

# Extrasolar Planet Finding via Optimal Apodized and Shaped Pupil Coronagraphs

N. Jeremy Kasdin

*Dept. of Mechanical and Aerospace Engineering, Princeton University, Princeton, NJ  
08544*

jkasdin@princeton.edu

Robert J. Vanderbei

*Dept. of Operations Research and Financial Engineering, Princeton University, Princeton,  
NJ 08544*

rvdb@princeton.edu

David N. Spergel

*Dept. of Astrophysical Sciences, Princeton University, Princeton, NJ 08544*

dns@astro.princeton.edu

and

Michael G. Littman

*Dept. of Mechanical and Aerospace Engineering, Princeton University, Princeton, NJ  
08544*

mlittman@princeton.edu

## ABSTRACT

In this paper we examine several different apodization approaches to achieving high-contrast imaging of extrasolar planets and compare different designs on a selection of performance metrics. These approaches are characterized by their use of the pupil's transmission function to focus the starlight rather than by masking the star in the image plane as in a classical coronagraph. There are two broad classes of pupil coronagraphs examined in this paper: apodized pupils with spatially varying transmission functions and shaped pupils, whose transmission values are either 0 or 1. The latter are much easier to manufacture to the needed

tolerances. In addition to comparing existing approaches, numerical optimization is used to design new pupil shapes. These new designs can achieve nearly as high a throughput as the best apodized pupils and perform significantly better than the apodized square aperture design. The new shaped pupils enable searches of 50% - 100% of the detectable region, suppress the star’s light to below  $10^{-10}$  of its peak value and have inner working distances as small as  $2.8 \lambda/D$ . Pupils are shown for terrestrial planet discovery using square, rectangular, circular, and elliptical apertures. A mask is also presented targeted at Jovian planet discovery, where contrast is given up to yield greater throughput.

## 1. Introduction

With over 80 extrasolar planets discovered to date, interest in planet finding is becoming intense. Both the scientific community and the public are increasingly enthusiastic for new ideas and efforts toward discovering and characterizing extrasolar earthlike planets — something that current ground based searches cannot accomplish. It is generally agreed that terrestrial planet discovery, at least in the foreseeable future, will be accomplished via direct imaging by a space based observatory.

In December, 2001 NASA completed a two year study by four industry teams of various architecture concepts for the Terrestrial Planet Finder (TPF). In a well published conclusion, NASA indicated the consensus that two approaches merit further study—an infra-red nulling interferometer and a filled aperture, visible light coronagraph. Two of the teams proposed apodization concepts for a coronagraph. One of them, the Ball Aerospace team, of which the authors are members, focused on optimal shaped pupil coronagraphs, where the transmission values are either 0 or 1.

In this paper we compare and contrast two different methods for achieving high contrast images. In particular, we restrict ourselves to examining apodized and shaped pupil coronagraphs, excluding for now classical Lyot coronagraphs that control the diffracted light by masking the star’s image. While in the literature apodization typically refers to any modification of the aperture to alter the point spread function, here, to avoid confusion, we use apodization to refer to smooth attenuation and reserve the terms “shaped aperture” or “shaped pupil” for modifications to the pupil shape. We describe the tradeoffs in evaluating various coronagraph options, developing the methodology in section 2 and various metrics for comparison in section 3, and conclude that these favor shaped pupils over apodized apertures. We then present recent results from numerical optimization approaches for both square and circular shaped pupil coronagraphs.

## 2. Pupil Apodization

The objective of an apodized or shaped pupil planet finding system is to tailor the PSF in such a way as to make planet discovery and characterization possible. In contrast to a classical coronagraph, where the bright stellar image is blocked in the image plane, an apodized or shaped pupil system modifies the pupil to create a PSF with the needed contrast at the planet location. In such a system, the PSF of the planet and star are identical, merely shifted to correspond to their relative location in the sky and scaled by the ratio of irradiances,  $I_s/I_p$ , where  $I_s$  is the irradiance of the star and  $I_p$  is the irradiance of the planet. In this section, we briefly derive the relevant relationships for the PSF of telescopes with different shapes and apodizations.

The principals of telescope apodization and diffraction theory are well established—see, e.g., (Hecht 1998) and (Born and Wolf 1999). We only review here the basic equations in order to establish notation and provide convenient references for the essential results. Our notation follows closely that of Hecht (1998). We consider only Fraunhofer diffraction theory, where the field in the image plane is the Fourier Transform of the entrance field. We also assume perfect, aberration free optics with a Strehl of one. Detailed discussions of sensitivity and error will be considered in a future paper.

We consider the electric field response from plane parallel light waves arriving at the entrance pupil as Huygen’s wavelets with source strength  $\varepsilon_A$  per unit area constant across the aperture, and passing through an arbitrary aperture apodization  $\tilde{\mathcal{A}}(\tilde{y}, \tilde{z})$  (here written in rectangular coordinates):

$$E(Y, Z) = \frac{\varepsilon_A e^{i(\omega t + kR)}}{R} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{\mathcal{A}}(\tilde{y}, \tilde{z}) e^{-ik(Y\tilde{y} + Z\tilde{z})/R} d\tilde{y} d\tilde{z} = \mathcal{F}\{\tilde{\mathcal{A}}(\tilde{y}, \tilde{z})\} \quad (1)$$

where  $Y$  and  $Z$  are positions on the image plane,  $k = \frac{2\pi}{\lambda}$  is the wavenumber of the incoming light,  $R$  is the focal length,  $\tilde{y}$  and  $\tilde{z}$  are the position across the pupil, and  $0 \leq \tilde{\mathcal{A}}(\tilde{y}, \tilde{z}) \leq 1$  (we assume here that there is no phase effect from the apodization). Note that the aperture can be open, which we refer to here as a *shaped aperture*, where  $\tilde{\mathcal{A}}(\tilde{y}, \tilde{z})$  takes on values of only 0 or 1, or it can include a smooth transmission mask, which we will refer to here as *apodization*, where  $\tilde{\mathcal{A}}(\tilde{y}, \tilde{z})$  takes on a continuum of values.

It is useful when comparing different designs to change independent variables to a dimensionless form. If we consider a rectangular telescope opening with length  $a$  in the  $y$ -direction and  $b$  in the  $z$ -direction, the electric field can be written in the dimensionless form:

$$E(\xi, \zeta) = \frac{A\varepsilon_A e^{i(\omega t + kR)}}{R} \int_{-b/2a}^{b/2a} \int_{-1/2}^{1/2} \mathcal{A}(y, z) e^{-2\pi i(\xi y + \zeta z)} dy dz \quad (2)$$

where  $A = a^2$  is the area of an open (unapodized) fiducial  $a \times a$  square aperture,  $y$  and  $z$  are dimensionless positions in the pupil, and  $\xi$  and  $\zeta$  are dimensionless angular positions in the image plane in units of  $\lambda/a$ , i.e.,

$$\xi = \frac{aY}{\lambda R}, \zeta = \frac{aZ}{\lambda R}$$

The image plane detector measures the irradiance of the arriving field,  $I = \frac{c\epsilon_o}{2h}|E|^2$ , in photon  $\text{sec}^{-1} \text{ m}^{-2} \mu\text{m}^{-1}$ , where  $c$  denotes the speed of light,  $h$  is Planck's constant and  $\epsilon_o$  is the permittivity of free space. Squaring the magnitude of the electric field Eq. (2) provides the irradiance equation of interest,

$$I(\xi, \zeta) = I_o P(\xi, \zeta) \quad (3)$$

where  $I_o$  is the normalized irradiance of the source (planet or parent star) entering an aperture of area  $A$  before apodization,

$$I_o = \frac{c\epsilon_o}{2h} \frac{A^2 \epsilon_A^2}{R^2}, \quad (4)$$

and  $P(\xi, \zeta)$  is the point spread function (PSF) for the apodization or embedded shaped pupil,

$$P(\xi, \zeta) = \left| \int_{-b/2a}^{b/2a} \int_{-1/2}^{1/2} \mathcal{A}(y, z) e^{-2\pi i(\xi y + \zeta z)} dy dz \right|^2.$$

Note that for an open square aperture,  $P(0, 0) = 1$ . Hence,  $I(0, 0) = I_o$  and the irradiance for the square aperture is:

$$I(\xi, \zeta) = I_o \left( \frac{\sin \pi \xi}{\pi \xi} \right)^2 \left( \frac{\sin \pi \zeta}{\pi \zeta} \right)^2.$$

We have used a slightly unconventional form for the irradiance and the PSF. The objective later in the paper is to compare different apodizations in throughput or integration time and other metrics. In order to have a meaningful comparison, we constrain all designs to have preapodized apertures of a given area. Thus the irradiance in Eq. (4) is the same for all designs—the same amount of light enters the telescope. The spatial frequency axis has been scaled by the length of the equivalent square aperture with area  $A$ . For apertures of other shapes (including the optimal shapes we present below), the point spread function must be properly scaled so that  $I_o$  is the same for each one. For example, an open rectangular aperture with area  $A$  but aspect ratio  $r < 1$  has the point spread function:

$$I(\xi, \zeta) = I_o \left( \frac{\sin \pi \xi / \sqrt{r}}{\pi \xi / \sqrt{r}} \right)^2 \left( \frac{\sin \pi \zeta \sqrt{r}}{\pi \zeta \sqrt{r}} \right)^2.$$

As expected,  $P(0,0) = 1$  as before since the area, and thus throughput, is unchanged but the PSF has become narrower by  $\sqrt{r}$  in the  $\xi$ -direction and wider by  $1/\sqrt{r}$  in the  $\zeta$ -direction.

Important special cases of shaped apertures are the common circular aperture (where the PSF in Eq. (2) reduces to the Hankel Transform) and stretched elliptical apertures. An elliptical aperture has the same effect as the rectangular one above—narrowing the PSF on one axis for the same area (it also has the benefit of being easier to fit into a typical launch vehicle shroud).

As an example, for the circular pupil the two-dimensional Fourier Transform in Eq. (1) reduces to the one-dimensional Hankel Transform. For a circular aperture with area  $A = a^2$  and thus diameter  $a' = 2a/\sqrt{\pi}$ , the irradiance becomes:

$$I(\rho) = I_o P(\rho) \quad (5)$$

where  $I_o$  is again given by Eq. (4) and the point spread function is dependent on the normalized image plane radius only,

$$P(\rho) = \left[ 2\pi \int_0^{1/\sqrt{\pi}} r \mathcal{A}(r) J_o(2\pi r \rho) dr \right]^2. \quad (6)$$

Again, this is slightly different from conventional usage because we are making an equal area comparison to the square result (in other words, a circular telescope with the same area as a square has a slightly larger diameter and thus slightly better resolution). As before, the image plane coordinate  $\rho$  is the angular position in the image plane in units of  $\lambda/a$ .

As with the square aperture, it is straightforward to calculate the PSF for an open circular pupil (where  $\mathcal{A}(r) = 1$ ) to find the modified Airy function:

$$P(\rho) = \frac{J_1^2(2\sqrt{\pi}\rho)}{(\sqrt{\pi}\rho)^2}. \quad (7)$$

Again, for the open pupil and the normalizations used here, the PSF at the origin is unity and  $I(0) = I_o$ . These forms of the PSF for the open apertures will form important benchmarks for comparison as they represent the maximum system throughput for a given size and shape of telescope.

Finally, in some cases it is useful to consider an elliptical aperture (such as the proposed TPF architecture from the Ball Aerospace studies (Kilston 2001)). Omitting the details for brevity, the irradiance for the equal area ellipse can again be put in the form of Eq. (3), where  $I_o$  is the same as Eq. (4),  $a' = 2a/\sqrt{\pi r}$  is twice the semi-major axis ( $y$ -direction) and  $b' = ra'$  is twice the semi-minor axis dimension ( $z$ -direction) (as for the rectangle,  $r < 1$  is

the aspect ratio of the ellipse). The point spread function is then given by:

$$P(\xi, \zeta) = \left| \int_{-1/\sqrt{\pi r}}^{1/\sqrt{\pi r}} \int_{-\sqrt{\frac{r}{\pi}(1-\frac{\pi r y^2}{4})}}^{\sqrt{\frac{r}{\pi}(1-\frac{\pi r y^2}{4})}} \mathcal{A}(y, z) e^{-i2\pi(\xi y + \zeta z)} dy dz \right|^2, \quad (8)$$

where  $\xi$  and  $\zeta$  are in units of  $\lambda/a$ . For a symmetric apodization, where  $\mathcal{A}$  varies with  $y$  only, this simplifies to:

$$P(\xi, \zeta) = \left| \int_{-1/\sqrt{\pi r}}^{1/\sqrt{\pi r}} \frac{\mathcal{A}(y) \sin(\pi \zeta \sqrt{\frac{r}{\pi}(4 - \pi r y^2)})}{\pi \zeta} e^{-i2\pi \xi y} dy \right|^2 \quad (9)$$

While it is possible to find a closed form for the PSF for the open elliptical pupil, the expression is complicated so we omit it here. We do note, however, that this normalization again results in  $P(0, 0) = 1$  for the open, elliptical pupil.

The main advantage of the elliptical pupil is that, for the same collecting area, it allows for a much higher resolution in one axis for planet finding, which we'll call the sensitive axis. It is also easier to fit in a launch vehicle and weighs less than a rectangular telescope with the same resolution in the sensitive axis.

### 3. Optimization Criteria

Qualitatively, it is possible to describe the basic features of an ideal planet finding system. The background at the planet location, either due to the star's PSF or due to scattered light from imperfections in the system, must be low enough to allow detection and characterization of the dim planet. The system throughput must be high enough such that sufficient light from the planet arrives at the detector in a reasonable integration time. In addition, we desire the ability to detect planets as close as possible to the star with a field large enough to find planets in the fewest exposures. Thus, the PSF of the system must be such as to maximize contrast while also maximizing planet throughput in some to-be-determined sense.

It is also useful to distinguish between statistical noise and systematic errors. In the ideal system, the PSF is designed to accomplish detection in a minimum integration time, considering only the statistical noise due to the planet and background. In addition to this, however, systematic effects must also be considered, such as scatter from poor optics ("speckles" in the image plane), manufacturing errors in the coronagraph, telescope pointing errors, thermal stability, and finite stellar image size, to name a few. In this paper, we consider only the statistical noise when designing an optimal coronagraph. We separate the

systematic errors and consider them as errors on the statistical performance, compensated for in either observatory design or by an adaptive optics system. More detailed considerations of these systematic effects will be presented in a later paper.

We also distinguish between planet detection and characterization. For detection, only sufficient signal in a pixel area is needed to determine the existence of the planet. In contrast, for characterization photometry is performed on the planet signal, thus requiring sufficient photons to estimate the planet irradiance to a desired accuracy. For the optimizations here we will focus only on the problem of planet detection, leaving a detailed description of planet photometry and its relationship to the point spread function to another paper.

There are three quantitative criteria we will use for measuring the effectiveness of a planet detection system—*contrast ratio*, *integration time*, and *inner working distance (IWD)*. In the following subsections we detail the relationships among these and the point spread function of the telescope. These can then be used for both comparisons of different designs as well as for direct optimization.

### 3.1. Contrast

Brown and Burrows (1990) introduce the *contrast quotient*,  $Q$ , defined as the ratio of the mean central irradiance of the planet image to the mean surface irradiance of the local background on the detector. The local background is dominated by the value of the diffracted starlight and any scattered light due to imperfections in the optics<sup>1</sup>. Thus, the contrast can be partitioned into two terms— $Q_{ss}$  (for suppressed star) and  $Q_{sc}$  (for scattered starlight). Using  $I_p$  to represent the peak irradiance of the planet and  $I_s$  to represent the peak irradiance of the star, located on the telescope axis,  $Q_{ss}$  is given by:

$$Q_{ss} = \frac{I_p P(0, 0)}{I_s P(\xi_p, \zeta_p)}$$

where  $P(0, 0)$  is the central value of the PSF and  $P(\xi_p, \zeta_p)$  is the value of the PSF at the location of the center of the planet image,  $(\xi_p, \zeta_p)$ . For the planet to be detectable in a reasonable integration time, it is desirable that the contrast ratio  $Q$  be greater than or equal to 1.

It is shown in Brown and Burrows (1990), Brown et al. (2002), and Traub (in preparation) that for the star systems of interest, the flux ratio of the star and planet,  $I_p/I_s$ , is as

---

<sup>1</sup>Here we are ignoring the smaller contribution due to the mean local and exozodiacal light.

small as  $1 \times 10^{-10}$ . In order to achieve a unit contrast ratio,  $Q = 1$ , for the dimmest planet, the final PSF of any planet detecting system must have an *inherent contrast*,

$$C = \frac{P(\xi_p, \zeta_p)}{P(0, 0)} \leq 10^{-10}$$

Note that we are making a distinction here between the commonly used contrast ratio,  $Q$ , and the PSF contrast,  $C$ . The first is an observed quantity dependent upon the particular planetary system being detected and can vary with each observation. The second is a fixed property of the point spread function of the telescope and is a design requirement in our optimizations. We show in the next subsection that the performance of the system is strongly dependent on  $Q$ , the contrast in the observation, with a premium placed on high  $Q$ , thus establishing a requirement on  $C$  in our design process.

The same definition for  $Q$  can be applied to the contrast due to scattered light. Phase and amplitude errors introduced because the primary mirror is not perfect produce a halo of scattered light in the image plane. The mean value of this halo must be less than the peak of the planet image as well, introducing the same requirement on  $Q_{sc}$ . This reduction in scattered light is accomplished via wavefront control using an array of actuators across the diameter of a reimaged pupil. The properties of such a system are described in Brown and Burrows (1990) and Brown et al. (2002). In this paper, we simply assume a fixed (and generally known) background due to both diffracted starlight and scatter, where the scatter has been reduced to the level of the stellar PSF and the system is assumed stable during an observation (some relaxations to these considerations will be discussed below). The total contrast ratio is then given by:

$$Q \equiv \left( \frac{1}{Q_{ss}} + \frac{1}{Q_{sc}} \right)^{-1}$$

We note that, in principal, planet detection is possible with  $Q$  much less than one, assuming a stable diffraction pattern that can be calibrated. However, as we will show next, statistical considerations show that the integration time in order to average the resulting photon noise due to the suppressed star background becomes unreasonably long and systematic considerations introduce severe requirements on system stability (Brown et al. 2002). Of course, a large  $Q$  can also be considered, increasing the requirement on  $C$  for the dimmest planets. However, producing a  $Q$  much larger than 1 offers little improvement in integration time (as this becomes dominated by the planet photon noise and detector noise) while it introduces tremendously difficult requirements on the optics.



### 3.2. Integration Time

The most important performance metric for a planet finding system is integration time. Integration time here is defined as the observing time necessary to guarantee a detection with some confidence. In general this is different than the time needed for photometry (characterization), i.e., for determining the planet irradiance with some relative accuracy. The optimizations we explore strive to minimize the integration time for a planet detection. Minimizing integration time allows for the most possible observations during the mission as well as relaxes requirements on system stability.

The difficulty is that there are many possible ways to define the integration time, each a function of the particular numerical process used to process the planet and star signal. While it is possible to consider many approaches to the planet detection problem (simple detection, photometry, least squares fitting, bayesian decision making, hypothesis testing, and maximum likelihood estimation to name a few), it is beyond the scope of this paper to explore these in detail. Instead, we present two methods that can be used to find an integration time formula for optimization and evaluation. These provide the advantages of being straightforward to compute, being in common use, and providing equations that are easily utilized in an optimization scheme. Both approaches involve defining how to estimate the signal  $S$  and the noise  $N$ . Planet detection is defined as the time at which the ratio  $S/N$  reaches a certain value, typically  $S/N = 5$ .

#### 3.2.1. Detection Time

In the first case, we consider the integration time needed to detect a planet, with irradiance  $I_p$ , in a known and uniform background (suppressed starlight and scatter) with irradiance  $I_b = I_{ss} + I_{sc}$ . Photons are collected in a region of normalized area  $\Delta S$  about the central peak of the PSF, with the arrival of the photons governed by a Poisson process with rate

$$I_p t \Delta P + I_b t \Delta S$$

where  $t$  is the integration time,  $I_p$  and  $I_b$  are the normalized peak irradiances of the planet and background respectively and  $\Delta P$  is the integral of the PSF over  $\Delta S$ :

$$\Delta P = \int_{\Delta S} P(\xi, \zeta) d\xi d\zeta \quad (10)$$

The signal-to-noise ratio is defined as the expected mean planet signal,  $I_p t \Delta P$ , divided by the residual background photon noise remaining after subtracting the mean level of the background,  $\sqrt{I_b t \Delta S}$ . We are confident in a detection when this signal-to-noise ratio achieves

a certain value, e.g.,  $S/N = 5$ . In other words, a bright spot in the image plane is considered a planet when its irradiance is greater than 5 standard deviations above the expected background noise.

The resulting equation for  $S/N$  can be rearranged to provide an equation for the integration time necessary to achieve the desired signal-to-noise ratio:

$$t_1 = \frac{(S/N)^2 P(0,0) \Delta S}{I_p Q \Delta P^2}$$

The integration time is dependent upon the size of the region of the PSF being collected, the shape of the PSF, and the contrast of the observation. There is also a minimum integration time for some collecting area—too small an area results in too few photons and too large an area results in too much background relative to the signal as the PSF is decreasing while the background remains uniform. For most PSF’s considered, this is usually a bit larger than the full-width-half-max (FWHM) point.

### 3.2.2. Characterization Time

A second approach to detection is to perform photometry on the signal, computing an estimate of the planet irradiance. This is equivalent to the planet characterization problem. A straightforward hypothesis test is used to examine the estimated value—if it is larger than the expected error in the estimate by some amount, then we accept the hypothesis that there is a planet at that location. We form a signal-to-noise ratio as the planet irradiance divided by the estimate error and accept the hypothesis for  $S/N$  greater than some amount, again usually  $S/N = 5$ .

There are a variety of ways to perform photometry on the measured signal. Perhaps the most straightforward is to collect photons in some region,  $\Delta S$ , as above. It can be shown that estimating the irradiance based on the total integrated intensity in a region, however, will always produce a longer integration time than the simple detection rule described previously because of the impact of the planet photon noise on the estimate, while providing no benefit to the detection problem. A more common approach to photometry is to fit the known PSF to the measured signal (demodulation) in the same bright central region. This is an efficient and accurate characterization method but again provides little benefit to detection in the case of a known background. In the case of an unknown background, however, the peak detection described above is impractical as it requires subtraction of a known background level. In many cases separate estimates of the background may not be available or we may not have confidence that the background is uniform across the discovery space. The only

alternative for detection in that case is to perform photometry, simultaneously estimating the background and planet irradiance by fitting the PSF, and testing for a planet.

The most straightforward estimation approach in this case is to use a simple unweighted least-squares fitting procedure of the expected planet PSF and a constant background. The details of the two-dimensional estimation are omitted for brevity. The resulting estimate for the planet irradiance is given by:

$$\hat{I}_p = \frac{\Delta S \Delta \alpha \sum_{i,j} \Delta P_{i,j} z_{i,j} - \Delta \alpha^2 \sum_{i,j} \Delta P_{i,j} \sum_{i,j} z_{i,j}}{t \left[ \Delta \alpha \Delta S \sum_{i,j} \Delta P_{i,j}^2 - \Delta \alpha^2 \left( \sum_{i,j} \Delta P \right)^2 \right]}$$

where  $z_{i,j}$  is the measured photon count at the  $i, j$  pixel in the image plane,  $\Delta \alpha$  is the normalized area of each pixel, and  $\Delta P_{i,j}$  is the integrated value of the PSF over the  $i, j$  pixel.

The integration time for this estimate is found by computing the mean square error and dividing by  $I_p$ . Here we need to consider the contribution to the error from both the planet photon noise and the background. The result one obtains, assuming small pixel sizes relative to PSF width, is:

$$t_2 = \frac{(S/N)^2 P(0)}{I_p [\Delta S \Delta^2 P - (\Delta P)^2]} \left\{ \frac{\Delta S^2 \Delta^3 P - (\Delta P)^3}{P(0,0) [\Delta S \Delta^2 P - (\Delta P)^2]} + \frac{\Delta S}{Q} \right\}$$

where  $\Delta P$  is the integral over  $\Delta S$  as in Eq. (10) and:

$$\begin{aligned} \Delta^2 P &= \int_{\Delta S} P^2(\xi, \zeta) d\xi d\zeta \\ \Delta^3 P &= \int_{\Delta S} P^3(\xi, \zeta) d\xi d\zeta \end{aligned}$$

This integration time equation is important both because it quantifies detection in an unknown and uniform background and because of the balance between the PSF and contrast. For a given size region,  $\Delta S$ , this equation will provide a shorter integration time for some  $Q$  than does  $t_1$ . In particular, for small  $Q$  where the background is larger than the planet irradiance, this will provide a shorter detection. However, as an optimization requirement it is much too complicated. We therefore use  $t_1$  in our optimizations and only examine  $t_2$  for comparisons, both because the nominal detection case is for  $Q = 1$  and because the computations are greatly simplified.

### 3.3. Inner Working Distance

Inner working distance (IWD) is defined as the smallest angular separation from the star for which the PSF contrast,  $C$ , reaches the required value of  $10^{-10}$ . In words, it is the closest position to the star at which a detection is possible. For TPF, the coronagraphic system must achieve an IWD at the inner edge of the habitable zone of the most distant candidate systems. A coronagraph with a smaller IWD allows discovery at more distant stars and, consequently, a larger sampling of systems. We note that IWD is not defined as the closest discovery with the integration times quantified above. We recognize that for planets at the very edge of the IWD the PSF may be distorted by the coronagraph, resulting in larger integration times. For the purposes of integration time comparison, we assume a location that is more than one FWHM from the IWD. We use IWD as a separate metric.

### 3.4. Discovery Space

In this context, discovery space refers to the region of the image plane within which a detection is possible—that is, the region where  $C = 10^{-10}$ . For an achievable IWD, this defines the azimuthal region where dark contrast is maintained as well as the outer working distance (OWD) or field of view of possible discoveries. A classical coronagraph, for example, has a full  $2\pi$  discovery space around the central star location. In contrast, the apodized square aperture (ASA) of Nisenson and Papaliolios (2001) has a narrow discovery space around the diagonals of the square (when near the IWD). The single Spersgel pupil (Spersgel 2000) has an even narrower discovery space about the axis in the image plane. For any system with a discovery space less than  $2\pi$ , some rotation of the telescope or optical system will be necessary for detection. This results in an effective increase in the integration time by the number of orientations required to fully scan the habitable zone around the target star.

It is worth observing again at this point that these various metrics are interdependent and a variety of tradeoffs are possible. For instance, shorter integration times can often be achieved by narrowing the search space or smaller inner working distances can be accomplished by greatly reducing the outer working distance or field-of-view of the telescope. It is likely that a visible light coronagraphic version of TPF will be equipped with a variety of stellar suppression masks allowing observations to be modified for each system.

#### 4. Apodized Pupils

In this section we review some proposed apodized apertures for achieving the needed contrast,  $C = 10^{-10}$ . There is a long history of apodization concepts, many dealing with the high contrast imaging problem (Indebetouw 1990; Jacquino and Roizen-Dossier 1964; Watson et al. 1991). A particularly comprehensive review can be found in Jacquino and Roizen-Dossier (1964). Recently, Nisenson and Papaliolios (2001) proposed an apodized square aperture for TPF with separate apodizations in each direction, that is,  $\mathcal{A}(y, z) = \mathcal{A}_1(y)\mathcal{A}_1(z)$ . This square, cross apodization results in a strong attenuation on the diagonals in the image plane. It also greatly simplifies the analysis as the resulting point spread function is the product of the Fourier transforms. For the apodization they proposed sonine functions:

$$\mathcal{A}_1(y) = (1 - 4y^2)^{\nu-1},$$

where  $\nu$  is either 3, 4, or 5. Larger  $\nu$  creates a smaller inner working distance at the expense of throughput.

Fig. 1 shows the image plane response and diagonal cross section of the ASA for  $\nu = 5$ . Unfortunately, the ASA is less than optimal on all the metrics described. Its inner working distance along the diagonal is  $5 \lambda/a$ , deteriorating rapidly for other azimuths. It has a narrow discovery space, and, as shown in Table 1, it has large integration times because of the low throughput, aggravated by the need for many rotations to cover fully all azimuths at the closest IWD.

Alternatively, Slepian (1965) found the compactly supported function that maximally concentrates energy in the frequency domain—the zero-order prolate spheroidal wave function. In other words, it maximizes  $\Delta P$  under a fixed throughput constraint. This is very close to directly optimizing  $t_1$ . Slepian showed how this function can be used to apodize a square aperture in one direction, resulting in the PSF in Fig. 2. This PSF has a better IWD and Table 1 shows that it has a considerably improved integration time. While, like the ASA, it must be rotated for detection, two rotations cover the discovery space and even at twice the integration time it is better. Also notable is its performance for unknown background. Because of its increased sharpness, it is dramatically better in integration time even when using only a small portion of the PSF.

The IWD can be improved off axis by using a circularly symmetric apodization at the slight expense of integration time. Here, Slepian (1965) proposed the use of the generalized prolate spheroidal functions. These functions solve the same optimization problem for circularly symmetric pupils (that is, maximally concentrating the energy in the Hankel trans-

form). Fig. 3 displays the PSF of a Slepian apodization in an equal area circular pupil.<sup>2</sup> This apodization has a full 360 deg discovery zone, so no rotations are necessary, and the smallest IWD. The integration time is considerably smaller than the effective integration time for the square apodizations, where rotations are necessary for a full detection, and on the same order as that for the 1-D prolate spheroidal apodization.

It is possible to further optimize the apodizations. For instance, we can numerically optimize the integration times directly or we can search for apodizations with narrower discovery regions and smaller IWDs. However, because of the fundamental limitations of apodized approaches, this effort is not warranted. Apodized pupils suffer from a severe sensitivity to the quality of the apodization.

In fact, very slight changes to the apodization function, particularly in the dark areas, result in significant degradations in performance. There is currently no available process that can achieve the required accuracy on graded masks. For example, Fig. 4 shows the point spread function for the one-dimensional prolate spheroidal apodization with an extremely small additive error of  $6 \times 10^{-12} \text{ Hz}^{-2}$  (where Hz is in units of  $\lambda/a$ ). Even with this slight error the contrast ratio increases by over an order of magnitude, directly affecting integration time and the ability to detect planets.

## 5. Shaped Pupils

Because of the limitations in manufacturability of apodized apertures, we turn to shaped pupils as an alternative. Spergel (2000) first introduced the concept of using a Gaussian shaped aperture to achieve high contrast on-axis. In Kasdin et al. (2002) an optimal shaped pupil was proposed using the prolate spheroidal wave function. This can easily be seen by modifying Eq. (2) for a shaped aperture rather than an apodization. If  $w(y) \geq 0$  is the height of the opening as a function of  $y$  across the aperture, then one of the transform integrals is easily computed resulting in:

$$E(\xi, \zeta) = \frac{A\varepsilon_A e^{i(\omega t + kR)}}{R} \int_{-1/2}^{1/2} \frac{\sin(2\pi\zeta w(y))}{2\pi\zeta} e^{-2\pi i \xi y} dy$$

The response on axis ( $\zeta = 0$ ) is identical to the one-dimensional apodization equation. Thus, on-axis the optimal solution is the same as the prolate spheroidal apodization discussed above

---

<sup>2</sup>Since the computation of the generalized prolate spheroidal functions is quite involved, these results were obtained by direct numerical optimization.

and shown in cross section in Fig. 2. Fig. 5 shows a single prolate spheroidal pupil of width  $a$  inscribed in a circular aperture and the corresponding point spread function.

The biggest advantage of using a shaped pupil is in the ease of manufacturing. For example, a 1 micron rms error in cutting a 1 cm pupil, something easily achieved in practice, has no effect on the point spread function and thus no effect on contrast. In addition, for the single shaped pupil, the integration time is substantially decreased, with a  $t_1$  of 4.6 and a  $t_2$  at full width of 8.

The most significant drawback of the single prolate spheroidal pupil is the rapid decrease in contrast off axis and thus the very small discovery space near the IWD. It was shown in Kasdin et al. (2002), however, that this discovery space can be increased by using multiple pupil openings with little to no effect on throughput. In what follows, we present a selection of optimal multi-opening pupils to achieve high contrast at minimum integration time. These pupils were computed by numerical optimization using the second author’s LOQO (Vanderbei 1999). The optimization problems are expressed in the AMPL (Fourer et al. 1993) language. AMPL models for any of the designs can be provided on request.

The apodization/mask functions to be found by optimization are represented by discrete approximation over a fine grid. The optimization requirements involve the electric field in the image plane, which is a Fourier integral of the entrance pupil. Those integrals that can be evaluated explicitly are. Those that can’t are approximated by their Riemann sum. After optimization, approximating grids are refined by a factor of 3, the optimal functions are extended to the refined grid, and then figures and metrics are produced from this refined representation. In the following subsections, we present a selection of five optimized pupils. While not necessarily definitive, these pupils indicate the utility of a shaped-pupil optimization approach.

### 5.1. Non-Azimuthally Symmetric Pupils

We begin with pupils that are symmetric with respect to reflection in the  $y$  and  $z$  axes but are azimuthally asymmetric. An example of such a pupil is the prolate spheroidal mask shown in Fig. 5. We consider such pupils with multiple openings. Let the  $j$ -th opening in the upper halfplane be given by two even functions  $v_j$  and  $w_j$ :  $\{y, z : v_j(y) \leq z \leq w_j(y)\}$ . The corresponding opening in the lower halfplane is then given by  $\{y, z : -w_j(y) \leq z \leq -v_j(y)\}$ . Of course, we impose constraints that ensure that these openings do not overlap. The

problem is to maximize throughput

$$4 \sum_{j=1}^n \int_0^{1/2} (w_j(y) - v_j(y)) dy$$

subject to the constraint that the contrast is at least  $10^{-10}$  throughout a pair of annular sectors covering the  $y$ -axis and extending from an IWD  $\rho_0$  to an OWD  $\rho_1$ . For example, in polar coordinates, a 90 deg annular sector over the positive  $y$ -axis is given by

$$\{\rho, \phi : \rho_0 \leq \rho \leq \rho_1, -\pi/4 \leq \phi \leq \pi/4\}.$$

We represent this continuum of constraints by a finite set layed out on a polar grid over this set. This problem is not precisely that of minimizing integration time but is more tractable and achieves essentially the same goal. In any case, we evaluate the quality of our pupils with integration time as the metric.

Using an IWD of 4.2, an OWD of 29.3, and a  $1.0 \times 0.7$  rectangular pupil design space, we get the pupil and PSF shown in Fig. 6. With this pupil, the entire 360-deg habitable zone can be explored in just two integrations (see Table 2). Another pupil designed for a  $1.0 \times 0.4$  elliptical mirror, an IWD of 2.8, and an OWD of 14.8, is shown in Fig. 7. While it requires four integrations to cover 360 degrees, this pupil has such high throughput that the total integration time is lower than for the rectangular one. The drawback of this mask is that the OWD is rather modest. As an alternative, we set the OWD to 67 and found the mask shown in Fig. 8. It has an IWD of 4.5. Integration times are shown in Table 2.

Lastly, in Fig. 9 we show an 8 pupil opening inscribed in a circular telescope. This pupil was designed for the Eclipse mission, a proposed Discovery class mission for detecting extrasolar Jupiter size planets (T.Hull et al. 2002). Because it is searching for brigher and larger planets, the contrast needs to be only  $10^{-8}$  rather than  $10^{-10}$ . This relaxation can be used to improve integration time. Table 2 shows that, as expected, the Eclipse mask has the lowest total integration time.

## 5.2. Azimuthally Symmetric Pupils

The generalized prolate spheroidal apodization function discussed earlier and illustrated in Fig. 3 is the solution to the following optimization problem:

$$\begin{aligned} &\text{minimize} && \int_{\rho_0}^{\infty} P(\rho) d\rho \\ &\text{subject to:} && \mathcal{A}(0) = 1, \\ &&& \mathcal{A}(r) \geq 0, \quad \text{for } 0 \leq r \leq 1/2. \end{aligned}$$



Here,  $\alpha$  denotes a given lower bound on throughput.

Since the habitable zone does not extend indefinitely, it is reasonable to consider changing the integral to one of finite extent, say from an inner working distance  $\rho_0$  to an outer working distance  $\rho_1$ . Furthermore, since apodizations are harder to manufacture than masks, we also can impose the constraint that the function  $A(r)$  be 0-1 valued. The mask then consists of concentric open rings. Using an IWD of 3.5, an OWD of 42.5, and 26 rings, we found the mask shown in Fig. 10. Of course, one must figure out how to support these concentric rings. Embedding them in a transmissive material is one possibility but raises concerns of scatter from imperfections in the material. Alternatively one can use spokes (i.e., a spider) to support them. Initial analysis seems to indicate that using lots of very fine spokes has less impact on the PSF than a smaller number of thicker spokes, however, more study is required.

Integration times for all of the binary masks are shown in Table 2.

## 6. Conclusions

There are several pupil apodization approaches that will achieve the needed contrast for discovery and characterization of extrasolar terrestrial planets—planets within the habitable zone of sunlike stars out to 10 or 20 pc. However, the differences in integration times and discovery space (including inner working distance) are notable and optimization can provide substantial benefit. Traditional pupil apodization is effective, with a circularly symmetric pupil using the generalized prolate spheroidal function of Slepian (1965) by the far the best choice—it has the closest inner working distance for a given area and the shortest integration time while searching a full 360 degrees in one observation. However, we showed that apodized techniques can be very problematic for terrestrial planet finding because of the extreme sensitivity of the PSF to errors in apodization. It is difficult to maintain a contrast of  $10^{-10}$  even with very precise manufacturing control. Instead, we proposed shaped pupils.

Optimal shaped pupils can achieve close to the same or better integration times while being far easier to manufacture using current technology. By taking advantage of elongated apertures, such as rectangles and ellipses, it is also possible to achieve extremely small IWDs for a given area with very little penalty on integration time. While shaped pupil apertures require multiple rotated exposures for planet discovery, the full 360 degrees at the smallest IWD can be searched with no more than 4 exposures at integration times comparable to the best apodizations. For characterization and spectroscopy, the planet location is known and shaped pupils with the narrowest discovery region are optimal as they provide the shortest

integration time for photometry (e.g., the single prolate spheroidal pupil has a normalized  $t_2$  for photometry of 8 and the elliptical shaped pupil a  $t_2$  of 9.5 compared with 34 for the generalized prolate apodization and 48 for the ASA). The results in this paper were only a small sample of the optimal shaped pupils we are studying. For example, we are continuing to pursue designs with additional openings to decrease integration times for the same IWD. It is also possible to decrease integration time for elongated pupils by sacrificing IWD. Because of their promise, further work is warranted to achieve the best possible shapes and to study in more detail the effects of manufacturing error and aberrations.

The authors would like to express their gratitude to our colleagues on the Ball Aerospace and Technology TPF team. We benefited greatly from the many enjoyable and stimulating discussions. This work was partially performed for the Jet Propulsion Laboratory, California Institute of Technology, sponsored by the National Aeronautics and Space Administration as part of the TPF architecture studies and also under contract number 1240729. The second author received support from the NSF (CCR-0098040) and the ONR (N00014-98-1-0036).

## REFERENCES

- M. Born and E. Wolf. *Principles of Optics*. Cambridge University Press, Cambridge, UK, 1999.
- R. A. Brown and C. J. Burrows. On the feasibility of detecting extrasolar planets by reflected starlight using the hubble space telescope. *Icarus*, 87:484–497, 1990.
- R. A. Brown, C. J. Burrows, N. J. Kasdin, S. Kilston, M. J. Kuchner, D. N. Spergel, W. A. Traub, J. T. Trauger, and E. L. Turner. New ability to study extrasolar planets in starlight. *submitted to Nature*, 2002.
- R. Fourer, D.M. Gay, and B.W. Kernighan. *AMPL: A Modeling Language for Mathematical Programming*. Scientific Press, 1993.
- E. Hecht. *Optics*. Addison Wesley Longman, Inc., Reading, Mass., 1998.
- G. Indebetouw. Optimal apodizing properties of gaussian pupils. *Journal of Modern Optics*, 37(7):1271–1275, 1990.
- P. Jacquinot and B. Roizen-Dossier. Apodisation. *Progress in Optics*, 3:29–186, 1964.

- N. J. Kasdin, D. N. Spergel, and M. G. Littman. An optimal shaped pupil coronagraph for high contrast imaging, planet finding, and spectroscopy. *submitted to Applied Optics*, 2002.
- S. Kilston. Terrestrial planet finder review final reports, 2001. [http://planetquest.jpl.nasa.gov/TPF/tpf\\_review.html](http://planetquest.jpl.nasa.gov/TPF/tpf_review.html).
- P. Nisenson and C. Papaliolios. Detection of earth-like planets using apodized telescopes. *The Astrophysical Journal*, 548(2):L201–L205, 2001.
- D. Slepian. Analytic solution of two apodization problems. *Journal of the Optical Society of America*, 55(9):1110–1115, 1965.
- D. N. Spergel. A new pupil for detecting extrasolar planets. *astro-ph/0101142*, 2000.
- T.Hull, J.T. Trauger, and S.A. Macenka. Eclipse architecture. In *Proceedings of SPIE: Astronomical Telescopes and Instrumentation*, 2002.
- W. A. Traub. Extrasolar planets and biomarkers. In K. R. Sembach, J. C. Blades, G. D. Illingworth, and R. C. Kennicutt, editors, *to appear in Hubble’s Science Legacy: Future Optical-Ultraviolet Astronomy from Space*, ASP Conference Series, in preparation.
- R.J. Vanderbei. LOQO user’s manual—version 3.10. *Optimization Methods and Software*, 12:485–514, 1999.
- S. M. Watson, J. P. Mills, S. L. Gaiser, and D. J. Diner. Direct imaging of nonsolar planets with infrared telescopes using apodized coronagraphs. *Applied Optics*, 30(22):3253–3260, 1991.

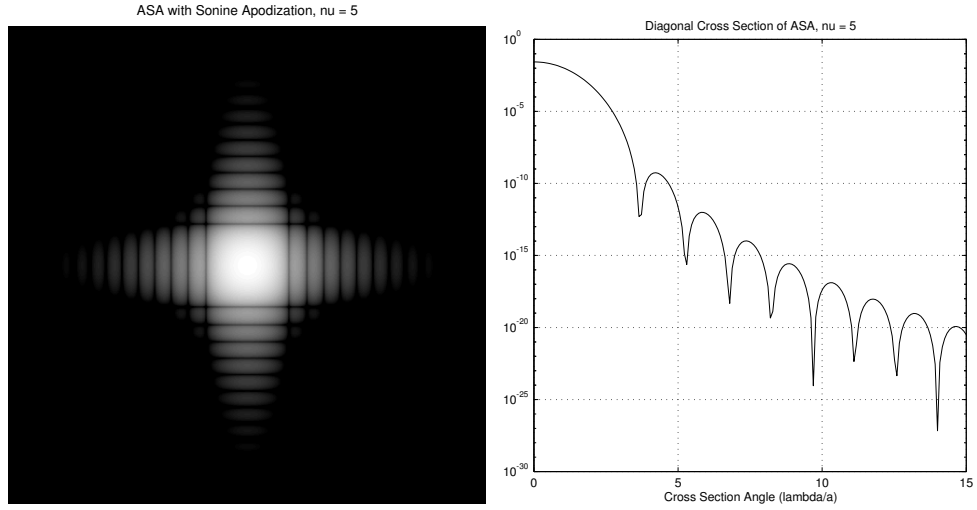


Fig. 1.— *Left* The PSF for a unit-area apodized square aperture (Nisenson and Papaliolios 2001) (ASA) with a sonine apodization ( $\nu = 5$ ) plotted on a logarithmic scale with black areas  $10^{-10}$  below brightest. *Right* Cross section of ASA on diagonal showing inner working distance of  $5\lambda/a$ . This mask has a single-exposure normalized discovery integration time of 25.5 with a small discovery space at the inner working distance (IWD).

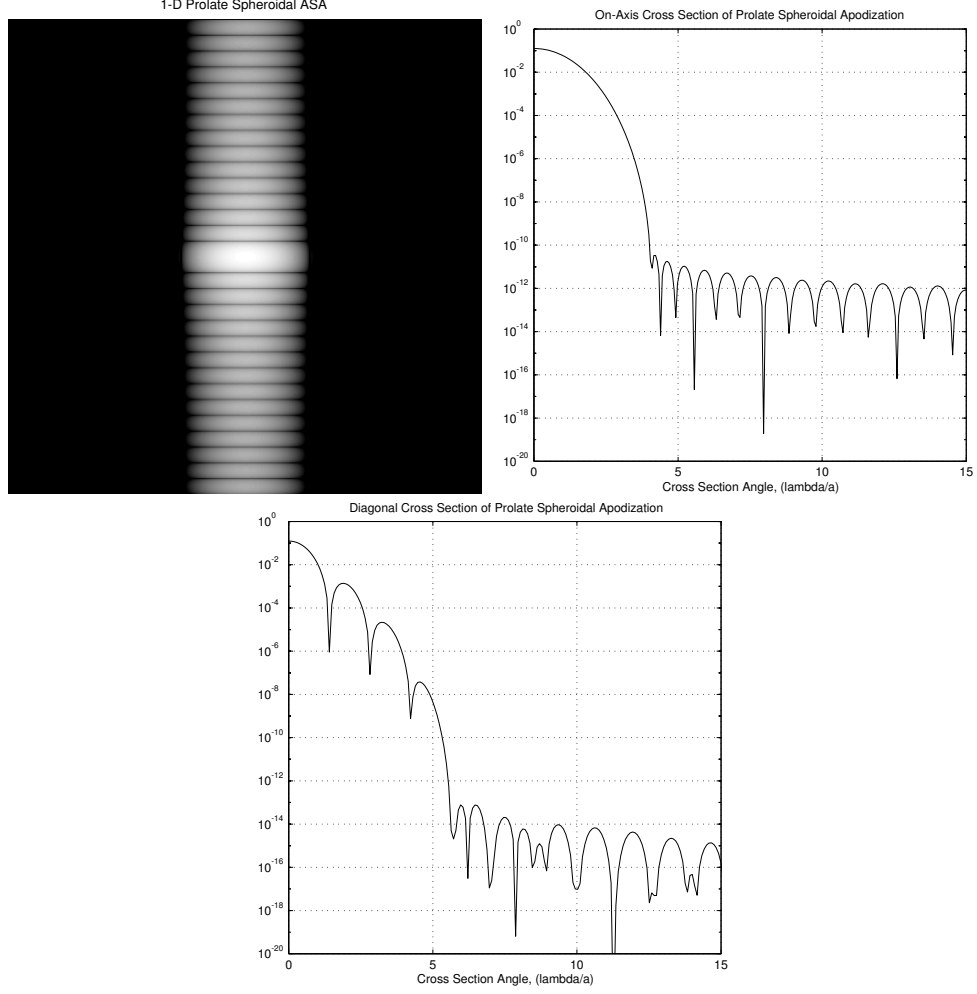


Fig. 2.— *Top Left* The PSF for the one-dimensional prolate spheroidal apodized square aperture (Slepian 1965) of unit area plotted on a logarithmic scale with black areas  $10^{-10}$  below brightest. *Top Right* On-axis cross section of the PSF showing an inner working distance of  $4\lambda/a$ . *Bottom* Diagonal cross section of the PSF showing an inner working distance of approximately  $5\lambda/a$ . This mask has a single-exposure normalized discovery integration time of 9.5 with a discovery space covered by two exposures.

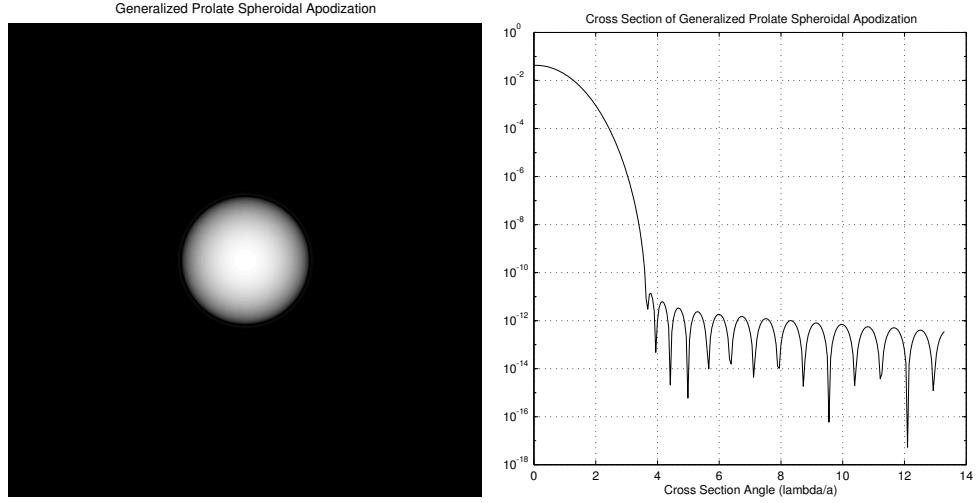


Fig. 3.— *Left* The PSF for the azimuthally symmetric generalized prolate spheroidal apodized circular aperture (Slepian 1965) of unit area plotted on a logarithmic scale with black areas  $10^{-10}$  below brightest. *Right* Cross section of the PSF showing an inner working distance of  $3.5\lambda/a$ . This mask has a single-exposure normalized discovery integration time of 17 that fully covers the discovery space.

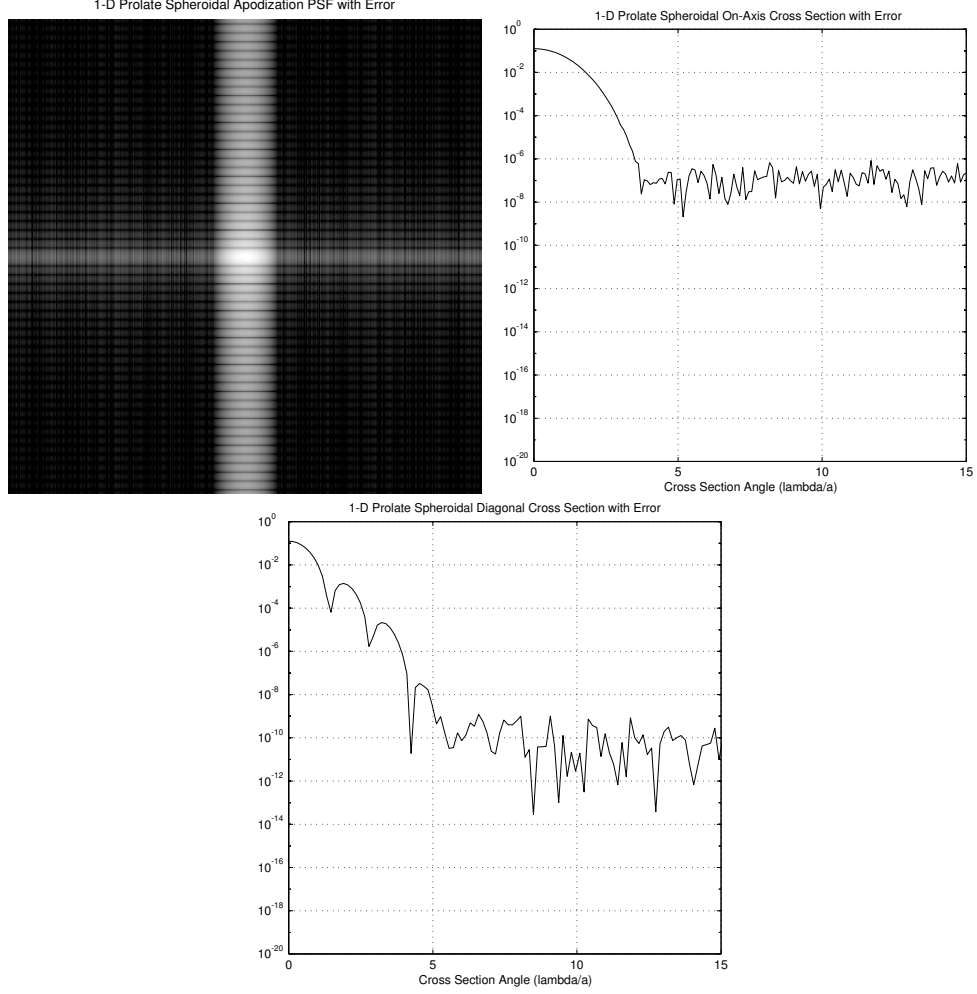


Fig. 4.— *Top Left* The PSF for the one-dimensional prolate spheroidal apodized square aperture (Slepian 1965) with a white apodization error of  $6 \times 10^{-12} \text{ Hz}^{-2}$  plotted on a logarithmic scale with black areas  $10^{-10}$  below brightest. *Top Right* On-axis cross section of the PSF showing four orders of magnitude degradation. *Bottom* Diagonal cross section of the PSF showing a similar degree of degradation.

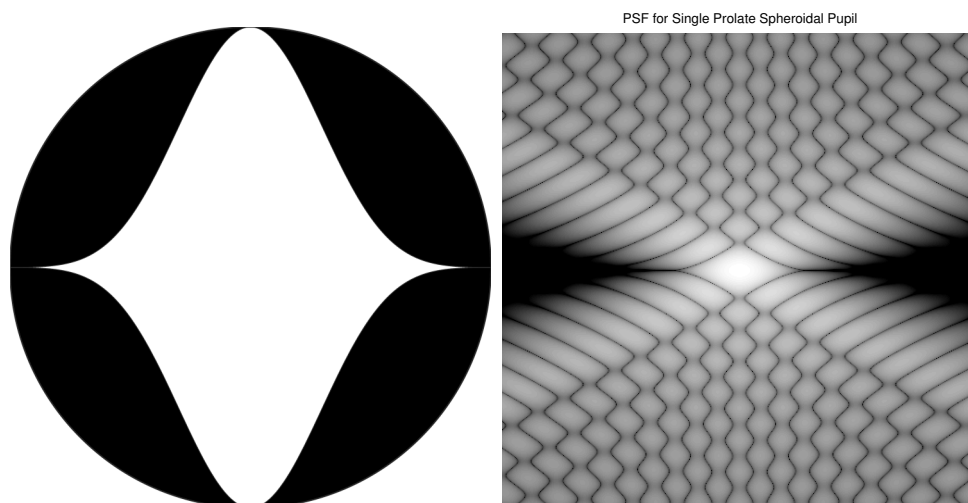


Fig. 5.— *Left* The single prolate spheroidal wave function shaped pupil aperture (Slepian 1965) inscribed in a circular aperture of unit area. *Right* The corresponding PSF plotted on a logarithmic scale with black areas  $10^{-10}$  below brightest. This mask has a single-exposure normalized discovery integration time of 4.6 with a small discovery space at the inner working distance (IWD).



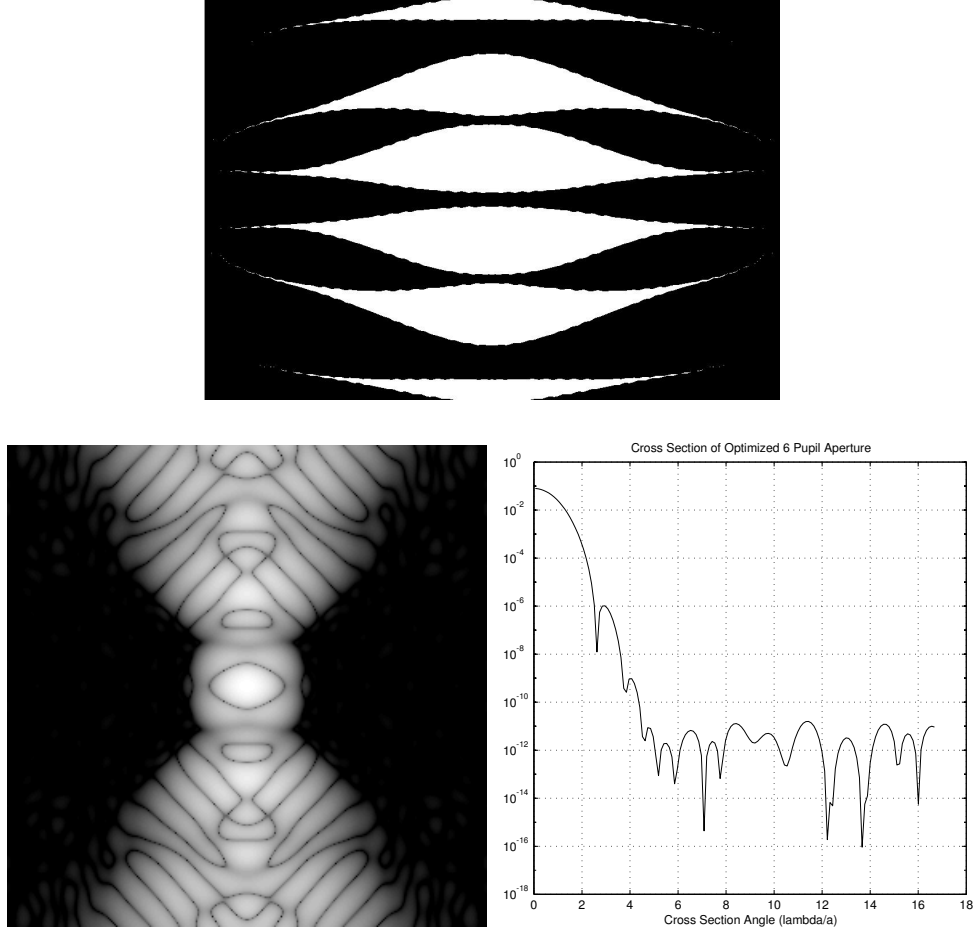


Fig. 6.— *Top* A six-pupil opening inscribed in a rectangular aperture of unit area. *Bottom Left* The corresponding PSF plotted on a logarithmic scale with black areas  $10^{-10}$  below brightest. *Bottom Right* On-axis cross section showing inner working distance of  $4.2\lambda/a$ . This mask has a single-exposure normalized discovery integration time of 15 and requires two exposures to cover the discovery space.

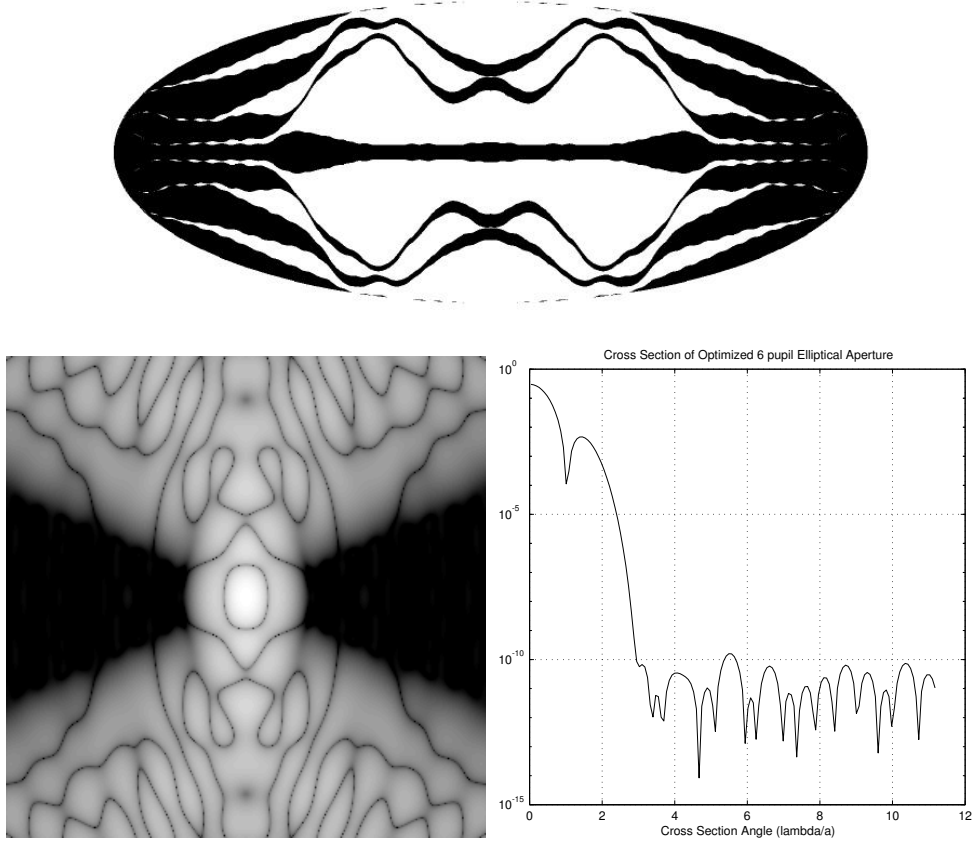


Fig. 7.— *Top* A six-pupil opening inscribed in an elliptical aperture of unit area. *Bottom Left* The corresponding PSF plotted on a logarithmic scale with black areas  $10^{-10}$  below brightest. *Bottom Right* On-axis cross section showing inner working distance of  $2.8\lambda/a$ . This mask has a single-exposure normalized discovery integration time of 6.5 and requires four exposures to cover the discovery space.

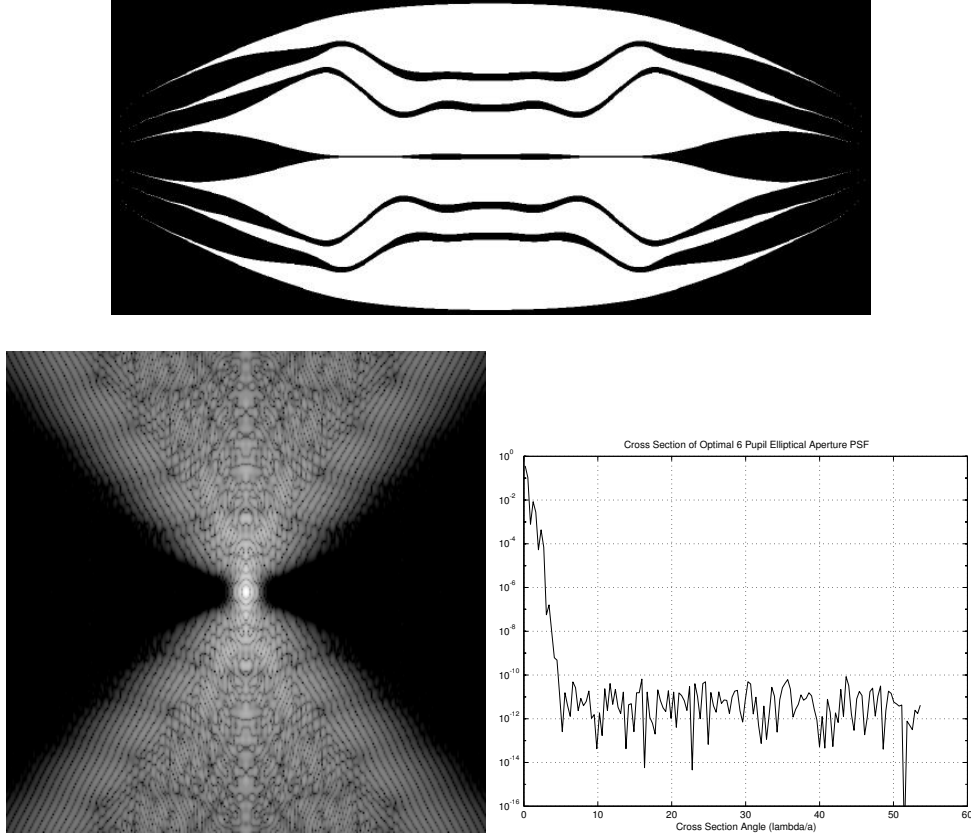


Fig. 8.— *Top* A six-pupil opening inscribed in an elliptical aperture of unit area with larger OWD than the mask in Fig. 7. *Bottom Left* The corresponding PSF plotted on a logarithmic scale with black areas  $10^{-10}$  below brightest. *Bottom Right* On-axis cross section showing inner working distance of  $4.5\lambda/a$ . This mask has a single-exposure normalized discovery integration time of 4.8 and requires four exposures to cover the discovery space.

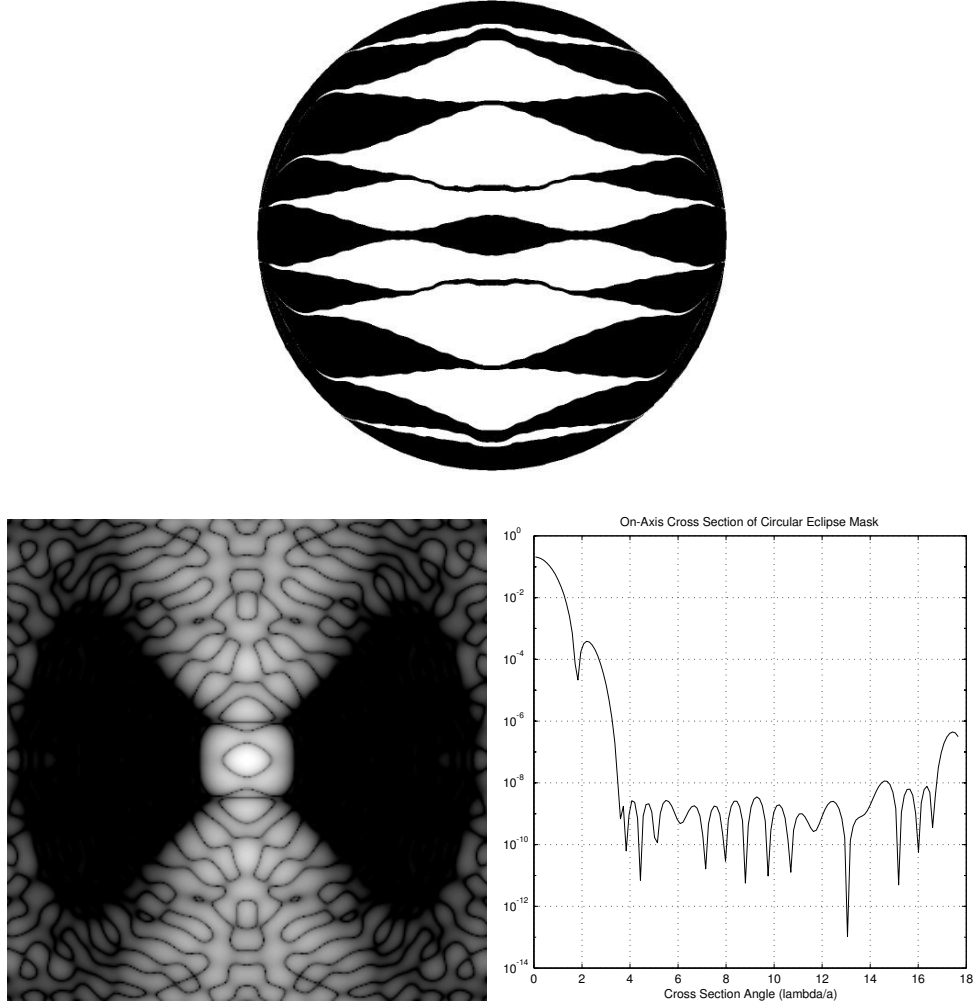


Fig. 9.— An Eclipse Telescope Mask. *Top* An eight-pupil opening inscribed in a circular aperture of unit area. *Bottom Left* The corresponding PSF plotted on a logarithmic scale with black areas  $10^{-8}$  below brightest. *Bottom Right* On-axis cross section showing inner working distance of  $3.5\lambda/a$ . This mask has a single-exposure normalized discovery integration time of 7.9 and requires two exposures to cover the discovery space.

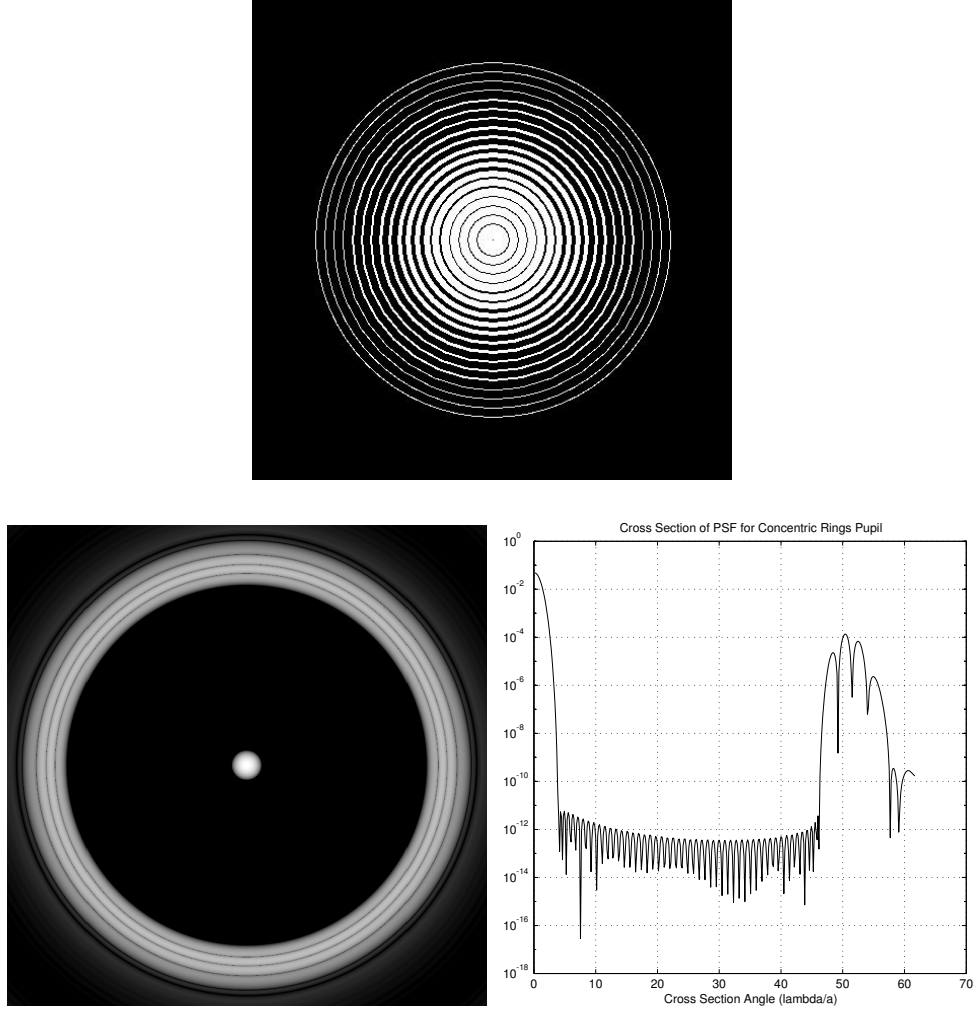


Fig. 10.— *Top* A 26-ring azimuthally symmetric pupil inscribed in a circular aperture of unit area. Black rings are opaque and white rings are open. *Bottom Left* The corresponding PSF plotted on a logarithmic scale with black areas  $10^{-10}$  below brightest. *Bottom Right* On-axis cross section showing inner working distance of  $3.5\lambda/a$ . This mask has a single-exposure normalized discovery integration time of 18 and requires one exposure to cover the discovery space.

<i>Type</i>	$\frac{t_1 I_p}{(S/N)^2}$	$\frac{t_2 I_p}{(S/N)^2}$	<i>IWD</i>	<i>Discovery Space</i>
Sonine, $\nu = 5$	25.5	48	5 (diag), 13 (on-axis)	$< 1/2$
Sonine, $\nu = 4$	19	35	7 (diag) , 15 (on-axis)	$< 1/2$
1-D Prolate	9.5	16.5	5 (diag), 4 (on axis)	$1/2$
Generalized Prolate	17	34	3.5	full

Table 1: Integration time comparisons for four different pupil apodizations in equal area apertures. Integration times are for a single exposure and have been normalized by planet irradiance and signal-to-noise ratio. The integration time  $t_1$  assumes a known background level and is for planet discovery only. The integration time  $t_2$  simultaneously estimates planet and background irradiance by using a region of the image plane slightly larger than the full width of the main lobe.

<i>Type</i>	$\frac{t_1 I_p}{(S/N)^2}$	$\frac{t_2 I_p}{(S/N)^2}$	<i>IWD</i>	<i>Discovery Space</i>
single prolate	4.6	8	4	$\ll 1/2$
6 pupil rectangle	15	21	4.2	$1/2$
6 pupil ellipse small OWD	6.5	9.5	2.8	$1/4$
6 pupil ellipse large OWD	4.8	7	4.5	$1/4$
8 pupil Eclipse	7.9	11	3.5	$1/2$
Concentric Rings	18	25	3.5	<i>full</i>

Table 2: Integration time comparisons for three different shaped pupil apodizations in equal area apertures. Integration times are for a single exposure and have been normalized by planet irradiance and signal-to-noise ratio. The integration time  $t_1$  assumes a known background level and is for planet discovery only. The integration time  $t_2$  simultaneously estimates planet and background irradiance by using a region of the image plane slightly larger than the full width of the main lobe.