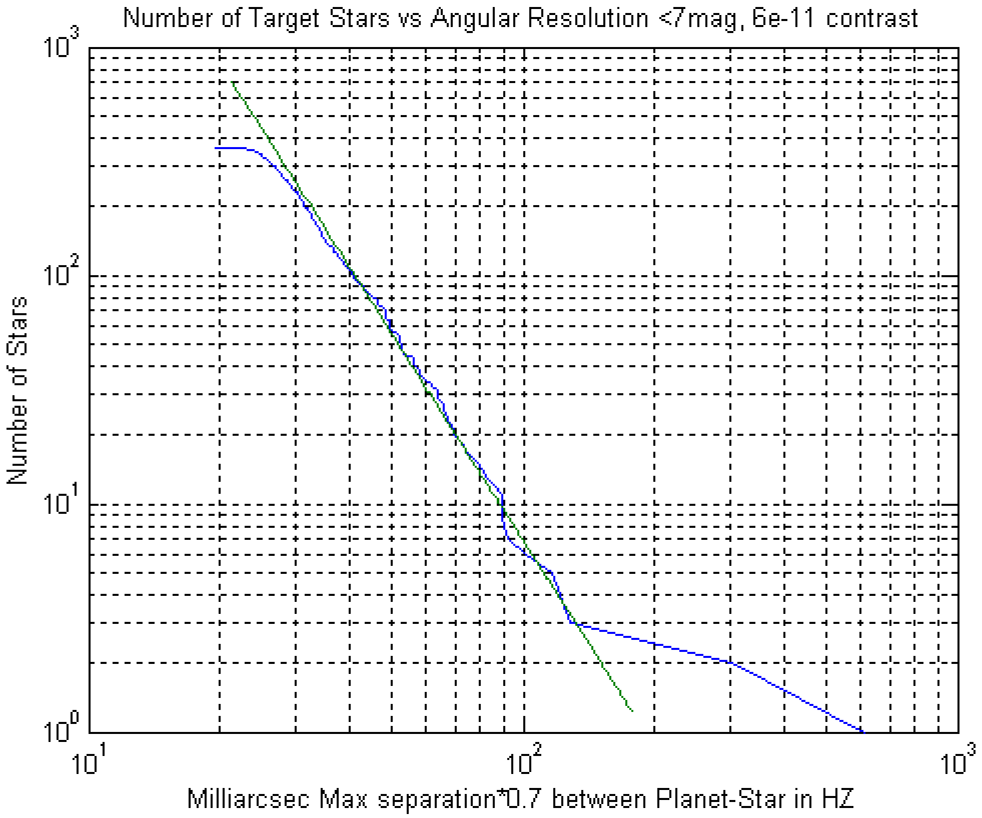
Direct Detection and Spectroscopy of Exo-Earths;
The Need for High Angular Resolution and Other Observational Requirements

Michael Shao[1], Sarah Bairstow[1], Elizabeth Deems[1], Leigh Fletcher[1], B. Martin Levine[1], Glen Orton[1], Gautam Vasisht[1], Feng Zhao[1], Mark Clampin[2] , Richard G. Lyon[2], Olivier Guyon[3] , Benjamin F. Lane[4] , Keith Havey[5] , Jeff Wynn[5], Rocco Samuele[6], Gopal Vasudevan[7], Robert A. Woodruff[8], Volker Tolls[9], Fabien Malbet[10], Alain Leger[11]



Number of target stars versus angular resolution of a coronagraph, for stars brighter than 7 magnitude and star planet contrast > 6e-11, for an Earth clone in mid-habitable zone.

---

[1] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA
[2] NASA/Goddard Spaceflight Center, Greenbelt, MD
[3] University of Arizona/Subaru Telescope, Tucson, AZ
[4] C.S. Draper Laboratory, Cambridge, MA
[5] ITT Industries, Rochester, NY
[6] Northrop Grumman Space Technology, Redondo Beach, CA
[7] Lockheed Martin Space Systems Co., Palo Alto, CA
[8] Lockheed Martin Space Systems Co., Louisville, CO
[9] Smithsonian Astrophysical Laboratory, Cambridge, MA
[10] University of Grenoble, France
[11] Institut d'Astrophysique Spatial (CNRS), Universite Paris-Sud, Orsay

Direct Detection and Spectroscopy of Exo-Earths; The Need for High Angular Resolution and Other Observational Requirements

Abstract

The study of planets outside our solar system has advanced by leaps and bounds in the last decade. Nearly 350 planets are known outside the solar system. Even spectra of transiting hot-Jupiters have been measured[1]. Our ability to measure the flux and spectra of transiting planets is ultimately limited by a combination of the shot-noise and the spectra-photometric stability of the star. When the star-planet flux ratio exceeds a million, we will need to use coronagraphic and interferometric techniques that spatially separate the starlight from the planet light. An Earth twin is about 10 billion times fainter than the Sun, when viewed from outside the solar system. All coronagraphs have a so-called "inner working angle", IWA. Planets inside the IWA are not observable. Often the number of habitable Earths that a coronagraphic mission can detect is calculated by the number of stars where the maximum star-planet separation is equal to or greater than the IWA, where the habitable planet is 1 AU*sqrt(luminosity) from the star. This simplistic view can overestimate the number of viable targets by as much as a factor of 4 to 8.

Direct imaging of an exo-Earth can be used to measure a number of important planetary parameters, 1) planetary orbit, is it in the habitable zone? 2) planet flux versus orbital phase, is the planet a Lambertian scatterer? 3) spectra of its atmosphere in the visible and Near IR, is there oxygen or water in the atmosphere? and 4) possible seasonal changes in albedo, and spectra. If the IWA is 2X smaller, on average we will be able to detect earth-like planets around stars twice as far away, or roughly an 8 fold increase in the number of targets. But if we need to be able to observe the planet over a large part of its orbit and we need the maximum star-planet separation to be twice the IWA, we've just reduced our list of potential targets by a factor of 8. This white paper examines in more detail what science is possible as a function of the max-star-earth/IWA ratio.

Introduction

When a potential exo-Earth is detected, the first thing we want to know is, "is this an Earth?" and is it in the habitable zone? Measuring the orbit of a planet in our solar system is pretty straight forward, that's because we can observe the planet over approximately 90% of its orbit. In Figure 1, the large blue circle is the IWA and the yellow arcs are the parts of the orbit when the planet is observable. The planet is not always observable even outside the IWA, when the bright side of the planet is facing away from us. With a coronagraph whose IWA is only slightly smaller than the max star-planet separation, some orbital parameters like orbit inclination can't be measured. The planets's apparent brightness can vary by a factor of 3 from the "full moon" phase to the half moon phase. In multiple planet systems two images with one planet in each image leaves open the possibility that there are two separate planets only one of which is outside the IWA at a time. The next section of this white paper describes in more detail, the IWA requirements for measuring planetary orbits.

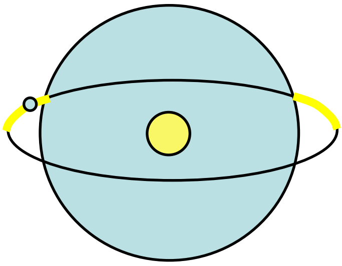


Figure 1: Observable IWA and planet orbit

In all coronagraph designs, internal, interferometric, and external, the IWA is a strong function of wavelength. But if we are able to measure the spectra of the planet at longer wavelengths, we will be able to identify a larger number atmospheric constituents. If the IWA is substantially smaller (e.g. 50%) of the maximum star-planet separation, the planet becomes observable over most of its orbit. In this case it will be possible to look for seasonal variations in albedo. Such variations may be because the surface is a non-lambertian scatterer. But there could also be real seasonal variations, such as the top layer of a water world freezing over in winter time.

Last of all this white paper examines the question of how many nearby stars and what types of stars should be searched in order to detect an Earth twin. Double stars make up slightly more than half the stars in the solar neighborhood, and most of these double stars have habitable zones where planet orbits would be stable.

We end this white paper with a summary of the observational capabilities a space observatory would need to spectroscopically characterize an Earth clone around reasonable number nearby stars, including double stars.

Planetary Orbits, with Imaging and/or Astrometry

When we image a planet as a dot in a sea of speckles, we want to know if this is a potentially habitable planet. We want to know its mass and the semi-major axis of its orbit. A 1 M_{earth} in a 1 AU orbit at 10pc, only astrometry at the sub-microarcsec level can measure the mass of the planet with reasonable precision (+/- 0.3 M_{earth}). But both astrometry and imaging can in theory measure the orbit. Astrometry, because it looks at the star, doesn't have an IWA limitation and the motion of the planet as inferred by its reflex motion on the star can be measured throughout the orbit. We consider 2 different scenarios; Scenario 1) where the planet is first discovered by an astrometric mission, the role of imaging is to a) confirm the discovery, and b) improve on its orbit determination. Scenario 2) is where the imaging mission must both detect and then characterize the discovered planet to the same level of precision as in the precursor astrometric mission followed by the imaging mission.

Astrometric Orbit precision

NASA conducted a double blind study for the astrometric detection of Earth like planets in multiple planet systems[2]. The result of the test was that the presence of multiple planets has a marginal to negligible impact on the astrometric mission's ability to detect terrestrial planets in the habitable zone. One of the side products of that study was a determination of the accuracy of the astrometric orbit at the "edge" of detectability. A mission with a Signal to Noise Ratio (SNR) of 5.8 was deemed necessary to detect planets with a false alarm probability of only 1%. At this SNR, the period of a 1 year planet would have a 1 sigma error of 3% and the mass 1 sigma error of 0.3 M_{earth}. In indirect detection, the semi-major axis of the orbit is derived from its period using Kepler's laws. The orbital phase at mid-mission has an error that was roughly 0.24 radian (±14 days in a 365 day year). If the astrometric data preceded the imaging search by 5 years, the uncertainty in the orbital period would cause the orbital phase error bar to grow linearly with time. Five years after the mid-epoch of the astrometry data, the orbital phase uncertainty would be roughly (± 50 days, or about 0.85AU).

If we start an imaging search of an Earth-Sun clone at 10pc that was previously found astrometrically, the astrometric error bar 5 yrs after the mean epoch of astrometric measurements would be 0.03AU in the radial direction and about 0.85AU in the circumferential direction. A single image of the planet taken five or more years after the astrometric data set could dramatically reduce the 1AU error bar in the circumferential direction.

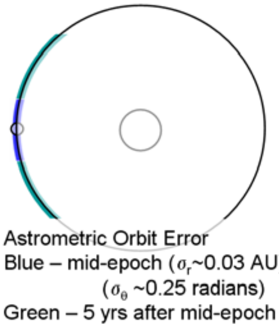


Figure 2: Astrometic Orbit Error

Imaging verification and refinement of the Orbit

If the coronagraph has an IWA that was 0.9 of the max star-planet separation, the planet would be observable for 26 days on either extreme of its orbit. With a 60 day 1 sigma error bar, the probability that the planet would be seen on the first attempt is 18%. Two attempts spaced 26 days apart would increase the probability to 33%. The situation is better for a planet where the IWA was 0.8 of the max star-planet separation. The first visit probability is 25% and the two visit probability is 48%. For planets that barely poke their heads beyond the IWA, one needs about 4-6 images to image the planet the first time with near 100% probability.

One image of the planet would significantly narrow the circumferential error bar in the orbit. If the coronagraph was working at $2\lambda/D$, planet-star separation = 1AU = $2\lambda/D$, then a single SNR=5 image would locate the planet to roughly 0.1AU. The major error in the astrometric orbit is the circumferential position of the planet (1AU) due primarily to the 5 year time delay between the astrometric survey and imaging follow up. That error is reduced by a factor of about 10 with one image. The one image would improve the period from 3% to 0.3% and the orbital phase uncertaintity would be 0.1AU at the time the image was taken and degrade to 0.14AU after 5 years after the first image..

Planet orbit from Imaging data alone

Astrometry can detect 1 M_{earth} planets in the habitable zone. But the “best” targets for astrometry and coronagraphic imaging are two distinct sets of stars that overlap at the 60% to 70% level. The most important result from astrometry is the detection and the “non-detection” of planets in the 70% overlap. If astrometry has detected a planet we know that the probability that a planet exists is 99%. But conversely if astrometry sees no periodic signal it can state with greater than 99% confidence the planet doesn’t exist. If 10% of stars have terrestrial planets in the habitable zone, this knowledge saves 90% of valuable mission time in “searching” for the planets and getting its orbit. Without this knowledge we need to image the star-planetary system many times to first detect one or more planets, then confirm the faint dots are not background objects and finally to get their orbits.

An important consideration is that a planet with uniform albedo in reflected light will exhibit factors of 3 changes in apparent brightness from zero phase (full moon) to 90° phase (half moon). If the system has multiple planets, you can’t use the apparent brightness (with a SNR=5 image) to identify a planet. Three observations of a planet are then needed to determine its orbit, if we can use photometry to identify a planet, but four images of a planet can be used to derive an orbit and verify proper identification of the planet. If the planet in only observable over 10% of its orbit then a total of 40 images need to be taken to get the required four images of the planet to get its orbit.

Direct imaging when only 10~20% of the orbit is visible has significant limitations for certain orbital parameters, notably orbit inclination and eccentricity. Measuring the orbit from imaging data alone, becomes a lot easier with a smaller IWA, when the planet is observable over 50% or more of its orbit. With 50% of the orbit visible, orbit inclination and eccentricity can also be measured. Eight images instead of 40 images would be sufficient to get an orbit and one image instead of four to six if the planet was previously detected astrometrically.

Science Enabled with Small IWA

In all coronagraphs, internal and external, the IWA is inversely proportional to the size of the instrument, the size of the telescope for internal coronagraphs and the size of the occulter for external coronagraphs. IWA is also proportional to wavelength, the longer the wavelength, the larger the IWA. A factor of two improvement in IWA means a 2X larger telescope for an internal coronagraph or a 2X larger starshade, and 16X larger solar panels and propellant mass for an external occulter. Small IWA is the principle cost driver of a coronagraphic mission, but it is also a principle science driver.

Spectroscopic science in the Near IR

Every coronagraphic mission plans to cover the visible spectral range 500-800nm, because of the Oxygen line at 760nm. But if the instrument can gain a factor of 2X improvement in IWA, the spectroscopic science between 800 and 1700nm becomes possible for comparable costs.

Habitability: In general terrestrial planets are considered habitable if they can maintain liquid surface water. This is not a sufficient condition because a planet may have poor levels of water at formation or insufficient mass to maintain an atmosphere or an ocean over time. To conclusively establish if a planet is habitable, two approaches are possible (1) to directly address the presence of an ocean via a phase dependent measurement of the ocean glint or (2) undertake a comprehensive study of the mass, size, orbit and atmospheric composition of the planet to determine whether the surface conditions on the planet are conducive to stable liquid-water.

Although the mass, density and orbital parameters of a planet serve as a zeroth order parameters of habitability, a definitive determination of whether a planet can support liquid water comes from a characterization of its atmosphere. The presence and abundances of greenhouse gases, coupled with a climate model, can determine the surface temperature. For instance in the Solar System, Earth and Venus have surfaces that are 17º and 430º K hotter than their global effective temperatures. This difference between surface and effective temperatures is due to the greenhouse effect, and leads to drastically different surface conditions. Although early characterization of the atmosphere (and surface) of a planet can be undertaken with time-resolved photometry at a few wavelengths, the potential diversity of planets, and the diversity of factors leading to similar photometric colors means that photometry alone is not sufficient and that spectroscopy (R ~ 80) to determine the content of the atmosphere is essential.

Low resolution spectroscopy in the visible and near infrared allows for unambiguous detection of many molecular band shapes. With sufficient wavelength coverage, ambiguities in absorption signatures can be resolved by searching for multiple bands for a give species. The three most dominant greenhouse gases, H_2O, CO_2 and CH_4 have strong absorption bands in the near IR. Water vapor also has strong bands in the visible part of the spectrum.

Searching for Biosignatures: Any biosphere on a targeted planet has a chance of being detected in disk-averaged spectra only if it is truly global, and has had sufficient span to influence atmospheric and surface conditions. First, observations of the atmosphere must show the presence of a biological source that exists in disequilibrium with the atmosphere and the environment (such as oxygen and its products on Earth). Second, observations of anomalous surface bio-signatures such as the red-edge of plant-life shows significantly enhanced reflectivity longwards of 700nm. Finally, temporal bio-signatures, where an atmospheric constituent or the surface albedo or signature shows marked seasonal variation due to seasonal fluctuation of its underlying biology. In the life-history of any planet, a range of environmental conditions, evolutionary and metabolic conditions may have a range of biospheric impact that is different from that seen on the Earth today or during any past period on the ancient Earth. Although on the modern Earth, the biosphere's signatures are dominated by oxygenic photosynthesis for the past 2 Gyr, alternate metabolisms still flourish.

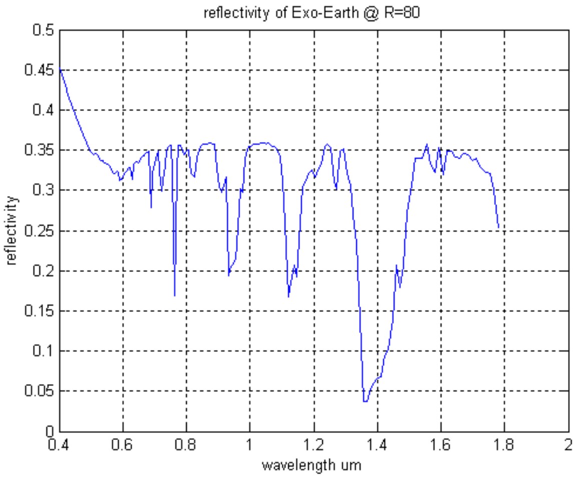
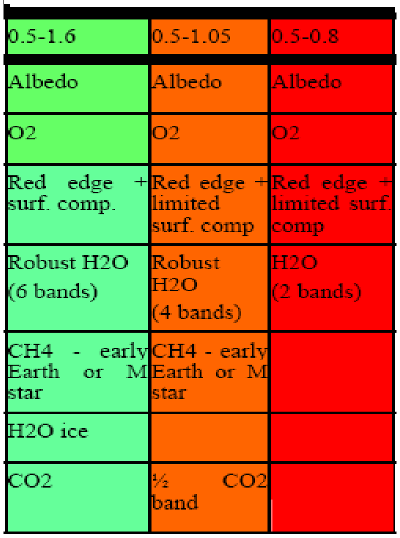


Figure 3: Theoretical spectrum of an exo-Earth at R=80.

On the Earth, photosynthesis produces the only global primary signature of life i.e. atmospheric oxygen and ozone in the presence of liquid surface water, and pigmentation that can be distinguished from an otherwise mineral surface. The possibility of false positives for life, given oxygen, exist and continue to be further explored[3]. However, given a paucity of hypotheses, the presence of an oxidizing atmosphere -- O_2, O_3, simultaneously with liquid water and reduced gases such as CH_4, are widely agreed as strong priority for spectroscopy for future exo-planet missions. A much more challenging observation would be the detection of a signature such as the vegetation red-edge, the strong contrast in red absorption and NIR reflectance, although this would have to be present at a higher level than the 2-3% signature seen in the Earth's disk averaged spectrum

The 900-1700nm range is essential to characterize a range of different terrestrial planets[5]--visible detection in the 500-800nm region is designed with only modern Earth in mind. The spectral region beyond 1000nm is essential for detecting the terrestrial gases CO_2 and CH_4. The numerous strong CO_2 features between 1000-1700nm will allow CO_2 identification. CH_4 is an important biomarker gas and is expected to be considerably more abundant in the early Earth due to the metabolism of methanogenic bacteria (e.g., 3Gyr ago CH_4 may have been 1000 times more abundant in Earth's atmosphere than it is today). Newly discovered terrestrial planets have the potential to be extremely different from Earth, Venus and Mars---planets that appear very similar over a short wavelength span in the visible may show substantially greater differences when observed over an extended wavelength region 500-1700nm.

Table 1: Benefits of detecting spectral signatures over a broad band in the visible. The column, 0.5-1.6, is our proposed spectral region. This is compared to the TPF-goal region between 0.5-1.05 and the TPF-goal required of 0.5-0.8. (Taken from Meadows, 2004, 2005)



0.5-1.6	0.5-1.05	0.5-0.8
Albedo	Albedo	Albedo
O2	O2	O2
Red edge + surf. comp.	Red edge + limited surf. comp	Red edge + limited surf comp
Robust H2O (6 bands)	Robust H2O (4 bands)	H2O (2 bands)
CH4 - early Earth or M star	CH4 - early Earth or M star	
H2O ice		
CO2	½ CO2 band	

Qualitatively, the benefits of using an extended wavelength range is given by Meadows' table (Table 1)[5]. Extended spectral sensitivity enables full detection of CO_2 and water-ice. Meadows has shown that many of the important biomarkers of planetary spectra lie within the 0.5 μm - 1.6μm spectral band. There are 3 much deeper H_2O absorption features between 0.9 1o 1.5μm. On a terrestrial planet covered with high clouds, the H_2O features will be obscured in the 0.5 to 0.9 region, and only the deeper H_2O features beyond 0.9 μm will be detectable[6]. O_2 also has an additional absorption feature at 1.27 μm.

Multi-color photometry and spectroscopy versus orbital phase

If instead of just seeing the planet over 10 to20% of its orbit we had access to 50 to60% of the orbit we can measure changes in the brightness, color and spectra of the planet as a function of orbital phase. If the planet has a lot of water, like our own, the polar ice caps would grow and shrink with the seasons. Ocean has an albedo approximately 20% while ice and snow 80 to 90%. If we are looking at primarily one of the polar ice caps, we might be able to see seasonal changes in albedo.

A small IWA will let us measure with high accuracy, the orbital inclination. A polar view of the orbit is correlated with polar view of the spin axis of the planet. If the planet has an eccentricity larger than the planets in our solar system, the seasonal changes may result more from the orbit eccentricity than the inclination of the spin axis to the orbit. But orbit inclination and eccentricity are much more easily measured with a small IWA.

In a planet composed of vast flat lands covered with grass, seasonal changes might not just be a change in albedo, if there is a season change of landcover from vegetation to snow, the red edge of vegetation would also exhibit a seasonal change.

Star Lists and Observational Capabilities

Observational requirements, number of target stars

The Exo-planet Task Force (EXOPTF) report recommended that an astrometric mission be able to survey 60 to100 nearby stars for Earths. We feel that this is an appropriate number for a direct imaging mission as well.

Until Kepler is launched, no one has any data on what fraction of stars have terrestrial planets in the habitable zone, and we are left to speculate. If we look at the fraction of stars that have Jovian planets, we find that the number of planets (per unit mass) increases dramatically at low masses. But a plot of density vs log Mass and log Period show the density only slowly varying with logM and logP. From periods of 3 days to 3 years and from 0.3 M_{jup} to 10 M_{jup}, about 15% of stars have planets and about 10% of stars that have planets have multiple Jovian planets. If we use this density in logM and log P and extrapolate to terrestrial planets about 1% of stars would have terrestrial planets in the habitable zone. The reason for the small number is the small volume of phase space of the habitable zone. More recently, the Swiss radial velocity team have predicted that between 3 days and 3 months and 5~50 Earths masses, up to 30% of stars have one such Neptune/super earth type planet. This is a dramatic increase in the density of planets. When extrapolated to terrestrial planets of 1~10 M_{earth} in the habitable zone, about 10% of stars are expected to have such planets.

While we won't have data on the prevalence of Earths in the habitable zone until Kepler data has been analyzed, an assumption of 10% seems a reasonable guess given with currently knowledge. A coronagraph capable of detecting an Earth in the habitable zone of 60 nearby stars seems like a reasonable "minimum" capability for a mission designed to characterize the spectra of an exo-Earth in the habitable zone. An Earth at 1AU from the Sun has a contrast of 1.2e-10 when the planet is at 90deg phase angle (½ moon). We can select candidate stars by assuming a 1 M_{earth} planet in the mid-habitable zone,1 AU*sqrt(Luminosity), satisfying the following criteria. 1) star planet contrast < 6e-11 at 90deg phase, 2) brighter than 7 magnitude, and 3) < 30pc away. Figure 4 shows the number of target stars versus maximum star to planet separation. There are a total of around 360 stars that fit the above three criteria. The curve follows the power law #stars = Sep^-3 as expected once we get past Alpha cen A and B, two stars very close to the Sun where a terrestrial planet would be markedly easier to image. The list stops at 360 because of the 7 mag and 6e-11 contrast cut offs.

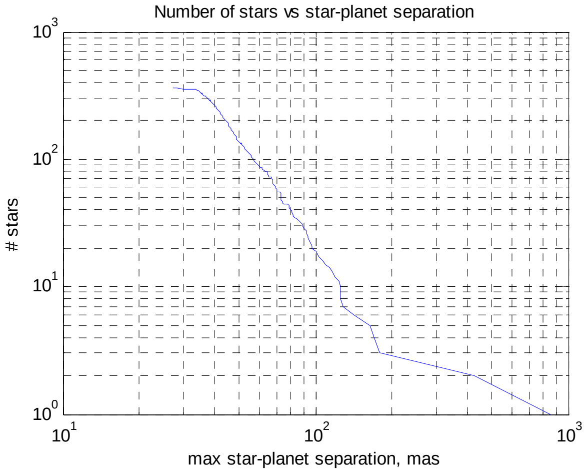


Figure 4: Number of stars vs. planet separation

While there are 20 candidate stars whose max star-planet separation is 100mas or larger, in practice if the IWA is only slightly smaller than the max star-planet separation measurement of its orbit is impossible. While "detection" of the planet with SNR=5 might only take in about 10hrs or so, R=80 spectroscopy with SNR=10 per spectral channel might take 500 hrs or approximately 3 weeks. Typically a planet will be outside the IWA twice year, as in Figure 1. The IWA must be smaller than 0.9 of the maximum separation if the planet is to spend 3 contiguous weeks (6% of an orbit) outside the IWA. Over a year the planet would be observable 12% of the time. If the IWA was 0.8 of the max star-planet separation, the planet would be observable about 20% of the time. If the planet had not been detected astrometrically, 20% observability means about 20 observations would be needed to image the planet four times to correctly identify the planet in a keplerian orbit. A search of 60 stars with 20 observations means ~1200 images would need to be taken with sufficient integration time to detect an Earth. If we want to be able to conduct the search in about 3 years, the IWA has to be around 30% less than the max-star planet separation. If we wanted to see if the planet's albedo change with seasons, or whether light scattered from the planet was Lambertian, we would want the planet to be observable over 40 to 60% of its orbit and the IWA should be about 0.5 of the max star-planet separation.

If we want to have 50 target stars where IWA = 0.7 max star planet separation, the coronagraph would need an IWA of the order 50mas. Ideally this small IWA would be available at 800nm wavelength.

Observational requirements binary stars

Approximately half of the stars in the solar neighborhood are binary stars. The vast majority of these are either very close binaries, where there are stable circumbinary habitable zone orbits or wide binaries where each star has a stable habitable zone orbit. It would be highly desirable for a major coronagraphic mission to be able to detect habitable planets around binaries. It is possible that the majority of habitable planets in the solar neighborhood are in binaries.

Summary and Conclusions

There is a wealth of science gained when we can see the planet over more than one half of its orbit. With an IWA of 40mas, there are up to 50 stars (including doubles) that have a maximum star-planet separation greater than 80mas. Further, additional science in is gained from spectroscopy when the observing wavelength range is extended to 1700nm. If the IWA is one half the max star planet separation at 850nm, then spectra at 1.7um wavelength is possible when the planet is furthest from the star.

Often, the number of stars around which an Exo-Earth can be detected is taken to be the number of stars where a 1AU orbit exceeds the IWA. In practice, the IWA has to be between 0.5 to 0.7 the max star-planet separation. At about 0.7 if imaging alone has to determine the orbit and 0.5 if we wish to explore changes in albedo and spectra with orbital phase.

Acknowledgments

The research described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

References and Further Reading:

---

[1] Richardson, L.J., Deming, D., Horning, K. Seager, S. Harrington, J., 'A spectrum of an extrasolar planet', Nature, v445, p892, 2007.

[2] http://planetquest.jpl.nasa.gov/SIM/researchOpps/roDoubleBlind

[3] Segura, A. et al. (2008). Abiotic formation of O_2 and O_3 in high-CO_2 terrestrial atmospheres. *Astronomy & Astrophysics* 472(2): 665-679

[4] 'TPF-C Instrument Concept Study: Visible and Infrared Nulling Coronagraph Spectrometer", A Report Submitted to the Terrestrial Planet Finder Coronagraph Science and Technology Development Team 7 March 2006, p4., Michael Shao, Principal Investigator.

[5] Meadows, V., "Synthetic Terrestrial Planet Spectra in the MIR", presented at the TPF Science Working Group Meeting, October 2004. Also see, Meadow's paper, 'Modelling the Diversity of Extrasolar Terrestrial Planets', p25 and Figure 2 in Direct Imaging of exoplanets: Science and Techniques, Proceedings IAU Colloquium No. 200, C. Aime and F. Vakili, eds., 2005, Cambridge University Press, New York, c.2006.

[6] Des Marais, D.J., et. al, Remote Sensing of Planetary Properties and Biosignatures on Extrasolar Terrestrial Planets, Astrobiology, 2, 153-181, 2002.