Planetary science beyond the boundaries of our Solar System is today in its infancy. Until a couple of decades ago, the detailed investigation of the planetary properties was restricted to objects orbiting inside the Kuiper Belt. Today, we cannot ignore that the number of known planets has increased by two orders of magnitude nor that these planets resemble anything but the objects present in our own Solar System. Whether this fact is the result of a selection bias induced by the kind of techniques used to discover new planets—mainly radial velocity and transit—or simply the proof that the Solar System is a rarity in the Milky Way, we do not know yet. What is clear, though, is that the Solar System has failed to be the paradigm not only in our Galaxy but even ‘just’ in the solar neighbourhood. This finding, although unsettling, forces us to reconsider our knowledge of planets under a different light and perhaps question a few of the theoretical pillars on which we base our current ‘understanding’. The next decade will be critical to advance in what we should perhaps call Galactic planetary science. In this paper, I review highlights and pitfalls of our current knowledge of this topic and elaborate on how this knowledge might arguably evolve in the next decade. More critically, I identify what should be the mandatory scientific and technical steps to be taken in this fascinating journey of remote exploration of planets in our Galaxy.

1. Introduction

If I had to select a single word to define the field of exoplanets, that word would be revolutionary. During the past years, over 1000 planets have been found around every type of star from A to M, including pulsars and binaries. Being the leftover of the stellar formation processes, planets appear to be rather ubiquitous and,
in reality, the presence of a host star is not even a mandatory circumstance. The current statistical estimates indicate that, on average, every star in our Galaxy hosts at least one planetary companion [1], i.e. our Milky Way is crowded with one hundred billion planets.

The most revolutionary aspect of this young field is the discovery that the Solar System does not appear to be the paradigm in our Galaxy, but rather one of the many possible configurations we are seeing out there. These include planets completing a revolution in less than 1 day, as well as planets orbiting two stars or moving on trajectories so eccentric as to resemble comets. This variety of stellar and orbital parameters converts into planetary temperatures that span over two orders of magnitude. Unexpectedly, planetary sizes and masses do not appear to be ‘quantized’, as happens in the Solar System, where the terrestrial planets are well separated from Neptune and Uranus, and those are, in turn, quite distinct from Jupiter and Saturn. Instead, a continuum of sizes and masses appear to exist, from the super-Jupiter down to the sub-Earth objects [2,3].

While the relative frequency of ‘odd’ planets compared to the ‘normal’ ones—assuming the Solar System planets represent the normality—might be the result of some selection effects caused by the detection techniques used so far—mainly radial velocity and transit—it is undoubtable that a great diversity of planets does exist around other stars. In the short term, we should be able to shed light on this issue. The European Space Agency’s GAIA mission is expected to find several thousand new planets through astrometry [4,5], a technique sensitive to planets lying in a different region of the parameter space compared to transit and radial velocity, in particular to planets at intermediate separation—typically a few astronomical units—from their mother star. The instruments ESO-VLT SPHERE [6], Gemini Planet Imager [7] and Subaru SCExAO [8] were built to detect young, massive planets at large separation from the stars, a regime not yet well explored till now.

With these numbers and premises, emphasis in the field of exoplanetary science must shift from discovery to understanding: understanding the nature of exoplanetary bodies and their history. The following fundamental questions need to be addressed:

— What is the origin of the observed exoplanet diversity?
— How and where do exoplanets form?
— What are the physical processes responsible for exoplanet evolution?

In all disciplines, taxonomy is often the first step towards understanding, yet we do not have, to date, even a simple taxonomy of planets and planetary systems. For planets transiting in front of their parent stars—of which over 400 are known today—the simplest observables are the planetary radius and, when combined with radial velocity, the mass. Mass and radius allow us to estimate the planetary bulk density. From figure 1, it is evident that even gas giants have a broad range of interior structures and core compositions, as shown from the different bulk densities observed [10,11]. While this has stimulated very interesting theoretical work on planetary interiors and equations of state of hydrogen at high pressure and temperature, the implications on, for example, planetary formation and evolution mechanisms are still unclear. Most probably, the different bulk densities reflect the different nature and size of the planets’ cores, which in turn will depend on both the formation mechanism and the ‘birth distance’ from the parent star. Objects lighter than 10 Earth masses (super-Earths/sub-Neptunes, figure 1b) are even more enigmatic, as they often can be explained in different ways [12–14]. Among those, Kepler-10 b, Kepler-78 b, CoRoT-7 b and 55 Cnc e all have high densities and orbit G stars like our Sun with periods of less than 1 day. By contrast, GJ 1214 b and Kepler-11 d, e, f have much lower densities and are subjected to less intense insolation because of their longer period or cooler parent star. In the next years, dedicated space missions, such as NASA TESS [15] and ESA CHEOPS [16], combined with radial velocity surveys, will measure the sizes and masses of a few thousand new planets, completing the current statistics of available planetary densities in the solar neighbourhood down to the terrestrial regime.

As explained earlier, density is a very important parameter, but alone it cannot be used as a discriminant of the variety of cases we are seeing out there. We need additional information
to proceed. The other key observable for planets is the chemical composition and state of their atmosphere. Knowing what atmospheres are made of is essential to clarify, for instance, whether a planet was born in the orbit it is observed in or whether it has migrated a long way; it is also critical to understand the role of stellar radiation on escape processes, chemical evolution and global circulation. To date, two methods can be used to sound exoplanetary atmospheres: transit and eclipse spectroscopy, and direct imaging spectroscopy. These are very complementary methods and we should pursue both to get a coherent picture of planets outside our Solar System (figure 2).

**Figure 1.** Masses and radii of currently confirmed transiting planets [9]. Extrasolar planets are denoted by circles (red online) and Solar System planets are represented by triangles (green online). The grey lines (green and red online, respectively) denote Earth-like composition (67% rock, 33% iron) and Mercury-like composition (40% rock, 60% iron). (Online version in colour.)

**Figure 2.** (a) Phase curve of the hot Jupiter HAT-P7b while orbiting around its mother star as observed by Kepler [17]. The transit and eclipse events occur at 10 h and 35 h, respectively. (b) Image of the planet β-Pic b, located 8–15 astronomical units from the star, as observed with instrument VLT-Naco [18]. (Online version in colour.)
2. Brief review of exoplanet spectroscopic observations

(a) Transit

When a planet passes in front of its host star (transit), the star flux is reduced by a few per cent, corresponding to the planet/star projected area ratio (transit depth). The planetary radius \( R_p \) can be inferred from this measurement. If atomic or molecular species are present in the exoplanet’s atmosphere, the inferred radius is larger at some specific wavelengths (absorption) corresponding to the spectral signatures of these species [19–21]. The transit depth \( \kappa(\lambda) \) as a function of wavelength \( \lambda \) is given by

\[
\kappa(\lambda) = \frac{R_p^2}{R_\star^2} + 2 \int_0^{z_{\text{max}}} (R_p + z)(1 - e^{-\tau(z,\lambda)}) \, dz,
\]

where \( R_\star \) is the stellar radius, \( z \) the altitude above \( R_p \) and \( \tau \) the optical depth. Equation (2.1) has a unique solution, provided we know \( R_p \) accurately; \( R_p \) is the planetary radius at which the planet becomes opaque at all \( \lambda \). For a terrestrial planet, \( R_p \) usually coincides with the radius at the surface. By contrast, for a gaseous planet, \( R_p \) may correspond to a pressure \( p_0 \sim 1–10 \text{ bar} \) depending on the transparency of the atmosphere.

(b) Eclipse

A measurement of the planet’s emission/reflection can be obtained through the observation of the planetary eclipse, by recording the difference between the combined star plus planet signal, measured just before and after the eclipse, and the stellar flux alone, measured during the eclipse. In contrast with the primary transit observations, the dayside of the planet is observed, which makes the two methods fully complementary. Observations provide measurements of the flux emitted and/or reflected by the planet in units of the stellar flux [22,23]. The planet/star flux ratio \( \phi(\lambda) \) is defined as

\[
\phi(\lambda) = \frac{R_p^2}{R_\star^2} \frac{F_p(\lambda)}{F_\star(\lambda)}.
\]

(c) Phase curves

In addition to transit and eclipse observations, monitoring the flux of the star plus planet system over the orbital period allows the retrieval of information on the planet emission at different phase angles. Such observations have to be performed from space, as they typically span over a time interval of more than a day [17,24–26].

(d) Direct imaging

The planet/star brightness contrast may typically range between \( 10^{-4} \) and \( 10^{-10} \) depending on many parameters of the system, i.e. age, distance, planetary size, temperature, etc., and of course spectral interval. To fix the ideas, Jupiter has a contrast of about \( 10^{-9} \) relative to the Sun in the visible and an angular separation of 0.5′′ at 10 pc. The use of a coronagraphic system [27,28] is therefore essential to extract the planetary signal out of the stellar light.

Wavefront aberrations and stellar speckles are another critical problem that needs to be attenuated. Deformable mirrors [29] and speckle calibration techniques, such as angular differential imaging [30], can be used effectively to address this issue.

3. Highlights and problems with current photometric and spectroscopic data

(a) Highlights

Water vapour appears to be ubiquitous in the atmospheres of transiting hot Jupiters with temperatures between 800 and 2200 K observed to date [31–38]. The additional presence of carbon-bearing species, such as methane, carbon monoxide and carbon dioxide, in those
atmospheres has been supported by both observations and spectral simulations [26,35,39–43], but their relative abundances are still unclear [40,44–47]. Nitrogen-bearing species (e.g. HCN, NH₃) are most probably also there [48,49], but current observations are not precise enough to indicate their presence. Ground-based observations in the L-band have been interpreted as bearing the signature of methane fluorescence in the atmosphere of one of these hot Jupiters [50,51]. This would be an important diagnostic of the physical structure of the upper atmosphere of these planets probed through a minor atmospheric constituent. In the atmosphere of very hot Jupiters, where temperatures approach 3000 K, exotic species commonly present in brown dwarfs, such as metal oxides (TiO, VO) or metal hydrides (CrH, TiH, etc.), have been suggested to explain observations by the Hubble STIS and WFC3 [31,52,53]. These species are important, as they may influence both the planetary albedo and the vertical thermal structure of the planet. Sodium and perhaps potassium are present in most hot Jupiters analysed [54–57]. Apart from these alkali metals, the spectra in the visible appear dominated by Rayleigh scattering or condensates/hazes [58,59].

Warm Neptunes are expected to be methane-rich [48,60,61], and indeed photometric observations of GJ 436b may point in this direction [62]. Spectroscopy will be needed to unravel the full picture of this and other objects, such as GJ 3740b [63,64]. The ~6 Earth-mass, warm planet GJ 1214b is the first super-Earth that has been probed spectroscopically [65]. The VLT observations were followed by other space and ground observations [66] that are suggestive of an atmosphere heavier than pure molecular hydrogen, but additional observations are needed to confirm its composition [67].

Information on the stability of the atmospheres of transiting planets has been collected through UV observations with Hubble [68–70]: hydrodynamic escape processes are likely to occur for most of the planets orbiting too close to their parent star [71–75]. Also, orbital phase curves in the IR [24,25,76] and eclipse mapping measurements [77,78] have provided first constraints on the thermal properties and dynamics of hot Jupiters’ atmospheres.

In parallel with transit studies, in the next decade direct imaging techniques are expected to allow observations of hot, young planets at large separations from their parent star, i.e. gaseous planets newly formed in the outer regions of their planetary disc and not (yet?) migrated inward.

Multiple-band photometry and spectroscopy in the near-IR (1–5 μm) have been obtained for a few young gaseous planets, such as β Pic-b [79,80], GJ 504 b [81] and the planets around HR 8799 [82]. These observations will be perfected and extended to tens of objects with dedicated instruments, such as SPHERE and GPI. The comparison of the chemical composition of these young gaseous objects with the composition of their migrated siblings probed through transit will enable us to understand the role played by migration and by extreme irradiation on gaseous planets.

(b) Issues and possible solutions

Although the field of exoplanet spectroscopy has been very successful in past years, there are a few serious hurdles that need to be overcome to progress in this area, in particular the following:

1. Instrument systematics are often difficult to disentangle from the signal. In the past, parametric models have extensively been used by most teams to remove instrument systematics. This approach has caused many debates regarding the use of different parametric choices to remove the systematic errors. Parametric models approximate systematic noise by fitting a linear combination of optical state vectors to the data (e.g. X- and Y-positional drifts of the star or the spectrum on the detector, the focus and the detector temperature changes, positional angles of the telescope on the sky). Even when the parametrization is sufficient, it is often difficult to determine which combination of these parameters may best capture the systematic effects of the instrument.

Unsupervised machine learning algorithms do not need to be trained prior to use and do not require auxiliary or prior information on the star, instrument or planet. The machine learning approach will ‘learn’ the characteristics of an instrument from
observations, allowing one to de-trend systematics from the astrophysical signal. This approach guarantees a higher degree of objectivity compared with traditional methods. In Waldmann [51,83], Waldmann et al. [84] and Morello et al. [85], independent component analysis (ICA) [86] has been adopted as an effective way to decorrelate the exoplanetary signal from the instrument systematics in the case of Hubble NICMOS and Spitzer IRS and IRAC data.

2. Especially for transit observations, stellar activity is the largest source of astrophysical noise. Stellar noise is an important source of spectral and temporal instability in exoplanetary time-series measurements [87,88]. This is particularly true for M dwarf host stars as well as many non-main-sequence stars. Correction mechanisms for fluctuations must be an integral part of the data analysis. The problem of stellar activity removal from time-series data is a very active field of research. Whereas most instrumental effects can be measured or calibrated to some degree, stellar and general astrophysical noise does not usually grant us this luxury. In Waldmann [51] and Danielski et al. [89], the same methods explained in point 1 to decorrelate the systematic noise were successfully used to eliminate/reduce the effects of the stellar activity in Kepler photometric light curves. These methods need to be applied to spectroscopic time series, to assess their validity and potential also in the spectral domain.

3. Data are sparse, i.e. there is not enough wavelength coverage and most of the time the observations were not recorded simultaneously.

4. Absolute calibration at the level of $10^{-4}$ is not guaranteed by current instruments, and therefore caution is needed when one combines multiple datasets not recorded simultaneously.

5. Transmission and emission spectra, as measured through transit, eclipse and direct imaging, are intrinsically degenerate. In transit spectroscopy, the degeneracy in the retrieval of molecular abundances may be caused by the imprecise knowledge of $R_p$ (equation (2.1)). In IR eclipse and direct imaging spectroscopy, the information on molecular abundances is entangled with the atmospheric vertical thermal profile; see for instance Tinetti et al. [90] for a more detailed discussion. For transiting planets, to remove the degeneracy between molecular abundances/planetary radius or molecular abundances/vertical thermal gradient, a broad wavelength coverage is needed together with adequate signal-to-noise ratio (SNR) and spectral resolving power (SRP) (see point 7). Direct imaging observations also suffer from the lack of knowledge of the planetary radius and sometimes mass. When the mass and the radius are not known, model estimates need to be invoked, increasing the source of degeneracy.

6. Accurate linelists are an essential element of radiative transfer models, and this fact is not always appreciated. As a result, the abundances for molecular species are often derived with linelists that are incomplete or extrapolated from measurements/calculations at low temperatures. This issue—especially together with point 3—may introduce large errors. For instance, all the current claims of carbon-rich or carbon-poor planets [91] published in the literature are unsubstantiated for this reason. This problem is well known to spectroscopists, and linelists at high temperatures are being calculated ab initio or measured in the laboratory [92].

7. We are dealing with very low SNR observations. While the adoption of new data analysis methods and models might address some of the issues listed above, the lack of good data is something we cannot solve in the short term. I indicate below the SNR per spectral resolution element and SRP that would be needed to guarantee a sound spectral retrieval. The reader should refer to Tinetti et al. [90] and Tessenyi et al. [93] for a more extensive discussion of these parameters.

(a) Basic. SNR $\sim 5$ and SRP $\sim 50$ for $\lambda < 5 \mu m$ and SRP $\sim 30$ for $\lambda > 5 \mu m$. Key molecular species can be detected and the main thermal properties are captured.

(b) Deep. SNR $\sim 10$ and SRP $\sim 100$ for $\lambda < 5 \mu m$ and SRP $\sim 30$ for $\lambda > 5 \mu m$. Molecular abundances (i.e. the abundance of one component relative to that of all other components) are retrieved and so is the vertical thermal structure.
new large, general-purpose observatories from space and the ground will come online, notably

In the next decade, a new generation of ground and space facilities is needed to tackle those. In the next decade, new large, general-purpose observatories from space and the ground will come online, notably

Table 1. Warm Neptune: minimum detectable abundances (mixing ratios) for a basic, deep and ultradeep retrieval. The bulk composition of the planetary atmosphere in this simulation is molecular hydrogen with a small fraction of helium. See Tessenyi et al. [93] for further details on the method.

<table>
<thead>
<tr>
<th></th>
<th>CH₄</th>
<th>CO</th>
<th>CO₂</th>
<th>NH₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>retrieval</td>
<td>3.3 μm</td>
<td>8 μm</td>
<td>2.3 μm</td>
<td>4.6 μm</td>
</tr>
<tr>
<td>ultradeep</td>
<td>10⁻⁷</td>
<td>10⁻⁶</td>
<td>10⁻⁴</td>
<td>10⁻⁶</td>
</tr>
<tr>
<td>deep</td>
<td>10⁻⁷</td>
<td>10⁻⁶</td>
<td>10⁻⁶</td>
<td>10⁻⁶</td>
</tr>
<tr>
<td>basic</td>
<td>10⁻⁷</td>
<td>10⁻⁵</td>
<td>10⁻⁴</td>
<td>10⁻⁷</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>PH₃</th>
<th>CH₃CH₃</th>
<th>H₂S</th>
<th>C₂H₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>retrieval</td>
<td>4.3 μm</td>
<td>10 μm</td>
<td>3.3 μm</td>
<td>12.2 μm</td>
</tr>
<tr>
<td>ultradeep</td>
<td>10⁻⁷</td>
<td>10⁻⁶</td>
<td>10⁻⁶</td>
<td>10⁻⁶</td>
</tr>
<tr>
<td>deep</td>
<td>10⁻⁷</td>
<td>10⁻⁶</td>
<td>10⁻⁵</td>
<td>10⁻⁵</td>
</tr>
<tr>
<td>basic</td>
<td>10⁻⁷</td>
<td>10⁻⁵</td>
<td>10⁻⁵</td>
<td>10⁻⁷</td>
</tr>
</tbody>
</table>

Table 2. Hot super-Earth, around a G-type star: minimum detectable abundances (mixing ratios) for a basic, deep and ultradeep retrieval. The bulk composition of the planetary atmosphere in this simulation is H₂O. See Tessenyi et al. [93] for further details on the method and the impact of other main atmospheric components on the results.

<table>
<thead>
<tr>
<th></th>
<th>H₂O</th>
<th>H₂O</th>
</tr>
</thead>
<tbody>
<tr>
<td>retrieval</td>
<td>—</td>
<td>2.8 μm</td>
</tr>
<tr>
<td>ultradeep</td>
<td>—</td>
<td>10⁻⁴</td>
</tr>
<tr>
<td>deep</td>
<td>—</td>
<td>10⁻⁴</td>
</tr>
<tr>
<td>basic</td>
<td>—</td>
<td>10⁻³</td>
</tr>
</tbody>
</table>

(c) Ultradeep. SNR ~ 20 and SRP ~ 300 for λ < 5 μm and SRP ~ 30 for λ > 5 μm. A very thorough spectral retrieval study can be performed.

In tables 1 and 2, I show the detectable molecular abundances at fixed SNR and SRP for a typical warm Neptune, for example, GJ 436 b, and a hot super-Earth, for example, 55 Cnc e. The results for hot Jupiters are very similar to the ones reported for warm Neptunes. The reader should refer to Tessenyi et al. [93] for the case of temperate super-Earths around late dwarfs.

4. The next decade and beyond

In §3b, I identified the hurdles that cannot be solved in the short term (in particular, points 3, 4 and 7): a new generation of ground and space facilities is needed to tackle those. In the next decade, new large, general-purpose observatories from space and the ground will come online, notably
JWST and E-ELT. It is understood that, among many other science goals, they will contribute significantly to exoplanet spectroscopic observations, in both transit and direct imaging [94–96]. More crucially for this field, dedicated instruments and missions are being studied or planned.

The idea of a dedicated IR observatory in space to study exoplanetary atmospheres is clearly not new: back in the 1980s Bracewell [97] and Angel et al. [98] proposed that exoplanets around nearby stars could be detected in the IR (6–17 μm) and their spectra analysed, searching for CO2, H2O, O3, CH4 and NH3 spectral features. The proposal to implement this idea under the form of an IR nulling interferometer in space came almost ten years later [99]. The concept, named DARWIN, was first proposed to ESA in 1993, when the only known planets were the nine in our Solar System (plus three around a neutron star). Its principal objectives were to detect Earth-like planets around nearby stars, to analyse the composition of their atmospheres and to assess their ability to sustain life as we know it. Similar mission concepts were proposed to NASA in the USA (Terrestrial Planet Finder-Interferometer [100]). The working hypothesis of an Earth-twin plus Sun-twin as the only cradle of life was too geocentric to survive the ‘exoplanet revolution’ and none of these very challenging missions have been implemented. A couple of decades of exoplanet discoveries have taught us that the pathways to habitable planets are multiple, but the most interesting ones are those able to cast light on a host of physical and chemical processes not entirely understood or missing altogether in our Solar System [101,102].

In past years, mission concepts for IR transit spectroscopy from space were proposed to and studied by both ESA and NASA, in particular THESIS [103], Finesse [104] and EChO [105]. The transit and eclipse spectroscopy methods allow us to measure atmospheric signals from the planet at levels of at least \(10^{-4}\) relative to the star. No angular resolution is needed, as the signals from the star and from the planet are differentiated using knowledge of the planetary ephemerides. This can only be achieved in conjunction with a carefully designed, stable payload and satellite platform. EChO, the Exoplanet Characterization Observatory, is currently one of the five M3 mission candidates being assessed by ESA, for a possible launch in 2022.1 If selected, EChO will provide low–mid resolution (\(R = 30–300\)), simultaneous multi-wavelength spectroscopic observations (0.55–11 μm, goal 0.4–16 μm) of a few hundred planets, including hot, warm and temperate gaseous planets and super-Earths around different stellar types. These measurements will allow the retrieval of the molecular composition and thermal structure of the atmospheres observed. The design of the whole detection chain and satellite will be optimized to achieve a high degree of photometric stability (i.e. approx. 100 ppm in 10 h) and repeatability: the telescope will have a collecting area of 1.13 m², will be diffraction-limited at 3 μm and positioned at L2. This Lagrangian point provides a cold and stable thermal environment, as well as a large field of regard to allow efficient time-critical observation of targets randomly distributed over the sky. I show in figure 3 the simulated performances achievable by EChO to observe the warm super-Earth GJ 1214 b. Planets that are much smaller (less than 1.5 Earth radii) and colder than this one (colder than 300 K) will be challenging for an EChO-like mission. Temperate super-Earths may be observable only around bright late M dwarfs.

Provided a small/medium-size transit spectroscopy mission is launched in the next decade, would it make sense to envisage a large spectroscopy mission later on? Probably not. To illustrate why, it is useful to first discuss a few basic concepts. The numbers of electrons per spectral element on the detector from the star (\(N_s\)) and planet (\(N_p\)) are

\[
N_s = \frac{\Delta F_s A_{\text{eff}} \eta Q}{n} \quad \text{and} \quad N_p = \phi N_s = \phi \frac{\Delta F_s A_{\text{eff}} \eta Q}{n} t,
\]

where \(\phi\) is the planet/star contrast defined in (2.2), \(\Delta F_s\) is the stellar flux in the spectral band observed (photons s\(^{-1}\) m\(^{-2}\)), \(A_{\text{eff}}\) is the effective collecting area (m\(^2\)), \(\eta\) is the instrumental throughput (dimensionless), \(Q\) is the detector quantum efficiency (e\(^-\)/photon) and \(t\) is the integration time (s). If we assume the observations to be dominated by the stellar photon noise,

---

Figure 3. (a) Transit spectroscopic observations of the super-Earth GJ 1214 b recorded with Hubble-WFC3 [66]. (b) Simulations of EChO performances compatible with a ‘deep retrieval’. In December 2013, about 160 known exoplanets could be observed with SNR and SRP corresponding to ‘basic’, a good fraction of them as ‘deep’ and about 10–20 as ‘ultradeep’; see http://sci.esa.int/echo/ and Varley et al. [106]. (Online version in colour.)

The SNR per spectral element is

$$\text{SNR} = \frac{N_P}{\sqrt{N_s}} = \phi \sqrt{\frac{\Delta F_s A_{\text{eff}} \eta Q_t}{n}}.$$  (4.2)

The SNR thus scales with $\sqrt{A_{\text{eff}}}$, i.e. it goes linearly with telescope diameter ($D$). The cost of a telescope scales, in the most optimistic cost models, as $D$ to the 1.2 power, with some models indicating an exponent of 2 [107]. Therefore, an increase in telescope diameter of a factor of 2 means a cost increase of a factor of 2 to 4, while doubling the SNR has a small to negligible impact on the science.

To be transformational, we should aim at an improvement of at least a factor of 10 in the SNR, and this would require the idea of an agile, highly stable platform to be abandoned in favour of a large, deployable structure, as monolithic space telescopes are limited by fairing size to about 4 m diameter. The said structure might represent an encumbrance when trying to reach the pointing stability required by transit observations and certainly might limit the ability to move and repoint agilely from one target to another in the sky. Note that a factor of 10 in SNR might not be sufficient in any case to observe the atmospheres of Earth-like planets around Sun-like stars. For those targets, in fact, transits are expected to occur once per year, and 5–6 transits (assuming a mission lifetime of approx. 5 years) will not be enough to collect the required photons.

Direct imaging from space is the expected next step to be taken in space after transit. Space telescopes with various types of coronagraphs are being studied in the USA, Europe and Japan [108–114]. A mission for direct imaging would be technically more challenging than a transit one and certainly more expensive—the telescope cannot be a light bucket, to start with. The said mission, though, would open up the spectroscopic exploration of planets at larger separation from the stars, a domain that is impracticable with transits.
In the past two decades, the field of exoplanets has spoiled us in terms of creativity and transformational ideas, so perhaps we should not be too surprised if a new technology or a new observing strategy comes online soon, making all the other techniques obsolete or just inefficient.

Acknowledgements. G.T. is a Royal Society University Research Fellow.

References


