Annu. Rev. Astron. Astrophys. 1998. 36:507–37

ASTRONOMICAL SEARCHES FOR EARTH-LIKE PLANETS AND SIGNS OF LIFE

Neville Woolf and J. Roger Angel
Steward Observatory, The University of Arizona, Tucson, Arizona 85721;
e-mail: nwoolf@as.arizona.edu

KEY WORDS: extrasolar, terrestrial, exobiology, origins

ABSTRACT

If Earth-like planets orbit nearby stars, they could be detectable with specially designed telescopes. Direct observations would be very revealing, particularly low resolution infrared spectra, which could establish habitability on the basis of temperature and atmospheric water. Abundant, primitive life based on organized molecular structure might reveal itself, as on Earth, by an atmospheric composition modified in ways unlikely to be from inorganic processes. The technical challenge is to detect and obtain spectra of an object with $M_{bol} \sim 28$ that is very close to a star and some 5×10^9 times less luminous. Indirect methods, used to detect Jupiter-mass planets, do not seem to offer an easy intermediate step to finding Earth-like planets. However, the direct detection techniques needed for spectroscopy also offer a viable method for discovery by imaging. Thermal infrared wavelengths, in which a planet emits most energy, are the most favorable. A robust search for planets of ∼100 nearby solar-type stars, with spectroscopic follow-up of Earth-like candidates, could be made with an interferometer ∼75 m in length. In visible light, the Next Generation Space Telescope (NGST) could, with the addition of a high resolution correction instrument, see Earth-like planets around a dozen or so of the nearest stars. Both infrared and optical instruments are possible within the range of current space agency plans.

1. INTRODUCTION

Stellar radial velocity measurements have shown that some nearby stars and pulsars are orbited by planets. In this volume, Marcy & Butler (1998) review

0066-4146/98/0915-0507$08.00

the evidence for giant planets in close orbits. Such planets are relatively common and probably exist around ∼5% of F, G, and K stars, with a preference for the earlier spectral type (F and G) of stars so far examined. Long-term acceleration measurements indicate that about as many large planets are in long-period orbits as are in short ones. Two ordinary stars, ρ^1 Cancri and Lalande 21185, show some evidence for the presence of more than one planet. Present indirect detection methods do not have the sensitivity to detect planets of the mass of Earth around normal stars, but Earth-mass planets are found around some pulsars (Wolszscan 1997). For one pulsar with two planets, the evidence suggests near coplanar orbits. An excellent complete and updated bibliography on the Internet about extrasolar planets, with special emphasis on terrestrial planets, is maintained by Jean Schneider at http://www.obspm.fr/department/darc/planets/encycl.html.

The small fraction of the Sun's energy intercepted by the Earth, $1/4(r_E/r)^2 = 4.5 \times 10^{-10} L_{\odot}$, is reemitted as a source with absolute bolometric magnitude $M_{bol} = 28.1$. Most of this energy appears as thermal radiation with temperature ∼270 K. The technical leap needed to detect a similar planet of even the closest solar-type stars at ∼5 pc requires the ability to deal with extreme contrast and dynamic range. Already the Hubble Space Telescope (HST) and the largest ground-based telescopes have the sensitivity to detect and study objects this faint. The reflected component from 0.4–3 μm would be at a level of ∼40 nJy (nano-Janskys), similar in flux to faint galaxies at a redshift of $z = 2.5$ identified in the Hubble Deep Field. However, the illuminating star would appear to be only 0.2 arcsec away and approaching 10^{10} times brighter than the planet. Only with large and extremely well-corrected optics in space could this difficulty be overcome. The contrast is more favorable in the mid infrared, where the planet is "only" 10^7 times fainter than the star, but here the angular resolution of a telescope is far poorer because of the long wavelength. Interferometric techniques can be used to suppress stellar emission, but as we discuss below, diffuse emission from zodiacal emission from dust in our system and that of the system observed is not similarly suppressed and presents special difficulties for infrared detection. Despite these problems, observations including spectroscopy should be possible with technology now being developed. That ability opens the possibility of exploring atmospheric composition and whether the atmosphere of a planet has been modified by the presence of life.

Until now, the only way to discover life outside the Solar System has been to look for radio signals from alien civilizations. Searches for life in this way were proposed (Cocconi & Morrison 1959), and the topic developed further (Drake 1965). Searches made to date have been summarized (Tartar 1997). No alien civilizations have been discovered. We should note that monochromatic radio signals have been characteristic of Earth for only $\sim 10^{-8}$ of its history, and we are unable to predict how long they will persist. By contrast, Earth has been inhabited by simple life forms for at least 80% of its history, and spectroscopic evidence of that life has probably been present for about 50% of its history. The search for similar evidence in other nearby planetary systems is a development with great promise and is the main theme of this review. We consider the radiation flux that would be expected from Earth-like planets, the potential for indirect and direct searches in the future, their interpretation, and the technological developments required for them.

2. PLANETS SEEN FROM A LARGE DISTANCE

In this section, we consider the characteristics of planets of other stars that might be investigated by direct methods, from the crudest measurements at poor signal-to-noise ratio to higher resolution spectroscopy.

During the initial discovery phases, even the spatial information may be ambiguous and perhaps so poor as to confuse planetary and zodiacal emission. The crudest information might be simply the presence of radiation that is not of stellar origin, whose planetary origin might only become apparent with the detection of periodic changes caused by orbital motion. Better imaging might then show directly the presence and motion of one or more planets, in specific orbits. Since the system distance is known, the true orbital separation could be determined, as with double star observations. Once the orbit and stellar luminosity are known, even crude, broadband photometric measurements of the thermal and reflected starlight components could be used to determine the approximate size and albedo of the planet. If the thermal emission spectrum can be measured at a few wavelengths, a color temperature could be determined for a better estimate of an effective temperature and diameter. Measurements over time could be used to search for photometric or polarimetric periodic fluctuation of the planet and to determine the rotation period.

2.1 *Crude Spectral Classification*

The next level of information, crude spectrophotometry, with resolving power $\sim$20, would be a byproduct from interferometric imaging in the thermal infrared. By analogy with the Solar System, with modest signal-to-noise ratio at this resolution, we might expect to be able to distinguish five spectral types classed according to chemical composition (Table 1). Four are classes of compact sources, and one (listed second) includes the extended emission of zodiacal dust and comet tails.

2.2 *Spectra of Terrestrial Planets and the Earth*

The infrared spectra of the Sun's terrestrial planets (Hanel et al 1975) are characterized by the strong absorption feature of carbon dioxide at 15 μm,

Table 1 Spectral classification of planets and dust clouds

Classification	Rocky	Dusty	Giant	Icy	Terrestrial
Spectral features	Continuum	Continuum and silicate emission	Methane absorption	Ice band absorption	CO^2 absorption (15 μm)
Solar System examples	Mercury, Moon, Io, most asteroids	Comet tails, zodiacal dust	Jupiter, Uranus	Jovian moons, Kuiper belt objects, comets	Venus, Earth, Mars

while only the Earth shows in addition water and ozone, at 9.7 μm (Figure 1). Our special interest is in finding new terrestrial planets with liquid water at their surface and hence in their atmospheres. We take this to define a habitable planet (Kasting 1996).

Considering their spectra from millimeter wavelengths on down, such planets' atmospheres will be opaque down to $\sim$20 μm because of the strong water vapor rotational transitions. The bands of other molecules in that spectral region would reveal only the small amount of gas in the upper atmosphere. The best observations of atmospheric constituents will be obtained in the region from 7.5–20 μm, where the water vapor is semitransparent, and a planet surface at $\sim$300 K has its peak blackbody emission. Analysis of the composition of the atmosphere of another Earth-like planet will probably start with strong bands in this spectral region.

We can infer atmospheric pressure on a wet planet even if the main atmospheric constituent is not spectroscopically apparent. The strength of the broad features is set by the wings of individual band lines, which are controlled by pressure broadening. The maximum possible amount of water causing a spectral feature is set by saturation vapor pressure, which depends on temperature. Thus, from the measured temperature and line profile, the minimum pressure of the broadening gas can be inferred.

We can also measure the strength of the greenhouse effect and the upper atmosphere thermal structure. The effective temperature in the 15-μm CO_2-band center is characteristic of the tropopause and stratosphere, while the out-of-band emission gives the temperature of the layer where the planet becomes opaque. This layer may be either the surface or a cloud deck. Distinguishing between these may be possible if the cloud deck is time variable, as it is on Earth.

The reflection spectrum, though informative about Earth's atmosphere (Sagan et al 1993), will be hard to observe for extrasolar planets. From 4.2 to $\sim$2 μm, it will be quite weak because of the weakness of the stellar emission in this

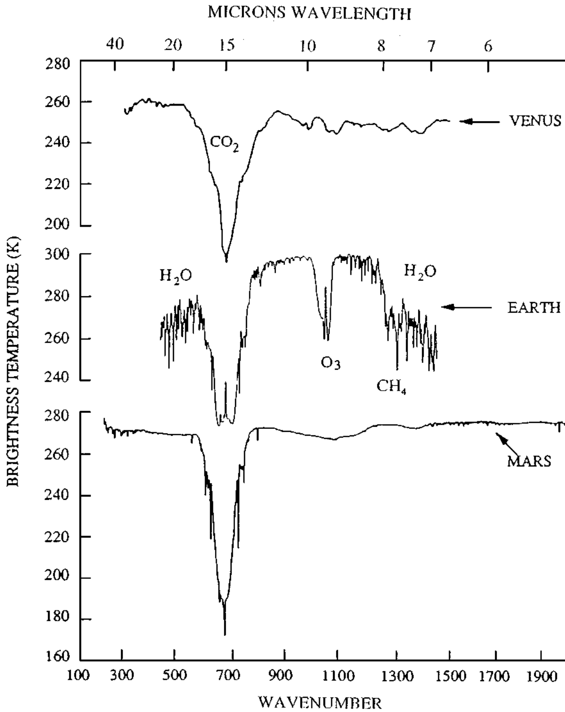

Figure 1 The mid-infrared spectra of the Solar System terrestrial planets adapted from Hanel et al (1975).

region. The most accessible region will be the near-infrared, visible, and near-ultraviolet region where the reflected spectrum of the star peaks. In this region, the overtones of the rotation vibration bands of gases will be useful for observing weaker features, as discussed above. On Earth, the most obvious feature that adds new information is the A band of oxygen at 0.76 μm, but if the atmosphere is warm and wet, the weaker near-infrared water bands may be a better measure of the water amount than the strong bands of the mid infrared. In the blue and ultraviolet regions of the reflection spectrum, strong scattering by the molecules of the atmosphere occurs on Earth, and relatively little new spectroscopic information is available.

2.3 *Studies Needing Higher Resolution*

The determination of the absolute quantities of gases with higher precision must wait for high resolution observations in the reflection spectrum, in order to find individual lines that are not saturated. High spectral resolution observations will also be needed to determine the amount of trace gases. On Earth, the trace gases of the atmosphere include CH_4 (7.65 μm), NO_2 (7.8 μm), and additional simple organic molecules. There are some possibilities of observing molecules in a relatively poor atmospheric window from 4.5 to 5.5 μm, between the 6.3-μm water band and the 4.2-μm band of carbon dioxide.

2.4 *Terrestrial Variants that Might Be Revealed by Spectroscopy*

The Solar System gives us a snapshot of planets after 5 billion years of evolution. From our understanding of this evolution, we can expect to find planets of other stars at earlier or later stages, formed with different masses and different amounts of stellar warming.

Water is seen in the atmospheres of cool giant stars (Kuiper 1963, Woolf et al 1964) and the brown dwarf Gliese 229B (Oppenheimer et al 1995). We can expect it to be generally abundant in the atmosphere of any object warmer than Mars, such as Earth. However, through evolutionary brightening of the star, sooner or later a runaway greenhouse effect and loss of water will occur (this is the Venus scenario). The loss of water from Venus is demonstrated by the unusually high deuterium/hydrogen ratio (Donahue et al 1982). We possibly could see a Venus before or during the water loss phase, with lots of water and photodissociated oxygen. However, this phase is estimated to have lasted only $\sim$20 million years (Kasting 1996); thus, only $\sim$0.5% of the life of Venus was spent in this state.

High mass terrestrial planets, or those with much younger age than the Solar System, might show carbon in the form of CH_4 in a mainly hydrogen atmosphere, with oxygen mainly in the form of H_2O. These are the dominant spectral features of giant planets and brown dwarfs from $\sim$1000 to 100 K. If a planet starting with such an atmosphere lost its hydrogen, methane would not survive long against ultraviolet decomposition, and a primarily CO_2 atmosphere would come into existence, still containing water. This is another possible phase that might be detected.

Another possibility is an earlier Mars, a planet with less stellar warming than the Earth. The small amount of water in Mars' atmosphere (Kaplan et al 1964) is a consequence of the coldness of the planet, and so most water must be in a solid state. We have evidence that Mars once had liquid water on its surface, presumably because of the presence of an additional greenhouse gas, perhaps CH_4 (Kasting 1996). Again, in a more massive planet, there would have been a potential for a stronger greenhouse effect and persistence of liquid water for a long period.

3. WHAT IS EVIDENCE OF LIFE?

3.1 *Atmospheric Changes Caused by Life*

Primitive life based on organized molecular structures could be evident from spectroscopic signatures of a planet, even if its chemistry is different from known life. We can expect that evolutionary pressures will force biological organisms to become proficient at causing chemical reactions. If life develops to a scale limited only by raw materials, as occurred on Earth, the result is that reactive chemicals come to be present in proportions that imply their constant large-scale production (Hitchcock & Lovelock 1967, Lovelock 1975). That is, for dominantly photosynthetic life, the negative entropy of the nearby stellar radiation field gets transformed into a negative entropy of living organism chemistry. The process of building the organisms leads to an excess of the waste products of the process. Then the destruction of the organisms upon death can give a different excess of death waste products. Lovelock's work initiated the consideration of atmospheres to explore the presence of life on a planet.

3.2 *The Spectroscopic Evidence of Life on Earth*

In Earth's atmosphere, oxygen was produced and is maintained by the presence of life. Its source is photosynthesis, the photodissociation of water and carbon dioxide by living organisms (Berkener & Marshall 1964).

$$nCO_2 + nH_2O + 2n \quad \text{photons} > (H_2CO)_n + nO_2, \tag{1}$$

where $(H_2CO)_n$ represents sugars and starches. Oxygen is highly reactive, and so in the absence of life, it is expected, without the presence of life, to be lost to combination with surface materials (Broecker & Teng 1982). However, the continuous production of oxygen has resulted in it becoming 20% of the atmosphere. Owen raised the possibility of using the 760-nm band of oxygen as a spectroscopic tracer of life on another planet (Owen 1980). The most prominent spectroscopic evidence of oxygen in the Earth's spectrum is the 9.7-μm band of ozone. This feature was pointed out as a tracer of oxygen and hence life (Angel et al 1986). Ozone will be in equilibrium with oxygen (Léger et al 1993), and its presence also implies a continuous source of oxygen production.

A second indicator of life on Earth is the presence of methane. Just as with oxygen, it is continuously being destroyed, in this case by interaction with atmospheric oxygen. Its presence implies a steady production mechanism, which is the anaerobic decomposition of organic matter. On Earth, the prime sources of production are by bacteria in termites and in the stomachs of grass eaters. Methane is much rarer than oxygen on Earth, and higher sensitivity and spectral resolution is needed to observe it in the spectrum (Figure 1). Spectral absorption by chlorophyll-like substances that trap the photons for photosynthesis is also seen (Sagan et al 1993).

3.3 *Variants in a Planet with Life We Might Expect to See by Spectroscopy*

There is also a potential for seeing life at various stages of evolution and on planets of different mass.

EARLIER EARTH Earth seems to have spent about half its life with rather little oxygen in its atmosphere, and half with an observable amount (Kasting 1996). Ozone observations can detect trace abundances of oxygen. Kasting has described how ozone abundance changes with time at various atmospheric heights. The associated band strengthening needs investigation. Large amounts of oxygen are best revealed by the 0.76-μm band. In all, there is a potential for finding interesting variants for the infrared spectra of Earth-like planets at different developmental stages. The earliest stage of life is currently believed to have been that of archaebacteria, in which planetary heat outflow was responsible for providing the negative entropy and in which conversion of carbon dioxide to methane was by intake of hydrogen, which was released by the interaction of water with rock magma. Atmospheric trace constituents would potentially show evidence of this era.

BIGGER EARTH An interesting possibility remains that a planet moderately larger than Earth might retain a more primitive atmosphere. It would be intermediate in character between a terrestrial and a giant planet. That is, there might be liquid water at an ocean-atmosphere interface, but the atmosphere could still contain some hydrogen with methane and water vapor. The evidence of liquid water would be the agreement between the surface temperature of the planet and the associated saturation vapor pressure with the amount of water seen in its atmosphere being near saturation vapor pressure. Such a planet would also have claims to being Earth-like, or at least like primitive Earth. If life developed there, it might be intriguingly different from life here.

3.4 *Ambiguity of Spectral Evidence Elsewhere*

While the presence of most ozone on Earth is attributable to life processes, might oxygen be produced on another planet without life, by direct photodissociation of water by the stellar flux in the ultraviolet? On Earth, this reaction is limited because the low temperature and water vapor pressure of the tropopause prevent much water from penetrating to the stratosphere, while solar ultraviolet cannot penetrate deeper. The consequence of a dry stratosphere is that most ultraviolet photons are absorbed by oxygen and ozone molecules before having a chance to dissociate water molecules. Only about 10^{-6} of the oxygen on Earth is produced inorganically (Walker 1977), and so on a lifeless planet, the temperature would continuously decline with height.

For a planet that is so cold that it is covered with ice, a small amount of ozone will be produced by decomposition of surface ice by the star's ultraviolet radiation, but such an object would be distinguished by its low temperature and the weak water absorption features. It has been argued that oxygen might be produced by photodissociation of water in comets, and cometary infall would have brought some oxygen to Earth (Noll et al 1997). However, rock oxidation shows that early Earth had negligible amounts of free oxygen, and that was when comet infall would have been most frequent (A Léger, M Ollivier, K Altwegg, N Woolf, in preparation).

We must recognize that the signature of life on Earth is not just ozone. A similar exoplanet would be identified by a similar (luminosity-scaled) distance from its star, a similar amount of atmospheric water vapor, and a temperature inversion above its tropopause, indicated by the shape of the bottom of its 15-μm CO_2 band. Then the presence of ozone with the provisos given above would be strongly suggestive of the presence of life.

More evidence could come from detection of methane and other rarer molecules out of equilibrium, but this would require higher resolution studies. If abundant ozone is found, it may be thought worthwhile to build a second generation instrument to make such studies and provide clearer evidence (see Section 7.3). Those who require the sight of organisms wiggling and reproducing as necessary evidence of life will have to wait for another era when probes are practicable.

4. INDIRECT DETECTION METHODS

4.1 *Indirect Observations*

In this section, we consider the potential for indirect detection of planets that are as small and with as low mass as the Earth. Our knowledge to date of the planets of other stars has been established indirectly, almost entirely from Doppler measurements of stellar reaction motion. Current sensitivity levels allow detection of only the relatively large motions caused by giant planets. Other indirect techniques developed in the study of binary stars are astrometry, occultation in an eclipsing system, and gravitational lensing. These clearly also have the potential to establish the presence of planets and give information about their mass, orbit, and (from occultation) diameter. Here we consider whether indirect methods, pushed to their intrinsic sensitivity limits, have the sensitivity to detect planets of the Earth's mass and size. We explore fundamental sources of noise rather than engineering difficulties. A discussion of developments in concepts for detection of planets has been given (Woolf 1997).

4.2 *Radial Velocity and Astrometry*

Radial velocity of a star has to be measured in the presence of convection currents on the star's surface. It seems likely that these will limit the precision to no better than ± 1 m/sec, even in older, less active stars (Marcy & Butler 1998). In contrast, the motion of the Sun caused by the Earth has a semiamplitude of 0.1 m/sec, and so detection of similar exoplanets would require precision an order of magnitude better than is achievable.

Astrometric detection has not yet been pushed to its fundamental limits, which on Earth are likely set by Earth's atmosphere. Currently (M Shao, personal communication), the Palomar Test Interferometer has an apparent short-term precision of 250 μas (microarcseconds) in a 5-min exposure, but there are slow variations whose origin is still being explored, and precision of 70 μas in 1 h is expected. The Keck Interferometer is expected to be able to measure to a precision of about 20 μas, and Stellar Interferometry Mission spacecraft to be able to measure to a precision of 1 μas for narrow angle work. Seen from 10-pc distance, the Sun would execute a yearly oscillation of 0.3 μas semiamplitude because of the presence of the Earth, and this would be undetectable.

The ultimate limit to detection of Earth-like planets from space may well be the nonuniformity of illumination over the disk of a star. Earth causes the Sun's center of mass to move with a semiamplitude of about 500 km, i.e. 0.03% of the stellar diameter. Some sunspots have areas up to 1% of the area of the Sun and so can cause the apparent center of light of the Sun to move by as much as 0.5% of the diameter, i.e. 7000 km. Faculae have even larger brightness changes associated with them. For visible/near-infrared observations, some improvement in centroiding may be possible by making two-color observations, but there is likely to be a relatively erratic relationship between color change and center of light motion. Unless other stars are more stable than the Sun, it would seem unlikely that planets of Earth's mass can be detected by astrometry or radial velocity.

4.3 *Occultation (Transit) Method*

For planetary systems oriented by chance with their orbital plane including the line of sight to the Earth, occultations will reveal the presence of planets. This method of detection has the advantage of not needing any special control of pathlengths or optical surfaces, as in imaging or interferometry, and is thus a different way to find planets. Further, the technique appears to be the only practical way for determining the frequency of planets in binary stars. There remains concern, though, that photometric variability, if any greater than for the Sun, will preclude photometry to the accuracy required for planets that are as small as Earth.

The highest sensitivity will be for larger planets in close orbits. Jupiter-sized planets will remove about 1% of a star's light during an occultation. An Earth-sized planet removes only 0.005% and would thus require photometric precision of about 1 part in 10^5 for detection (Borucki et al 1996, 1997). This is not much larger than the intrinsic noise of fluctuation of the Sun's brightness for likely occultation duration. Because potential periods can cover a wide range, there is some concern that even with measurements free from photon noise, occasional random stellar fluctuations at the few-sigma level will be confused with true occultations. Borucki et al estimated, therefore, that regularly repeated transits each at the 7σ level will be required. This sets the intrinsic limit for detection close to the size of Earth-like planets if the star is similar to the Sun in size and variability.

Dedicated spacecraft missions are needed to avoid the photometric limitations imposed by the Earth's atmosphere, which are at about the 10^3 level. One small French satellite mission, COROT, is currently planned to both study astroseismology and search for stellar occultations by large planets (Deleuil et al 1997). Borucki et al suggested that a wide-field meter-aperture camera in space could monitor all stars continuously to $\sim$14-mag limit over a 12-deg field. Since the probability of seeing an occultation is about 1/100 for a planet like the Earth about the Sun, and only about 1/300 of the sky is surveyed, typical stars found would be found $(30{,}000)^{1/3} \sim 30$ times farther away than the nearest examples. Thus, if the closest Earth-like planet is 10 pc away, the nearest Earth-like planet found by this method would be about 300 pc away, too difficult for follow-up by direct detection methods.

The probability of finding planetary occultations would be greatly increased by observing eclipsing binary stars, under the assumption that the orbital plane of planets is likely to be similar to that of the stars. Also, if stars of small diameter are selected, for larger amplitude occultations, a program of sequential pointings by a photometric space telescope could be productive (Schneider & Doyle 1995). The sample of candidate systems is very limited, but if they are adequately stable photometrically, this could be a way to see it they have Earth-mass planets.

Farmer (1997) has suggested that the atmospheric absorption spectra of planets in the reflection-spectrum range would show as a difference between the spectrum of a star after and during a planetary occultation. However, the spectrometric accuracy needed seems beyond reach. The atmosphere of a planet can produce absorption features over a few optical depths; thus when the Earth occults the Sun, in addition to the depression of the Solar continuum by 0.005%, there is an additional $\sim$0.00005% of the continuum that will have absorption features imposed. Even if the stellar spectrum were stable to this level through the occultation, to recover the difference spectrum to a precision of $\pm 3\%$ would require collection of 3×10^{15} photon events during the occultation in each resolution band, an impossibly large number.

4.4 *Gravitational Lensing*

Gravitational lensing might be a possible way to detect Earth-mass planets (Mao & Paczynski 1991). Since then, gravitational lens events caused by stars have been detected, but the detection of smaller mass objects requires very high time resolution observations, with dedicated telescopes for the task. The planet detection process has been reconsidered by Paczynski (1996) and Peale (1997, and S Peale, private communication). Neither is very supportive of the idea that it should be developed for detection of Earth-like planets, though its use for studying the frequency of giant planets around different types of stars is attractive and being attempted. Peale suggested that in every 30 stellar microlensing events, occurring at a roughly 1 per day rate toward the galactic center, one will show the presence of a giant planet. Earth-like planets would be observed between 10 and 100 times more rarely than that, and the detection rate is very model dependent, with most of the events due to planets around M dwarf stars. Thus if Earth-like planets are mainly around warmer stars, we mistakenly might assume that they are absent from the results of this technique.

Peale (1997) suggested that the detection of large numbers of giant planets would be a better indication of a high frequency of Earth-like planets than the detection of such planets themselves. That is, giant planets are considered more difficult to form than terrestrial mass planets. Therefore if giant planets are observed (and have not swept up the smaller planets during giant planet migration), Earth-like planets are to be expected. But expectation is not detection. If Earth-like planets were observed, they would be at a large distance from the Sun, without the possibility of follow-up by direct detection.

5. DIRECT DETECTION

Direct detection methods, in which we distinguish radiation that comes from the planet, seem to offer the best way to discover Earth-like planets. Such methods are also the only way to determine physical characteristics such as temperature and chemical composition through spectroscopy.

We distinguish the goal of direct detection, which simply requires resolving the planet's radiation from that of the star, from the far more ambitious goal of resolving the disk and mapping the surface of an Earth-like planet. This is in principle possible, given a telescope with high enough spatial resolution, and is stated as a long-term goal for the National Aeronautics and Space

Administration (NASA), through its Planet Mapper mission. However, the technical difficulties are stupendous. For example, if the planet is at 10-pc distance, to obtain optical surface detail similar to that obtained by the HST for Mars (~100 pixels across the diameter), an effective aperture a million times HST's is required, i.e. the size of the Moon. Such an aperture could be synthesized by an interferometric array, but fundamental considerations of surface brightness and integration time lead to a minimum requirement for many square kilometers of mirror surface, corrected to a wavefront quality better than the HST. We do not consider resolved imaging further here. For the foreseeable future, the study of extrasolar planets must parallel that of stars, for which essentially all information has been derived from position and spectroscopy, without imaging information.

5.1 *Basic Considerations*

The fundamental challenge for direct detection is to distinguish the planet against the halo of the adjacent and enormously brighter star. From consideration of both angular resolution and signal photon noise, in order to obtain detection or spectroscopy at given signal-to-noise ratio, telescope size must be increased in proportion to system distance. Thus, the problem can be framed only when the distance of the systems to be studied is known, and this in turn depends on the type and number of stars to be sampled, which increases as the cube of distance.

Based on the fact that the Sun is a G star, and that the giant planets found so far seem to favor G stars, a search for Earth-like planets should include a number of stars of type early K through A. These are bright enough to place Earth-temperature planets 0.5 AU or more from the star. We estimate the number of single stars of these types, seen to distances of 5, 10, and 15 pc, in Table 2. Since there are only 4 near–solar-type stars within 5 pc that do not have an intermediate separation companion (α Aql, τ Ceti, ε Eri, ε Ind), we would be lucky to find an Earth-like planet so close. The ~30 stars out to 10 pc would constitute the minimum number for a meaningful survey. A robust

Table 2 Requirements for surveys to different distances

Distance	5 pc	10 pc	15 pc
Number of stars sampled	4	30	100
Angular separation θ_p (arcsec) for planet in 1-AU orbit	0.2	0.1	0.067
Telescope scale size λ/θ_p (m)			
$\lambda = 1\ \mu$m	1	2	3
$\lambda = 10\ \mu$m	10	20	30

survey would include the ∼100 stars found out to 15 pc. Lists of candidates need to be developed and the stars studied in detail.

Our first priority in devising schemes for direct detection is an optical system that will resolve objects at the angular separation appropriate to the distance surveyed, cleanly enough to allow planet detection. Angular separations, θ_p, appropriate to a 1-AU orbit are listed in Table 2, along with the instrument scale sizes, $\lambda\theta_p$. These sizes are approximately equal to the filled aperture diameter for resolving θ_p at the Rayleigh criterion and also the separation of interferometer elements with fringe spacing θ_p.

The criterion for resolving and detecting objects of enormously different intensity needs some discussion. The planet does not have to be brighter than the local halo of scattered starlight, but its detection does require that, at a minimum, the random fluctuations in the halo due to photon noise be smaller than the planet signal. We characterize an optical system by its gain, G, defined as the ratio of the star peak intensity to the star intensity at the image of the planet. Suppose detection with signal-to-noise ratio S is desired. We can show that the gain required at the angular separation of the planet is given by

$$G > \frac{S^2}{A\eta t}\frac{R_*}{R_p^2}, \tag{2}$$

where R_* and R_p are the photon fluxes of the star and planet, respectively; A is the collecting area; η the end-to-end quantum efficiency; and t the integration time. (The collecting area and scale size of the optics are not necessarily coupled, since elongated or separated apertures can provide size scales larger than $(A)^{1/2}$.) As examples, simply to get a feel for the quantities involved, suppose we set as goals signal-to-noise ratio $S = 5$ in 10 h for imaging and $S = 25$ in 10 days for spectroscopy, with $A = 10$ m^2 (about twice the area of HST) and $\eta = 0.25$. With these parameters, $S^2/A\eta t = 1$ h^{-1} m^{-2} for both cases. For the bandwidths given, the minimum required gains for imaging or spectroscopy of different spectral features are as listed in Table 3.

Table 3 Continuum fluxes for an Earth-like planet and Sun-like star[a]

			Distance								
			5 pc			10 pc			15 pc		
Species	λ (μm)	$\Delta\lambda$	F_p	F_*	G	F_p	F_*	G	F_p	F_*	G
O_2	0.80	0.024	70	6.8×10^{11}	1.5×10^8	17	1.7×10^{11}	6×10^8	7.5	7.5×10^{10}	1.4×10^9
O_3	9.7	1.0	4,600	4.6×10^{10}	2,200	1,150	1.1×10^{10}	9,000	512	5×10^9	20,000
CH_4	7.7	0.04	90	1.2×10^9	1.5×10^5	23	3×10^8	6×10^5	10	1.1×10^8	1.4×10^6

[a]Continuum fluxes in photons m^{-2} h^{-1} at the wavelengths and bandwidths listed, for an Earth-like planet (F_p) and Sun-like star (F_*) at the distances tabulated. G is the minimum gain for imaging and spectroscopy, as described in the text.

We note the striking differences in requirements for the reflected light and thermal observations. The scale size at optical wavelengths is 1–2 m, but very high gains are needed, a hundred million and more. As we show below, such gains can only be realized in optical systems that are $\sim$5 times larger that the scale size. For the thermal emission, scale sizes are large, 20 m and more, but the minimum gain requirements of $\sim$10,000 or so are modest. For the highest sensitivity, the gain should be higher. The photon noise background should not be dominated by the stellar halo. Further, stability of gain against pointing errors will be important to avoid systematic errors. In practice, we must aim for gains much higher than those in Table 3.

To understand how these requirements for imaging at very high levels of contrast can be met, it is helpful to recall the fundamentals of light propagation, as given by the Huygens-Fresnel principle (Born & Wolf 1980). The probability that a photon emitted by the star will be detected at some point at the detector is given by the square of the sum of phase vectors over all possible paths. Normally, we are concerned that for each star, a small area exists in the telescope's focal plane where the probability is high, "the image," with all the phase vectors through the telescope aperture adding to make a "whopping vector," in Feynman's words (Feynman 1985). For planet detection, we add an additional constraint: At the same region on the detector where the phase vectors from a planet add coherently, those from the nearby star must cancel nearly perfectly. The challenge is having the phase vectors sum change from a "whopping" to a "weenie" vector, for a very small angular shift in source position.

The requirement for a gain of 10^8 for optical detection is that the star vector be reduced by 10^4 times in amplitude, placing a far more severe constraint than is usual on optical design and fabrication accuracy. Normally, we are not too worried if the peak amplitude is only 90% of its ideal value—this is the normal criterion for diffraction limited imaging and allows for 20% of the energy lost from the diffraction peak. For planet detection, such loss would be a disaster, especially if the lost energy is found near where we want to see a planet millions or billions of times fainter than the star. We consider in this section the constraints set by diffraction in ideal systems, with perfectly smooth optics, and then look at issues of accuracy and practical realizations (Sections 6.2 and 7). Many of the techniques we review were anticipated in an early paper on imaging methods of detecting extrasolar planets by Kenknigh (1977).

5.2 *Star Suppression for Single Aperture Telescopes*

The point-spread function for the usual circular telescope aperture, given by the Fresnel integral, drops only slowly with angle. The gain, given by the ratio of the peak intensity to the average intensity of the Airy rings at angle θ, far from the central peak, is given by

$$G(\theta) = 4\pi^2 (D\theta/\lambda)^3, \tag{3}$$

where D is the telescope diameter. For example, for $\theta = 10\lambda/D, G(\theta) = 5 \times 10^4$. From the minimum G values of Table 2, we find that the diameter required for infrared detection ($D = 10\lambda/\theta_p \sim 100$–300 m) is too large to be practical, while for any realistic diameter, the gain is far too low for optical planet detection.

To reduce the diffracted halo, a technique called apodization is used (Jacquinot & Roizen-Dossier 1964). Strong suppression of the diffraction rings is achieved by tapering off the transmission at the edge of the aperture. For example, reduction at angle $\theta = 6\lambda/D$ by 4 orders of magnitude to an average flux of $\sim 10^{-8}$ was calculated by Jacquinot & Roizen-Dossier for a 40% transmitting mask, with only 20% broadening of the central peak. This approach can be further improved by coronagraphic optics, in which amplitude and/or phase are modified also at a field stop (Roddier & Roddier 1997). It appears that gain of 10^8 at $\theta \geq 5\lambda/D$ and beyond should be achievable. From the scale sizes and gains in Tables 2 and 3, we see that an 8-m telescope with perfect optics and coronagraphic correction has the potential for optical imaging and spectroscopy of Earth-like planets of stars at 5–10 pc, but 8 m is too small an aperture to resolve thermal emission (Angel & Woolf 1998).

To obtain strong suppression closer to the star image, a special form of apodization that reduces the flux in a close-in annular zone is possible (Angel et al 1986). An annular pupil mask has the effect of deepening and broadening the first dark ring of the diffraction pattern. Gain greater than 10^5 is achievable over the range of radii $\theta = 1.7$ to 2.3 λ/D. A complete search for planets, extending to larger radii, would require several images obtained with smaller pupil masks to produce dark rings at larger radii. A telescope designed to sense a 1-AU planet at 5 pc at 10-μm wavelength would require a diameter of 20 m, still uncomfortably large.

Further reduction in the size of a single aperture able to resolve given angular separation (needed for the infrared) can be realized through the use of an interferometric device with beamsplitters incorporated into the focal plane instrumentation. The basic idea is to divide the incoming wavefront (spatially or by amplitude) in two and to superpose these parts with π phase difference to null out the flux from a source along the line of sight. The simplest version, based on the concept of Bracewell (1978; see below), calls for dividing the pupil spatially into two large, side-by-side elliptical sections, with mirrors and a single beamsplitter to accomplish the translation and superposition (Angel 1990). The center-to-center separation between the sections is approximately $D/2$, so that the superposition will become constructive for a planet that is at angular separation λ/D in the direction of the line joining centers. The diffraction-limited images of the telescope will be enlarged to $2\lambda/D$ in this direction because the pupil width has been halved. The signature of a planet at the optimum angle λ/D will thus be a shift by half the resolution width in the residual flux when the star is nulled.

Another form of interferometric suppression is based on a rotation shearing interferometer (Roddier et al 1978). As in a Michelson interferometer, a single beamsplitter is used both to derive two half-intensity but full-sized wavefronts, then to recombine them. Before recombination, one is rotated 180° with respect to the other. The pupil rotation is accompanied by an inversion of the electromagnetic wave, providing the half-wave achromatic retardation needed for nulling. Diner (1990) proposed an implementation in which the 180° rotation is achieved by retroflection at roof mirrors at 90° to each other. Gay & Rabbia (Gay & Rabbia 1996, Gay et al 1997) obtained the rotation with one arm reflected at a plane mirror, the other at a cat's eye retroreflector. The advantage of the rotation shearing method is that amplitude balance and π retardation for star nulling are achieved automatically, and the angular separation for the brightest planet images is the smallest of all single aperture methods, about $0.8\,\lambda/D$. The disadvantage compared to the pupil division method is that planet images are doubled, appearing symmetrically on either side of the star.

A common feature of all interferometric supression methods is that star suppression is limited by the spatial incoherence of the starlight, arising from the finite size of the stellar disk (see Section 6.1 below). These methods are most practical when used with separated smaller elements, as we describe below.

5.3 *Star Suppression with a Two-Element Nulling Interferometer*

The direct detection method devised by Bracewell (1978) is to combine the light from two separated elements by means of a semitransparent beamsplitter. Each path through one aperture is brought to superposition with the corresponding path from the other aperture, with a 180° phase shift, independent of wavelength. The result is detailed cancellation across the common overlapped wavefront. The entire focal plane Airy pattern of an unresolved star on axis is canceled, with all the energy appearing in the complementary second output from the beamsplitter. The cancellation does not depend on the size of the two interferometer elements.

In this configuration, the interference fringes are localized on the plane of the sky, as in radio interferometry, and not in the image planes, as in Young's fringes. The star is located in the central, dark interference fringe, but for any planet lying in a bright fringe, the wavefronts combine constructively. The minimum element separation, s, for constructive interference at angular separation θ_p is given by $\theta_p = \lambda/2s$.

In general, the interferometer will be used with elements whose sizes are much less than their separation. The detected images will thus be much larger than the angular separation of the fringes, and the residual flux from a planet when the star is canceled will be displaced only a small fraction of the image width. A way must be found to distinguish residual flux from a planet from that due to imperfect star cancellation or thermal emission from dust around the star being studied.

To mark this distinction, Bracewell (1978) suggested rotating the interferometer about the line of sight, thus rotating the fringe pattern on the sky while the null fringe remains centered on the star. Any close unresolved companion will show as an intensity modulation of the residual star image, at frequency 2ω, where ω is the rotation frequency. More generally, objects at larger separations from the star will appear as a more complex modulation, with the number of maxima per revolution measuring the separation of the planet and the time of most extended minimum measuring position angle with a 180° ambiguity. This pattern must be deconvolved in order to map the positions of individual planets in a system. Such a deconvolution was proposed in unpublished notes by RN Bracewell and R McPhie and has now been demonstrated in computer simulation (Angel & Woolf 1997a). The Bracewell interferometer thus performs three distinct functions: 1. It nulls the starlight and its associated photon noise; 2. by rotation, it modulates the planet signal to distinguish it from imperfect star cancellation; and 3. it codes the separation and azimuth angle of the planet in frequency and phase.

6. PROBLEMS AND SOLUTIONS FOR DIRECT DETECTION

We have seen that exoplanets will be detectable only if the underlying background from the stellar halo is adequately reduced. In this section, we consider how to minimize the effects of backgrounds from imperfect star cancellation and zodiacal emission, from both the Solar System and the system under observation. For nulling interferometers, the geometry, structure, and orbits to minimize the background are reviewed. The effects of imperfect optical manufacturing and of systematic errors as well as photon noise are also examined.

6.1 *Improved Interferometric Nulls*

Starlight suppression in Bracewell's method is perfect only if the star is unresolved at the baseline used. The wavefronts arriving at the two separated apertures must be identical if they are to cancel each other out completely. In reality, the star has finite angular size, setting a limit to the degree of cancellation. For $\sin^2$ fringes that are zero at the center of the stellar disk and reach the full transmission peak ($T = 1$) at the planet separation angle θ_p, the convolution of transmission across a uniformly illuminated stellar disk of diameter θ_s will result in transmission

$$T = \pi^2(\theta_s/\theta_p)^2/64. \tag{4}$$

This would be negligible for thermal detection of a giant planet with Jupiter's radius and contrast ratio about a star like the Sun (the case studied by Bracewell), but not for the Earth-like case. Here, $\theta_s/\theta_p = 9.3 \times 10^{-3}$, so the best that can be done, with a baseline tailored for the specific angular separation, is a stellar transmission of 1.3×10^{-5}, still 100 times the 10-μm flux of an Earth-like planet. In practice, it will be desirable to search over a range of wavelengths around stars with a considerable range of angular diameters. The spacing of a two-element interferometer would have to be adjustable over the range 2–20 m to achieve even 10^{-5} nulling in a search for planets in 1-AU orbits over a range of distances of 3–15 pc.

Fortunately, this "star leak" problem can be remedied by using interferometers with multiple fixed elements. The first proposal for blocking to a higher power than $\sin^2\theta$ (Angel 1990) used four elements in a diamond shape, combined with multiple beamsplitters for transmission varying as $\sin^4\theta$ near the null. In this case, the leak for an optimally separated interferometer is 10^{-8}, and for a star 4 times closer (larger) than optimum, the cancellation is still to better than 10^{-6}. The longest array dimension is twice that of a Bracewell interferometer with the same fringe spacing. Another two-dimensional solution with five elements is under consideration (Mennesson & Mariotti 1997).

An even more perfect null with transmission varying as θ^6 near the null has been found (Angel & Woolf 1997a). The interferometer, with four colinear elements, is 2.67 times longer than the equivalent Bracewell pair, but the added length has the advantage of yielding sharper fringes. Various configurations of linear interferometers have been discussed that yield deeper and deeper nulls (Woolf & Angel 1997). Interferometers whose individual elements were themselves nulling interferometers have the potential for very high resolution and also allow signal modulation by phase instead of rotation (see Section 6.5 below).

6.2 *Effect of Figure Errors*

In practice, residual figure errors will cause slight random rotations in the stellar phase vectors and prevent their summing to zero. To analyze this effect, consider an optical system, either an apodized single aperture or a nulling interferometer, which, if perfect, would result in zero stellar flux at the planet's image. Let us suppose that in a practical realization, the wavefront phase errors are measured and are actively corrected, but that the correction servo is subject to random errors at scale l, which might correspond to the actuator spacing. The uncorrectable residual errors will be predominantly on this scale l and uncorrelated on larger spatial scales. From the Huygens-Fresnel principle, the gain is found to be approximately uniform out to angular radius l/l and is given by

$$G \sim (D/l\sigma)^2. \quad (5)$$

There is the possibility of using a correction servo with a limited number of actuators to improve gain locally in a region of the focal plane close to the star. Wavefront corrections would be applied to correct the residual phases and amplitudes in just this region. If the wavefront is corrected by n independent variables, it should in principle be possible to null amplitudes over a grid of n points (in round numbers) in the focal plane by the "dark spot" algorithm (Malbet et al 1995). In any event, however, there will be errors; thus, for the present analysis we adopt Equation 5 to derive gain values.

An example of the limiting performance that could be expected for a large space telescope like the Next Generation Space Telescope (NGST) at optical wavelengths is given by Angel & Woolf (1998). Assuming correction by a deformable mirror with $\sim 10^5$ degrees of freedom, we estimate the residual wavefront error is likely to be $\sim$2 nm rms, on a spatial scale of 2 cm, resulting in a gain of 3×10^8 for $D = 6$ m and $\lambda = 0.8\ \mu$m. Observing a twin of the Solar System at 8-pc distance, a space telescope with these properties will have a scattered starlight halo in the form of speckles of apparent luminosity 300 nJy, while the Earth's twin would be at 6 nJy.

Detection should be possible against the photon noise of the halo speckles, provided a way can be found to average out the speckles, which in a snapshot would appear like a cloud of bright planets. Azimuthal averaging would be achieved by rotating the telescope about the line of sight through a full circle, provided the optical aberration and speckle structure remains unchanged. Radial averaging is achieved by observations over a range of wavelengths. The planets will show up as diffraction limited point sources against this smoothed background. If the rotation averaging and wavefront correction can be realized with the NGST, the photon noise limit of the smoothed background would allow detection of the Earth at 8 pc at about 25 standard deviations in a 50,000-sec integration, while study of the oxygen 0.76-μm band would be possible for the closest systems, within 4 pc.

Substantial advances in the practical implementation of wavefront correction and coronagraphy are needed to reach such performance limits. These could be developed and tested by using ground-based telescopes and the HST to image Jupiter-like planets, which would be about 10 times brighter than Earth-like ones and with larger angular separation. The limiting wavefront error for a high resolution adaptive optics system correcting atmospheric aberration is set by photon statistics in the wavefront sensor; the error is about 20 nm rms for a solar-type star at 8 pc, corrected on scale $l = 5$ cm at intervals of 0.5 msec (Angel 1994). The resulting gain for a 6.5-m telescope at 1.2-μm wavelength is $\sim 10^6$. This should be enough to detect Jupiter-sized planets, given an image averaged over $\sim 4 \times 10^7$ independent wavefront corrections made during several hours (Stahl & Sandler 1995). On this same basis, it is conceivable that Earth-like planets, 10^{10} times fainter than the star, could be detected from the ground, but only given a telescope much larger than any now existing. A 10-fold increase in gain would require a telescope of $D \sim 20$ m, equipped with a similarly accurate adaptive optics system. For the HST with $D = 2.4$ m, if the wavefront error were corrected to 1 nm on a scale of 6 cm, the gain at 800-nm wavelength would be 10^7 (Malbet et al 1995). A coronagraph with high gain at $5\lambda/D$ would allow resolution of Jupiter-like planets at 5 AU to 15 pc distance.

For the case of a nulling interferometer, the halo of speckles surrounding the star is of no consequence; all of the imaging information is contained within the central core of the diffraction-limited star image. The information in this core is characterized by two parameters only, amplitude and phase, which can be isolated by passing radiation through a spatial filter, such as a single mode fiber in the focal plane. At any given wavelength, only the phase and amplitude need be adjusted to get a null (Coude de Foresto et al 1997). However, both quantities depend on the details of the image structure entering the fiber and thus will be wavelength dependent. It is not clear that any form of spatial filter will be needed. It may be preferable to record the combined image with an imaging infrared array that adequately resolves the Airy core.

The residual star intensity in the diffraction core from an imperfect wavefront is given by $1/G$ (Equation 3), where σ is now the error in the wavefront differential. For a representative small-element aperture $D = 1.5$ m, a gain of 10^6 would be achieved at 10-μm wavelength for differential wavefront errors of 25 nm rms on scales of 10 cm. The accuracy requirements for lower order aberrations, such as tilt and coma, are more stringent and are most severe for piston phase match between the nulled elements—this must be π out of phase to an accuracy of 10^{-3} radians, or $\sim$1–2 nm rms, and would require adaptive path length control. We imagine that in practice, active control of the two wavefronts to obtain good cancellation will be achieved using measurements of their interference at wavelengths <7 μm.

6.3 *Achromatic Beam Combination*

To maintain a destructive interference spectral range of more than an octave, the beam combining optics for a nulling interferometer must create a phase shift π that is accurate and wavelength independent. Amplitudes also must be matched to 0.1% over the same range. Elegant solutions based on the rotation shearing interferometer discussed in Section 5.2 above have been proposed. Shao & Colavita (1992) showed how the interferometer based on roof prism rotation may be used. In two passes through the beamsplitter, once in reflection and once in transmission, amplitudes are balanced regardless of wavelength dependence of the beamsplitter transmission. Differential rotation of 180° between the two element pupils gives the required π phase shift, also independent of wavelength. The same properties are realized in the recent similar concept for beam combination of Gay & Rabbia (1996) that uses the cat's eye reflector for the rotation shearing. A different beam combiner concept with refractive path length correctors, analogous to achromatic lenses, yields the required phase shift achromatic to an accuracy of better than 10^{-3} radians over bandwidths of 20% (Angel et al 1997). Recently, J Burge (private communication) has found solutions that achieve a full octave of bandwidth with a null of $\sim 10^{-5}$.

6.4 *Solar System Zodiacal Background*

Once stellar emission is effectively canceled, the next strongest source of radiation detected by a space interferometer will be thermal emission of the Solar System zodiacal dust. The flux for angles $>$60 deg from the Sun has been measured by the IRAS and COBE satellites (Good 1994, Reach et al 1995). The photon count rate in the diffraction-limited beam is independent of telescope size, while the signal from a planet increases in proportion to collecting area. To obtain an acceptable signal-to-noise ratio for Earth-like planet detection, Angel et al (1986) proposed the use of a 16-m telescope, for which the zodiacal background signal at 10 μm is 2.3×10^{-5} of a Sun-like star at 10 pc. Bracewell raised the possibility of taking advantage of lower background farther from the Sun for exoplanet searches. The 10-μm brightness is expected to be about 200 times reduced at 5 AU from the Sun, where the density and temperature of zodiacal dust are both lower. The first practical interferometer concept for Earth-like planet detection took advantage of this lower background, allowing the use of small ($\sim$1 m) elements (Léger et al 1996). At 1 AU, these would accept a very large background at $\sim 6 \times 10^{-3}$ of a 10-pc star, adequate for imaging but not for spectroscopy; they are acceptable for both at 5 AU.

6.5 *Zodiacal Dust Around the Star*

If the system under study were a twin of the Solar System seen face-on, by far the strongest source of thermal emission around the star would be from its disk of zodiacal dust, not its planets. At 10 μm, the total flux from the cloud would be about 10^{-4} of that from the star, coming mostly from within 0.4 AU of the star. At a distance of 10 pc, most of the cloud lies within the broad central null pattern of a multielement interferometer, with up to $\sim$90% suppressed

(Angel 1998). Even so, an Earth-like planet at its peak transmission through the interferometer will be 100–300 times dimmer. In principle, large aperture interferometer elements with narrow beams could be used to reject zodiacal emission, but the apertures required for a beam width smaller than the cloud, >20 m, are prohibitive with current technology.

To distinguish a planet, we therefore must rely on sensing the interferometer's characteristic pattern of modulation for a point source, which is much stronger than that of a smooth, diffuse cloud. Spatial fluctuations of the dust could be problematic (Beichman 1998), but in the Solar System, it seems that in the vicinity of Earth, the dust is probably quite smooth, with variations in intensity of more than 1% rarely occurring. The strongest variation seems to be an excess dust tail associated with Earth. The distribution of dust around the closest nearby stars could be studied in reflected light at shorter wavelengths with the NGST, given suitable correcting optics.

Supposing that the zodiacal clouds of other stars are smooth, they could still present a problem if their emission is much brighter than the Solar System. Even if they are at Solar System level, the dominant noise for an interferometer at 5 AU would be photon noise of the exozodiacal cloud. IRAS and unpublished ISO observations show the 10-μm emission from dust around a number of early nearby main sequence stars to be orders of magnitude brighter than for the Solar System (Backman & Paresce 1993). If the cloud levels prove to be typically higher than solar, then an interferometer with small elements that takes advantage of the low local zodiacal background of a 5-AU orbit may not be viable because of increased source noise. Larger apertures may be required to overcome this noise, and operation in a 1-AU orbit may be preferred because local zodiacal background may no longer represent an overwhelming background source.

Thus, before a space interferometer can be designed to observe Earth-like planets, the exozodiacal dust emission that it will see in the mid infrared should be measured. This should be possible with a ground-based Bracewell interferometer, operating in the atmospheric window at 10 μm. The most powerful instrument for this task, with almost the same null pattern and angular response as the space interferometer, will be the Large Binocular Telescope (Woolf & Angel 1995, Angel & Woolf 1997b, Angel 1998). The Keck Interferometer (Shao 1997) will also be used for measurements of the zodiacal cloud component close to the star. Measurements to Solar System levels around stars to 10 pc will take integrations of several hours if systematic noise can be controlled so that telescope background photon noise is dominant. Clearly, zodiacal dust will be an important factor to understand. A workshop has been held to discuss zodiacal dust and its relationship to searching for terrestrial planets (Backman et al 1998).

6.6 *Geometries, Structures, and Orbits for Nulling Interferometry*

The interferometer should ideally maximize the difference in the signal modulation produced by an exozodiacal cloud and planet. The designs reviewed and discussed above all have rotation reflection symmetry about the line of sight, yielding ambiguities in source position. The linear configurations also have the disadvantage that the 2ω signal modulation from an edge-on cloud will be similar to that from a close-in planet. Designs that avoid the linear 180° symmetry but create worse ambiguity have been developed. In Europe, Mennesson & Mariotti (1997) explored a regular pentagonal configuration, in which a point source with a 5ω signal would be clearly distinct from an elliptical cloud with a 10ω signal.

A more powerful configuration that avoids both problems is the double nulling interferometer (Woolf & Angel 1997). Here, two linear nulling arrays of two, three, or four elements each are interdigitated, and their signal is combined with adjustable phase delay. The first constructive fringe is asymmetric and can be shifted according to phase setting from one side of the star to the other. Not only does this remove any imaging ambiguity, it allows the response of the interferometer on the sky to be rapidly modulated without rotation.

We have seen that the optimum size of interferometer elements is dictated by the levels of zodiacal flux both locally and from the system under study. With the HST and SIRTF telescopes as the only cost models for precision space optics, an array with elements much larger than 1 m seemed prohibitively expensive. Recently, new ideas for manufacture of low-cost large optics for the NGST (Angel et al 1997) have encouraged us to rethink this balance. A local interferometer with ~4-m elements would have performance comparable to one with 2-m elements at 5 AU for anticipated exozodiacal signals. The pros and cons for these options have been discussed (Woolf 1998). Figure 2 shows how the necessary interferometer mirror size changes between a 1- and 5-AU device for different spectral resolution and star distances.

An industrial study by three US companies focused primarily on structures, and deployment has been just completed. Concepts reveal a variety of options for linking telescopes with a truss, for having small mirrors at 5 AU, and for having larger monolithic or segmented mirrors at 1 AU (Figure 3).

The value of a weak mechanical link between the interferometer elements is still to be evaluated. Free-flyer versions are being considered in both the United States and Europe. Each element would be an independent spacecraft, and precision station keeping would be used to point and change the array orientation and configuration. A potential drawback is the demand on propellant for motions to map out the U-V plane and give adequate signal modulation. A solution studied by J-M Mariotti (private communication) has a six-element

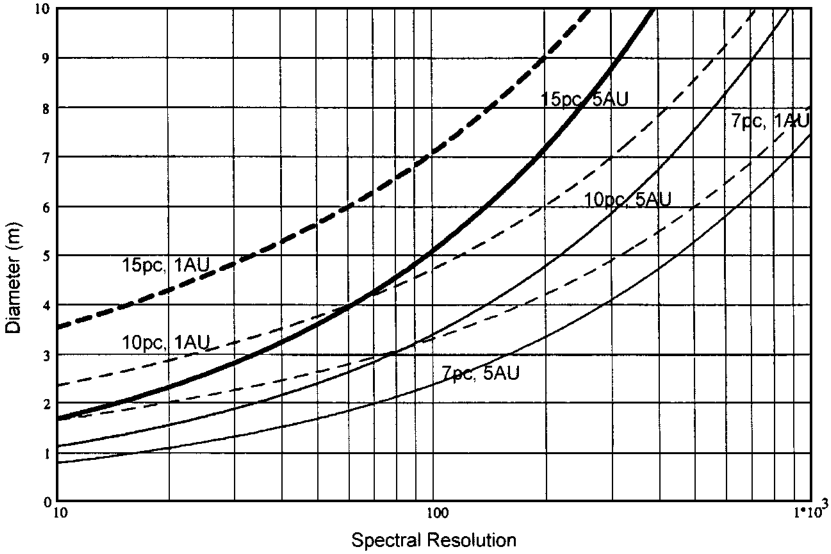

Figure 2 Aperture diameter versus spectral resolving power. The diameter of the inner mirrors of a four-element interferometer required for a changing spectral resolving power *R* (abscissa) at a constant signal-to-noise ratio at a standard continuum signal strength. Typical required resolving power is 20 for observing ozone and 200 for observing methane. The planetary systems have modeled three times as much dust as the Solar System and are considered at 15 pc (*thick lines*), 10 pc, and 7 pc (*thin lines*). Interferometers are considered at 5 AU (*solid lines*) and 1 AU (*dashed lines*).

free-flyer interferometer, with the telescopes at the vertices and side centers of an equilateral triangle. Each side is a θ^4 nulling linear array, the 1-2-1 arrangement (Woolf & Angel 1997). Signals from two sides are combined in and out of phase to chop the planet signal. The device needs rotation through only 60° to cover all angles. The problem suffered by free flyers is the danger of collisions or losing an element. There is no passive safe mode. Large space systems, whether linked or free flying, suffer from the problem of verifying their performance in a terrestrial environment.

7. PLANS AND OPTIONS

Below we elucidate current and future plans; please see Table 4 for a summary.

7.1 *Mid-Infrared Interferometry*

Both NASA and the European Space Agency (ESA) have candidate interferometric missions to find and obtain spectra of Earth-like planets. NASA's interest

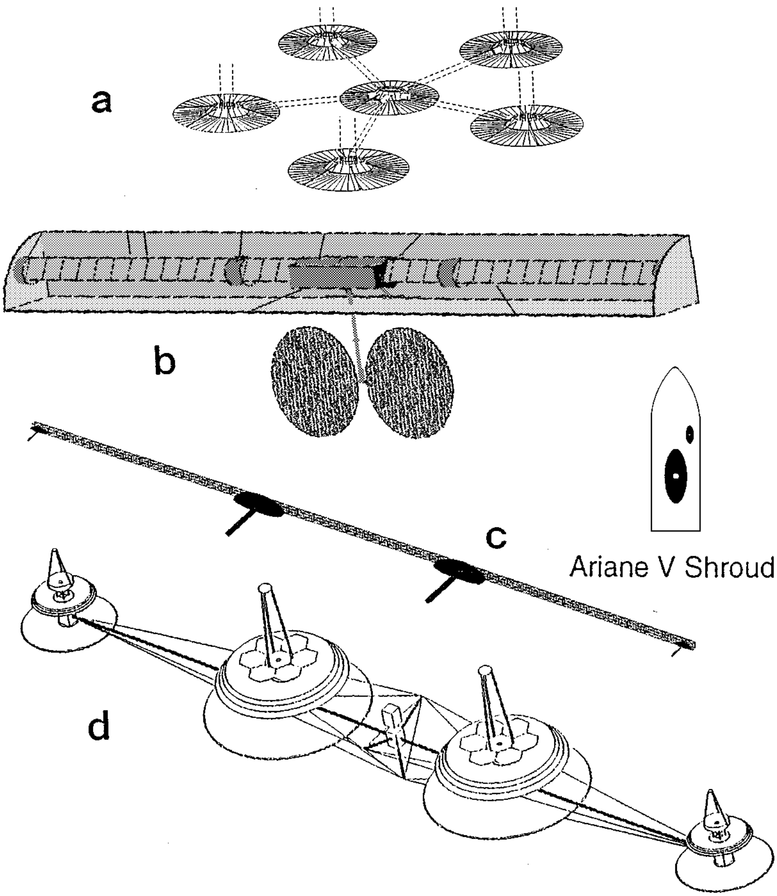

Figure 3 Four versions of an interferometer for seeking terrestrial planets. From *top* to *bottom*: (*a*) A DARWIN/IRSI concept for a two dimensional arrangement of five 1-m class mirrors as free flyers at 5 AU. (*b*) Ball Brothers concept for a four-element linear interferometer at 5 AU. The truss is extruded from canisters and the package can fit an Ariane-sized rocket. (*c*) Lockheed-Martin Corporation and the University of Arizona combined concept for a four-element linear interferometer at 1 AU. The mirrors are elliptically shaped monoliths. The truss unfolds and also fits an Ariane-sized rocket as shown. (*d*) This Thompson Remo Woolridge (TRW) concept for a linear interferometer at 1 AU uses Next Generation Space Telescope (NGST)-sized mirrors, deployed as with the TRW concept for NGST. It would have the sensitivity to observe methane. The truss is thin but is strengthened by tensioning outriggers.

Table 4 Summary of plans

Mission	Spectral range	Optical system	Orbit	Performance limit	Distance; species
Terrestrial Planet Finder (TPF) (representative design)	Mid infrared	Linear interferometer, 4-m elements	1 AU	Local zodiacal background	15 pc; O_3, CH_4, H_2O, CO_2
IRSI-DARWIN (representative design)	Mid infrared	Two-dimensional interferometer, 1.5-m elements	5 AU	Exozodiacal cloud	15 pc; O_3, CH_4, H_2O, CO_2
Next-Generation Space Telescope (NGST) (with figure correction)	Optical and near infrared	Filled aperture 6–8 m	1 AU	Optical figure	4–8 pc; O_2 H_2O

was developed during a 1996 study, "Exploration of Neighboring Planetary Systems" (Ex-NPS) (Beichman 1996). The primary recommendation for a space mission was to discover and obtain infrared spectra of Earth-like planets with a nulling interferometer, dubbed the Terrestrial Planet Finder (TPF). NASA plans a launch about 2010 as the culmination of its current Origins program. Ground-based observations were recognized as likely to lead the exploration for the decade before TPF and to find the first planets, a prediction that was confirmed almost immediately. The Ex-NPS report was endorsed by a blue ribbon panel committee chaired by C Townes. An independent study of astronomy sponsored by the Associated Universities for Research in Astronomy (AURA) identified the search for Earth-like planets as one of the two major goals in the post-HST era (Dressler 1997).

ESA is working with a concept of a two-dimensional interferometer called DARWIN (Léger et al 1996) (e.g. Figure 3), which it has recently renamed Infrared Space Interferometer (IRSI). One ESA study has concluded that a mission with each telescope as a separate spacecraft is possible. A second European industrial study was due to be initiated at the time of writing. IRSI is one of two contenders for a "cornerstone" mission to be launched circa 2010.

7.2 *Near Infrared Imaging*

There are at present no specific plans for imaging or spectroscopy of the reflected light of terrestrial planets. The NGST is baselined with optical quality inferior to HST, aiming at diffraction-limited operation only in the infrared beyond 1 μm (Stockman 1997). However, the techniques being developed for adaptive correction of atmospheric wavefront errors on the ground could be used in a space instrument to correct to the quality needed for planet detection. To avoid excessive complexity, we have suggested that consideration be given to building two 6-m class telescopes at once, both covering the crucial 1- to 5-μm range, but with one extended for precision optical work including planet detection, the other for longer infrared observations (Angel & Woolf 1998).

7.3 *Beyond Terrestrial Planet Finder and Next Generation Space Telescope*

If terrestrial planets are discovered, and crude spectra find another planet with abundant water or signs of life, then there is the possibility of building a more powerful (and expensive) second generation space interferometer (Figure 4). This would be able to make a meaningful search for trace constituents, such

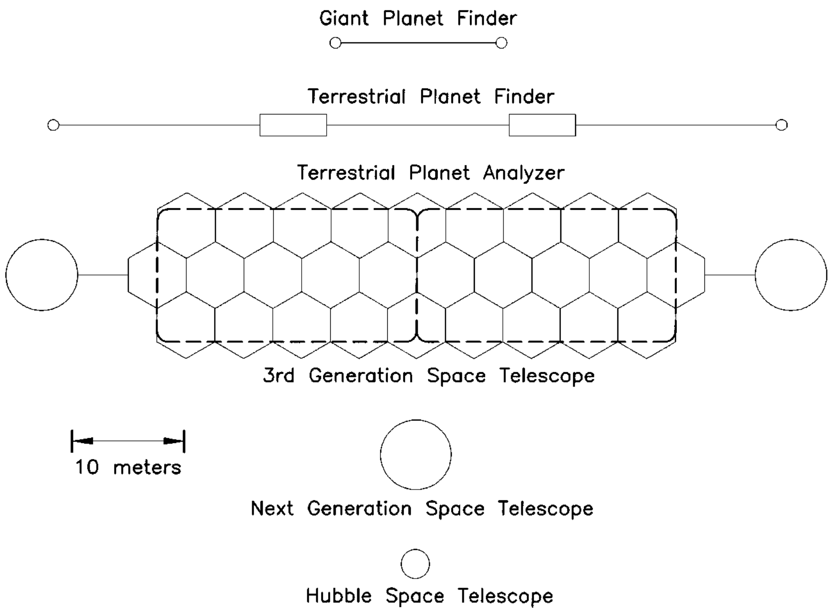

Figure 4 Terrestrial Planet Analyzer concept. A to-scale comparison of how a second generation space interferometer might evolve from both a Terrestrial Planet Finder (TPF) linear interferometer and the Hubble Space Telescope (HST)/NGST to be a device capable both of high resolution spectroscopy in the mid infrared to study trace molecules and of coronagraphic active optics observations of the reflection spectra of planets to observe absolute abundances from unsaturated lines. It would have a visible wavelength resolution of about 2 mas and a high sensitivity for low surface brightness features.

as methane in the planet's atmosphere, which could make the presence of life more certain. It could also make observations of the reflection spectrum and observe unsaturated lines.

8. CONCLUSION

Remarkably, it is now possible and even realistic to plan for telescopes in space that could detect life on planets of other stars. In the past, we could search for alien life only in the following forms: fossils or tenuous life under extreme conditions on other planets of our system or as alien civilizations communicating between stars by radio. Now we can add the possibility, arguably more likely, of finding spectroscopically new planets with abundant, primitive life, characteristic of life on Earth for the past billion years. We live in a fortunate time when space exploration could realize the dreams of past generations, of finding Earth-like planets and life independent of the Earth. Of those devices currently in the planning stage, infrared interferometers alone seem capable of making measurements around a substantial number of stars and of testing for the first evidence of life. The NGST, if equipped with suitable active optics, could survey a smaller number of nearer stars and search for signs of life around the four closest stars. The extension to making detailed observations of any planets discovered will require much larger areas of mirrors in space.

ACKNOWLEDGMENTS

We are indebted to many colleagues who have provided assistance for this article, especially to Borucki, Burge, Butler, Kasting, Léger, Marcy, Mariotti, Martin, Olivier, Paczynski, Peale, Penny, and Schneider. Also, assistance has been given by all the members of the industrial teams for Ball Brothers, Lockheed-Martin, and Thompson Remo Woolridge, who are currently studying TPF concepts. This research has been supported by the NASA Jet Propulsion Laboratory under contracts 960427 and 961286.

Literature Cited

Angel JRP. 1990. In *The Next Generation Space Telescope*, ed. PY Bely, GD Illingworth, CJ Burrows, p. 81. Baltimore: Space Telescope Sci. Inst.
Angel JRP. 1994. *Nature* 368:203
Angel JRP. 1998. See Backman 1998. In press
Angel JRP, Burge J, Woolf N. 1997. *Proc. Soc. Photo Instrument. Eng.* 3216. In press
Angel JRP, Cheng AF, Woolf NJ. 1986. *Nature* 322:341
Angel JRP, Martin B, Miller S, Burge J, Hinz P, et al. 1997. NASA white paper. Steward Observatory, Tucson, AZ
Angel JRP, Woolf NJ. 1997a. *Ap. J.* 475:373

Angel JRP, Woolf NJ. 1997b. See Soderblom 1997, p. 207
Angel JRP, Woolf NJ. 1998. In *Science with the Next Generation Space Telescope*, ed. E Smith, A Koreans, *Conf. Ser. 133*, p. 172. San Francisco: Astron. Soc. Pac.
Backman DE, Caroff LJ, Sandford SA, Wooden DH. 1998. *Extrasolar Zodiacal Dust Workshop, NASA Ames*. Mountain View, CA: NASA Ames Res. Center. In press
Beichman C. 1996. *A Road Map for the Exploraration of Neighboring Planetary Systems (Ex-NPS)*. Jet Prop. Lab. Publ. 96-22, Pasadena, CA
Beichman C. 1998. *Bull. Am. Astron. Soc. Meet.* 29:In press
Berkener LV, Marshall LC. 1964. *Discuss. Faraday Soc.* 37:66
Born M, Wolf E. 1980. *Principles of Optics.* New York: Pergamon. 370 pp.
Borucki WJ, Dunham E, Koch D, Cochran W, Rose J, et al. 1996. *Astrophys. Space Sci.* 241:111
Borucki WJ, Koch D, Dunham E, Jenkins J. 1997. See Soderblom 1997, p. 153
Bracewell RN. 1978. *Nature* 274:780
Broecker WF, Teng TH. 1982. *Tracers in the Sea.* New York: Columbia Univ. Press
Cocconi G, Morrison P. 1959. *Nature* 184:844
Coude de Foresto V, Mariotti J-M, Perim G. 1997. See Soderblom 1997, p. 269
Deleuil M, Barge P, Léger A, Schneider J. 1997. See Soderblom 1997, p. 259
Diner DJ. 1990. In *The Next Generation Space Telescope*, ed. PY Bely, GD Illingworth, CJ Burrows, p. 133. Baltimore: Space Telescope Sci. Inst.
Donahue TM, Hoffman JH, Hodges RR. 1982. *Science* 216:630
Drake F. 1965. In *Current Aspects of Exobiology*, ed. G Mamikunian, MH Briggs, Jet Prop. Lab. Tech. Rep. 32-428, Pasadena, CA
Dressler A. 1997. In *HST and Beyond*, ed. A Dressler, pp. 17, 60–64. Washington, DC: Assoc. Univ. Res. Astron.
Farmer CB. 1997. In *The Blue Dot Workshop: Spectroscopic Search for Life on Extrasolar Planets*, ed. D Des Marais, NASA CP 10154, p. 25. Mountain View, CA: NASA Ames Res. Center
Feynman RP. 1985. *QED: The Strange Theory of Light and Matter.* Princeton, NJ: Princeton Univ. Press
Gay J, Rabbia Y. 1996. *CR Acad. Sci. Paris* 322:265
Gay J, Rabbia Y, Manghini C. 1997. In *Infrared Space Interferometry: Astrophysics and the Study of Earth-Like Planets*, ed. C Eiroa, A Alberdi, H Thronson, T deGraaw, C Schalinski, p. 187. Dordrecht: Kluwer
Good J. 1994. *An Introduction to IRAS Sky Survey Atlas.* Jet Prop. Lab. Publ. 94-11:G-1, Pasadena, CA
Hanel RA, Conrath BJ, Jennings DE, Samuelson RE. 1975. *Exploration of the Solar System by Infrared Remote Sensing.* Cambridge: Cambridge Univ. Press
Hitchcock DR, Lovelock JE. 1967. *Icarus* 7: 149
Jacquinot P, Roizen-Dossier B. 1964. *Prog. Optics* 3:29
Kaplan LD, Munch G, Spinrad H. 1964. *Ap. J.* 139:1
Kasting J. 1996. *Origins Life Evol. Biosph.* 27:291
Kenknigh CE. 1977. *Icarus* 30:422
Kuiper GP. 1963. *Commun. Lunar Planet. Lab.* 1:179
Léger A, Mariotti JM, Menesson B, Puget JL, Rouan D, Schneider J. 1996. *Icarus* 123: 249
Léger A, Pirre M, Marceau FJ. 1993. *Astron. Astrophys.* 277:309
Lovelock J. 1975. *Proc. R. Soc. London B* 189,167
Mao S, Paczynski BE. 1991. *Ap. J. Lett.* 374: L37
Malbet F, Yu JW, Shao M. 1995. *PASP* 107: 386
Marcy G, Butler P. 1998. *Annu. Rev. Astron. Astrophys.* 36:57–97
Mennesson B, Mariotti J-M. 1997. *Icarus* 128:202
Noll K, Roush T, Cruikshank D, Johnson R, Pendelton Y. 1997. *Nature* 388:45
Oppenheimer BR, Kulkarni SR, Matthews K, Nakajima T. 1995. *Science* 270:1478
Owen T. 1980. In *Strategies for the Search for Life in the Universe*, ed. M Papagiannis, p. 177. Dordrecht: Reidel
Paczynski BE. 1996. *Annu. Rev. Astron. Astrophys.* 34:419
Peale S. 1997. See Soderblom 1997, p. 281
Reach WT, Franz BA, Welland JL, Hauser MG, Kelsall TN, et al. 1995. *Nature* 374:521
Roddier FC, Roddier C. 1997. *PASP* 109: 815
Roddier FC, Roddier C, DeMarcq J. 1978. *J. Optics Paris* 9:145
Sagan C, Thompson WR, Carlson R, Gurnett D, Hord C. 1993. *Nature* 365:715
Schneider J, Doyle LR. 1995. *Earth Moon Planets* 71:153
Shao M. 1997. In *The Search For Extra-Solar Planets*, ed. JM Shull, HA Thronson, SA Stern, p. 107. Dordrecht: Kluwer
Shao M, Colavita M. 1992. *Annu. Rev. Astron. Astrophys.* 30:457
Soderblom D, ed. 1997. *Planets Beyond the Solar System and the Next Generation of Space Missions, Conf. Ser. 119.* San Francisco: Astron. Soc. Pac.

Stahl S, Sandler D. 1995. *Ap. J.* 454:L153

Stockman HS. 1997. *Next Generation Space Telescope*, p. 37. Baltimore: Space Telescope Sci. Inst.

Tartar J. 1997. *Summary of SETI Observing Programs.* Menlo Park, CA: SETI Inst.

Walker JC. 1977. *Evolution of the Atmosphere.* New York: Macmillan

Woolf NJ. 1997. In *Infrared Space Interferometry: Astrophysics and the Study of Earth-Like Planets*, ed. C Eiroa, A Alberdi, H Thronson, T deGraaw, C Schalinski, p. 283. Dordrecht: Kluwer

Woolf NJ. 1998. *Bull. Am. Astron. Soc* 29:In press

Woolf NJ, Angel JRP. 1995. *Adapt. Optics Optic. Soc. Am. Tech. Dig. Ser.* 23:44

Woolf NJ, Angel JRP. 1997. See Soderblom 1997, p. 285

Woolf NJ, Schwarzschild M, Rose WK. 1964. *Ap. J.* 140:833

Wolszczan A. 1997. See Soderblom 1997, p. 135