Pupil Configurations for an Interferometric TPF Mission

Eugene Serabyn et Bertrand Mennesson
Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Drive
Pasadena, CA 91109
818-393-5243
gene.serabyn@jpl.nasa.gov

Abstract—The primary goal of the proposed Terrestrial Planet Finder (TPF) mission [1] is the direct detection of radiation from potential terrestrial planets orbiting around nearby stars. However, the faintness of the expected signals leads to very stringent technical requirements on the optical performance. As a result, both of the candidate approaches currently under active development by the TPF project, infrared nulling interferometry with separated aperture telescopes, and optical coronagraphy with a large single aperture telescope, require complex optical systems which push the state of the art. Identification of the simplest optical approach, as well as laboratory demonstrations of the basic capabilities, are therefore essential. In this paper, the range of proposed interferometric architectures is briefly summarized and compared. On this basis, it is concluded that a fairly simple, compact linear array of telescopes on a structure, such as a chopped dual-Bracewell system, can attain the bulk of the goals set out for the TPF mission.

TABLE OF CONTENTS

1. INTRODUCTION 1
2. RESOLUTION REQUIREMENTS 2
3. INPUT PUPIL CONFIGURATIONS 3
4. SUMMARY 6
REFERENCES 6

1. INTRODUCTION

[1]Current observational capabilities still fall short of being able to directly resolve extra-solar planets from their parent stars, due to the small angular scales and large contrast ratios involved. In fact, both of the successful extra-solar planet detection techniques that have been successfully applied to date (stellar radial velocity variations and planetary occultations of the parent stars) are indirect methods, relying on observations of the parent stars rather than of the planets themselves. While hot-Jupiter type planets may soon be amenable to direct detection with the new generation of ground-based infrared nulling interferometers currently under development, and jovian planets may be amenable to detection in the future with coronagraphs behind active optics correctors, the direct detection of terrestrial (Earth-like) planets, with contrast ratios ranging from several million in the infrared to several billion in the optical, remains well beyond present observational capabilities, and very likely will require high-contrast observations from space.

To enable the direct detection of very faint planets in close proximity to vastly brighter stars, current observational limitations such as finite telescope diameters and atmospheric fluctuations need to be overcome. Given that telescope diameters cannot be increased arbitrarily, the stellar light destined for the inner region of the focal plane needs to be suppressed to a high degree. This suppression can be brought about by means of such starlight rejection techniques as coronagraphic masking and nulling interferometry between separate telescopes. Since both of these high dynamic-range detection techniques require substantial development before deployment on a space mission can be considered, both are receiving attention as possible enabling technologies for NASA's proposed Terrestrial Planet Finder (TPF) mission.

The TPF project is currently considering a range of architectures, which fall into two classes: large-aperture actively-corrected coronagraphs operating at optical wavelengths, and cryogenic nulling interferometers operating in the thermal infrared. The selection of the starlight suppression approach thus has tremendous implications for all aspects of the mission architecture, from the wavelength range and temperature of the optical system, to all aspects of the technological development plan. Since complexity and technological maturity are cost drivers, the primary near-term goal of the TPF project is to identify the simplest, most technologically ready, and most cost effective approach to achieving the detection of earth-like planets around other stars. The goal of this paper is to provide a brief overview of the range of interferometric mission architectures. The possible architectures identified to date can be categorized most simply in terms their pupil functions, but consideration of the various pupil configurations leads immediately to a host of related issues, such as the baseline size scale, the stellar rejection levels, the individual telescope aperture diameters, the beam-combination and modulation strategies, as well as the strategies for separating the various signals and backgrounds which will be present. The interplay of all these parameters leads to a rather complex parameter space in which to look for solutions.

[1] 0-7803-7651-X/03/$17.00 © 2003 IEEE
[2] IEEEAC paper #1281, Updated December 20, 2002

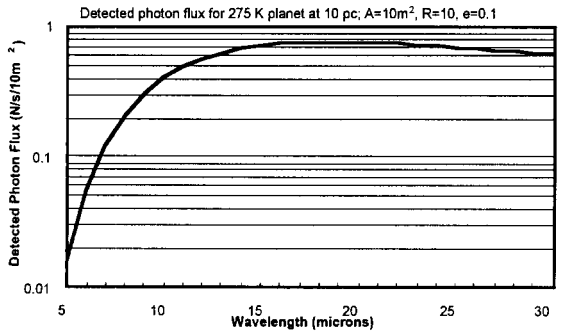

Figure 1. Detected photon flux from a terrestrial planet located 10 pc away. A collecting area of 10 m^2, a spectral resolving power of 10, and an optical efficiency of 0.1 are assumed.

2. Resolution Requirements

The thermal emission from potential terrestrial planets around nearby stars is expected to be quite faint, as is indicated in Figure 1. However, the more serious impediment to the detection of such sources is the enormous dynamic range between the planetary and stellar signals (Figure 2), in the face of the minute angular scales involved. Indeed, at 10 pc, a body 1 AU away from it parent star will subtend an angle of only 0.1 arc seconds (Figure 3).

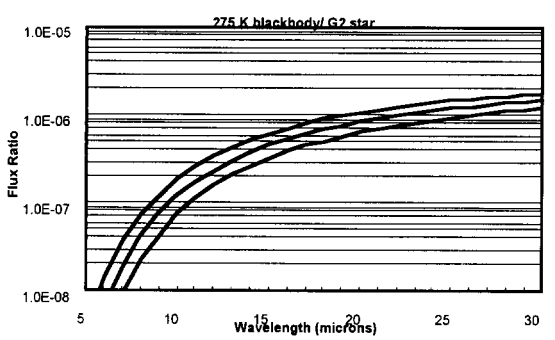

Figure 2. Flux ratio between an Earth-size "blackbody" planet at 250, 275 and 300 K and a solar-type G2 star.

Assuming that terrestrial analogs are located in "habitable zones" or "target annuli" at somewhat similar offsets (0.5 to 2 AU) from their parent stars (to allow for the possibility of liquid water being present), a rather small angular inner working distance is needed by the optical system. Figure 3 plots the inner working distance, defined here as the innermost angle, θ_{IWD}, at which the fringe response drops to _ that at the first constructive maximum, or $\theta_{IWD} = \lambda/4b$ for the case of a single baseline nulling interferometer (the Bracewell case [2]) of length b. This inner working distance is 0.05 arc seconds for a 10 m baseline at a wavelength, λ, of 10 μm, and decreases to 0.01 arc seconds for a 50 m baseline. The result is easily to remember, because $d_{pc} = b_m$, i.e., the maximum distance, d_{pc}, in parsecs, for which the inner working distance is 0.5 AU out from a star, is equal to the baseline length in meters. The main question is then how to balance the opposing needs of longer baselines to provide for larger stellar samples, and shorter baselines to allow for a more realistic mission in the near term.

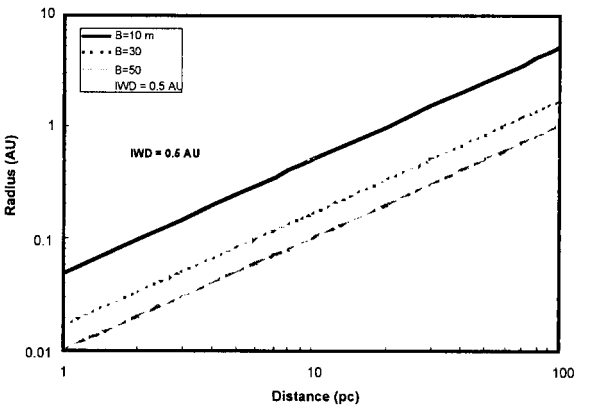

Figure 3. Sloped lines: half power response angles for baselines of 10, 30 and 50 m. The horizontal line is at 0.5 AU. The intersection of the sloped and flat lines gives the maximum distance for which the inner working distance (IWD) can be 0.5 AU, i.e., $d_{pc}=b_m$.

Figure 4 shows the stellar-diameter-limited null depths vs. wavelength and baseline length for the single baseline case. As both Figures 3 and 4 intimate, the performance of an interferometer will have a steep distance dependence, since the stellar null depth will vary at least as the inverse square of the distance, while angular scales only vary inversely with the first power of distance. Thus, a transition will naturally occur from the few nearest stars, where the stellar rejection won't be as deep, but the planets will be better resolved interferometrically, to the case of the more distant stars, where the stellar nulls will be deeper, but the angular resolution poorer. In one sense this complementarity works in the right direction, since with better stellar rejection at distance, angular resolution becomes somewhat less important. Thus, one should plan for a wide range in performance, especially because longer wavelengths (in the foreseen range of about 6-16 μm) will also have degraded angular resolution, and so the inner working distance, or the distance at which the inner working distance limit is reached, will vary by about a factor of three across the waveband of interest. Thus, the number of stars accessible to long wavelength investigation will necessarily be smaller than at shorter wavelengths.

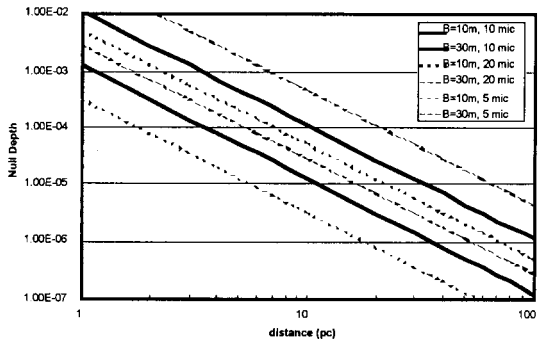

Figure 4. Single-baseline null depth vs. distance for solar-type stars and baselines of 10 and 30 m, and wavelengths of 5, 10 and 20 μm.

3. Input Pupil Configurations

A variety of interferometric configurations can be considered for space-based nulling interferometers, beginning with the simplest case of the two-telescope, single-baseline nuller already discussed. However, a single-baseline nuller does not by itself sufficiently attenuate the stellar flux from nearby stars unless the baseline is very short, because of finite stellar angular diameters. As seen in Figure 4, shorter baselines suppress starlight to a greater extent, with the null depth, N, given by $N = (\pi b \theta_D / 4\lambda)^2$, where θ_D is the stellar angular diameter. However, near-in planets would then be more attenuated by the resultant broader central null fringe. A fringe transmission of 50% at the inner working distance then requires $b = \lambda / 4\theta_{IWD}$.

The main cause of this apparent mismatch between the angular resolution and null depth requirements for the single baseline case is the fairly slow, θ^2, off-axis transmission rise near the axis. To overcome this limitation, off-axis transmissions, $T = \theta^n$, which rise as higher powers of θ than the single-baseline n=2 case are needed, and this can in general be achieved with a larger number of telescopes. Indeed, a number of one and two dimensional configurations for small nulling arrays have been proposed over the years. However, the choice of pupil configuration immediately impacts numerous parameters in addition to the stellar null depth and angular resolution, and in particular, the ability to distinguish planetary from exo-zodiacal emission. Thus, the modulation strategy is of fundamental importance. While the baseline can be rotated slowly to modulate the residual, non-nulled signal (from the star, exozodiacal emission and the planets), stability concerns prevent this slow modulation from being the primary modulation strategy, and so we now briefly discuss the more rapid alternatives.

Modulation approaches

One basic limitation of any nulling interferometer is its inability to phase modulate the off-axis signals. In both radio and optical/infrared interferometers, signal detection by phase modulation is a fairly standard technique, but the fixed phase relationships between the component telescopes in a nulling interferometer prevents doing this. Phase modulation between multiple nullers is a viable possibility, but this of course immediately doubles the number of optical components. Dual-nuller phase modulation would however be capable of separating (removing) azimuthally symmetrical exozodiacal signals from the decidedly asymmetric planetary signal.

Simultaneous spatial chopping is another modulation option, but this likely will cause interruptions in the stabilization loops. This approach does not separate the exozodiacal and planetary signals in the normal sense of spatial chopping, but instead merely removes long term drifts from their sum. If at the same time the baseline is rotated, the rotation-modulated signals from planets located outside of the first fringe maximum will show harmonics higher than twice the rotation rate of the baseline. The exozodiacal signal will also contain higher harmonics, but it's power spectrum will likely not extend very high in frequency. Thus, planets located several fringes off-axis might be detectable with this approach. However, TPF's primary goal is the inner planets, so it is not clear that this modulation approach is compatible with TPF's goals, without requiring a rather long baseline.

One dimensional configurations

The basic one-dimensional nulling interferometer configurations proposed to date are shown in Figure 5. The relative aperture diameters in the Figures represent the

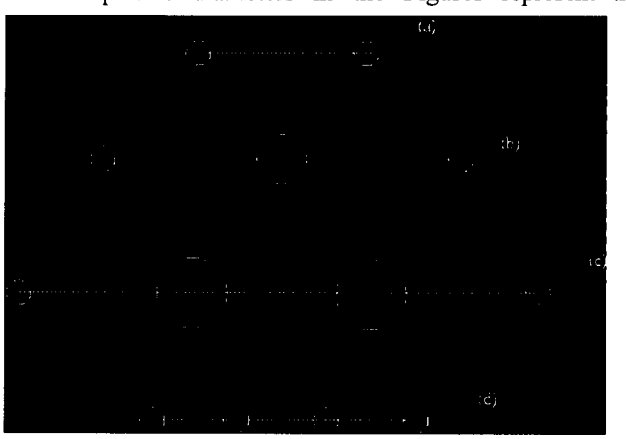

Figure 5. Simple linear nulling configurations. All of these configurations have their lengths scaled to yield first maxima at the same angular offset (same basic fringe spacing): a) single-baseline BW case with separation b, b) DAC case of total length 2b, c) OASES case with spacing b and total length 3b, and d) chopped dual-Bracewell case with total length 1.5b, where b is the 1-3 and 2-4 nulling baseline length.

relative amplitudes assigned to the corresponding beams at the recombination stage. This does not imply any assumptions about the actual recombination scheme, so that the relative telescope diameters may be different from that shown. For example, for the DAC configuration, the amplitude ratios needed are -1:2:-1, which corresponds to telescopes with diameters $1:\sqrt{2}:1$. In addition, for the purposes of this comparison, the baselines in each configuration were chosen so that the first transmission

maxima are all at the same angular offset from the star. Even so, the half-power points and resolutions differ from system to system, as will be discussed below.

Beginning with the one dimensional cases, the simplest case, that show in Figure 5(a), is the 2-element nuller ("Bracewell" or "BW" case) discussed above. Of course this has the best resolution for a given total length, but as already discussed, no phase chopping is possible, and the θ^2 null implies limited null depth on the nearest stars. In addition, the exo-planetary and exo-zodiacal signatures are mingled.

Next is the three-element array referred to commonly as the degenerate Angel cross (or DAC), which combines beams with amplitudes of -1:2:-1. Three is the minimum number of telescopes needed to produce a θ^4 central null, and with such a system, much better nulls are possible on nearby stars. However, again phase chopping is not possible, because of the need to maintain fixed phases between the telescopes contributing to the null (unless the array is instead decomposed into a pair of single-baseline θ^2 nullers). In addition, telescopes of different diameters are needed, which is sure to add to the cost.

Next in figure 5 is (c): the 4-element interferometer referred to as OASES [3], with weightings -1:3:3:-1. Four telescopes is the smallest number that can provide a θ^6 null, but this is not vital, as θ^4 is already broad enough to reduce the stellar flux below that of the backgrounds. Again, no phase chopping is possible, and either telescopes of different sizes or beamsplitters with very specific coating properties are required, unless the telescopes are instead reconfigured into a pair of DACs, each with -1:2:-1 weighting. Although this would enable dual-DAC chopping,, such an approach likely requires rather Baroque beamcombiner layouts.

Finally, Figure 5(d) shows another four-telescope arrangement, an interleaved dual-BW [1]. Here it is possible to chop between two single-baseline subnullers, producing θ^3 suppression of asymmetric emission near the axis (if the two sub-nullers attain equal nulls). The advantages of this approach include telescopes of a single size, and the need for only simple 50/50 beamsplitters, but the most significant advantage is the fact that phase-chopping between the two single-baseline nullers is possible, which allows the removal of centro-symmetric exo-zodiacal (and stellar residual) signals. Thus, this configuration seems to enable all of the needed signal separations, as long as the individual nulling baselines are long enough, and as long as appropriate symmetry is maintained between the two sub-nullers. In any case, this system does provide an existence proof.

Another dual-nuller one-dimensional configuration [4] is shown in Figure 6. This layout consists of a pair of well-separated DACs, each of which provides a θ^4 null. The pair of nullers is then used to resolve and image the remaining flux at high resolution by means of phase modulation. The issues of null depth and angular resolution are thus very effectively decoupled here, as are the problems of distinguishing exoplanetary and exozodiacal emission. However, this solution brings with it both additional telescopes and complexity. In fact, assuming that the DACs are structurally connected, but that the pair of DACs is not, solutions to both structurally connected interferometry and formation-flying interferometry are then needed for this approach.

Angular resolution requirements

How long do simple linear systems need to be? This depends on the exact criterion used to define the required angular resolution, and so is worth addressing in some detail. The first criterion to consider is of course the classical "resolution" of half the fringe spacing, or $\lambda/2b$ (or the off-axis distance to the first fringe peak), which is applicable in the simple imaging case. However, the flip side to the broader nulls achieved with higher values of n is narrower constructive peaks, as shown in Figure 7. Thus a second criterion might be the full width at half maximum (FWHM) of the constructive fringe. Finally, and especially in the case of good stellar rejection, a third criterion should be considered, that being the innermost half power point (HP), which we earlier identified with the inner working distance (IWD). Table 1 compares the values for these three criteria for the four linear interferometer cases considered.

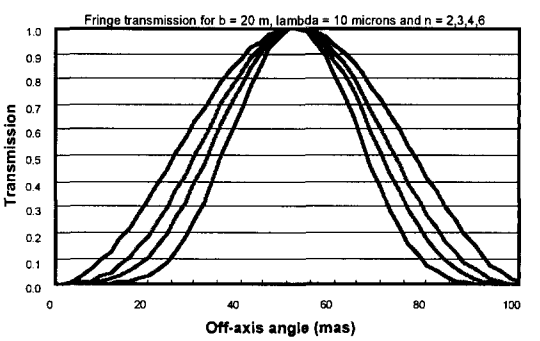

Figure 7. Fringe patterns for θ^n nullers with n = 2,3,4,6.

As can be seen in Figure 6 and columns 2 through 4 in Table 1, the three resolution criteria lead to non-negligibly different performance estimates for a given system. Likewise, noteworthy performance differences between the various configurations can also be seen. Selection of the optimal resolution criterion is thus very important, as the nulling baseline length and total array length, s, will depend on this choice. For example, in comparing the BW and dual-BW cases, the classical criterion leads to an overall length ratio of 1.5 between the systems (column 5 in Table 1). On the other hand, if the inner HP point is the criterion, then the

Figure 6. Dual DAC configuration.

length ratio is 1.5×(4/3.43)=1.75. Yet again, if the fringe FWHM is the criterion, the length ratio is only 1.5×(2/2.4)=1.25. Thus, the length of the dual-BW case can

vary by about ±1/6 of its classically-defined length, depending on the resolution criterion selected.

One further complication worth considering is the fact that the resolution criterion may differ in different observing scenarios. Thus, for more distant stars, where the null itself is deeper, resolving a planet from a fainter central source in the final image has somewhat less stringent requirements than for nearby stars, where a fairly bright stellar residue remains. Of course in the nearby case, the planets themselves would be at larger angular offsets, and so easier to resolve in the final image from that point of view.

	$\theta_{MAX}/(\lambda/b)$	$\theta_{HP}/(\lambda/b)$	$\theta_{FWHM}/(\lambda/b)$	Length/b	Mean transmission
BW	1/2	1/4	1/2	1	1/2
Dual-BW	1/2	1/3.43	1/2.4	1.5	$4/3\pi$
DAC	1/2	$1/\pi$	1/2.75	2	3/8
OASES	1/2	1/2.86	1/3.33	3	5/16

Table 1. Resolution criterion, relative system lengths, and mean transmissions of one-dimensional nullers.

Related to the fringe width is the question of instantaneous sky coverage, or in other words, the mean transmission of the fringe pattern. This is given in column 6 of Table 1, where it can be seen that narrower constructive fringes lead to lower mean transmissions, or in other words, to smaller duty cycles for a given fringe to intercept a given exoplanet.

Two dimensional configurations

Two dimensional configurations have also received attention, predominantly in the context of the Darwin project. Possible advantages include easier beam recombination for configurations in which there is a point equidistant from all telescopes. In addition, one may not need to rotate such arrays as far in order to provide equivalent uv-plane coverage. Thus only 60 or 72 degrees of rotation might be needed. However, once configuration rotation is enabled, it is not clear that it is advantageous to stop and rotate back instead.

An overview of two dimensional configurations is provided in Figure 8. The first configuration is (a), the Angel Cross (AC) configuration [5]; a diamond shape guarantees that transmission zeros get washed out by baseline rotation. Since all four telescopes are needed to maintain the θ^4 null, phase chopping is not possible.

Layout (b), that in the 1993 Darwin proposal [6,7], has 5 telescopes on a circle and a θ^4 null pattern. With an odd number of telescopes, the transmission map is asymmetric, which results in different responses to exoplanet emission and centro-symmetric exozodiacal emission. However, no provision for phase chopping is included, some transmission holes can survive array rotation, and the beam combiner needs to provide phase shifts of $2\pi/5$ rather than the simpler field reversals.

Placing the five telescopes on an ellipse [8], as in layout (c) in Figure 8 is similar, but with an improved transmission pattern, as the transmission zeros wash out with array rotation. However, now long delay lines are needed to make up for unequal spacings, and phase chopping is still not provided for.

The triangular layout in figure 8(d) is the Mariotti interferometer, in which each side of the triangle is used to form a DAC with a θ^4 null. Phase chopping between any pair of DACs is then possible, but at the cost of having six collectors, and a complex beam combiner.

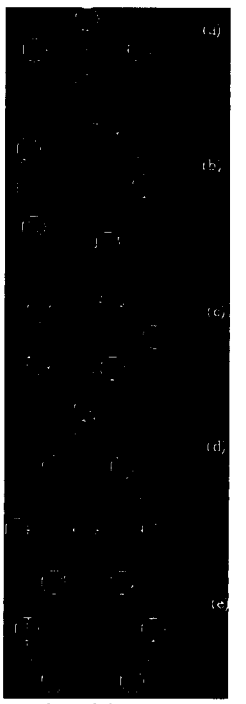

Figure 8. Two dimensional layouts: a) Angel cross (AC), b) five on a circle (DARWIN), c) five on an ellipse (M&M), d) six on a triangle (Mariotti), and e) six on a circle or hexagon (Laurance).

Finally, layout 8(e) is the so-called Laurance configuration [9,10], with six telescopes on a hexagonal pattern, which makes the recombination in a central hub equidistant from all the telescopes straightforward. There are several variants, using either 3 subsets of 4 telescopes (but with odd achromatic amplitude ratios), or 2 subsets of 4 telescopes (in a bent OASES-like configuration), which seems much more workable. The main problems with many of these two-dimensional configurations is the large number of telescopes (and spacecraft if these are free flyers), the odd amplitude ratios required in a few cases, and the phasing problems inherent in sharing a telescope between several sub-nullers. However, the bent double-OASES configuration does seem to provide an existence proof for the two dimensional case.

Deployment

It is beyond the scope of this paper to select an optimal configuration or to discuss the tradeoff between structurally-connected and free-flying telescopes. However, existence proofs for viable configurations do seem to be present in both the one-dimensional and two-dimensional cases. It is also worth presenting an existence proof that at least one such system is easily deployable. In fact, it is possible to show that two configurations are deployable from a single folded configuration. Figure 9 shows a "Swiss army knife"-like deployment, in which four closely packed telescopes are arranged on three parallel linear sub-booms. Upon opening two sets of hinges, the telescopes necessarily pass through an Angel cross configuration before reaching the final dual-Bracewell configuration. The packaged length is of order 1/3 of the fully extended length, implying that structures of length up to about 30 m can be considered.

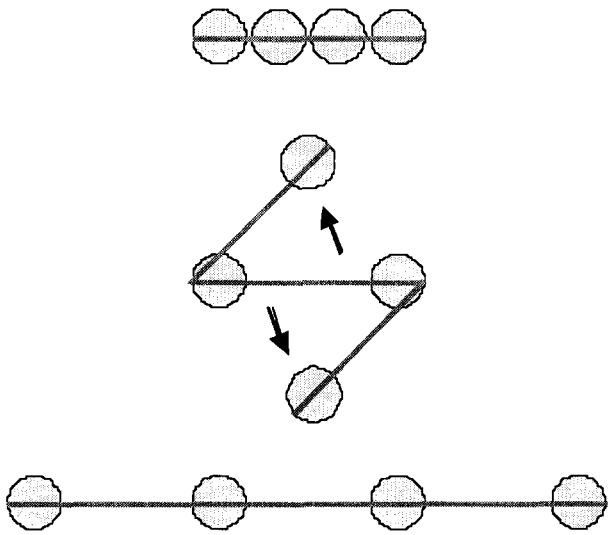

Figure 9. A four-telescope, two hinge configuration which allows for compact packaging and a linear dual-BW configuration, while also passing through an AC configuration during the unfolding.

4. Summary

The potential approaches for enabling the TPF mission require highly complex and costly optical systems which push the state of the art. Identification of the simplest possible system consistent with the mission goals is therefore highly desirable. In this paper a few of the basic interferometer configurations are surveyed and contrasted. Conceptual existence proofs are available for both one- and two-dimensional configurations, and the process of understanding the detailed implications of such designs is underway. In particular the issues related to the angular resolution requirement are explored in some detail here, as this relates directly to the overall system scale size. One-dimensional configurations lend themselves naturally to a folded deployable structure, and it may be possible to meet the mission angular resolution requirement in the near term with e.g., a linear dual-Bracewell structure which unfolds to a length of about 30 m. Two-dimensional configurations are also viable, but as these are more consistent with free-flying telescopes, they may require longer timescales.

References

[1] "The Terrestrial Planet Finder (TPF): A NASA Origins Program to Search for Habitable Planets." Written by the TPF Science Working Group. JPL Publication 99-3, 1999, eds. C.A. Beichman, N.J. Woolfand C. Lindensmith.

[2] Bracewell R.N. and McPhie R.H. 1979, "Searching for Non-solar Planets," Icarus 38, 136

[3] Angel, J.R.P & Woolf, N.J. 1997, "An Imaging Nulling Interferometer To Study Extrasolar Planets," Ap.J. 475, 373

[4] Velusamy, T., Beichman C.A., and Shao M. 1999, "A dual 3-element nulling interferometer for TPF." In "Optical and Infrared Interferometry from Ground and Space," ASP Conf. Ser. Vol. 194, 430, eds. S.C. Unwin & R.V. Stachnik

[5] Angel, R. 1990, "Use of a 16-m Telescope to Detect Earthlike Planets," Proceedings of the Workshop on The Next Generation Space Telescope, eds. P. Bely and C. Burrows, Space Telescope Science Institute, 81

[6] Leger A. et al, 1993, "The DARWIN mission concept. Proposal to the ESA Horizon 2000 Plus Planning Process"

[7] Leger A., Mariotti J.M., Mennesson B. et al., 1996, "Could we search for primitive life on extrasolar planets in the near future? The DARWIN project" Icarus 123, 249

[8] Mennesson, B. and Mariotti, J.M. 1997, "Array Configurations for a Space Infrared Nulling Interferometer Dedicated to the Search for Earthlike Extrasolar Planets", Icarus, 128, 202

[9] Karlsson A & Mennesson B. 2000, "The Robin Laurance Nulling Interferometers," SPIE Conf. Series Vol. 4006, 871

[10] Absil O. 2001, "Nulling Interferometry with DARWIN: detection and characterization of Earth-like Exoplanets," Master's Thesis, Universite de Lieges.