Part V

Discovery and study of extrasolar planets - future

Planetary Systems in the Universe - Observation, Formation and Evolution
Proceedings IAU Symposium No. 202, ©2004 IAU
Alan Penny, Pawel Artymowicz, Anne-Marie Lagrange, & Sara Russell, eds.

Planetary System Imaging and Spectroscopy from Earth and Space

N.J. Woolf and J.R.P. Angel

Steward Observatory, The University of Arizona, Tucson Arizona 85721, USA

J.I. Lunine

Lunar and Planetary Laboratory, The University of Arizona, Tucson Arizona 85721, USA

Abstract. We describe the two main techniques by which planetary systems may be imaged, and by which spectra of the planets may be obtained. There is a discussion of how planetary systems will appear, and of the extreme differences in brightness between stars and their planets Coronography is an extension of Lyot's technique by addition of adaptive optics. Nulling is an interferometric technique originally proposed by Bracewell. Differences between the techniques are emphasized. The paper concludes with a discussion of the problem of training people to develop these technically difficult tasks.

1. Introduction

This paper is mainly concerned with the techniques that can be used to observe directly external planetary systems from Earth and Space. We discuss the reasons for wanting direct observations. Then, because the observations require astronomers who have a more substantial technical training than is common these days, we discuss the training of astronomers for this type of work.

The view of a planetary system from far off would differ with the wavelength of observation, and the spatial resolution. At most wavelengths for systems like our own, the scattered or emitted light from the interplanetary dust would be the next most luminous object to the star. At the shortest wavelengths, where the reflected light from both dust and planets is seen, the inner components of the planetary system show well, and the ill-illuminated outer parts show weakly. But at longer wavelengths where the thermal emission of dust and planets can be seen, the star and inner dust show up less well, and it should be possible to see the planets by their own emission. In the long run, observations of planets will be needed in every spectral region to fully understand them. Graphs of emission versus wavelength for the star, planets and dust have been presented by Angel & Woolf (1998).

Direct observations of planetary systems are desirable because the spectral features of planetary atmospheres are good diagnostics of the physical state of a planet's outer parts. Also, although parts of a planetary system may be only weakly interacting today, during formation they were interacting very strongly, and the detailed observations will show the effects of this. The inner planets of our solar system would seem to have been formed late, in an environment set by whatever happened in giant planet formation. From the way dust seems to form in most unpromising circumstances and rocky planets are found around pulsars, coagulation should have been no problem. We do not know whether Jupiter and Saturn were the only giant planets formed in our system, or whether other early giant planets formed, and migrated into the sun or were lost to space. We know so little about other planetary systems that we do not know whether the solar system is typical of other planetary systems or is a rarity. Similarly while we know that about 1/3 of all forming stars seem to have a planetary system or at least a dusty disk forming around them, only about 5% have close in giant planets seen spectroscopically by the reflex motion of their star. Did the other ~25% form systems like our own, or form something very different - we do not yet know, and direct observations seem to be the best way to find out.

Another reason to examine entire systems is that the transport of volatiles into the inner parts of a planetary system is believed to be strongly affected by giant planets, and the very collision process that gives rise to terrestrial planets may be caused by the disturbance of orbits by giant planets, (Lunine, Owen,& Borwn 2000). Our Earth, its volatiles and its living organisms arose out of the same system that created the other planets

Indirect observations are slow and insensitive. The only planet in our solar system that there would be any opportunity to detect by indirect spectroscopic observations is Jupiter. And from astrometric observations, there would be a long slow perturbation from Neptune with a 164 year period that would make it difficult to discover Saturn without about a century of observations.

Table 1. Properties Of The Solar System As Seen By Indirect Observations From 10pc.

Planet	Induced velocity of star meters/sec	Induced radius of star orbit m.arcsec	Period(years)
Earth	0.09	0.3	1
Jupiter	12	503	12
Saturn	2.6	274	30
Uranus	0.2	84	84
Neptune	0.3	156	164

In contrast, table 2 shows that direct observations, even separated by as little as 1 month are likely to show the differential proper motion of planets due to their orbital motion - provided that the planets can be seen. Both NASA and ESA have developed plans to reconnoiter nearby planetary systems for terrestrial planets using nulling interferometry, (Beichman, Lindensmith & Woolf 1999; Fridlund & Karlsson 2000). Recently however NASA JPL has begun a comparison of nulling with alternate options, especially coronography to determine the optimum form for TPF.

Table 2. Potential Direct Observation Of The Solar System From 10pc.

Angular motion around Sun	Arcsec/year seen from 10pc	Arcsec/month seen from 10pc
Venus	0.74	0.062
Earth	0.63	0.052
Jupiter	0.27	0.023
Saturn	0.20	0.017
Uranus	0.14	0.012
Neptune	0.11	0.009

2. Seeing the planets

The difficulty in seeing the planets can best be appreciated by imagining the study of our solar system from a distance of 10pc. The corresponding visual magnitudes would be: The Sun $m_v = 4.7$, Zodiacal Cloud $m_v \sim 22$, Jupiter $m_v = 27.2$ radius from star <0.5 arcsec, Earth $m_v = 29.5$ radius from star <0.1 arcsec.

There are various ways of comprehending these numbers. First, an Earth-like planet is as faint as a Hubble Deep Field galaxy. And second, its distance from its parent star is about the width of the HDF image of one of those galaxies. The planet's parent star is about 4000 times brighter than the brightest star in the Hubble Deep Field!

In the infrared, in the 10 micron region where Earth is brightest, the contrast ratio is improved by about 1000. However, there is also the background from the glow of zodiacal dust in both the solar system and the external system. The dust glow/planet ratio is about the same in the visible as in the infrared, but in the visible the angular resolution to spread the dust radiation thinly can be obtained with a physically smaller device.

We may also consider the problem in terms of photons. Within the complete visible plus near-UV spectrum, and without any instrumental losses one square meter of aperture would collect 1 photon every 7 seconds from an earth-like planet at 10pc. If we wanted to observe the Oxygen A-band, we would need a spectral resolving power of ~200, and we would have instrumental losses of ~ a factor 10. So within that band we would collect 1 photoelectron every ~4 hours for each square meter of aperture. With an 8-meter aperture there would be one photoelectron every 5 minutes, as compared with over 10 million per second from the star.

In the infrared the planet would give 17 times as many photons, and the ozone band is broader and stronger than the oxygen A band so that a resolving power of 20 would be adequate. Within that narrow band and again assuming 10% efficiency we would receive 1 photoelectron every 1.6 seconds. Now the star would be giving somewhat fewer than 1million per second, but there would also be energy collected from our zodiacal dust cloud and the zodiacal cloud of the external system. Both tasks are daunting!

Table 3. Counting rate for an 8m mirror within an oxygen/ozone spectral feature for an Earth-like planet at 10pc.

	Planet/sec	Star/sec
Visible	0.0036	14,000,000
Mid-IR	0.6	6,000,000

3. Choices of Technology

There are two main ways of separating the radiation of the star from the planet. The first of these is "coronography" developed by Lyot (1930) who used it to detect the solar corona out of eclipse. The second is an interferometric technique, "nulling" invented by Bracewell (1978), and owes its concept to radio astronomy.

3.1. coronography

In coronography, the telescope aperture must be large enough to resolve the planet from the star. This is a more difficult condition than is normally considered in resolution. The "feet" of a star diffraction pattern are normally huge compared to a planet signal, and even at 5 times the standard "resolution of the telescope," the rings are still about 10^{-4} of the Airy peak of the pattern. "Apodization" can remove the feet of the diffraction pattern. In this technique the aperture has a pupil plane mask, graded down in transmission towards the aperture edge. This both suppresses the "feet," and broadens the core of the star and planet patterns. The angular resolution is degraded by a factor a of about 3. In a distance ~5 normal resolution widths, and with a loss of about a factor 3-4 in transmission it is possible - at least in principal - to reduce the diffracted light enough to permit the star and planet images to be separated. With even greater transmission losses, the separation may be improved even more.

The second problem is scattering either by phase irregularities or amplitude irregularities. If the rms phase error of the diffraction limited optic is σ, the fraction of starlight scattered is $(2\pi\sigma/\lambda)^2$. This light is spread over an area around the star that varies with the correlation length of the phase errors. If a square adaptive optic with n by n independent actuated segments is used, this length is $1/n\surd D$. The light is mainly spread over $(\mathrm{n/a})^2$ areas each as large as the apodized diffraction core. However, if all phase error correlation over larger lengths can be eliminated, there is in principle a "black hole" that appears around the star image (Malbet, Yu & Shao 1995). Its depth is set by the zero-order ghosts of the AO correction - that is if we imagine the adaptive optic to be a sort of diffraction grating. If we ignore the black hole effect at first, the fraction of starlight appearing on a planet image core will be $(2a\pi\sigma/n\lambda)^2$. The necessary values of σ and n are coupled by the structure function that describes the AO performance and the quality of optical polish. But if for example $\sigma = \lambda/3000$, a = 3, the n to make the scattered starlight less than the planet signal we need $n^2 = 400,000$. This number could be reduced by the "black hole" factor, perhaps by a factor 10. One perhaps minimum estimate of n is that if the surface without active correction has an rms error of λ/j, and we need λ/k, then the number of corrective elements n^2, is at least as large as $(2k/j)^2$, and could be considerably larger.

There is also scattering both by obscuration and by amplitude errors. Obscuration will occur at the edges of mirror segments if the primary is segmented, and in spiders to hold the secondary mirror. The amount of scattered energy is equal to the fractional area that is obscured. Some of this can be removed by pupil-plane masking. The energy removed by scattering is proportional to the reflectivity errors. Again this is spread over a dimension set by the correlation length, but this time since there is as yet no device to correct for amplitude irregularities we must live with the resulting scatter. The amount is something to be determined in practice. If for example there are reflectivity fluctuations of ±0.1% rms, then the corresponding scattered light is $2.5\ 10^{-7}$. Even if this light is scattered over as many as 100 resolution elements it will be somewhat brighter than the planet. Phase and amplitude terms will interact as in scintillation, where phase errors in the upper atmosphere result in amplitude variations at the ground.

If there is any permanent leaked signal from starlight in the focal plane, it may be removed (but not its noise) by rotating the telescope while keeping it pointed at the star. The planet maintains its position angle, while in a 360° rotation, the azimuthal scattered light becomes averaged in all directions relative to the planet. Indeed if the signal from a rotated focal plane is then further examined for radial errors, it is also possible to correct for the radial component of scattered radiation. If the scattered light is fairly smooth, the amount of telescope area needed to accomplish a given observation is proportional to $\surd$(1 + scattered light/planet light). Table 4 gives an example. Part of the difficulty of designing a coronagraph is that this number N of table 4 must be predicted in advance, and demonstrated on the ground. Then there is also a question whether it can be maintained in flight.

Table 4. N is the ratio of scattered light on the planet image divided by planet light. It shows a particular analysis of the required mirror area for looking for the oxygen A-band in an earth-like planet at 16.5pc

N	Square meters at at unit efficiency	Square Meters at 0.072 efficiency	Equivalent Primary mirror diameter
1	3.6	50	8m
2	4.3	60	8.7m
5	7.0	97	11.1m
10	11.9	165	14.5m
20	22.0	306	19.7m

A strange feature of coronagraph design is that the optical performance can be reduced if the telescope is made larger! The amount of scattered starlight falling within a planet image core will decrease as 1/(telescope mirror area) for a given optical quality, and so if the mirror is huge, the optical quality can be relaxed. This makes the coronagraph the natural device for use with giant low precision gossamer optic telescopes to study planet signals in great detail both in the visible and the infrared, after they have been discovered.

Ground based coronagraphs have been used, but to date there is not much experience of coupling them to adaptive optics. Analysis by Stahl & Sandler (1997) suggest that at visible wavelengths, seeing will be the dominant problem, and feed-forward will be needed to deal with the lag between measuring the seeing and correcting for it.

3.2. nulling

The alternative process is nulling in which starlight, collected from two separate apertures or two parts of one aperture, is interfered with a 180 ° phase shift. The phase shift will be different for light coming in at different angles, and so the result is to place a series of parallel zones of transmission and rejection on the sky, with a rejection zone through the star. The angular spacing of these zones varies with wavelength, and so it is essential to collect the radiation with spectrographic equipment.

Now the device is rotated while keeping the star light rejected. As the transmission and rejection bands rotate, they modulate the radiation from every other object so as to encode information about the object's position angle and radial distance from the star. So when the signals are collected as a function of position angle and wavelength, the position of every bright source and its spectrum can be obtained by analysis of the one data set. If more than two apertures are used it is possible to make the transmission bands asymmetric around the star and to see the radiation from the two sides alternately so as to permit phase sensitive detection. An additional use is to broaden the rejection zone so as to exclude the disk of a star (because its finite diameter prevents the starlight all being nulled).

It is possible to use nulling within the diffraction core of the individual telescope apertures, and so the angular resolution is set by the spacing of apertures, not by their size. This is helpful, because if we have two 1meter apertures separated by 10meters, the angular resolution of the apertures at a wavelength of 10μm is 2.4 arcsec but the central low transmission is separated from its closest high transmission by 0.1 arcsec. This becomes the resolution in the reconstructed image.

Hinz et al (1998) have made a demonstration of IR nulling (at 10μm) on the old MMT, and has made a cryogenic nulling device now in use on the 6.5m MMT. In the laboratory it has produced a null $\sim 10^{-4}$. Serabyn et al(1999)also demonstrated nulling in the visible, and have shown that it is possible to preserve a null depth for a long period.

The precision of the null is set by the precision of phase and amplitude matching - just as it is in the coronagraph. But it is not worth reducing the star radiation substantially below the zodiacal radiation collected by the telescope. If the solar system is typical, we will find the external zodiacal dust will become about 200 times as bright as an Earth-like planet. On the other hand it will look like a disk only 0.2 arcseconds across, and so for interferometer apertures smaller than 8m for a space device at 1 AU from the sun, the local zodiacal emission will dominate. If we set a requirement that the star null must reduce starlight to 1/2 local zodiacal emission we can examine the nulling factor needed. This calculation assumes interferometers with 4 apertures each of the size shown on the left in table 5.

Table 5. Nulling factors to examine a system at 8pc.

Apertures	7μ	8μ	9μ	10μ	12μ	14μ	20μ
0.85m	64,000	28,000	15,000	8,700	4,000	2,300	1,100
4m	$1.4\ 10^6$	600,000	340,000	190,000	90,000	50,000	24,000

In table 5 we see that the nulling need increases markedly towards short wavelengths. If a planet were very bright at short wavelengths, then such large nulling factors are not needed, but because the thermal emission from both earth-like planets and dust both decrease towards short wavelengths, the nulling increase is needed. Also, small telescopes do not require such good nulling. This is opposite to the trend we saw with the coronagraph where the larger devices can have relaxed tolerances. It tends to make the nuller the preferred discovery device, and the coronagraph the appropriate device for follow-up missions.

The nuller will use the short wavelength radiation of the star to measure and control the phase. And perhaps shorter wavelengths yet will be used to control the image positioning. Thus in general we will expect nulling mirrors to be diffraction limited near 1μm, while they operate in the 10μm spectral region. Although it has been suggested that spatial filtering will be needed for beam combining in the infrared, the high Strehl ratio that will automatically accompany a device that adequately controls pointing and phase is likely to make this unnecessary for a precursor mission. But spatial filtering will definitely be needed if nulling is pushed to high precision.

Infrared nulling will be accomplished on the ground at the LBT and Keck interferometers. Those studies will discover the strength of exozodiacal radiation around nearby main sequence stars, and will probably also result in direct detection of some shorter period giant planets.

It is possible to get a comparison of the dimensions of a coronagraph and a nulling interferometer for the same spectral region. The nuller will probably need to have at least 4 elements, so as to properly suppress a star disk, and if its linear dimension is L between centers of the outer elements, then $L = \lambda/\theta$. The coronagraph will probably have a resolution of ~5 times poorer than a non-apodized round aperture, and so will require a mirror diameter of $D = 6\lambda/\theta$. If we imagine that a device is required to observe planets at 15pc, and it must operate to 1.25 microns for short wavelengths and 20 microns for long wavelengths, with a resolution of 1/15 arcsec, then the sizes are ~60m for a nuller, and ~23m for a coronagraph.

4. Providing future technical expertise

The development of technical expert astronomers who will be able to implement the new technology of planet detection is a major concern. Traditionally, astronomy departments have been very poor at generating the technical expertise required for new instrumentation, and there has been a reliance on physics departments to do that. It is not obvious whether that mechanism will remain. It may well be that we are not requiring new devices to take advantage of new or different physical principles. It is quite likely that we are instead requiring them to build on earlier technological advances.

Technology is becoming increasingly more expensive, and gets focussed on performing tasks for which there is either a large market, or customers with very large resources. Astronomy fits neither of these categories. And so we are requiring a mechanism to generate new astronomical technologists who are clever at using developed technologies in different ways than was originally conceived. In the past such people have been trained by closely working with researchers of that same kind. University faculties operating on publish or perish are now either discarding such people, or moving them out of tenure track positions so that they do not teach.

5. Acknowledgments

This work has been supported by NASA through Jet Propulsion Laboratory contracts to study nulling (contract 961286), through a contract to Lockheed Martin to study Terrestrial Planet Finder Architectures (contract HC01N2001F), and by a contract from the NASA Institute for Advanced Concepts to explore very large optical systems for the study of extrasolar planets (contract 760036).

References

Angel, J.R.P., & Woolf, N.J. 1998, in ASP Conf. Series 133,172. ed E. Smith and Anuradha Koreans (San Francisco: ASP) 172

Beichman, C.A., Lindensmith, C., & Woolf, N.J. 1999, The Terrestrial Planet Finder NASA, JPL publications 99-3 Pasadena CA

Bracewell, R.N. 1978, Nature 274, 780

Fridlund, M., & Karlsson A. 2000, DARWIN the InfraRed Space Interferometer ESA-SCI(2000)12 European Space Agency

Hinz, P., Angel, R., Hoffmann, W., McCarthy, D., McGuire, P., Cheselka, M., Hora, J., & Woolf, N. 1998, Nature, 395, 251

Lunine, J.I., Owen, T.C., & Borwn R.H. 2000, in Protostars and Planets IV (ed. V. Mannings, A. Boss, and S. Russell), Tucson: University of Arizona Press, 1055

Lyot, B. 1930, C. R. Acad. Sci., 191, 834

Malbet, F., Yu, J. W., & Shao, M., 1995, PASP, 107, 386

Serabyn, E., Wallace, J. K., Hardy, G. J., Schmidtlin, E.G.H., & Nguyen H.T. 1999, Appl.Optics, 38, 7128

Stahl, S. & Sandler, D.E. 1997, ApJ, 454, L153