

ONE-SIDED ACHROMATIC PHASE APODIZATION FOR IMAGING OF EXTRASOLAR PLANETS

WEIDONG YANG AND ALEXANDER B. KOSTINSKI

Physics Department, Michigan Technological University, 1400 Townsend Drive, Houghton, MI 49931

Received 2003 October 14; accepted 2004 January 7

ABSTRACT

We propose a new approach to direct imaging of extrasolar planets: one-sided phase apodization. It is based on a discovery that an antisymmetric spatial phase modulation pattern imposed over a pupil or a relay plane causes diffracted starlight suppression sufficient for imaging of extrasolar planets. Calculations with specific square-pupil (side D) phase functions, such as

$$\phi(x, y) = a \ln \left[\frac{(1 + \epsilon) + 2x/D (1 + \epsilon) + 2y/D}{(1 + \epsilon) - 2x/D (1 + \epsilon) - 2y/D} \right],$$

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 $\epsilon = 10^{-3}$). 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 extrasolar planets has generated great interest and shows no signs of subsiding. While extrasolar 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 Marais et al. 2002). Yet, to date, there have been no reports of direct extrasolar 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 extrasolar planet detection. Among the current approaches to overcoming such difficulties, two general directions appear particularly promising: (1) imaging based on infrared interferometers, along the lines originally proposed by Bracewell & MacPhie (1979) and (2) visible-light imaging based on the coronagraph concept (originally introduced by Lyot 1939 in the solar physics context). Because of the longer wavelength, decreased angular resolution in the infrared region must be overcome by spatial separation, but in this paper we confine ourselves to monolithic telescopes.

Imaging in 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 sidelobe levels below 10^{-9} (Davies 1980; KenKnight 1977), and the need for a small inner working angles of a few λ/D , are daunting. Nevertheless, recent developments in phase sensing and control technology will enable figure errors of up to 100 cycles per aperture to be controlled under 1 Å per 100 hr (Trauger et al. 2003b; Shaklan, Moody, & Green 2002; Green et al. 2003). Therefore, coronagraphic imaging in the visible-light region is feasible (Malbet, Yu, & Shao 1995).

As to the diffraction sidelobe removal, in addition to the Lyot coronagraph, 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 either the phase or amplitude modulation of the star diffraction pattern on a focal plane, and includes the phase mask of Roddier & Roddier (1997), the four-quadrant phase mask of Rouan et al. (2000), the band-limited mask of Kuchner & Traub (2002), and the notch filter mask of 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 that has included several exciting results, such as the ringlike binary mask of Angel et al. (1986), the “eye-shaped” binary mask of Kasdin et al. (2003), the spider binary mask of Vanderbei, Spergel, & Kasdin (2003), the amplitude-apodized pupil mask of Nisenson & Papaliolios (2001), and amplitude pupil apodization by beam reshaping (Guyon 2003). Besides the specific mask designs mentioned above, the papers by Gonsalves & Nisenson (2003) and Aime, Soummer, & Ferrari (2002) addressed a general optimization analysis for coronagraph-type systems.

While pupil shape and amplitude 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 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 sidelobe optimization problems yields real functions. Nevertheless, in this paper we reexamine the question. Why?

High-contrast imaging through pupil phase-only spatial modulation, if realizable, might have the following advantages over the other techniques: (1) there is no loss of light energy going through the pupil, which shortens the integration time; (2) 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 (i.e., at least slowly adaptive) optics; and (3) 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 interpret phase apodization more broadly than usual, that is, as a general spatial phase modulation across the pupil plane that yields significant “improvement” in the focal plane energy distribution for imaging of a faint companion. Thus, we allow main-lobe reduction and shift, asymmetric apodization, etc. A brief, preliminary account of asymmetric phase apodization can be found in Yang & Kostinski (2003).

Let us begin by asking whether phase apodization patterns exist that can remove the sidelobes 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 § 3. We were further motivated by the recent work in Nisenson & Papaliolios (2001), who “revived” the square pupil for extrasolar planet detection. In addition, the pupil shape optimization work reported in Kasdin et al. (2003) suggested to us the idea of *partial* sidelobe removal in a focal plane at any given time. This led to consideration of antisymmetric (odd) phase functions, which proved to be the key, as detailed in § 3. In summary then, our proposal consists of the following elements:

1. consider phase-only spatial modulation pattern across the pupil;
2. consider a square pupil and assume separation of variables, so that an effectively one-dimensional problem can be examined in a semianalytic manner;
3. use odd (antisymmetric) 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 two dimensions (square pupil) in later sections. Let us denote the spatial phase over the (one-

dimensional) pupil as $\phi(x)$ and the transmission function as $T(x)$. Then the light on the diffraction plane would be

$$E(\eta) = \mathcal{F}\left(T(x)e^{i\phi(x)}\right) \quad (1)$$

(Goodman 1996), where x is the coordinate in the pupil plane, η is the coordinate in the diffraction plane, and $\mathcal{F}(\dots)$ denotes the Fourier transform operation.

An equivalent expression is

$$E(\eta) = \mathcal{F}(T(x)) \otimes \mathcal{F}(\cos[\phi(x)] + i\sin[\phi(x)]), \quad (2)$$

where \otimes denotes convolution. For a one-dimensional pupil of width D , the transmission 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, $\mathcal{F}(i\sin[\phi(x)])$ yields an even and imaginary function. The light intensity on the diffraction plane would be simply the sum of the squares of two real functions:

$$I(\eta) = [\text{sinc } \eta \otimes \mathcal{F}(\cos[\phi(x)])]^2 + [\text{sinc } \eta \otimes \mathcal{F}(\sin[\phi(x)])]^2. \quad (3)$$

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 $\mathcal{F}(i\sin[\phi(x)])$ 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):

$$I(\eta) = [\text{sinc } \eta \otimes \mathcal{F}(\cos[\phi(x)]) + \text{sinc } \eta \otimes \mathcal{F}(i\sin[\phi(x)])]^2. \quad (4)$$

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 $T(x)\cos[\phi(x)]$ and $T(x)\sin[\phi(x)]$ in this region. Guided by this argument, we explored a set of odd (antisymmetric) phase functions and found some that yield sufficiently deep light reduction on half of the x -axis.

4. ONE-DIMENSIONAL EXAMPLES OF ONE-SIDED PHASE APODIZATION

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

$$\phi_1(x) = a \tan[(0.5 - \epsilon)2\pi x/D] \quad (5)$$

and

$$\phi_2(x) = a \ln \frac{(1 + \epsilon) + 2x/D}{(1 + \epsilon) - 2x/D}, \quad (6)$$

where ϵ is a small (positive) parameter, defined so that the phase value is finite at the edge of the pupil. The shapes of

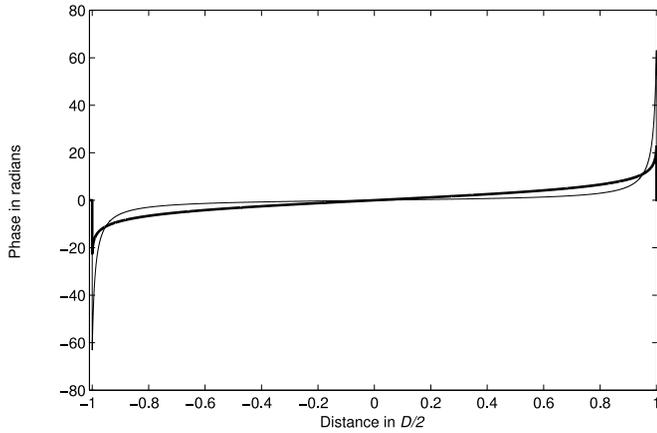


FIG. 1.—Shapes of the one-dimensional phase delay functions over the pupil. The thin line represents phase $\phi_1(x)$ (eq. [5]), with $a = 1$ and $\epsilon = 0.005$; The thick line represents the phase $\phi_2(x)$ (eq. [6]), with $a = 3$ and $\epsilon = 0.001$.

these two phase functions are illustrated in Figure 1, where the parameters are set as $a = 1$ and $\epsilon = 0.005$ for $\phi_1(x)$, while for $\phi_2(x)$, $a = 3$ and $\epsilon = 0.001$. (We defer discussion of computational devices needed to avoid severe aliasing in the computations and the selection of the a and ϵ parameters to § 5.)

The annulling effect by the destructive interference described in equation (4) is illustrated in Figures 2 and 3. The example of a diffraction pattern caused by phase function ϕ_1 with $a = 1$ and $\epsilon = 0.005$ is shown in Figure 2. Similarly, Figure 3 displays the diffraction pattern for the case of phase function ϕ_2 with $a = 3$ and $\epsilon = 0.001$. In order to “zoom in” and see the details, in Figure 4 we demonstrate the annulling effect on a \log_{10} scale. It can be seen that for ϕ_1 , a reduction level region of lower than $10^{-4.5}$ can be obtained at a distance of about $3.5\lambda/D$ from the shifted peak, and 10^{-5} can be obtained at about $4.5\lambda/D$ (thick solid line). The second example, ϕ_2 , produces an even sharper annulling effect, shown by the dashed line. We see that the level of lower than $10^{-5.5}$

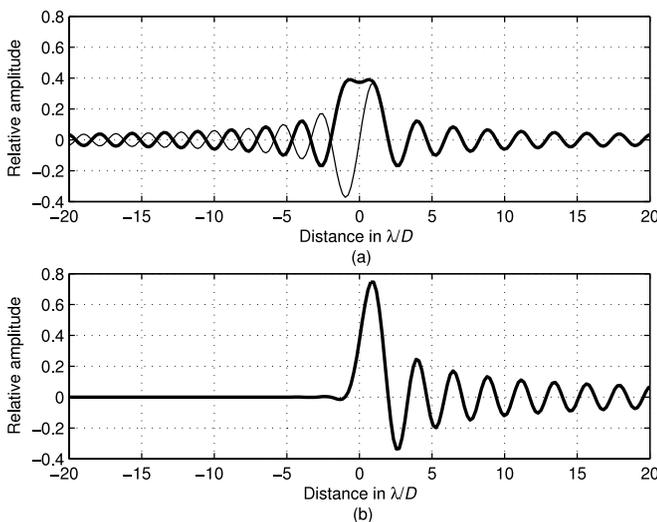


FIG. 2.—One-dimensional annulling effect due to the phase function $\phi_1(x)$ (eq. [5]), with $a = 1$ and $\epsilon = 0.005$. In the computation, 12,800 pixels are sampled over the pupil D , and the amplitude is normalized to peak amplitude without phase modulation. In (a), $\text{sinc } \eta \otimes \mathcal{F}(\cos[\phi_1(x)])$ is represented by the thick line, while $\text{sinc } \eta \otimes \mathcal{F}(i\sin[\phi_1(x)])$ is represented by the thin line. In (b), the 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.

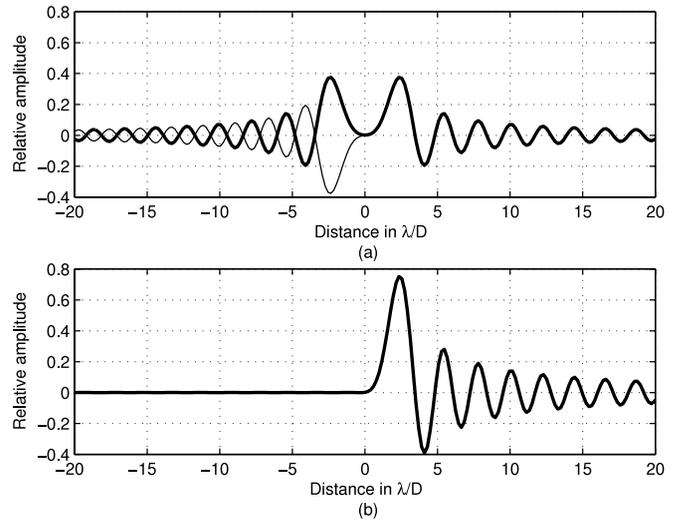


FIG. 3.—Same as Fig. 2, but for the annulling effect due to the phase function $\phi_2(x)$ (eq. [6]), with $a = 3$ and $\epsilon = 0.001$. Again, the destructive interference takes place on the left half-axis while constructive interference is seen on the right.

can be reached at a distance of 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. ONE-DIMENSIONAL 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

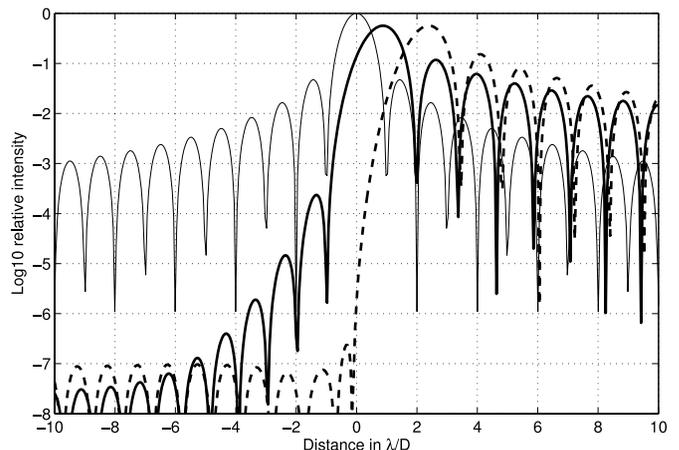


FIG. 4.—One-dimensional \log_{10} relative intensity. The intensity is normalized to the peak intensity without phase modulation. The thick solid line represents the logarithmic relative intensity of the annulling results due to $\phi_1(x)$ (eq. [5]), with $a = 1$ and $\epsilon = 0.005$. The dashed line represents the \log_{10} relative intensity of the annulling results due to $\phi_2(x)$ (eq. [6]), with $a = 3$ and $\epsilon = 0.001$. The thin solid line represents, as a reference, the diffraction intensity without phase modulation. One can see a sharper reduction from employing ϕ_2 than from employing ϕ_1 .

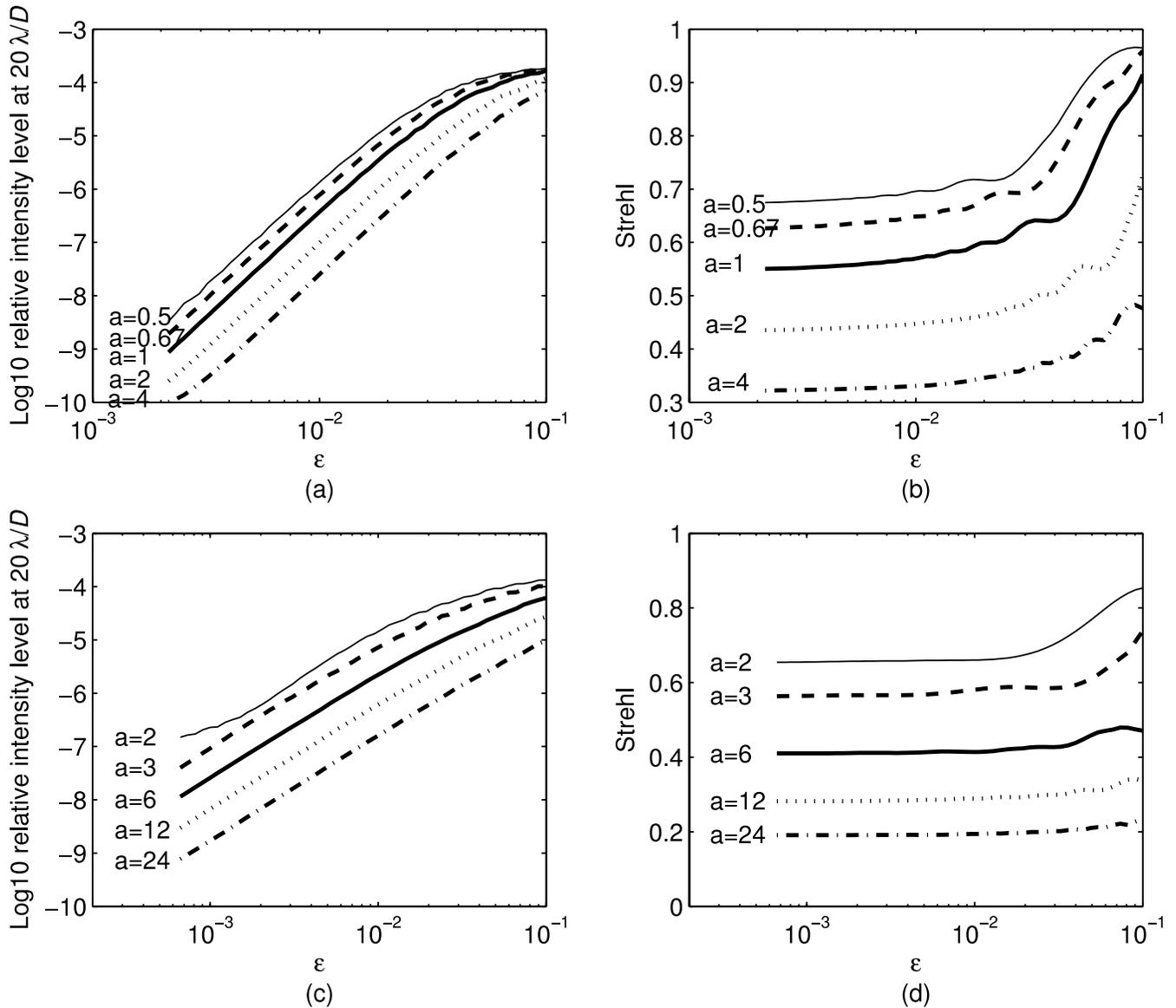


FIG. 5.—One-dimensional performances in reduction level and Strehl ratio and their relations with parameters. (a, b) Cases with phase modulation $\phi_1(x)$ (eq. [5]), in which (a) shows the relation of \log_{10} relative intensity level around $20\lambda/D$ vs. ϵ at $a = 0.5, 0.67, 1, 2,$ and 4 , while (b) shows the relation of the Strehl ratio vs. ϵ at the same a -values. (c, d) Cases with phase modulation $\phi_2(x)$ (eq. [6]), in which (c) shows the relation of \log_{10} relative intensity level around $20\lambda/D$ vs. ϵ at $a = 2, 3, 6, 12,$ and 24 , while (d) shows the relation of the Strehl ratio vs. ϵ at the same a -values.

intensity value between the 25th and the 20th λ/D . The corresponding curves are shown in Figures 5a and 5c. Figure 5a corresponds to ϕ_1 , while Figure 5c corresponds to ϕ_2 . It can be seen that the reduced intensity level is strongly dependent on the parameters a and ϵ . 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. This relation is shown for ϕ_1 in Figure 5b and for ϕ_2 in Figure 5d.

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 ϵ , within the selected ranges. The curves in Figures 5 and 6 can be used in phase design based on given specifications and are used in § 6.1 to give qualitative analysis on the broad-bandwidth performance.

In addition to antisymmetry, the phase functions in equations (5) and (6) have another common qualitative feature: 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 computation, 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 of width D . Then sampling theory suggests choosing the sampling rate N/D , of at least, twice the maximal spatial frequency of our phase functions. If local frequency $f = (1/2\pi)[d\phi(x)/dx]$ is used to estimate the frequency, then it can be expected that the N will have to satisfy $N \geq (D/\pi)[d\phi(x)/dx]_{\max}$. Figure 7 shows the relationship between the estimated minimum number of sampled pixels N_{\min} and the parameter ϵ for the phase functions. For smaller ϵ , more sampling pixels are required over

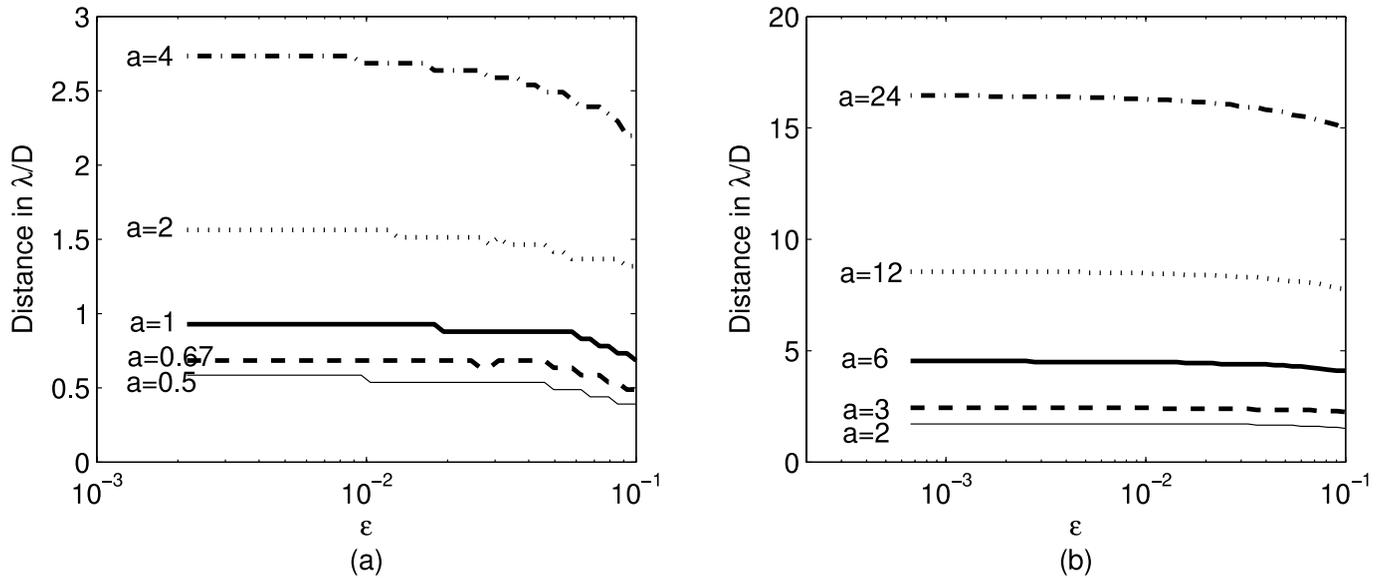


FIG. 6.—One-dimensional relation between the shifted peak and the parameters. (a) Shifted distance of the peak in units of λ/D vs. ϵ at $a = 0.5, 0.67, 1, 2,$ and 4 , with phase modulation $\phi_1(x)$ (eq. [5]). (b) Same as (a), but for $a = 2, 3, 6, 12,$ and 24 , with phase modulation $\phi_2(x)$ (eq. [6]). The steplike shape in the curves is due to finite number of pixel points within each λ/D interval. In this calculation, there are 20.56 pixels within each λ/D interval.

the pupil. To insure the stability of the results in a rather large range, e.g., 2000 diffraction rings in one dimension, and make the code applicable for all a and ϵ used in all of the above computations, at least 12,800 pixels were sampled over the pupil of width D . (This is not to be confused with the issue of the number of elements needed to implement the results in, e.g., active optics—as we 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 one-dimensional phase functions, along with the sidelobe reduction results, to the square pupil, which, as we mentioned earlier, has already generated a great deal of interest in the field of extrasolar planet imaging (Nisenson &

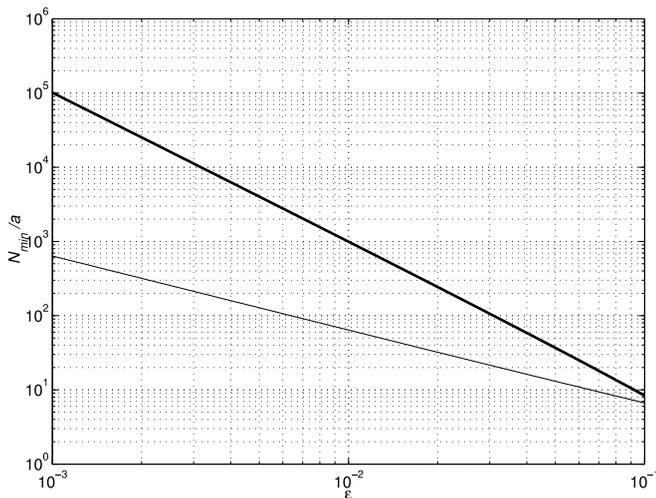


FIG. 7.—One-dimensional relationship between parameter ϵ and N_{\min}/a , where N_{\min} is the estimated minimum number of pixels sampled over the pupil D and a is the parameter in the phase functions. The thick solid line is for the case with phase modulation $\phi_1(x)$ (eq. [5]). The thin solid line is for the case with phase modulation $\phi_2(x)$ (eq. [6]).

Papaliolios 2001; Kasdin et al. 2003). There are two ways to proceed to the square pupil. One way is to apply the one-dimensional phase function along one axis only, where the light field over the two-dimensional square pupil would be $T(x)T(y)e^{i\phi(x)}$. The diffraction light intensity would then be given by $I(\eta, \xi) = I(\eta)\text{sinc}\xi$, where $I(\eta)$ is the one-dimensional diffraction field intensity determined by equation (4), in which the odd phase delay is applied. For the phase functions ϕ_1 and ϕ_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 the x - and y -axes, in which case the light field over the two-dimensional square pupil becomes $T(x)T(y)e^{i[\phi(x)+\phi(y)]}$. In this case, the diffraction light field intensity is given by $I(\eta, \xi) = I(\eta)I(\xi)$. The essential advantage gained is that *one quadrant of the diffraction plane can experience twice the one-dimensional reduction and do so at a closer separation angle*. This is illustrated in Figure 8a, where we show a logarithmic intensity image produced by the phase function $\phi_1(x) + \phi_1(y)$. Figure 8b displays relative intensity along the diagonal. It can be seen that a deep reduction region is obtained in the second quadrant and that the 10^{-9} level can be reached at distance of $4.5\lambda/D$ and the 10^{-12} level at about $7\lambda/D$. Figures 8c and 8d demonstrate that the level of 10^{-12} can be reached at a 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.

6.1. Bandwidth Tolerance

Let us consider the case in which 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 $\phi(x) = (2\pi/\lambda)[n(\lambda) - 1]d(x)$. If the reflecting mirror has a geometric shape $h(x)$, then it generates a phase delay of $\phi(x) = -(2\pi/\lambda)2h(x)$. For notational simplicity, we use a common term $G(\lambda)$ to represent either the term $[n(\lambda) - 1]$ for

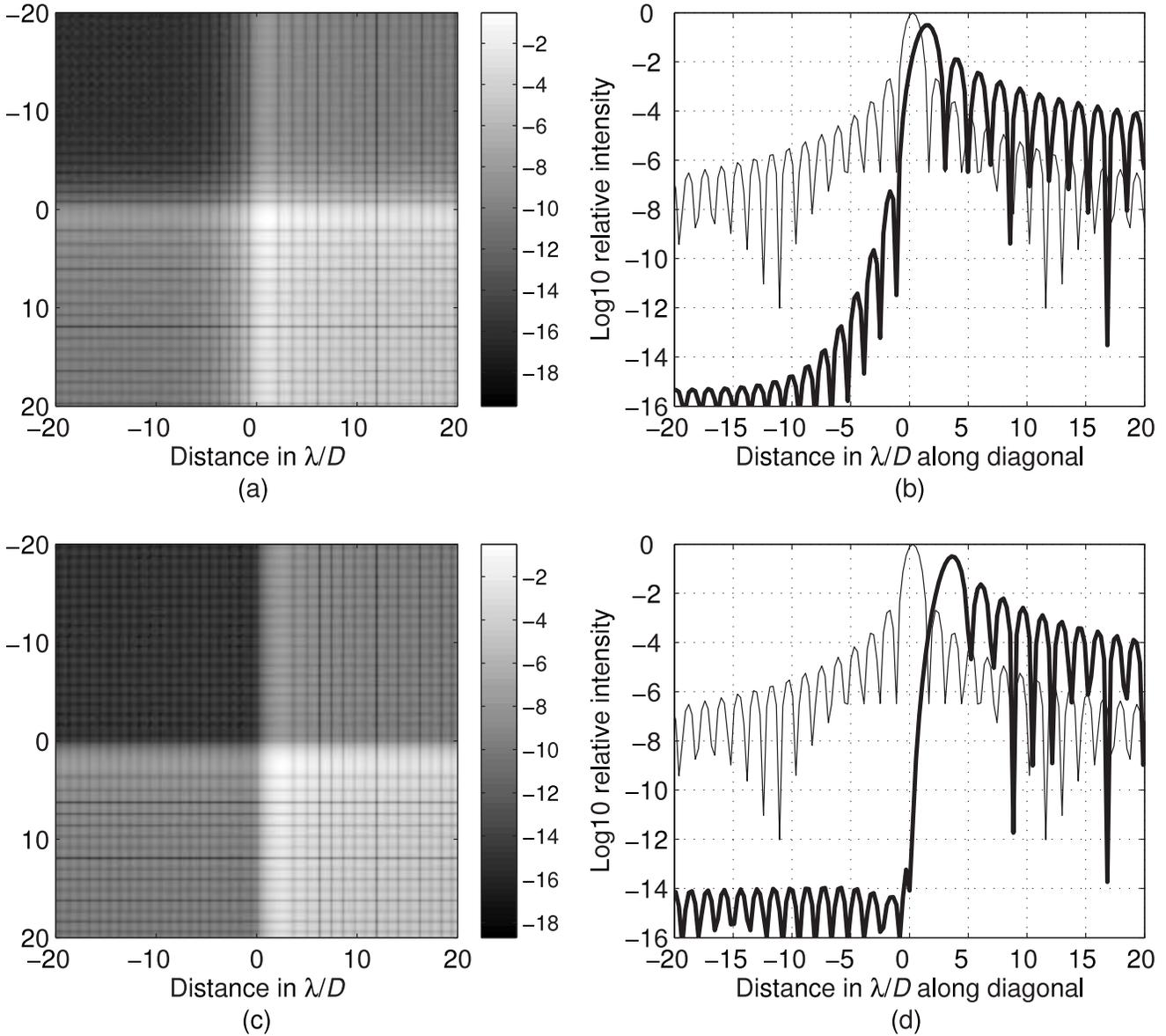


FIG. 8.—Light reduction effect on one quadrant of the focal plane when a phase function is applied along the x - and y -directions for the square pupil. (a) \log_{10} relative intensity image when phase $\phi(x, y) = a \tan[(0.5 - \epsilon)2\pi x/D] + a \tan[(0.5 - \epsilon)2\pi y/D]$, with $a = 1$ and $\epsilon = 0.005$, is applied to a square pupil. (b) The thicker line represents the \log_{10} relative intensity along the diagonal line crossing the second and the fourth quadrants in (a). The thinner line represents the one without phase modulation. (c) Same as (a), but when phase $\phi(x, y)$ from eq. (11), with $a = 3$ and $\epsilon = 0.001$, is applied to a square pupil. (d) Same as (b), but for the quadrants in (c). One can see that light in the first, second, and third quadrants has been greatly reduced and that the reduction level of 10^{-12} can be reached at the distance of about $3.5\lambda/D$ in the second quadrant.

the 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 $\phi(x) = (2\pi/\lambda)G(\lambda)H(x)$. To generate the phase function that works on central wavelength λ_0 , we set the geometric function to $H(x; \lambda_0) = \{\lambda_0/[2\pi G(\lambda_0)]\}\phi(x)$, where the second argument in $H(x; \lambda_0)$ indicates that the geometric function is designed for the central wavelength λ_0 . When light of wavelength λ goes through this phase delay element, the geometric function $H(x; \lambda_0)$ generates the phase delay given by

$$\phi(x, \lambda) = \frac{\lambda_0}{\lambda} \frac{G(\lambda)}{G(\lambda_0)} \phi(x), \quad (7)$$

and we see that the phase delay at the new wavelength equals the original phase function $\phi(x)$ multiplied by a factor of

$(\lambda_0/\lambda)[G(\lambda)/G(\lambda_0)]$. This factor is, in fact, equivalent to the parameter a in the phase function formulae (eqs. [5] and [6]). For a phase plate with positive dispersion materials or for the reflecting mirror, the term $(\lambda_0/\lambda)[G(\lambda)/G(\lambda_0)]$ decreases with increasing λ . Therefore, based on parameter relations of Figure 5, we see that a 50% “redshift” in wavelength causes a less than 1 order of magnitude increase in the reduction level, while a 50% “blueshift” in wavelength causes a less than 1 order of magnitude decrease in the reduction level. This is why our phase modulations tolerate a $0.6\lambda_0$ bandwidth in the simulations shown in Figure 9 and still keep a small inner working angle and low reduction level.

6.2. Phase and Shape Errors

The errors in the phase-only spatial modulation scheme are likely to come from two sources: imperfections in the phase

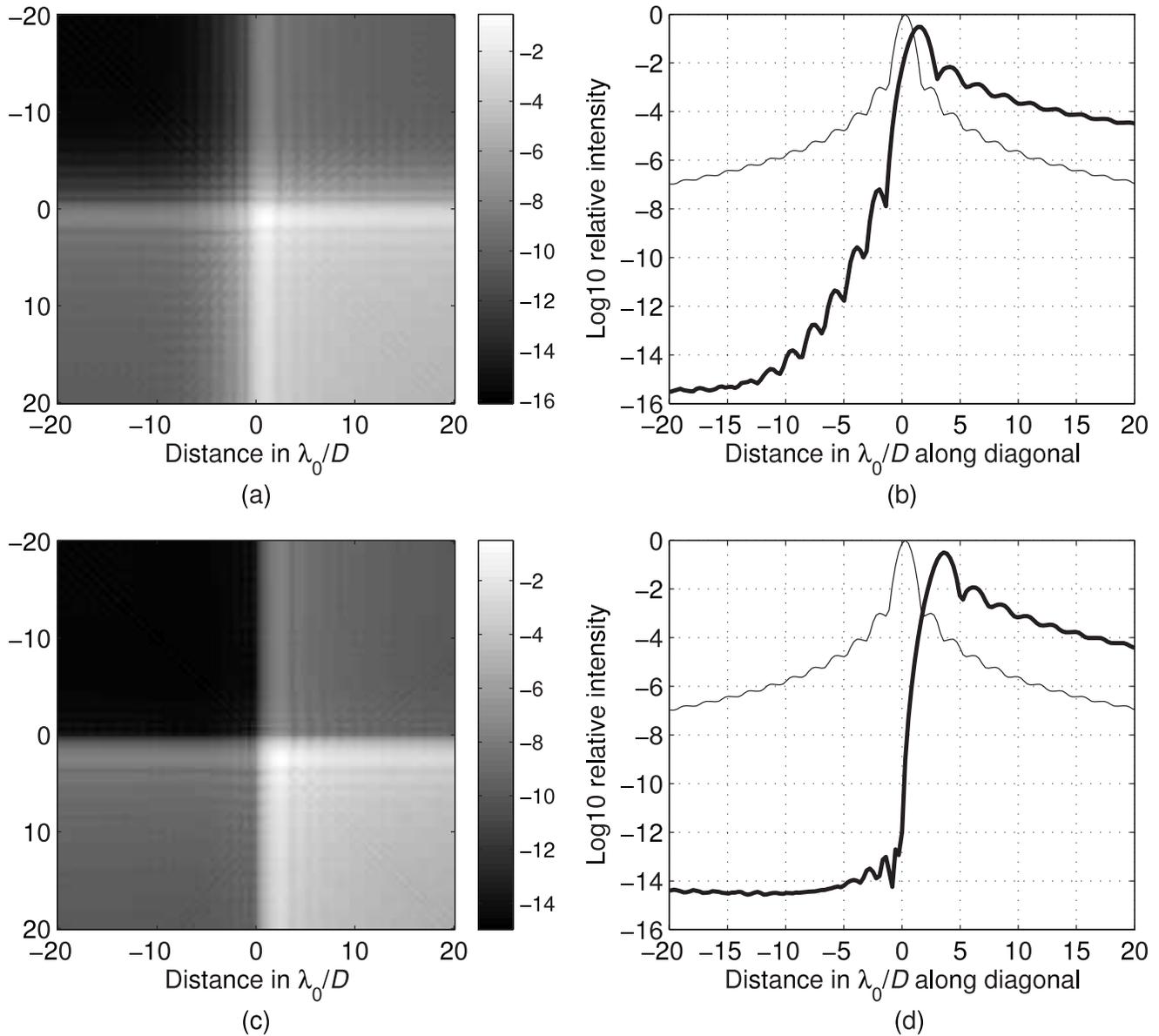


FIG. 9.—Broad-bandwidth light reduction effect on one quadrant of focal plane. The simulation is based on a rectangular spectrum distribution with total bandwidth of $0.6\lambda_0$. (a) \log_{10} relative intensity image when phase $\phi(x, y) = a \tan[(0.5 - \epsilon)2\pi x/D] + a \tan[(0.5 - \epsilon)2\pi y/D]$, with $a = 1$ and $\epsilon = 0.005$, is applied to a square pupil. (b) The thicker line represents the \log_{10} relative intensity along the diagonal line crossing the second and the fourth quadrants in (a). The thinner line represents the one without phase modulation. (c) Same as (a), but with phase $\phi(x, y)$ from eq. (11), with $a = 3$ and $\epsilon = 0.001$, applied to a square pupil. (d) Same as (b), but for the quadrants in (c). One can see that the 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 $0.6\lambda_0$ in the second quadrant.

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

$$E(\eta, \xi) \approx \mathcal{F}\left(T(x, y)e^{i\phi(x, y)}[1 + i\delta\phi(x, y)]\right), \quad (8)$$

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. There the light intensity in the dark region is given by

$$\delta I(\eta, \xi) \approx |E_{\text{ideal}}(\eta, \xi) \otimes \mathcal{F}(\delta\phi(x, y))|^2, \quad (9)$$

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]:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta I(\eta, \xi) d\xi d\eta \approx \int_{-D/2}^{D/2} \int_{-D/2}^{D/2} |\delta\phi(x, y)|^2 dx dy. \quad (10)$$

Spectral content can be important in considering the phase error tolerance; e.g., phase errors of spatial frequencies from 0.03 to 0.5 cycles cm^{-1} (about 5.4–90 cycles per aperture for the Eclipse design) are considered critical for imaging Jovian planets (Trauger et al. 2003a, 2003b), 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 the function

$$\phi(x, y) = a \ln \left[\frac{(1 + \epsilon) + 2x/D (1 + \epsilon) + 2y/D}{(1 + \epsilon) - 2x/D (1 + \epsilon) - 2y/D} \right], \quad (11)$$

($a = 3$ and $\epsilon = 0.001$), we assume a flat noise intensity level in the critical spatial frequency region. This results in integration from $5.4\lambda/D$ to $90\lambda/D$ (left-hand side of eq. [10]) and yields $10^{-9} I_0 (\lambda/D)^2 (90^2 - 5.4^2)$, where I_0 is the light peak intensity with phase modulation. But from Figure 5c we obtain $I_0 \sim 0.3 [D^2 (D/\lambda)^2]$, where $D^2 (D/\lambda)^2$ is the peak power of the square-pupil case without phase modulation (Born & Wolf 1999). This results in phase errors within the critical spatial frequency region below 15.5×10^{-4} rad rms, or below $2.5 \times 10^{-4} \lambda$ rms. This is feasible with current technology, as reviewed in § 1.

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 widespread

much larger than that due to the ideal pupil, and 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 area of the ideal pupil be A and the 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 $\sim Ll^3/\lambda^2$. Based on this scaling, the background noise n relative to the peak power of the ideal-pupil diffraction peak is

$$n \sim Ll^3/A^2, \quad (12)$$

and, on the basis of equation (12), we estimate that for a square pupil of, e.g., width 0.1 m, a requirement on the relative noise background of about 10^{-11} can be satisfied by confining boundary errors to less than $10 \mu\text{m}$ —a quite feasible task.

6.3. Diffraction-limited Planet Imaging by Combining a 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 for 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 pupil relay plane to compensate (reverse) the phase. For example, consider a schematic layout as shown in Figure 11, where two pupil relay planes contain the conjugate pair of phase plates. In addition, 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 the occulting mask and the 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 three quadrants of the image plane is used. It can be seen that the conjugate phase element moves the light energy in the strong sidelobes back into the image main lobe and that 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 fact that the mask also covers some amount of the broad sidelobes from the planet, near the mask boundary. Redesigning the mask may increase the restored peak power.

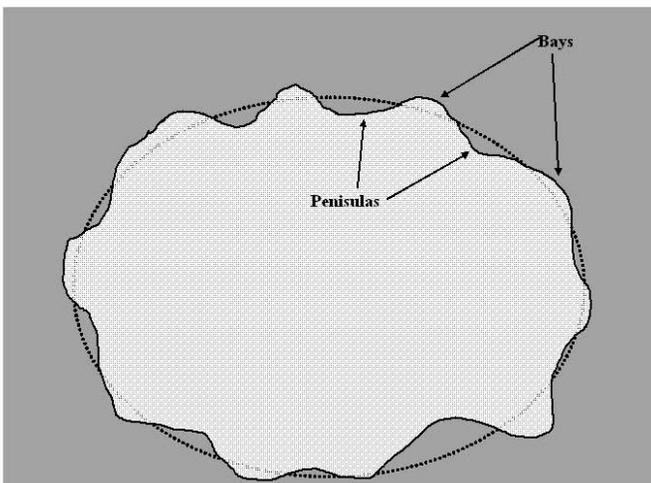


FIG. 10.—Illustration of the actual pupil formed by the sum of the ideal pupil and the peninsulas and bays around the ideal-pupil boundary. The ideal pupil is the region encircled by the dotted lines; the actual pupil is the white region encircled by the solid lines.

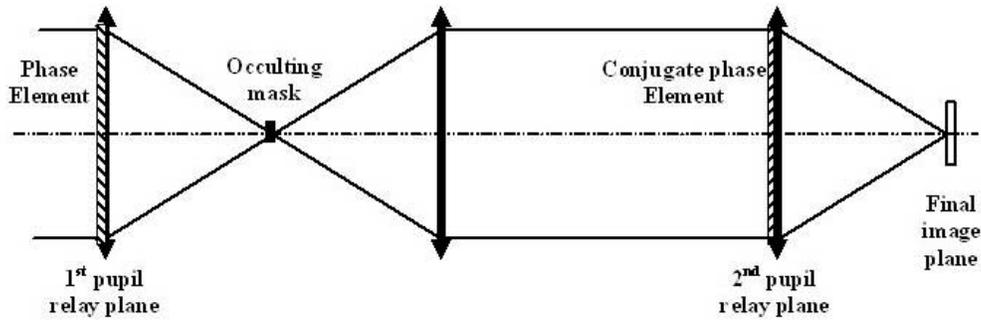


Fig. 11.—Schematic layout for restoring the diffraction-limited images of the extrasolar planets

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 directly imaging extrasolar 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

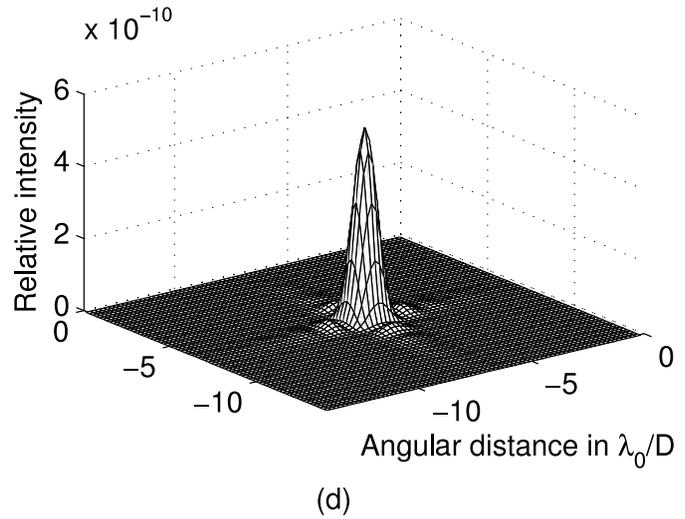
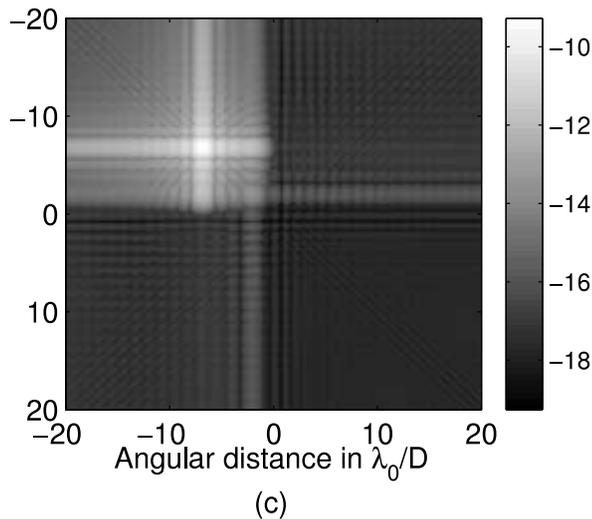
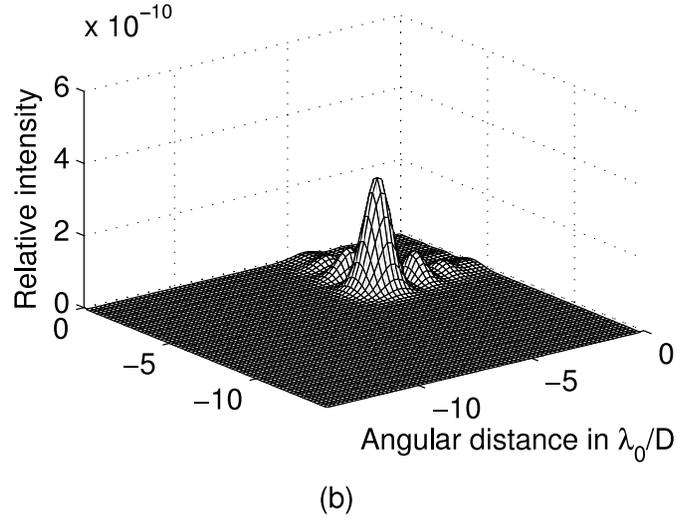
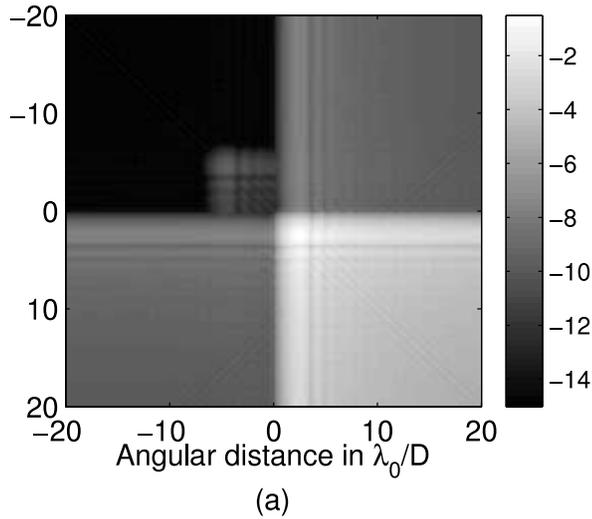


FIG. 12.—Comparison of the images of the planet before (a, b) and after (c, d) combination use of an occulting mask and a conjugate phase element. The occulting mask covers the light in the first, third, and fourth quadrants. The pair of conjugate phases is based on $\phi(x, y)$ from eq. (11), with $a = 3$ and $\epsilon = 0.001$, and the bandwidth is $0.6\lambda_0$. The planet is located at a $10\lambda_0/D$ angular distance with respect to the star along the diagonal line in the second quadrant and is 10^{-9} times fainter than the star. (a) and (b) show the \log_{10} relative intensity image and linear relative intensity plot of the detected planet before use of the occulting mask and the conjugate phase element, while (c) and (d) show the cases after use of the occulting mask and the conjugate phase element. One can see that the diffraction-limited image of the planet can be restored.

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 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 m phase plate or mirror. Our phase curve of Figure 1 corresponds to edge heights of a few microns and edge slopes on the order of 10^{-7} and 10^{-3} for ϕ_1 and ϕ_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, van Brug, & Frankena 1998), ion beam milling (Drueding et al. 1995), and wet-etch figuring (Rushford et al. 2003). 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 of extrasolar planets by a moderate-sized 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. 2003a, 2003b; Hull et al. 2003), one expects their deformable mirror driven by 96×96 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. 2003). It is feasible, then, to expect such a correction scheme to yield a 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 rotation and repeated collimation might not be necessary. Instead, one can reset the actuator stroke values, and the dark region will move to another quadrant.

This work was supported in part by the NSF grant ATM 01-0106271. W. Y. 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

- Aime, C., Soummer, R., & Ferrari, A. 2002, *A&A*, 389, 334
 Angel, J. R. P., Cheng, A. Y. S., & Woolf, N. J. 1986, *Nature*, 322, 341
 Born, M., & Wolf, E. 1999, *Principles of Optics* (7th ed.; Cambridge: Cambridge Univ. Press)
 Bracewell R. N., & MacPhie, R. H. 1979, *Icarus*, 38, 136
 Brown, R. A., & Burrows, C. J. 1990, *Icarus*, 87, 484
 Davies, D. W. 1980, *Icarus*, 42, 145
 Des Marais, D. J., et al. 2002, *Astrobiology*, 2, 153
 Drueding, T. W., Fawcett, S. C., Wilson, S. R., & Bifano, T. G. 1995, *Opt. Eng.*, 34, 3565
 Fähnle, O. W., van Brug, H., & Frankena, H. J. 1998, *Appl. Opt.*, 37, 6771
 Gonsalves, R., & Nisenson, P. 2003, *PASP*, 115, 706
 Goodman, J. W. 1996, *Introduction to Fourier Optics* (2nd ed.; New York: McGraw-Hill)
 Green, J. J., Redding, D. C., Shaklan, S. B., & Basinger, S. A. 2003, *Proc. SPIE*, 4860, 266
 Guyon, O. 2003, *A&A*, 404, 379
 Hull, T., Trauger, J. T., Macenka, S. A., Moody, D., Olarte, G., Sepulveda, C., Tsuha, W., & Cohen, D. 2003, *Proc. SPIE*, 4860, 277
 Jacquinot, P., & Roizen-Dossier, B. 1964, *Prog. Opt.*, 3, 29
 Kasdin, N. J., Vanderbei, R. J., Spergel, D. N., & Littman, M. G. 2003, *ApJ*, 582, 1147
 KenKnight, C. E. 1977, *Icarus*, 30, 422
 Kuchner, M. J., & Spergel, D. N. 2003, *ApJ*, 594, 617
 Kuchner, M. J., & Traub, W. A. 2002, *ApJ*, 570, 900
 Lyot, B. 1939, *MNRAS*, 99, 580
 Malbet, F., Yu, W. J., & Shao, M. 1995, *PASP*, 107, 386
 Mori, Y., Yamauchi, K., & Endo, K. 1987, *Precision Eng.*, 9, 123
 Nisenson, P., & Papaliolios, C. 2001, *ApJ*, 548, L201
 Roddier, F., & Roddier, C. 1997, *PASP*, 109, 815
 Rouan, D., Riaud, P., Boccaletti, A., Clénet, Y., & Labeyrie, A. 2000, *PASP*, 112, 1479
 Rushford, M. C., Britten, J. A., Dixit, S. N., Hoaglan, C. R., Aasen, M.D., & Summers, L. J. 2003, *Appl. Opt.*, 42, 5706
 Shaklan, S. B., Moody, D., & Green, J. J. 2003, *Proc. SPIE*, 4860, 229
 Trauger, J. T., Moody, D., Gordon, B., Gursel, Y., Ealey, M. A., & Bagwell, R. B. 2003a, *Proc. SPIE*, 4854, 1
 Trauger, J. T., et al. 2003b, *Proc. SPIE*, 4854, 116
 Vanderbei, R. J., Spergel, D. N., & Kasdin, N. J. 2003, *ApJ*, 590, 593
 Yang, W., & Kostinski, A. B. 2003, *Phys. Lett. A*, 320, 5