One-sided Achromatic Phase Apodization for Imaging of Extra-solar Planets

Weidong Yang and Alexander B. Kostinski

Physics Department, Michigan Technological University, Houghton, MI 49931

ABSTRACT

We propose a new approach to direct imaging of extra-solar planets: one-sided phase apodization. It is based on a discovery that an anti-symmetric spatial phase modulation pattern imposed over a pupil or a relay plane causes diffracted starlight suppression sufficient for imaging of extra-solar planets. Numerical simulations with specific square pupil (side D) phase functions, such as \( \phi(x, y) = a[\ln(1+\varepsilon+2x/D \cdot (1+\varepsilon)-2y/D)] \), demonstrate annulling in at least one quadrant of the diffraction plane to the contrast level of better than \( 10^{-12} \) with an inner working angle down to \( 3.5\lambda/D \) (with \( a = 3 \) and \( \varepsilon = 10^{-3} \)). Furthermore, our computer experiments show that phase apodization remains effective throughout a broad spectrum (60% of the central wavelength) covering the entire visible light range. Phase-only modulation has the additional appeal of potential implementation via active segmented or deformable mirrors, thereby combining compensation of random phase aberrations and diffraction halo removal in a single optical element.

Subject headings: instrumentation: adaptive optics; planetary systems; techniques: interferometric

1. Introduction

The search for extra-solar planets has generated great interest and shows no signs of subsiding. While extra-solar planets may be detected by indirect methods such as an observation of a small wobbling motion of a parent star, direct imaging can do so unambiguously. Furthermore, direct imaging would allow spectroscopic analysis of the planetary/atmospheric composition, possibly, leading to information about life, once the planet is detected (Angel,
Cheng, Woolf 1986; Des Maraise et al 2002). Yet, to date, there have been no reports of direct extra-solar planet imaging.

The difficulty of such imaging stems from the close proximity of planets to their parent stars, resulting in their faint signals being lost in the local “bright stellar halo”, which may be millions (infrared) or billions (near infrared and visible) of times brighter. Brown & Burrows (1990) studied a figure of merit, Q, which is the contrast ratio between a best-case planet and the background of scattered starlight and concluded that even in the case of the Hubble Space telescope (HST), the “halos” caused by the light, scattered from figure errors of the primary mirrors and diffracted from the pupil edge, render the HST unsuitable for extra-solar planet detection. Among the current approaches to overcoming such difficulties, two general directions appear particularly promising: (i) imaging based on infrared interferometers, along the lines originally proposed by Bracewell and MacPhie (Bracewell & MacPhie 1979); (ii) visible light imaging based on the coronagraph concept (originally introduced by Lyot in 1939 in the solar physics context (Lyot 1939)). Because of the longer wavelength, decreased angular resolution in the infrared region must be overcome by spatial separation but in this paper we shall confine ourselves to monolithic telescopes.

Imaging in the visible light allows a monolithic modest-sized telescope to reach a satisfactory resolution angle. However, strict tolerances on the mirror surface figure errors ($\leq 10^{-4}\lambda$) over a wide dynamic range, the required diffraction side lobe levels below $10^{-9}$ (Davies 1980; Kenknight 1977) and the need for a small inner working of a few $\lambda/D$, are daunting. Nevertheless, recent developments in phase sensing and control technology will enable the figure errors of up to 100 cycles/aperture be controlled under $1 \AA$ per 100 hours (Trauger et al. 2002a; Shaklan, Moody & Green 2002; Green et al 2002). Therefore, coronographic imaging in the visible light region is feasible (Malbet, Yu & Shao 1995).

As to the diffraction side lobe removal, in addition to the Lyot coronograph, several promising approaches have been developed and intensively studied in the past few years. These can be broadly classified into two types. The first is based on the modulation of the phase or amplitude of the star diffraction pattern on a focal plane, which includes: Roddiers’ phase mask (Roddier, F. & Roddier, C. 1997), Rouan et al’s four quadrant phase mask (Rouan, et al. 2000), the Kuchner & Traub’s band-limited mask (Kuchner & Traub 2002), and the Kuchner & Spergel’s notch filter mask (Kuchner & Spergel 2003). The second type is based on the modulation of the amplitude transmission function of the pupil (pupil amplitude apodization) (Jacquinot & Roizen-Dossier 1964) as well as pupil shape “apodization”. This field has undergone an explosion of activity recently, which has included several exciting results including: Angel et al’s ring-like binary mask (Angel, Cheng, Woolf 1986); Kadsdin et al’s “eye-like” shaped binary mask (Kasdin et al. 2003); Vanderbei et al’s spider binary
mask (Vanderbei, Spergel & Kasdin 2003); Nisenson et al’s amplitude-apodized pupil mask (Nisenson & Papaliolios 2001); Guyon’s amplitude pupil apodization by beam reshaping (Guyon 2003). Besides the specific mask designs mentioned above, the papers by Gonsalves and Nisenson (Gonsalves & Nisenson 2003) and by Aime et al (Aime, Soummer & Ferrari 2002) addressed a general optimization analysis for coronagraph-type systems.

While pupil shape and amplitude’s spatial distribution as well as focal plane amplitude and phase masks have all been explored, to the best of our knowledge, the possibility of using pupil phase apodization for high contrast imaging has not been considered. This is understandable. For example, an early influential review of apodization techniques by Jaquinot and Roizen-Dossier (Jacquinot & Roizen-Dossier 1964) contains a section, entitled “Impossibility of Apodising by a Pupil Phase Plate”. The argument presented there seems eminently reasonable and the proof relies on the fact that a class of side-lobe optimization problems yields real functions. Nevertheless, in this paper we re-examine the question. Why?

High contrast imaging through pupil phase-only spatial modulation, if realizable, might have the following advantages over the other techniques: (i) no loss of light energy when going through the pupil, which shortens the integration time; (ii) it is easier with current technology to sense and control the phase than it is to control the transmission rate, and the phase errors caused by the phase modulation element can be corrected by active (that is, at least slowly adaptive) optics; (iii) the strict tolerances on precise fabrication of specific shapes and/or transmission control can be relaxed and the entire imaging system design simplified.

2. The Proposal

Henceforth, we shall interpret phase apodization more broadly than usual, that is, as a general spatial phase modulation across the pupil plane which yields significant “improvement” in the focal plane energy distribution for imaging of a faint companion. Thus, we shall allow main lobe reduction and shift, asymmetric apodization, etc. A brief preliminary account of the asymmetric phase apodization can be found in Yang & Kostinski (2004).

Let us begin by asking whether phase apodization patterns exist which can remove the side lobes only approximately but down to a sufficiently low level. In order to conduct a systematic search with reasonable computational time, we chose the square pupil case because the separation of variables assumption renders the problem effectively one-dimensional as detailed in the following section. We were further motivated by the recent work in Nisenson & Papaliolios (2001) who “revived” the square pupil for extra-solar planet detection. In
addition, the pupil shape optimization work reported in Kasdin et al. (2003) suggested to
us the idea of *partial* side lobe removal in a focal plane, at any given time. This led to con-
sideration of anti-symmetric (odd) phase functions which proved to be the key as detailed
in the next section. In summary then, our proposal consists of the following elements:

- consider phase-only spatial modulation pattern across the pupil
- consider square pupil and assume separation of variables so that effectively one-dimensional
  problem can be examined in a semi-analytic manner
- use odd (anti-symmetric) phase functions so that half of the one dimensional focal
  plane pattern can be suppressed (the diffraction pattern can then be “switched” to
  another quadrant in a sequential manner.)

The last “ingredient” is based on the mathematical observation concerning symmetry
of a certain class of Fourier transforms, as discussed next.

3. Theoretical Motivation

For the sake of simplicity, we begin with the one-dimensional case and generalize to the
two dimensions (square pupil) in later sections. Let us denote the spatial phase over the
(1D) pupil as \( \phi(x) \) and the transmission function as \( T(x) \). Then the light on the diffraction
plane would be (Goodman 1996)

\[
E(\eta) = \mathcal{F}\{ T(x) e^{i \phi(x)} \}
\]

where \( x \) is the coordinate in the pupil plane, \( \eta \) is the coordinate in the diffraction plane,
and \( \mathcal{F}\{ \} \) denotes the Fourier Transform Operation.

An equivalent expression is

\[
E(\eta) = \mathcal{F}\{ T(x) \} \otimes \{ \mathcal{F}\{ \cos[\phi(x)] + \sin[\phi(x)] \} \}
\]

where \( \otimes \) denotes the convolution. For a one-dimensional “pupil” of width \( D \), the trans-
mission function \( T = 1 \) for \( |x| \leq D/2 \) and \( T = 0 \) otherwise. The Fourier Transform of \( T(x) \)
is \( \text{Sinc}(\eta) \equiv \sin(\pi \eta)/\pi \eta \) which is an even and real function. As is well known from Fourier
analysis, if \( \phi(x) \) is an even and real function, \( \mathcal{F}\{ \cos[\phi(x)] \} \) is also an even and real function
and, therefore, $\mathcal{F}\{\sin[\phi(x)]\}$ yields an even and imaginary function. The light intensity on the diffraction plane would be simply the sum of square of two real functions:

$$I(\eta) = \{\text{Sinc}(\eta) \otimes \mathcal{F}\{\cos[\phi(x)]\}\}^2 + \{\text{Sinc}(\eta) \otimes \mathcal{F}\{\sin[\phi(x)]\}\}^2$$  \hspace{1cm} (3)

Here one observes, that to generate a desired dark region (region of interest in a focal plane) by using an even phase function requires such $\phi(x)$ that both of the Fourier Transforms ($T(x)\cos[\phi(x)]$ and $T(x)\sin[\phi(x)]$) in this region will have to nearly vanish separately. However, when $\phi(x)$ is an odd function, $\mathcal{F}\{\cos[\phi(x)]\}$ is still be an even and real function but $\mathcal{F}\{\sin[\phi(x)]\}$ yields an odd and real function. In this case, the light intensity in the diffraction plane would be square of the sum of two real functions (as opposed to the sum of the squares):

$$I(\eta) = \{\text{Sinc}(\eta) \otimes \mathcal{F}\{\cos[\phi(x)]\}\} + \text{Sinc}(\eta) \otimes \mathcal{F}\{\sin[\phi(x)]\}\}^2$$  \hspace{1cm} (4)

This expression suggests that it might be possible to generate a dark region by using odd phase functions and attempting “destructive interference” between Fourier Transforms of $T(x)\cos[\phi(x)]$ and $T(x)\sin[\phi(x)]$ in this region. Guided by this argument, we explored a set of odd (anti-symmetric) phase functions and found some that yield sufficiently deep light reduction on half of the $x$-axis.

4. 1-D Examples of One-sided Phase Apodization

Because of exceptionally deep reduction ability and smaller inner working angles, let us consider the following odd (anti-symmetric) phase functions

$$\phi_1(x) = a \cdot \tan[(0.5 - \varepsilon) \cdot 2\pi x/D]$$  \hspace{1cm} (5)

$$\phi_2(x) = a \cdot \ln\left(\frac{1 + \varepsilon + 2x/D}{(1 + \varepsilon) - 2x/D}\right)$$  \hspace{1cm} (6)

where $\varepsilon$ is a small (positive) parameter, defined so that the phase value is finite at the edge of the pupil. The shapes of these two phase functions are illustrated in Figure 1 where the parameters are set as $a = 1$ and $\varepsilon = 0.005$ for $\phi_1(x)$ while for $\phi_2(x)$, $a = 3$ and $\varepsilon = 0.001$. (We shall defer discussion of computational devices needed to avoid severe aliasing in the simulations and the selection of the $a$ and $\varepsilon$ parameters to a later section.)
The annulling effect by the destructive interference described in equation 4 is illustrated in Figure 2 and Figure 3. The example of a diffraction pattern caused by phase function $\phi_1$ with $a = 1$ and $\varepsilon = 0.005$ is shown in Figure 2. Similarly, Figure 3 displays the diffraction pattern for the case of the phase function $\phi_2$ with $a = 3$ and $\varepsilon = 0.001$. In order to “zoom in” and see the details, in Figure 4 we demonstrate the annulling effect on $\log_{10}$ scale. It can be seen that for $\phi_1$, a reduction level region of lower than $10^{-4.5}$ can be obtained at about $3.5\lambda/D$ distance from the shifted peak and $10^{-5}$ can be obtained at about $4.5\lambda/D$ (thicker solid line). The second example, $\phi_2$, produces an even sharper annulling effect, shown by the dashed line. We see that the level of lower than $10^{-5.5}$ can be reached at distance only about $2.5\lambda/D$ from the shifted peak. (Jumping ahead a bit, we inform the reader that the reduction effect is squared in two dimensions). It should be pointed out that the parameters used in the above examples are not necessarily optimal and we now proceed to discuss the relations between performance and the parameters.

5. 1-D Focal Plane Performance, Parameter Selection, and Sampling

The performance of the above phase functions is described by the point spread functions (PSFs), they produce. For example, it can be seen from Figure 4 that the peak is broadened, shifted, and lowered, while the light on the left half axis is reduced. The reduced intensity level along the negative half axis is not constant. To investigate the relation between this level and the function parameters more generally, we define the relative reduced intensity level as the maximum relative intensity value within the $25th \lambda/D$ and the $20th \lambda/D$. The corresponding curves are shown in Figure 5, labelled (a) and (c). The curve (a) corresponds to $\phi_1$ while (c) corresponds to $\phi_2$. It can be seen that the reduced intensity level is strongly dependent on the parameters $a$ and $\varepsilon$. To examine the relation between the lowered peak power and the parameters, we still use the term “Strehl ratio” to represent the ratio of the lowered peak power to the peak power when no phase function is applied, despite the fact that the peak shifts a bit. In Figure 5, the corresponding curves are labelled as (b) and (d), where (b) is associated with $\phi_1$ and (d) with $\phi_2$.

Figure 6 shows relations between the shifted distance of the peak and the function parameters. It can be seen that the shifted distance has a strong and almost linear dependence on the parameter $a$ and a relatively weak dependence on the parameter $\varepsilon$, within the selected ranges. The curves in the above Figure 5 and Figure 6 can be used in phase design based on given specifications and will be used in a later section to give qualitative analysis on the broad bandwidth performance.

In addition to anti-symmetry, the phase functions in equation 5 and equation 6 have
another common qualitative features: a substantial and rapid rate of change near the edge of the pupil, as shown in Figure 1. This change of the phase near the edge contributes to the annulling effect along the negative half of the diffraction axis. In order to adequately capture this effect in the simulation, a large number of pixels over the pupil may have to be sampled. We can obtain a crude bound by the following argument.

Let $N$ pixels be sampled over the pupil $D$. Then, the sampling theory suggests choosing the sampling rate $N/D$ of, at least, twice the maximal spatial frequency of our phase functions. If local frequency $f = \frac{1}{2\pi} \cdot \frac{d\phi(x)}{dx}$ is used to estimate the frequency, then it can be expected that the $N$ will have to satisfy $N \geq \frac{D}{\pi} \cdot \frac{d\phi(x)}{dx}|_{max}$. Figure 7 shows the relationship between the estimated minimum number of sampled pixels $N_{\text{min}}$ and the parameter $\varepsilon$ for the phase functions. For smaller $\varepsilon$, more sampling pixels are required over the pupil. To insure the stability of the results in a rather large range, say, 2000 diffraction rings in 1-D and make the code applicable for all $a$ and $\varepsilon$ used, in all of the above simulations, at least 12800 pixels were sampled over the pupil $D$. (This is not to be confused with the issue of number of elements needed to implement the results in, say, active optics - as we shall see below, a much smaller number of elements is sufficient to attain required performance).

6. Phase Modulation for the Square Pupil

Separation of variables permits a straightforward application of our 1-D phase functions along with the side lobe reduction results to the square pupil, which as we mentioned earlier, has already generated a great deal of interests in the field imaging of extra-solar planet imaging (Nisenson & Papaliolios 2001; Kasdin et al. 2003). There are two ways to proceed to the square pupil. One way is to apply the 1-D phase function along one axis only, where the light field over the 2-D square pupil would be $T(x) \cdot T(y) \cdot e^{i\phi(x)}$. The diffraction light intensity would then be given by $I(\eta, \xi) = I(\eta) \cdot \text{Sinc}(\xi)$, where $I(\eta)$ is the 1-D diffraction field intensity determined by equation 4 in which the odd phase delay is applied. For the phase functions $\phi_1$ and $\phi_2$, the contrast level of $10^{-10}$ is reached when the observation position is about $15\lambda/D$ away from the optical axis as can be estimated from Figure 4.

A better alternative, however, is to apply the phase function along both of the $x$ and $y$ axes, in which case the light field over the 2-D square pupil becomes $T(x) \cdot T(y) \cdot e^{i[\phi(x)+\phi(y)]}$. In this case, the diffraction light field intensity is given by $I(\eta, \xi) = I(\eta) \cdot I(\xi)$. The essential advantage gained is that one quadrant of the diffraction plane can experience twice the 1-D reduction and do so at a closer separation angle. This is illustrated in panel (a) of Figure 8 where we show logarithmic intensity image produced by the phase function $\phi_1(x) + \phi_1(y)$. The panel (b) of Figure 8 displays relative intensity along the diagonal. It can be seen that a
deep reduction region is obtained in the second quadrant and the $10^{-9}$ level can be reached at distance $4.5\lambda/D$ and $10^{-12}$ at about $7\lambda/D$. Panels (c) and (d) of Figure 8 demonstrate that the level of $10^{-12}$ can be reached at the distance of about $3.5\lambda/D$ when $\phi_2(x) + \phi_2(y)$ is applied to the square pupil. These results are quite good but do they only hold for a single wavelength? Fortunately, the method is robust as we now demonstrate.

**Bandwidth Tolerance**

Let us consider the case where the required phase delay is realized by a reflecting mirror or a transmission phase plate. If the phase plate has a uniform and homogenous refractive index $n(\lambda)$ and a geometric thickness $d(x)$, then it generates a phase delay of $\phi(x) = \frac{2\pi}{\lambda} \cdot [n(\lambda) - 1]d(x)$. If the reflecting mirror has a geometric shape $h(x)$, then it generates a phase delay of $\phi(x) = -\frac{2\pi}{\lambda} \cdot 2h(x)$. For notational simplicity, we shall use a common term $G(\lambda)$ to represent either the term $[n(\lambda) - 1]$ for phase plate case or the term $-2$ for the reflecting mirror, and $H(x)$ to represent generally the geometric functions $d(x)$ or $h(x)$. Then, $\phi(x)$ is expressed simply as $\phi(x) = \frac{2\pi}{\lambda} G(\lambda) H(x)$. To generate the phase function that works on central wavelength $\lambda_0$, we set the geometric function to $H(x; \lambda_0) = \frac{\lambda_0}{2\pi G(\lambda_0)} \phi(x)$, where the 2nd argument in $H(x; \lambda_0)$ indicates that the geometric function is designed for the central wavelength $\lambda_0$. When light of wavelength $\lambda$ goes through this phase delay element, the geometric function $H(x; \lambda_0)$ generates the phase delay given by

$$
\phi(x, \lambda) = \frac{\lambda_0}{\lambda} \frac{G(\lambda)}{G(\lambda_0)} \phi(x)
$$

and we see that the phase delay at the new wavelength equals the original phase function $\phi(x)$ multiplied by a factor $\frac{\lambda_0}{\lambda} \frac{G(\lambda)}{G(\lambda_0)}$. This factor is, in fact, equivalent to the parameter $a$ in the phase function formulae 5 and 6. For phase plate with positive dispersion materials or for the reflecting mirror, the term $\frac{\lambda_0}{\lambda} \frac{G(\lambda)}{G(\lambda_0)}$ decreases with increasing $\lambda$. Therefore, based on parameter relations of Figure 5, we see that a 50% ”red-shift” in wavelength causes less than 1 order of magnitude increase in the reduction level, while a 50% ”blue-shift” in wavelength causes less than 1 order of magnitude decrease in the reduction level. This is why our phase modulations tolerate $0.6\lambda_0$ bandwidth in the simulations shown in Figure 9 and still keep a small inner working angle and low reduction level.
Phase and Shape Errors

The errors in the phase-only spatial modulation scheme are likely to come from two sources: imperfections in the phase function and perturbations in the pupil boundary. Let us begin with the former.

The phase errors caused, for example by imperfect manufacturing, scatter light into the dark region and, in doing so limit the reduction level. Assuming that phase errors satisfy \( \delta\phi(x, y) \ll 1 \), the light field \( E(\eta, \xi) \) on the focal plane is given by

\[
E(\eta, \xi) \approx \mathcal{F}\{T(x, y) \cdot e^{i\phi(x, y)}[1 + i\delta\phi(x, y)]\}
\]

and the light intensity is a square of the sum of the ideal field and the noise field caused by phase errors. However in the dark region, the ideal field is extremely low and the noise field dominates. Then, the light intensity in the dark region is given by

\[
\delta I(\eta, \xi) \approx |E_{\text{ideal}}(\eta, \xi) \otimes \mathcal{F}\{\delta\phi(x, y)\}|^2.
\]

and the integration of the noise intensity over the focal plane yields the phase error variance over the pupil as follows (the transmission function \( T(x, y) \) over the pupil is a rectangular unit step function):

\[
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta I(\eta, \xi) d\xi d\eta \approx \int_{-D/2}^{D/2} \int_{-D/2}^{D/2} |\delta\phi(x, y)|^2 dxdy.
\]

Spectral content can be important in considering the phase error tolerance, e.g., phase errors of spatial frequencies from 0.03-0.5 cycles per centimeter (about 5.4-90 cycles per aperture for Eclipse design) are considered critical for imaging Jovian planets (Trauger et al. 2002a,b) and the expected noise intensity level (relative to the peak power) is under \( 10^{-9} \).

Therefore, to estimate the phase error requirement for the phase modulated square pupil (side \( D \)) by function \( \phi(x, y) = a \cdot \ln\left[\frac{(1+\varepsilon)2x/D}{(1+\varepsilon)2y/D} \cdot \frac{(1+\varepsilon)2y/D}{(1+\varepsilon)2x/D}\right] \) (\( a = 3 \) and \( \varepsilon = 0.001 \)), we assume a flat noise intensity level in the critical spatial frequency region. This results in integration from 5.4\( \lambda /D \) to 90\( \lambda /D \) (LHS of equation 10), and yields \( 10^{-9} \cdot I_0 \cdot (\lambda /D)^2 \cdot (90^2 - 5.4^2) \) where \( I_0 \) is the light peak intensity with phase modulation. But, from panel (c) of Figure 5 we obtain \( I_0 \sim 0.3 \cdot [D^2 \cdot (D/\lambda)^2] \) where \( D^2 \cdot (D/\lambda)^2 \) is the peak power of the square pupil case without phase modulation (Born & Wolf 1999). This results in phase errors within the critical spatial requency region below \( 15.5 \times 10^{-4} \text{rad rms} \) or below \( 2.5 \times 10^{-4} \lambda \text{ rms} \). This is feasible with current technology as reviewed in the Introduction.
Let us next address the precision requirements for the pupil shape. The rough edge (pupil boundary) is illustrated in Figure 10 where the actual edge is formed by small concave and convex perturbations around the ideal boundary, forming randomly sized and shaped “peninsulas” and “bays”. The “bays” let more light pass through than in the ideal case. The “peninsulas”, of course, block the light. The blocked light can be regarded as a superposition of straight light and π-shifted light. Hence, one can view the peninsulas as letting more “π-shifted” light through the pupil. Then, one can argue that light exiting the actual pupil is a sum of the ideal pupil light and that due to a chain of “peninsulas” and “bays”. Therefore, because of the Fourier transform additivity, the light field in the image plane is the sum of the associated individual Fourier Transforms. To make further progress, let us invoke a probabilistic argument.

Since the size of each of the chain elements is much smaller than the size of the pupil, the diffraction cores are wide-spread, and are much larger than that due to the ideal pupil, and are randomly shifted in phase and position. Therefore, it is reasonable to treat the core light from each of the “peninsulas” and the “bays” as uncorrelated noise sources. Then, this noise intensity due to the chain can be estimated as the sum of the diffraction intensities from each of the chain elements, neglecting the interference cross-terms between elements of the chain because of the randomness in shape, size and π phase shift. This picture allows deduction of a scaling rule by the following, rather general, argument.

Let the ideal pupil area be $A$ and a circumference $L$. Then, the ideal pupil diffraction peak power scales as $\sim A^2/\lambda^2$. If the characteristic length of the chain elements is $l$, the background noise, similarly, scales as $\sim L \cdot l^3/\lambda^2$. Based on this scaling, the background noise $n$ relative to the peak power of the ideal pupil diffraction peak is

$$n \sim L \cdot l^3/A^2$$

(11)

and, based on equation 11, we estimate that for a square pupil of, say, width 0.1 meter, a requirement on the relative noise background of about $10^{-11}$ can be satisfied by confining boundary errors to less than 10\(\mu\)m - a quite feasible task.

**Diffraction-limited Planet Imaging by Combining Coronagraph and a Conjugate Phase Plate**

As in most of the coronagraph and pupil amplitude apodization techniques, our phase-only pupil modulation also lowers and broadens the core of the on-axis stellar image. Since the phase modulation is applied in the pupil or relay plane, the image from an off-axis
planet will have the same structure as the on-axis stellar image. However, unlike the other techniques, phase-only modulation conserves the light energy. Indeed, the light energy is not absorbed or blocked but is spread into the quadrant where the constructive interference occurs as shown in Figure 9. Is it possible to remove the unwanted parts and then use the principle of wave front reversal (phase conjugation) to restore the desired parts of the image?

One possibility of restoring the diffraction-limited image of the planet, is to use an occulting mask to block the star image first, and then add a conjugate phase modulation element on the next relay pupil plane to compensate (reverse) the phase. For example, consider a schematic layout as shown in Figure 11, where two pupil relay planes are used to put the conjugate pair of phase plates. Also, an occulting mask is placed in the first image plane so that it can block the image of the star. The restored images can be obtained from the final image plane. In Figure 12 we simulate and compare the detected planet’s image before and after the combination of occulting mask and conjugate phase element. In the simulation, the planet-to-star ratio is $10^{-9}$ and the planet image is located at $10\lambda_0/D$ away from the star along the diagonal. An occulting mask, covering the 3 quadrants of the image plane is used. It can be seen that the conjugate phase element moves the light energy in the strong side lobes back into the image main lobe and the broadened image is restored. It should be noted that the peak power is not completely restored (a complete restoration should have the peak power returned to $10^{-9}$ in this example). This is due to the mask that also covers some amount of the broad side lobes from the planet, near the mask boundary. Redesigning the mask may increase the restored peak power.

7. Concluding Remarks

We have proposed theoretically and provided specific numerical examples and simulations to demonstrate the possibility of using phase-only spatial pupil modulation to reach the goal of direct imaging the extra-solar planets. The results show that phase modulation can provide an alternative method for high contrast imaging within a rather large dynamic range in terms of both the observational field-of-view angle and the spectral bandwidth. It should be emphasized that there might be a variety of odd phase delay functions that can be used for this purpose. This flexibility implies likely tolerance of phase shape deviation and could help reduce the difficulties in phase realization or manufacturing.

There may be several ways of implementing the phase delay function in a real system. In previous sections, for the convenience in analysis, it is assumed that a fixed optical element, such as a reflecting mirror or a phase plate, is used to generate the phase delay function. However, the main difficulty is likely to be one’s ability to manufacture the shape precisely.
Consider, for example, a 0.1 meter phase plate or mirror. Our phase curve of Figure 1 corresponds to the edge heights of a few microns and edge slopes on the order of $10^{-7}$ and $10^{-3}$ for $\phi_1$ and $\phi_2$, respectively (for visible light). Such a shape can be manufactured with computer-controlled surface figuring techniques, e.g., Elastic Emission Machining (Mori, Yamauchi & Endo 1987), Fluid Jet Polishing (Fähnle, Brug & Frankena 1998), Ion Beam Milling (Drueding et al 1995), Wet-Etch Figuring (Rushford et al. 2003), etc. However, the precision of the surface shape is, at best, on the order of a few nanometers and scattering due to figure errors then limits the contrast to about $10^{-6}$. This is not sufficient for direct imaging extra-solar planets by a moderate telescope in visible light.

Thus, as in all other coronagraph or pupil apodization approaches, reaching lower levels will require employment of high density active mirrors. Based on experiments conducted at JPL (Trauger et al. 2002a,b; Hull et al. 2002), one expects their deformable mirror driven by $96 \times 96$ actuators to provide $10^{-9}$ contrast within the critical spatial frequency region. In our case of the modulating phase plate, the high density active mirror, such as that of JPL, can, perhaps, be used to correct the combined figure errors of the primary mirror and those of the phase modulating element. The combined figure errors could be precisely sensed by subtracting off the theoretical phase function from the actual one, retrieved by an iterative method, e.g., (Green et al 2002). It is feasible, then, to expect such a correction scheme to yield the reduction level sufficient for imaging Jovian planets.

An appealing alternative to the phase plate design, is to use an active mirror itself to induce the odd phase modulation pattern. Indeed, active mirrors are necessary for correcting the random phase aberrations in all high contrast imaging telescopes. If the task of generating the phase delay function can be integrated with that of correcting the random phase aberrations of the primary mirror, then the whole system will, likely, require no phase plate on a pupil relay plane, resulting in considerable simplification. In addition, such integration can ease the sequential space searching because the rotation and repeated collimation might not be necessary. Instead, one can reset the actuator stroke values and the dark region will move to other quadrant.

8. Acknowledgments

This work was supported in part by the NSF grant ATM01-0106271. WY also wishes to thank Christ Ftaclas for prior financial support through JPL (Contract # 1234467). We thank Miguel Levy for discussions concerning phase plate manufacturing.
REFERENCES

Bracewell R. N. & MacPhie, R. H. 1979, Icarus 38, 136
Davies, D. W. 1980, ICARUS 42, 145
Des Marais D. J. et al 2002, Astrobiology 2, 153
Jacquinot, P. & Roizen-Dossier, B. 1964, Prog. Optics, 3, 29
KenKnight, C. E. 1977, ICARUS 30, 422
Lyot, B. 1939, MNRAS 99, 580
Mori, Y., Yamauchi, K., & Endo, K. 1987, Precision Eng. 9, 123
Fig. 1.— Shapes of the 1-D phase delay functions over the pupil. The thinner solid line represents the phase $\phi_1(x) = a \cdot \tan[(0.5 - \varepsilon) \cdot 2\pi x / D]$ with $a = 1$ and $\varepsilon = 0.005$; The thicker solid line represents the phase $\phi_2(x) = a \cdot \ln\left(\frac{1+\varepsilon+2x/D}{(1+\varepsilon)-2x/D}\right)$ with $a = 3$ and $\varepsilon = 0.001$. 
Fig. 2.— 1-D annulling effect due to the phase function $\phi_1(x) = a \cdot \tan[(0.5 - \varepsilon) \cdot 2\pi x / D]$ with $a = 1$ and $\varepsilon = 0.005$. In the simulation, 12800 pixels are sampled over the pupil $D$ and the amplitude is normalized to peak amplitude without phase modulation. In (a), the $Sinc(\eta) \odot F\{\cos[\phi_1(x)]\}$ is represented in thicker solid line, while the $i \cdot Sinc(\eta) \odot F\{\sin[\phi_1(x)]\}$ is represented in thinner solid line. In (b), the solid line represents the interference results of two terms in (a). One can see that the “destructive interference” occurs on the negative half axis while “constructive interference” happens on the positive one.
Fig. 3.— 1-D annulling effect due to the phase function $\phi_2(x) = a \cdot \ln\left(\frac{1+\varepsilon+2x/D}{1+\varepsilon-2x/D}\right)$ with $a = 3$ and $\varepsilon = 0.001$. In the simulation, 12800 pixels are sampled over the pupil D and the amplitude is normalized to peak amplitude without phase modulation. In (a), the $\text{Sinc}(\eta) \otimes \mathcal{F}\{\cos[\phi_2(x)]\}$ is represented in thicker solid line, while the $i \cdot \text{Sinc}(\eta) \otimes \mathcal{F}\{\sin[\phi_2(x)]\}$ is represented in thinner solid line. In (b), the solid line represents the interference results of two terms in (a). Again, the “destructive interference” takes place on the left half axis while “constructive interference” is seen on the right.
Fig. 4.— 1-D logarithmic (base 10) relative intensity. The intensity is normalized to the peak intensity without phase modulation. The thicker solid line represents the logarithm (base 10) relative intensity of the annulling results due to $\phi_1(x) = a \cdot \tan \left((0.5 - \varepsilon) \cdot 2\pi x / D \right)$ with $a = 1$ and $\varepsilon = 0.005$. The dashed line represents the logarithm (base 10) relative intensity of the annulling results due to $\phi_2(x) = a \cdot \ln \left( \frac{(1+\varepsilon) + 2\pi x / D}{(1+\varepsilon) - 2\pi x / D} \right)$ with $a = 3$ and $\varepsilon = 0.001$. The thinner solid line is, as a reference, the diffraction intensity without phase modulation. One can see a sharper reduction in employing $\phi_2$ than that in employing $\phi_1$. 
Fig. 5.— 1-D performances in reduction level and Strehl ratio and their relations with parameters. Panels (a) and (b) are for the cases with phase modulation $\phi_1(x) = a \cdot \tan[(0.5 - \varepsilon) \cdot 2\pi x / D]$, in which (a) shows the relation of logarithmic (base 10) relative intensity level around $20\lambda / D$ vs. $\varepsilon$ at different values of $a = 0.5, 0.67, 1, 2$ and $4$ while (b) shows the relation of the Strehl ratio vs. $\varepsilon$ at different values of $a = 0.5, 0.67, 1, 2$ and $4$. Panel (c) and (d) are for the cases with phase modulation $\phi_2(x) = a \cdot \ln \left( \frac{(1+\varepsilon)+2x/D}{(1+\varepsilon)-2x/D} \right)$, in which (c) shows the relation of logarithmic (base 10) relative intensity level around $20\lambda / D$ vs. $\varepsilon$ at different values of $a = 2, 3, 6, 12$ and $24$ while (d) shows the relation of the Strehl ratio vs. $\varepsilon$ at different values of $a = 2, 3, 6, 12$ and $24$. 


Fig. 6.— 1-D relation between the shifted peak and the parameters. Panel (a): the shifted distance of the peak in units of $\lambda/D$ vs. $\varepsilon$ at different values of $a = 0.5, 0.67, 1, 2$ and $4$ with phase modulation $\phi_1(x) = a \cdot \tan[(0.5 - \varepsilon) \cdot 2\pi x/D]$. Panel (b): the shifted distance of the peak in units of $\lambda/D$ vs. $\varepsilon$ at different values of $a = 2, 3, 6, 12$ and $24$ with phase modulation $\phi_2(x) = a \cdot \ln[(1+\varepsilon)2x/D/(1+\varepsilon)-2x/D]$. The step-like shape in the curves is due to finite number of pixel points within each $\lambda/D$ interval. In this calculation, there are 20.56 pixels within each $\lambda/D$ interval.
Fig. 7.— 1-D relationship between parameter $\varepsilon$ and $N_{\text{min}}/a$, where $N_{\text{min}}$ is the estimated minimum number of pixels sampled over the pupil $D$ and $a$ is the parameter in the phase functions. The thicker solid line is for the case with phase modulation $\phi_1(x) = a \cdot \tan[(0.5 - \varepsilon) \cdot 2\pi x/D]$. The thinner solid line is for the case with phase modulation $\phi_2(x) = a \cdot \ln \frac{(1+\varepsilon)+2x/D}{(1+\varepsilon)-2x/D}$. 
Fig. 8.— Light reduction effect on one quadrant of focal plane when phase function is applied along \( x \) and \( y \) directions for the square pupil. Panel (a): Logarithmic (base 10) relative intensity image when phase \( \phi(x, y) = a \cdot \tan[(0.5 - \varepsilon) \cdot 2\pi x/D] + a \cdot \tan[(0.5 - \varepsilon) \cdot 2\pi y/D] \) with \( a = 1 \) and \( \varepsilon = 0.005 \) is applied to a square pupil. Panel (b): The thicker solid line represents the logarithm (base 10) relative intensity along the diagonal line crossing the second and the fourth quadrants in (a). Thinner solid line represents the one without phase modulation. Panel (c): Logarithmic (base 10) relative intensity image when phase \( \phi(x, y) = a \cdot \ln[(1+\varepsilon+2x/D)/(1+\varepsilon-2x/D) \cdot (1+\varepsilon+2y/D)/(1+\varepsilon-2y/D)] \) with \( a = 3 \) and \( \varepsilon = 0.001 \) is applied to a square pupil. Panel (d): The thicker solid line represents the logarithm (base 10) relative intensity along the diagonal line crossing the second and the fourth quadrants in (c). Thinner solid line represents the one without phase modulation. One can see that light in the 1st, 2nd and 3rd quadrants has been greatly reduced and the reduction level of \( 10^{-12} \) can be reached at the distance of about \( 3.5\lambda/D \) in the second quadrant.
Fig. 9.— Broad bandwidth light reduction effect on one quadrant of focal plane. The simulation is based on a rectangular spectrum distribution with total bandwidth of $60\% \lambda_0$.

Panel (a): Logarithmic (base 10) relative intensity image when phase $\phi(x, y) = a \cdot \tan[(0.5 - \varepsilon) \cdot 2\pi x/D] + a \cdot \tan[(0.5 - \varepsilon) \cdot 2\pi y/D]$ with $a = 1$ and $\varepsilon = 0.005$ is applied to a square pupil. Panel (b): The thicker solid line represents the logarithm (base 10) relative intensity along the diagonal line crossing the second and the fourth quadrants in (a). Thinner solid line represents the one without phase modulation. Panel (c): The logarithm (base 10) relative intensity image when phase $\phi(x, y) = a \cdot \ln[(1 + \varepsilon) + 2x/D] - (1 + \varepsilon) - 2y/D]$ with $a = 3$ and $\varepsilon = 0.001$ is applied to a square pupil. Panel (d): The thicker solid line represents the logarithm (base 10) relative intensity along the diagonal line crossing the second and the fourth quadrants in (c). Thinner solid line represents the one without phase modulation. One can see that reduction level of $10^{-12}$ with an inner working distance of about $3.5\lambda_0/D$ can still be kept with a broad bandwidth of $60\% \lambda_0$ in the second quadrant.
Fig. 10.— Illustration of the actual pupil which is formed by the sum of the ideal pupil and the “peninsulas” and “bays” around the ideal pupil boundary. The ideal pupil is the region encircled by the dotted lines, the actual pupil is the white region encircled by the solid lines.
Fig. 11.— Schematic layout for restoring the diffraction limited images of the extrasolar planets.
Fig. 12.— Comparison of the images of the planet before and after combination use of an occulting mask and a conjugate phase element. The occulting mask covers the light in the 1st, 3rd and 4th quadrants. The pair of conjugate phase is based on $\phi(x, y) = a \cdot \ln\left[\frac{(1+\varepsilon)+2x/D}{(1+\varepsilon)-2x/D} \cdot \frac{(1+\varepsilon)+2y/D}{(1+\varepsilon)-2y/D}\right]$ with $a = 3$, $\varepsilon = 0.001$ and the bandwidth = $0.6 \lambda_0$. The planet locates at $10\lambda_0/D$ angular distance with respective to the star along the diagonal line in the 2nd quadrant and is $10^{-9}$ fainter than the star. Panels (a) and (b) show the logarithm (base 10) relative intensity image and linear relative intensity image of the detected planet before using the occulting mask and the conjugate phase element while (c) and (d) show the cases after using the occulting mask and the conjugate phase element. One can see that diffraction limited image of the planet can be restored.