was launched with the intent to address these issues (see above). In April 2007, the first radial velocity detection of a planet likely to be “rocky” was discovered close to the HZ of its central star (Udry et al., 2007). In February 2009, CoRoT reported its first “rocky” planet.


The European Space Agency’s Cosmic Vision Science Programme for the period 2015–2025 identifies four main themes, each of which is centered on a fundamental question that currently remains unanswered. These questions are such that the answers would significantly impact the scientific community and our understanding of the Universe.

Among the themes, the first stands out in this context in that it is centered on the question “What are the conditions for life and planetary formation in the universe?” Thus, already in its formulation it has unusual cross-disciplinary aspects.

The other themes in Cosmic Vision, namely “How does the Solar System work?” “What are the fundamental physical laws of the Universe?” and “How did the Universe originate?” are of course closely related to the issues asked in the first theme, and they can be viewed together as a set of interrelated issues, the answering of which will improve mankind’s understanding of himself and his world.

The central issue in Theme 1 can be subdivided into the following three topics:

- From gas and dust to stars and planets
- From exoplanets to biomarkers
- Life and habitability in the Solar System

By addressing these subtopics, researchers will be able to place Earth and the Solar System into the proper setting with regard to how worlds like our own form and evolve. Advances made by way of such a quest would undoubtedly lead to new fields of science, which would involve topics as widely disparate as biology, geophysics, and astronomy. Our newly acquired understanding would also allow for true comparisons to be made, and it would enhance our ability to draw conclusions as to how different circumstances lead to different evolutionary scenarios. The ultimate element of study, of course, would be the origin of life and its subsequent evolution, which could, in time, lead to a paradigm shift in our understanding of our genesis and the planetary environment within which we exist. Such studies may even one day reveal specifics as to our future destiny. Previously, science was restricted to the study of objects in our Solar System alone, though it should be said that this did not preclude our use of meaningful comparisons. As an example, many comparisons have been made between Venus and Earth, which are located relatively close together in the Solar System (both currently orbiting within the HZ, i.e., where life can exist). The two planets are of the same size and have roughly the same average density. Nevertheless, they are completely different in appearance; and, to date, we have little understanding as to why this is so. The comparative planetology on which we are currently set to embark will allow such questions to be addressed in a statistical way.

In ESA’s outline of the Cosmic Vision plan (ESA BR-247, 2005), a proposed strategy for Theme 1 is also given:

- First: In-depth analysis of terrestrial planets.
- Next: Understanding the conditions for star, planet, and life formation.
- Finally: Image terrestrial exoplanet.

Abbreviations

CNES, Centre National d’Études Spatiales; CoRoT, Convection, Rotation and Planetary Transits; HZ, habitable zone;
TE-SAT, the Terrestrial Exoplanet Science Advisory Team; TPF, Terrestrial Planet Finder; TPF-SAG, the TPF Science Advisory Group.

References


Address correspondence to:
Malcolm Fridlund
RSSD, ESA, ESTEC
Noordwijk
the Netherlands

E-mail: malcolm.fridlund@esa.int