MichiganTech
Create the Future

Michigan Technological University
Digital Commons @ Michigan Tech

Dissertations, Master's Theses and Master's Reports - Open

2004

Pupil phase apodization for achromatic imaging of extra-solar planets

Weidong Yang
Michigan Technological University

Follow this and additional works at: https://digitalcommons.mtu.edu/etds

Part of the Physics Commons

Recommended Citation

Yang, Weidong, "Pupil phase apodization for achromatic imaging of extra-solar planets", Dissertation, Michigan Technological University, 2004.
https://digitalcommons.mtu.edu/etds/120

Follow this and additional works at: https://digitalcommons.mtu.edu/etds

Part of the Physics Commons

Pupil Phase Apodization for Achromatic Imaging of Extra-solar Planets

by

Weidong Yang

A Dissertation

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

Physics Department

Michigan Technological University

This dissertation, *Pupil Phase Apodization for Achromatic Imaging of Extra-solar Planets*, is hereby approved in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the field of Physics.

DEPARTMENT: Physics

_____

Thesis Advisor: Alexander B. Kostinski

_____

Department Head/Chair: Ravi Pandey

Date:_____

Acknowledgment

Abstract

Direct imaging of extra-solar planets in the visible and infrared region has generated great interest among scientists and the general public as well. However, this is a challenging problem. Difficulties of detecting a planet (faint source) are caused, mostly, by two factors: sidelobes caused by starlight diffraction from the edge of the pupil and the randomly scattered starlight caused by the phase errors from the imperfections in the optical system. While the latter difficulty can be corrected by high density active deformable mirrors with advanced phase sensing and control technology, the optimized strategy for suppressing the diffraction sidelobes is still an open question.

In this thesis, I present a new approach to the sidelobe reduction problem: pupil 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

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

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.

In addition to the specific phase functions that can yield deep sidelobe reduction on one quadrant, we also found that a modified Gerchberg-Saxton algorithm can help to find small sized (101×101 element) discrete phase functions if regional sidelobe reduction is desired. Our simulation shows that a 101×101 segmented but gapless active mirror can also generate a dark region with Inner Working Distance about $2.8\lambda/D$ in one quadrant.

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.

Contents

List of Figures

3

Chapter 1

General Introduction

The nine planets of our solar system, including our own Earth, revolve around the Sun, which is one of the hundreds of billions of stars in our galaxy. It is natural for one to ask: are there any other planets revolving around their Suns beyond our solar system? Speculation about the existence of other planets beyond our solar system can be traced back to the ancient Greeks and has further triggered scientists' investigations regarding the existence of life outside our solar system. In 1950, Fermi asked the famous question "where is everybody?" and started debates about the prevalence of extraterrestrial civilizations. In 1961, astronomer Frank Drake presented his approach, the Drake equation, to estimate the number of technological civilizations that may exist in our galaxy.

However, the existence of the extra-solar planets – which is an essential factor in Fermi's debate and in Drake's estimation – was a mere speculation based on the non-uniqueness of our solar system in the universe and has lasted for several centuries without being confirmed by observations.

Recently, a series of detections of planet-like objects around a pulsar and Sun-like stars have been reported [1, 2, 3, 4]. Till today there have been over 100 detection reports. The burst of these new discoveries in the last 14 years resulted from the development of detection ideas and technologies [1].

One of the principal techniques being applied in the search for extrasolar planets (and the most successful to date in terms of the number of confirmed detections) is the so-called

---

[1]There are many academic websites introducing the basic concepts of these technologies, one of them being http://planetquest.jpl.nasa.gov/

Radial Velocity Measurements method, sometimes also called Doppler Spectroscopy method. Consider a planet and Sun system which has an orbit plane aligned in such a way that it is not perpendicular to our line of sight. Then, the Sun would have chance to move towards or away from us periodically while it is circulating around their mass-center. If the Sun's motion toward or away from us is fast enough, then we would have chance to measure the periodic Doppler effect (blue shift or red shift) in the star's spectral lines. The magnitude of this effect can enable us to estimate the unseen planet's mass and orbit information. Jupiter, for example, causes a movement of the Sun with an amplitude of 13 m/s and a period of 12 years while the earth can cause a small movement of about 10 cm/s amplitude over a period of one year. Based on current technology with precision in the range 3-10m/s, Jupiter-like planets would have a better chance to be detected by this method.

There are also many other promising detection methods that are being developed, such as

- Astrometry - this method is based on the precise measurement of the periodic and slightly wobbling motion of the star. The unseen companions' gravity will drag the star to wobble around their mass center , so the star's motion would not follow a smooth trajectory across the sky if there are unseen companions around it. The precise measurement of the position of star can be accomplished by measuring the relative distances to the nearby background stars.

- Photometry - the brightness of the star will be slightly dimmed when a planet passes between the star and the observer. The photometry method is a method to detect and characterize extrasolar planets through the accurate quantitative measurement of the periodic change of light. From the period and depth of the transit changes, the orbit and size of the unseen companions can be calculated.

- Microlensing - This method is based on Einstein's theory of general relativity. The light beam will be bent and the propagating direction will be changed when the light beam is passing through the space with a massive object nearby. When a planet is passing in front of the star, its gravity will "attract" the star light passing by and behave like a focusing lens and thus can cause a transitory change in the brightness and apparent position of the star. By accurately measuring these changes, the properties of the planets

can be calculated.

This brief sketch is not complete and a more comparative review of the methods for extrasolar planet detection can be found in papers by Schneider [5] and Perrymann [6]. In the remainder of this thesis, I will concentrate on the direct imaging method.

Direct imaging method

Along with the development of various indirect detection technologies, direct imaging methods are also being studied intensively. Unlike the indirect methods introduced above, in which the existence of a planet is demonstrated by the observation and measurement of the "wobbling" motion or the periodic brightness change of their host stars, direct imaging of the planets will allow us to capture the weak light signals from the planets and give a direct and visualized proof of their existence. The discovery of the first Brown Dwarf Gliese 229B in 1995 was a good demonstration of the power of the direct imaging method. Although there haven't yet been any successful imaging report of extrasolar planets, scientists' efforts have not diminished. On the contrary, this has become one of the hottest areas in astronomy and astrophysics. Why?

- Direct imaging can give an unambiguous answer to the presence of a planet once it is detected.

- Direct imaging needs shorter detection time. To observe the periodic variations of measured quantities (such as the doppler shift, the position, and the brightness), the observation time for comparative analysis will have to last at leat one period which could be a few days to a few years or even longer. However for direct imaging with new technologies, the entire process could be complete in a few hours or less.

- Unlike the indirect methods which can provide only information about the mass and orbits, direct imaging can provide additionally spectral information about the planetary atmosphere's chemical composition and biomarkers life right after the planet is imaged [7, 8, 9, 10, 11].

The earlier strategy for imaging extra-solar planets was first proposed by Bracewell and MacPhie in 1978 and 1979 [12, 13] whose method was based on the infrared interferometers

and has been treated as one of the two most promising methods to date. Of the two methods proposed, the infrared interferometer method has less optical requirements but its long baseline and cost for deployment is making it still difficult. The other promising method is the visible-light coronagraph in which only a single mid-sized monolithic telescope is needed. This thesis will concentrate only on the latter approach and contribute a new method, pupil phase-only modulation method, for achromatic imaging of extra-solar planets.

Chapter 2

Technical Background

2.1 Challenges in direct imaging

Though promising, imaging extrasolar planets is difficult because their light signals, compared with that of the nearby host stars, are too weak to be detected by regular space or ground based telescopes. To speak more specifically, as studied in papers such as [14, 15, 7, 16, 17], there are two main noise sources from the host star that are much larger than the signal of planets and make the planet signal lost in the noise. One is the noise due to the scattering effect of the phase errors induced by the imperfection of the optical mirrors or air turbulence. The other is the shot noise induced by the strong diffraction side lobes of the star nearby.

2.1.1 Wave optics diffraction fundamentals

According to the scalar diffraction theory, the energy of a plane light wave, incident on a telescope aperture, will be spread into a wide solid angle and form sidelobes. Specifically for the near field, the Fresnel diffraction equation for propagation between two planes P0 and P1 separated by a distance z is given by:

$$\tilde{E}(\eta, \xi) = \tilde{C} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{T}(x, y) \cdot e^{i\frac{\pi}{2\lambda z}[(x-\eta)^2 + (y-\xi)^2]} dx dy \qquad (2.1)$$

where the (x, y) are the coordinates in the aperture (pupil) plane P0, (η, ξ) are the coordinates in the observational plane (diffraction plane) P1 and the $\tilde{(\cdot)}$ indicates a complex function.

9

Adding a thin lens with focal length τ over the pupil plane is equivalent to adding a phase shift $-\frac{\pi}{2\lambda\tau}(x^2 + y^2)$ to the incident wave. Then on the focal plane where $z = \tau$, the diffraction field would have the form of

$$\tilde{E}(\eta, \xi) = \tilde{C} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{T}(x, y) \cdot e^{-i\frac{2\pi}{\lambda\tau}(x\eta + y\xi)} dx dy \tag{2.2}$$

which indicates that the diffraction field on the focal plane of a thin lens is just the Fourier Transform of the light field on the aperture (pupil) plane. The intensity on the focal plane then becomes

$$I(\eta, \xi) = |\tilde{E}(\eta, \xi)|^2 \tag{2.3}$$

2.1.2 Diffraction sidelobes of the star

For a round aperture telescope, the pattern of the diffraction field on the image plane can be described by a Jinc function. The Jinc function [18] is defined as $Jinc(x) \equiv \frac{2 \cdot J_1(x)}{x}$ where $J_1(x)$ is a Bessel function of the first kind and the factor 2 is used here so that $Jinc(0) = 1$. The light intensity I_0 of the ideal circular pupil diffraction pattern can be written as:

$$I_0(\rho) = Jinc(\pi\rho)^2 \tag{2.4}$$

where ρ is the dimensionless field angle θ/θ_0, $\theta_0 = \lambda/D$ (D is the diameter of the ideal circular pupil) and the central intensity is normalized to 1. This pattern has some ring like structures described by the Left panel in Figure 2.1.

About 84% of the total energy is focused within the main lobe and the other 16% of the light energy is spread into the sidelobes (rings).

Using the asymptotic form for the Bessel function, the ratio of the average sidelobe intensity to the central intensity can be written as:[19]

$$\frac{I_0(\rho)}{I_0(0)} \approx \frac{4}{\pi} \frac{1}{(\pi\rho)^3} \tag{2.5}$$

which means that on average the sidelobe intensity drops off as an inverse cube of the dimensionless distance ρ. Due to the far distance between the star and the telescope, the
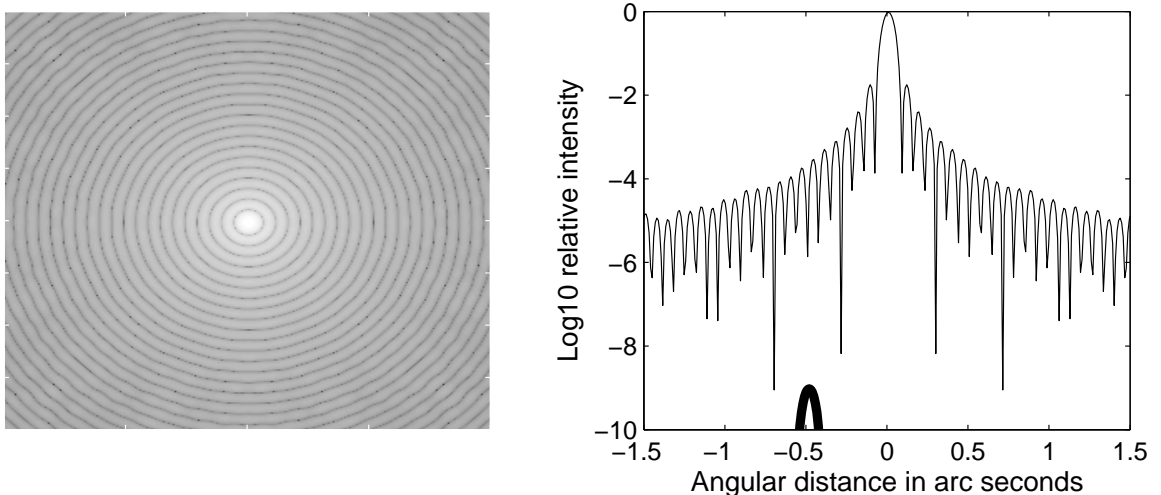
Figure 2.1: Circular aperture diffraction pattern and comparison of relative intensity of an extra-solar planetary system. Left: the ring-like diffraction pattern from a circular aperture, in log_{10} scale. Right: comparison of the log_{10} relative intensity distribution of an Jupiter-Sun extra-solar planetary system, in which the thin solid line represents the intensity distribution from the Sun and the thick solid line represents the intensity distribution from the Jupiter. This comparison is based on a simulated observation made from distance $10pc$ at wavelength 0.8μ through a space telescope similar to HST. One can see that the Jupiter image is much fainter than the local diffraction sidelobes of the Sun.

star light entering the pupil can be treated as a plane wave and it will form a image pattern described by equation 2.4 and 2.5, and so does the light from the planets.

Assume the planet's brightness is q times of the brightness of the star and located at dimensionless distance ρ' from the star (which is located at $\rho = 0$). Then the light intensity at the planet's location would be estimated as the sum of the peak intensity of the planet and the local sidelobe intensity due to the star [1], which is $I_0(0) \times (q + \frac{4}{\pi} \frac{1}{(\pi\rho')^3})$. Based on the Poisson statistical properties of the photon detection [20], the photon noise will be at the level of $\sqrt{t \times I_0(0) \times (q + \frac{4}{\pi} \frac{1}{(\pi\rho')^3})/(h\nu)}$ on per unit area, where t is the imaging time and ν is the frequency of light. Therefore, for the unit detection area, the average signal to noise ratio S/N can be estimated as (for q very small compared with the local sidelobe level):

$$S/N \sim \frac{\pi^2}{2} \times \sqrt{t \cdot \frac{I_0(0)}{h\nu}} \times q \times \rho'^{3/2} \qquad (2.6)$$

Now let's use the Sun-Jupiter system as an example. The brightness of the planet is only

---

[1]The presence of the host star image in the telescope is due to the fact that the planets are always associated with a star nearby. We will have to use the star to guide us to places where we search for the planets.

a factor of 10^9 less than the brightness of the star in the visible light [15]. If we observe such a system at a distance about $10pc$ [2], the angular separation between the star and the planet is about 0.5 arc seconds [3]. In the case of observation such a planet around the wavelength 0.8μ using a space telescope which is similar in size to Hubble Space Telescope, the planet will be located at about the 7th sidelobe ring of the star and the planet would be 10^{-5} fainter than the sidelobe intensity, as shown in the Right panel in Figure 2.1. Assuming a $3.3m^2$ collecting area of the telescope with a 50% of bandwidth at the 0.8μ wavelength, the telescope would receive 2×10^9 photons per second from a Sun-like star and about 2 photons per second from the Jupiter-like planet. From equation 2.6, we can estimated that for reaching a signal to noise ratio better than $S/N = 1$, the imaging time would have be at least 28 hours; for achieving a detection with S/N better than 5, the imaging time would have to be at least 28×5^2 hours or 29 days !

Apparently, in order to obtain an planet's image with confidence and within a reasonable time, the local sidelobes from the nearby star will have to be reduced at region where the planets might be located. In practice, the local intensity level of the star diffraction sidelobes should be equal or less than the intensity of the planet image core, therefore relative to the intensity of the star image core, the sidelobes level should be reduced to 10^{-10} for imaging earth-like planets and to 10^{-9} for imaging Jupitor-like planets. The physics and technologies for reducing the sidelobes will be the main topic of this thesis, and the detailed discussions will be presented in later chapters.

2.1.3 Scattering of phase errors

Phase errors are induced by the turbulence of atmosphere and the imperfections of the optical elements. The phase errors will lead part of the star light energy to deviate from the original propagation direction and form randomly distributed visible or invisible grain-like speckles in a wide region on the image plane.

---

[2] pc, also called $parsec$, is an astronomical unit for large distance. $1pc$ is the distance at which 1 AU would subtend a second of arc. AU, a short form the Astronomical Unit, is another distance unit for astronomy, in which $1AU$ is the average distance between the Earth and the Sun. $1AU = 1.495978 \times 10^{11} meters, 1pc = 3.085678 \times 10^{16} meters = 3.261633ly = 206265AU$.

[3] 1 arc second is one 3600th of a degree, or equivalently 4.848×10^{-6} radians

Imaging extra-solar planets places a much more stringent requirement on the phase errors even in the cases that satisfy the Maréchal's criteria used for so called "diffraction limited" in practice. Marréchal's criteria defines the Strehl ratio[4] $S > 0.8$ as the condition of "diffraction limited" and this is equivalent to phase errors less than $\lambda/14$ [21]. For example, HST (Hubble Space Telescope) is diffraction-limited at wavelengths longer than $\lambda > 0.33\mu$, but it is still not sufficient to image Jupiter-like extra-solar planets even if the shot noise from the diffraction sidelobes is not considered. [17]. Now let's see an example of how a small phase error that is from the imperfections of the reflecting mirror surface can affect the imaging of extra-solar planets. Consider a optical system with small phase error $\triangle\phi(\vec{x})$, which $\triangle\phi(\vec{x}) \ll 1$ and the average of $\triangle\phi(\vec{x})$ over pupil is zero, $\overline{\triangle\phi(\vec{x})} = 0$. The light field on the pupil plane can be written as $T(\vec{x})\cdot e^{i\triangle\phi(\vec{x})} \approx T(\vec{x}) \cdot [1 + i\triangle\phi(\vec{x})]$, where $T(\vec{x})$ is the pupil function in the domain of $\vec{x}$. The light intensity $I(\vec{\eta})$ on the focal (image) plane would be

$$I(\vec{\eta}) \approx |\mathcal{F}\{T(\vec{x})\} + i\mathcal{F}\{T(\vec{x}) \cdot \triangle\phi(\vec{x})\}|^2 \tag{2.7}$$

where $\mathcal{F}\{\cdot\}$ denotes the Fourier Transform.

If we choose the sinusoidal deformation along the y direction as an example of the small phase error, $\triangle\phi(\vec{x}) = \frac{2\pi h}{\lambda} \cdot \cos(\frac{2\pi y}{l_0})$ where h is the amplitude of the deformation of the reflecting mirror and l_0 is the spatial period of the deformation along the y direction, and if we assume that the pupil function $T(\vec{x})$ is real and even (which is true for most cases), then the light intensity expression, equation 2.7, would be simplified to

$$I(\vec{\eta}) \approx \widehat{T}^2(\eta,\xi) + (\frac{\pi h}{\lambda})^2 \cdot [\widehat{T}^2(\eta, \xi - l_0) + \widehat{T}^2(\eta, \xi + l_0) + 2 \cdot \widehat{T}(\eta, \xi - l_0) \cdot \widehat{T}(\eta, \xi + l_0)] \tag{2.8}$$

where $\widehat{T}(\eta, \xi)$ is the Fourier Transform of $T(\vec{x})$ which is real and even here. The $\widehat{T}^2(\eta, \xi)$ becomes the $I_0(\rho)$ described in equation 2.4 if the pupil is a round pupil.

It can be seen that there are two small speckles at distance $\pm l_0$ with coefficient $(\frac{\pi h}{\lambda})^2$ added to both side of the ideal diffraction pattern symmetrically . Assume that $h \sim \lambda/80$(which is equivalent to $h \sim 10nm$ when $\lambda = 800nm$), then, as shown in Figure 2.2 the two speckles will

---

[4]Strehl ratio S is defined as the ratio of the central intensities of the aberrated point spread function and the diffraction-limited point spread function. It is a quantity that relates phase aberration and image quality in a high quality imaging system.
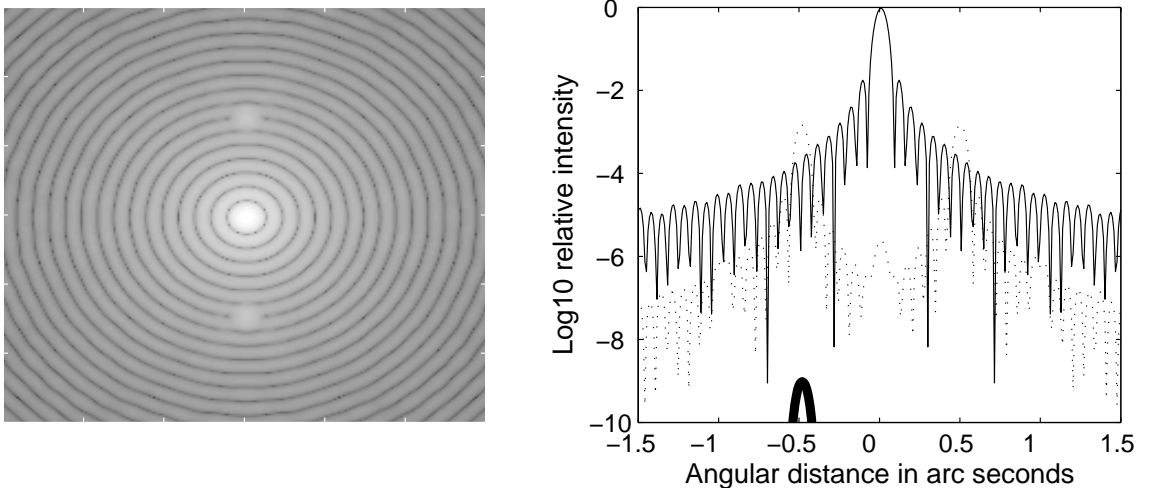
Figure 2.2: Scattering effect from a small phase error. Left: the log_{10} PSF with sinusoidal phase error $\triangle\phi(\vec{x}) = \frac{2\pi h}{\lambda}\cdot\cos(\frac{2\pi y}{l_0})$ in which the sinusoidal error period l_0 is one seventh of one aperture and error amplitude h is $\lambda/80$. Right: comparison of the log_{10} relative intensity of an Jupiter-like planet with the scattering background generated from the sinusoidal phase error as in the Left panel. The dotted line represents the scattering background due to the phase error. The thick solid line represents the relative intensity distribution of Jupiter. The thin solid line, as a reference, represents the relative intensity distribution of the Sun without phase errors. This comparison is based on a simulated observation made from distance $10pc$ at wavelength 0.8μ through a space telescope similar to HST. One can see that even a small phase error, as the one illustrated in the Left panel, can generate a strong scattering background and make the detection of a Jupiter-like planet difficult.

have peak power of only 3 orders less than the brightness of the star and are one million times brighter than the image of the Jupiter-like planets! So we can see that even small phase errors can add a large (compared with the signal of the planets) background and make imaging of even the Jupiter-like planets difficult. (Even in the case that these speckles are static and can be removed by subtracting in the data processing, the remaining shot noise generated by the speckles is still large within a reasonable time). Generally in practice for imaging extra-solar planets, the local scattered light level is expected to be equal or less than the intensity of the planets.

A more detailed and general description of the small surface error's scattering effect on Point Spread Function (the round pupil as an example)can be described as: [22, 23, 17, 24, 19]

$$I(\rho) \approx [1 - (\frac{2\pi}{\lambda}\sigma_{rms})^2] \times I_0(\rho) + (\frac{2\pi}{\lambda})^2 P(\rho) \qquad (2.9)$$

where $I_0(\rho)$ is the ideal PSF(Point Spread Function), $I(\rho)$ is the perturbed PSF with small phase errors, $P(\rho)$ is the PSD(Power Spectrum Distribution) and σ_{rms} is the root-mean-squared phase error in unit of distance. The PSD of the surface small errors is the Fourier Transform of the autocovariance function which Glenn [25] determined as the form an exponential function characterized by correlation length of the surface errors.

The phase errors caused by the atmospheric turbulence are large and vary with both time and space over a very wide dynamic range. Long-exposure point object images obtained through the ground-based telescopes will be blurred and have a wide halo around it. The mechanism of the turbulence was first revealed by Kolmogorov and its influences on the wave propagation were studied by Tatarski [26], Hufnagel and Stanley [27] and Fried [28] et al. A comprehensive review can be found in Hardy's book [29] and Roggemann and Welsh's book [30].

Obviously to reduce the scattered light, adaptive optics will have to be used to correct these phase errors to some extent. Because of the high dynamic range of the turbulence varying in time and space, it is extremely difficult to use a ground-based telescope to image the extra-solar planets in the visible light although Angel [16], Sandler, Stahl and Angel [31] and Labeyrie [32] have suggested different ways either using complex adaptive optics strategies and fast control loops to take the extra-solar planets' long-exposure image or finding the image of extra-solar planets in multiple short-exposure images.

Space telescopes are not affected by atmospheric turbulence. Consequently, the phase errors – due principally to polishing errors in manufacturing and strains from gravity and temperature fluctuations – remain static. This allows use of a slow control loop and high density active mirrors to make corrections for the phase error.

However, successful correction of the phase errors for the space telescope relies on the accurate knowledge of the phase front information and precise control of the deformable mirrors.

Many phase sensing techniques have been developed such as Shack-Hartmann sensors [33] which measure the slope (gradient) of the wavefront, curvature sensors [34, 35] which measure the curvature of the wavefront, interferometers and phase-retrieval methods which include phase-diversity based [36, 37] parametric methods [38, 39, 40] and iterative methods

[41, 42, 43], etc. These techniques have been widely used for estimating the optical misalignment and calibration. Current studies are examining their potential to sense the small phase errors at a precision of below $\lambda/10,000$. Currently, two methods proposed in JPL have been demonstrated by simulation to have such ability [43, 44].

There are two main issues regarding the ability of the deformable mirrors to compensate for phase errors. The first one is the size of actuators required. Hudgin [45] studied the relationship between the residue wavefront compensation error and the spacing of the finite corrector-element and concluded that, due to the finite size of the corrector-element, corrections can not be perfect and the residues are strongly related to the ratio of the size(spacing) to the coherence length of the wavefront: a higher ratio would yield lower residues. Further simulations [46, 47] show that the correction dynamic range (in spatial frequency) is limited by the Nyquist limit determined by the number of actuators per aperture. 96×96 elements/aperture would be sufficient to provide the required dynamic range (in spatial frequency) and sufficiently low residue error for imaging Jupiter-like planets if the elements(actuators) can be controlled precisely. The second issue is the stability of controlling the elements (or actuators). JPL has demonstrated with an experiment that its deformable mirror driven by the 96×96 actuators can be controlled with the precision and stability at $0.01 \mathring{A}/hr$ and $1 \mathring{A}/100hr$ [48] which will ensure the phase error requirement for imaging Jupiter-like planets in visible light in a wide dynamic range.

2.2   Review of the coronagraphic solution

In the preceding section, the difficulties in direct imaging of extra-solar planets have been explained and the current technologies in sensing and correcting the small phase errors for space telescopes have been introduced. As briefly described, the background from the phase error scattering can be reduced to the level of allowing imaging of the Jupiter-like extra-solar planets. In this section, we will review some of the methods that can be potentially used for reducing the diffraction sidelobes of the stars.

2.3 Coronagraph

The notion of a coronagraph was a technique originally invented for observing the corona of the Sun through the telescope at times other than during a solar eclipse. The inventer was Lyot and he published his work in 1939 [49]. Figure 2.3 shows the basic configuration of the Lyot coronagraph in which there are two key elements: the opaque mask (sometimes called density mask or amplitude mask) on the first focal (image) plane and the Lyot stop on the re-imaged pupil plane (sometimes called pupil relay plane).
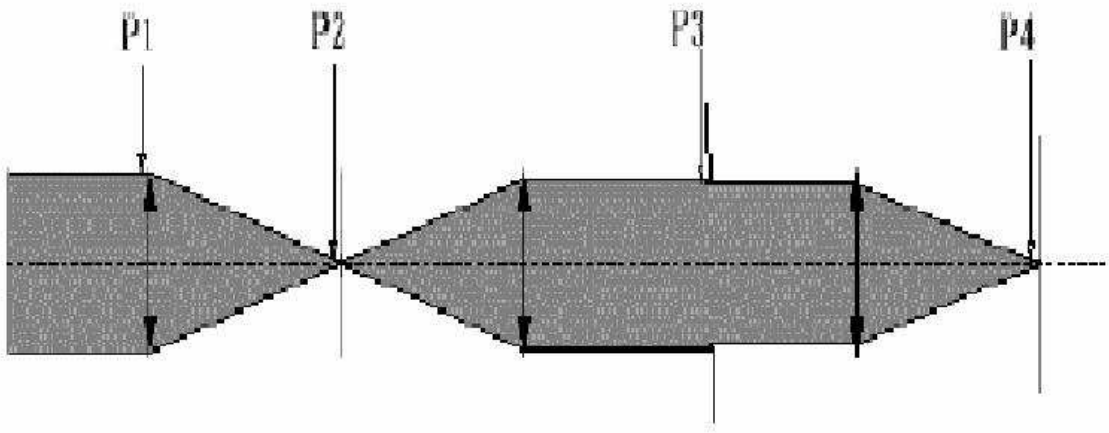
Figure 2.3: Schematic coronagraph layout. P1 is the pupil plane where the aperture is located. P2 is the first focal plane where the occulting mask is located. P3 is the re-imaged pupil plane, or called pupil relay plane, where the Lyot stop is located. P4 is the final focal plane where the faint source images are obtained.

When the light energy of an on-axis star is focused on the first focal plane, the opaque mask will cover the star image and part of the halo of it, and the remaining star halo light that is not blocked by the mask will continue to propagate and form a ring like pattern (around the edge of the re-imaged pupil) on the re-imaged pupil plane. The Lyot stop(an opening slightly smaller than the image of the pupil) will block the ring, thus here the star light experiences a second blockage and the starlight energy can be greatly reduced, forming a very dim star image on the second focal plane. Light from any off-axis faint object whose image is located outside the mask will not be blocked and will continue to propagate, forming a complete image of the pupil on the re-imaged pupil plane. Because most of the energy can pass the Lyot stop in this case, the image of these off-axis objects can be reproduced on the second focal plane.

Currently this technique has been refined and adapted for the purpose of imaging extrasolar planets that are very close to the star. For example, instead of using the hard edge opaque mask, an opaque mask tapered by, say, a Gaussian function will be used in new designs and a tapered Lyot stop is also found to be helpful for deeper reducing and smoothing the starlight halo background on the second focal plane. Some optimized designs of the mask and Lyot stop size can lead to reduction of starlight by about 9 orders of magnitude [47] [5].

The simulation fundamentals in the coronagraph design

In recent years there are some other coronagraph techniques that have been developed and can be classified into two types: the density mask coronagraph and the phase mask coronagraph. The analysis of their performance are based on the following:

Assume the light field on plane P_1, P_2, P_3 and P_4 are $\tilde{E}_1$, $\tilde{E}_2$, $\tilde{E}_3$ and $\tilde{E}_4$ respectively, then based on the Fourier Transform relations between the two adjacent planes in above, we have:

on P_2

$$\tilde{E}_2 = \mathcal{F}\{\tilde{E}_1\} \tag{2.10}$$

on P_3

$$\tilde{E}_3 = \mathcal{F}^{-1}\{\tilde{M} \cdot \tilde{E}_2\} \tag{2.11}$$

where $\tilde{M}$ is the mask transmittance function, $\mathcal{F}^{-1}\{\cdot\}$ denotes the inverse Fourier Transform operation.

on P_4

$$\tilde{E}_4 = \mathcal{F}\{\tilde{L} \cdot \tilde{E}_3\} \tag{2.12}$$

where $\tilde{L}$ is the Lyot stop transmittance function.

Density mask coronagraph

The density mask is the mask that the transmittance rate is modulated by some functions without modulating the phase. The Lyot coronagraph is a Density mask coronagraph in which the mask transmittance rate is modulated by a Gaussian function or simply a binary function. Other density mask coronagraphs include the Band-Limited mask coronagraph developed by

---

[5]Plots can also be found in Lyon's work on:

$http://code935.gsfc.nasa.gov/cube\,Folder/OSCAR/pdf/Coron\_PR.pdf$

Kuchner and Traub in 2002 [50]. In their design, the mask transmittance function is designed as a set of 1D or 2D functions, such as $\sin^4(\pi\epsilon D_\lambda x/2)$ where D_λ is the diameter of the pupil normalized to wavelength λ and ϵ is a non-zero parameter, so that the multiple pupil images with specific phase relations can be generated. In effect, the nulling interference can happen in the overlapped region and the starlight can be removed by adding a stop that has the shape of the overlapped region.

Phase mask coronagraph

An interesting alternative to using the density mask is to use phase mask. The phase mask is a mask that modulates only the phase, rather than the intensity transmittance rate. Here, the mask transmittance function is a complex function in equation 2.12 with intensity transmittance rate being 1 .

The first phase mask coronagraph was invented by Roddier in 1997 [51] in which the phase mask is a round π phase shift retarder with diameter less than $0.53\lambda/D$. The second phase mask design [6] was proposed by Roaun et al in 2000 [52] in which the phase mask is divided into four quadrants, each of which has a π phase shift relative to its neighbor quadrant. The diagrams of the Rouan's four-quadrant phase mask and the nulling effect on the pupil image are shown in figure 2.4. The idea of reducing the starlight by these phase masks is realized by separating the starlight into two beams and flipping the phase of one of the beam on the first focal plane and then combining destructively the two light beams on the re-imaged focal plane.

---

[6]A history fact here: exactly the same design was done at the end of year 1999 by Weidong Yang and Christ Ftaclas of Michigan Tech. Unfortunately our work was not submitted for journal publication in time. Our results were first posted at the departmental colloquium graduate poster section held around Feb. of year 2000 (see http://www.phy.mtu.edu/colloquiums/gradposters/deptgradposter2000.html). Our work was formally posted in the 197th AAS meeting of January 2001 (see http://www.aas.org/publications/baas/v32n4/aas197/1172.htm), but it was about one and half month later than the publication of the Rouan et al's work. This mask design has now become one of the candidates for the NASA's TPF mission (see http://planetquest.jpl.nasa.gov/TPF/TPFrevue/FinlReps/JPL/tpfrpt1a.pdf).
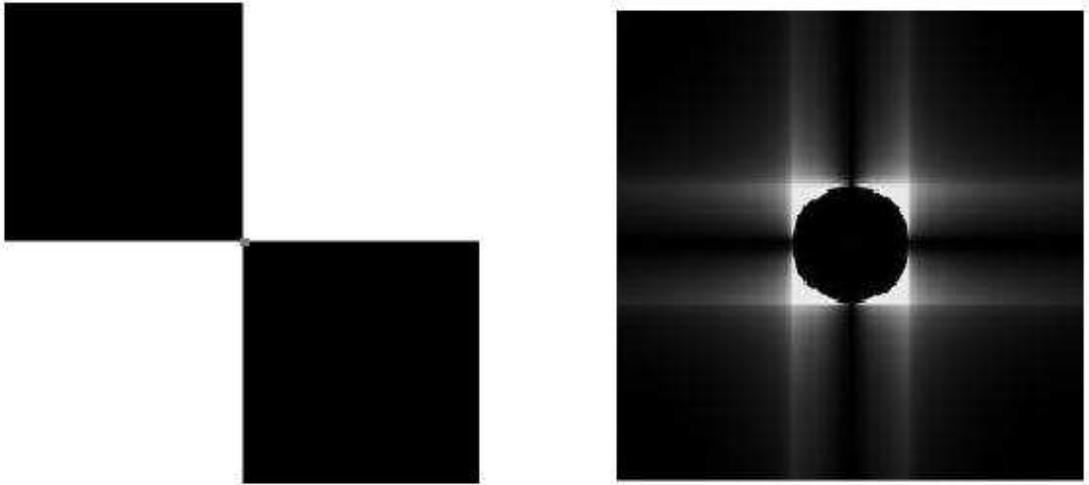
Figure 2.4: The four-quadrant phase mask and the annulling effect on the on-axis star light that can be observed on the re-imaged pupil plane. Left: gray-scale illustration of the phase shift distribution on the phase mask, the white quadrants have 0 phase shifts while the dark quadrants have π shifts. This mask divides the on-axis star light into 2 beams which have π phase difference. Right: the light intensity image on the re-imaged pupil plane, it can be seen that the destructive interference (annulling) happens inside the pupil image while constructive interference happens outside the pupil image.

Discussion

All of these coronagraph designs have sufficient star image suppression effects for the purpose of imaging extra-solar planets. However they have also many disadvantages:

1. They all are sensitive to the pointing errors. All the simulation is based on the ideal case when the center of the star image core is right in the center of the mask. Breaking the symmetry will induce the leakage of the star energy into the final focal plane and contaminate the background. A pointing error sensing and correcting system will have to be used and it should have the challenging control ability of about 1 milli arcsecond [53].

2. For the phase mask coronagraph, though it can provide a smaller inner working distance, the π phase shift can only be realized for a certain wavelength. The reduced background will be $\propto \pi^2 \cdot (\frac{\Delta\lambda}{\lambda})^2$

3. For the density mask coronagraph, the intensity transmittance noise on the scale of $\sim \lambda/D$ near the image core of the star is required to be less than 10^{-8} in order to avoid the background's being contaminated by the leaked starlight [54, 55], which is hard to realize.

2.4 Pupil apodization

In the past a few years, pupil apodization has been intensively studied for the purpose of imaging extra-solar planets. This method is based on the Fourier Transform relations between the light field on the pupil plane and the light field on the focal plane.

In Greeks, "a-"means to take away and "podis-" means the foot [56]. The process of reducing the diffraction sidelobes is called apodization. Various methods were studied intensively and comprehensively by Jacquinot about 40 years ago regarding the apodization in imaging and radar radiation patterns. Recently apodization methods have been further developed so that they can meet the stringent requirements in imaging extra-solar planets. Together with coronagraph, apodization has been regarded as an important approach in the task of imaging extra-solar planets and they will be briefly reviewed in this chapter.

Apodization can be further classified into two types also: Pupil Amplitude Apodization and the Pupil Shape Apodization

2.4.1 Amplitude apodization

Pupil Amplitude Apodization is used to reduce the starlight sidelobes by modulating the amplitude transmission function over the pupil. In 2001, P. Nisenson and C. Papaliolios proposed using a sonine function to modulate the amplitude transmission function along x and y direction for the square pupil, in which pupil amplitude transmission function $T(x, y)$ is

$$T(x,y) = \begin{cases} (1-(x/D)^2)^{\nu-1}(1-(y/D)^2)^{\nu-1}, & |x/D| \leq 1, |y/D| \leq 1 \\ 0, & else \end{cases} \tag{2.13}$$

where the D is the side of the square pupil. Their simulations show that the maximum reduction happens along the diagonal line with IWD (Inner Working Distance[7]) about $4\lambda/D$ for reduction level 10^{-10} when $\nu = 4$. Larger ν can create smaller IWD at the expense of throughput.

---

[7]The Inner Working Distance–IWD is defined, by Kasdin et al., as the smallest angular separation from the star for which the sidelobes contrast with the PSF core reaches the required value, for example 10^{-10}. It is the closest position to the star at which a detection is possible.

In 2003, Kasdin et al [57] and Gonsalves and Nisenson [37] described the effect of amplitude modulation for the square and round pupil by prolate spheroidal functions. Their results show that round pupil with amplitude modulated by the prolate spheroidal function can yield full discovery space with IWD$\sim 3.5\lambda/D$ and reduction level below 10^{-10}. It is especially noteworthy that Gonsalves and Nisenson [37] gave a general way to calculate the amplitude apodiser for various pupils.

2.4.2   Binary pupil shape apodization

Binary pupil shape apodization is used to realize the sidelobe reduction by modulating the shape of the pupil opening within which the transmission rate is 1.

The sidelobe reduction effect by different pupil shapes was described by Jacquinot and Roizen-Dossier in 1964 [56]. Angel et al, in 1986, proposed using a ring-like binary pupil to reduce the sidelobes at the position that a earth-like planet may be located [7]. The most interesting binary pupil was the "eye-like" pupil which was first suggested to apply to imaging extra-solar planets by Spergel in 2000 and then developed by Kasdin et al in 2003.

The "eye-like" pupil and the corresponding PSF are shown in figure 2.5. By Guyon's notation [58], the corresponding pupil function is defined as: the transmission function is 1 if and only if the following relations are satisfied

$$y_0(x) < |y| < y_0(x) + y_w(x) \tag{2.14}$$

with

$$y_w(x) = R(e^{-(\frac{\alpha x}{R})^2} - e^{-\alpha^2}) \tag{2.15}$$

and

$$\int_{y_0(x)}^{y_0(x)+y_w(x)} y^2 dy = \beta y_w(x) \tag{2.16}$$

where $\alpha = 2$ and $\beta = 0.4R^2$ and transmission rate is 0 elsewhere. This binary pupil can help reduce the sidelobes to below 10^{-10} along the x-axis with a IWD of about $4\lambda/D$.

For the purpose of expanding the discover space, further developments on using multiple "eye-like" openings and using "star-like" or multiple "ring-like" openings were also discussed by Kasdin et al.
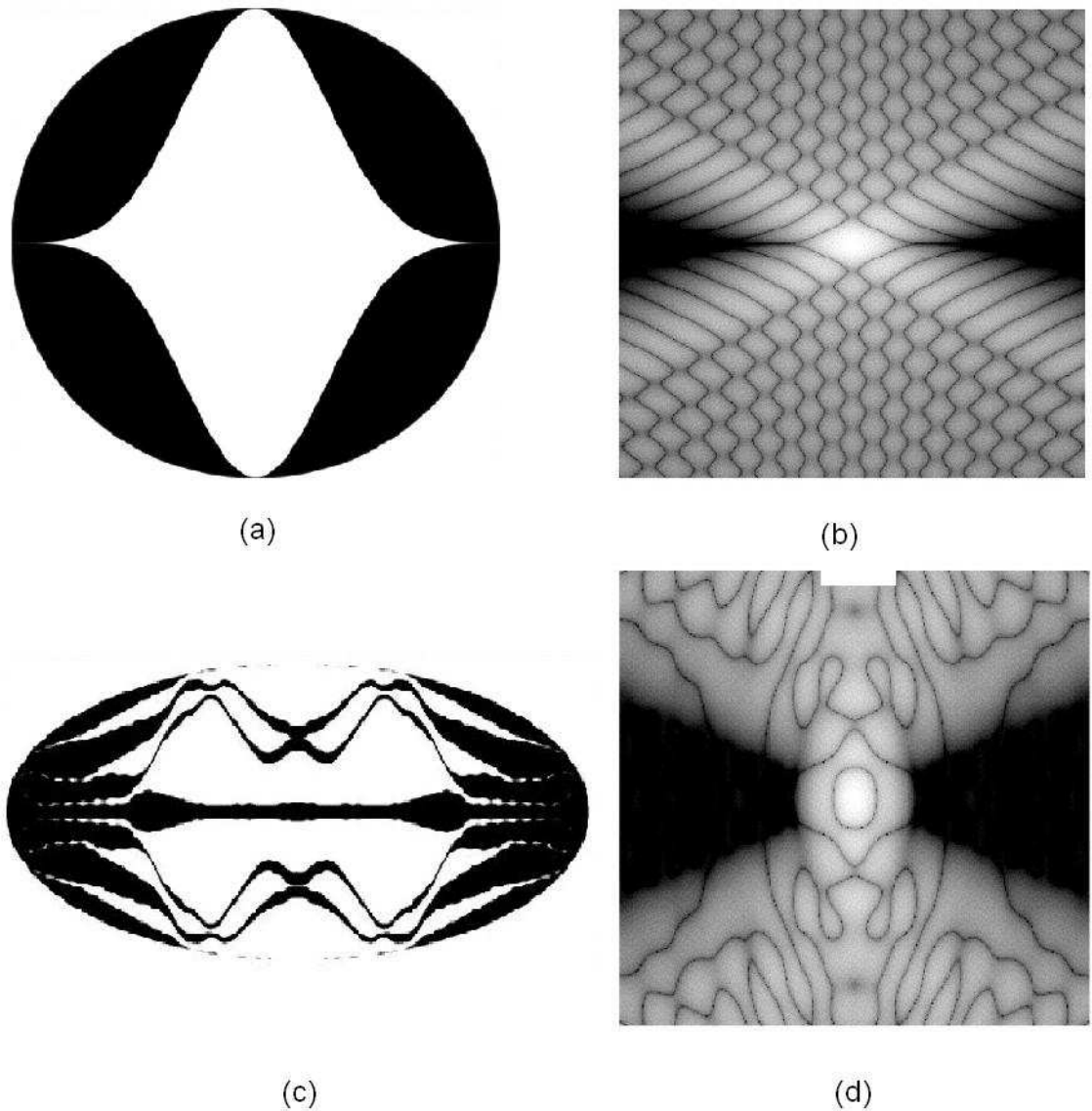
Figure 2.5: "eye-like" binary pupil mask and the corresponding PSFs. (a)the "eye-like" binary pupil mask inscribed in a round pupil. (b)the PSF of the mask in (a). (c)Six binary openings inscribed in an elliptical pupil. (d)the PSF of the mask in (c).

Discussion

Compared with the density or phase mask coronagraph, pupil apodization methods won't suffer from the pointing errors. In addition, the pupil shape apodization method avoids the stringent requirement on the errors in the amplitude transmission rate as it has for the case of amplitude modulation.

However, making the pupil edge shape might be difficult. Besides, both of the amplitude apodization and the shape apodization will cause the planet energy loss from about $50\% - 70\%$

which is also expensive for detecting the planet's signals and they can not be recovered. In addition, it will require rotating the telescope in order to have wider discovery space.

Chapter 3

One-sided achromatic phase-apodization

3.1 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 [7, 9]. 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 and Burrows [17] 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 [13]; (ii) visible light imaging based

on the coronagraph concept (originally introduced by Lyot in 1939 in the solar physics context [49]). Because of the longer wavelength, decreased angular resolution in the infrared region must be overcome by spatial separation but in this chapter we shall 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 side lobe levels below 10^{-9} [15, 14] and the need for a small inner working of a few λ/D, are daunting. Nevertheless, recent developments in phase sensing and control technology will enable the figure errors of up to 100 cycles/aperture to be controlled under 1 $\mathring{A}$ per 100 hours [47, 44, 43]. Therefore, coronographic imaging in the visible light region is feasible [46].

As to the diffraction side lobe 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 the modulation of the phase or amplitude of the star diffraction pattern on a focal plane, which includes: Roddiers' phase mask [51], Rouan et al's four quadrant phase mask [52], the Kuchner & Traub's band-limited mask [50], and the Kuchner& Spergel's notch filter mask [54]. The second type is based on the modulation of the amplitude transmission function of the pupil (pupil amplitude apodization) [56] as well as pupil shape "apodization". This field has undergone an explosion of activity recently, which has included several exciting results incuding: Angel et al's ring-like binary mask [7]; Kadsdin et al's "eye-like" shaped binary mask [57]; Vanderbei et al's spider binary mask [59]; Nisenson et al's amplitude-apodized pupil mask [60]; Guyon's amplitude pupil apodization by beam reshaping [58]. Besides the specific mask designs mentioned above, the papers by Gonsalves and Nisenson [37] and by Aime et al [61] 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 [56] 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 chapter 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.

3.2   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 our paper in Physics Letter A[62].

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 by Nisenson and Papaliolios [60] who "revived" the square pupil for extra-solar planet detection. In addition, the pupil shape optimization work reported by Kasdin et al.[57] suggested to us the idea of *partial* side lobe removal in a focal plane, at any given time. This led to consideration 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.

3.3   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 [63]

$$E(\eta) = \mathcal{F}\{T(x)e^{i\phi(x)}\} \qquad (3.1)$$

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

An equivalent expression is

$$E(\eta) = \mathcal{F}\{T(x)\} \otimes \{\mathcal{F}\{\cos[\phi(x)] + i\sin[\phi(x)]\}\} \qquad (3.2)$$

where $\otimes$ denotes the convolution. For a one-dimensional "pupil" of width D, the transmission function $T = 1$ for $|x| \leqslant D/2$ and $T = 0$ otherwise. The Fourier Transform of $T(x)$ is $Sinc(\eta) \equiv \sin(\pi\eta)/\pi\eta$ 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 square of two real functions:

$$I(\eta) = \{Sinc(\eta) \otimes \mathcal{F}\{\cos[\phi(x)]\}\}^2 + \{Sinc(\eta) \otimes \mathcal{F}\{\sin[\phi(x)]\}\}^2 \qquad (3.3)$$

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 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):

$$I(\eta) = \{Sinc(\eta) \otimes \mathcal{F}\{\cos[\phi(x)]\} + Sinc(\eta) \otimes \mathcal{F}\{i\sin[\phi(x)]\}\}^2 \tag{3.4}$$

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.

3.4   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

$$\phi_1(x) = a \cdot \tan[(0.5 - \varepsilon) \cdot 2\pi x/D] \tag{3.5}$$

$$\phi_2(x) = a \cdot \ln \frac{(1 + \varepsilon) + 2x/D}{(1 + \varepsilon) - 2x/D} \tag{3.6}$$

where ε is a small (positive) parameter, defined so that the phase value is finite at the edge of the pupil. The shapes of these two phase functions are illustrated in Figure 3.1 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 ε parameters to a later section.)

The annulling effect by the destructive interference described in equation 3.4 is illustrated in Figure 3.2 and Figure 3.3. The example of a diffraction pattern caused by phase function ϕ_1
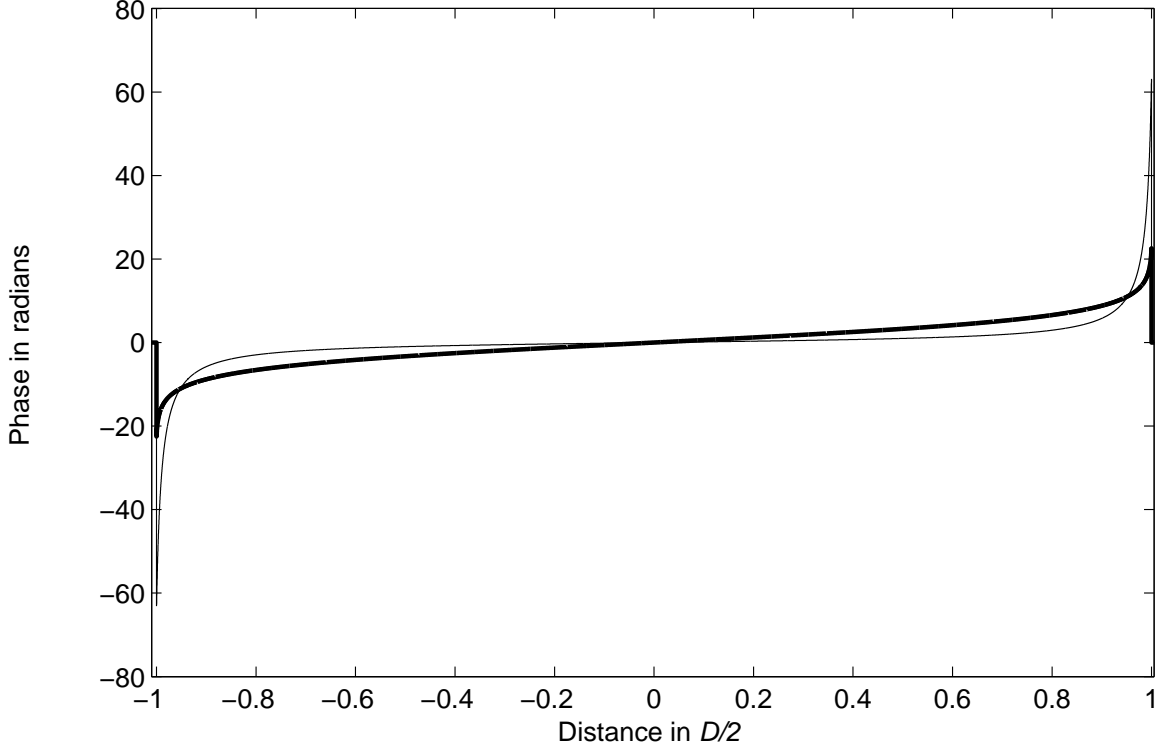
Figure 3.1: Shapes of the 1-D phase delay functions over the pupil. The thinner solid line represents the phase $\phi_1(x) = a \cdot \tan[(0.5 - \varepsilon) \cdot 2\pi x/D]$ with $a = 1$ and $\varepsilon = 0.005$; The thicker solid line represents the phase $\phi_2(x) = a \cdot \ln \frac{(1+\varepsilon)+2x/D}{(1+\varepsilon)-2x/D}$ with $a = 3$ and $\varepsilon = 0.001$.

with $a = 1$ and $\varepsilon = 0.005$ is shown in Figure 3.2. Similarly, Figure 3.3 displays the diffraction pattern for the case of the phase function ϕ_2 with $a = 3$ and $\varepsilon = 0.001$. In order to "zoom in" and see the details, in Figure 3.4 we demonstrate the annulling effect on 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 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, ϕ_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.
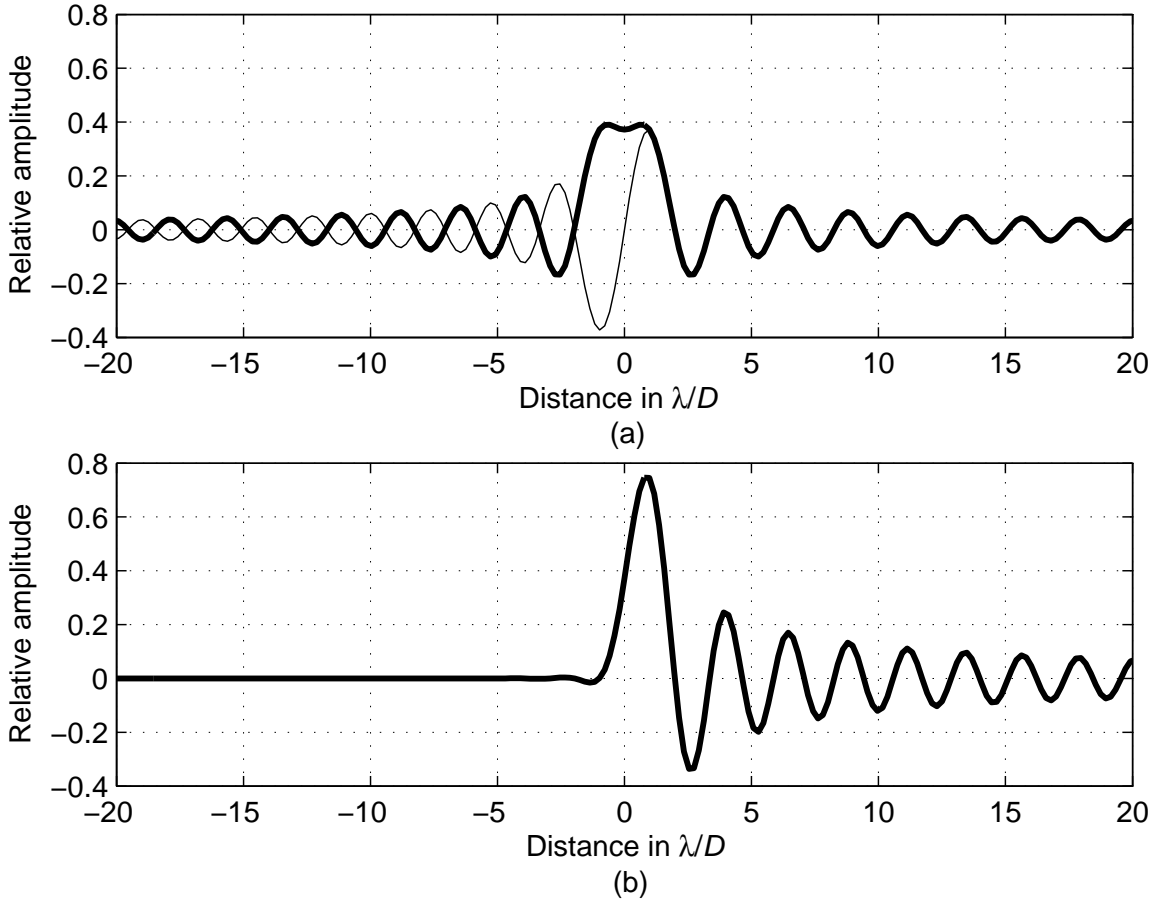
Figure 3.2: 1-D annulling effect due to the phase function $\phi_1(x) = a \cdot \tan[(0.5 - \varepsilon) \cdot 2\pi x/D]$ with $a = 1$ and $\varepsilon = 0.005$. In the simulation, 12800 pixels are sampled over the pupil D and the amplitude is normalized to peak amplitude without phase modulation. In (a), the $Sinc(\eta) \otimes \mathcal{F}\{\cos[\phi_1(x)]\}$ is represented in thicker solid line, while the $i \cdot Sinc(\eta) \otimes \mathcal{F}\{\sin[\phi_1(x)]\}$ 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.
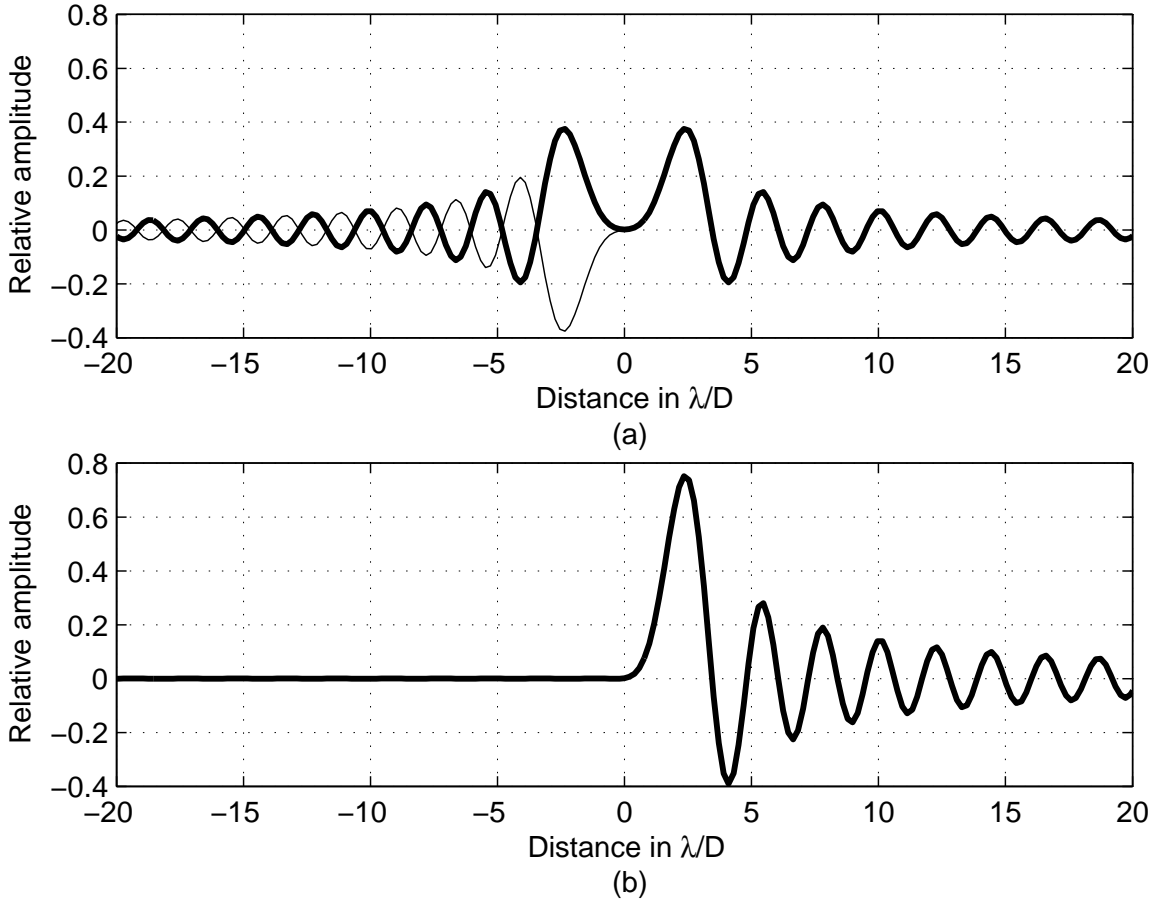
Figure 3.3: 1-D annulling effect due to the phase function $\phi_2(x) = a \cdot \ln \frac{(1+\varepsilon)+2x/D}{(1+\varepsilon)-2x/D}$ with $a = 3$ and $\varepsilon = 0.001$. In the simulation, 12800 pixels are sampled over the pupil D and the amplitude is normalized to peak amplitude without phase modulation. In (a), the $Sinc(\eta) \otimes \mathcal{F}\{\cos[\phi_2(x)]\}$ is represented in thicker solid line, while the $i \cdot Sinc(\eta) \otimes \mathcal{F}\{\sin[\phi_2(x)]\}$ 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.

3.5 1-D Focal Plane Performance, Parameter Selection, and Sampling

Figure 3.4: 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 $\phi_1(x) = a \cdot \tan[(0.5 - \varepsilon) \cdot 2\pi x/D]$ with $a = 1$ and $\varepsilon = 0.005$. The dashed line represents the logarithm (base 10) relative intensity of of th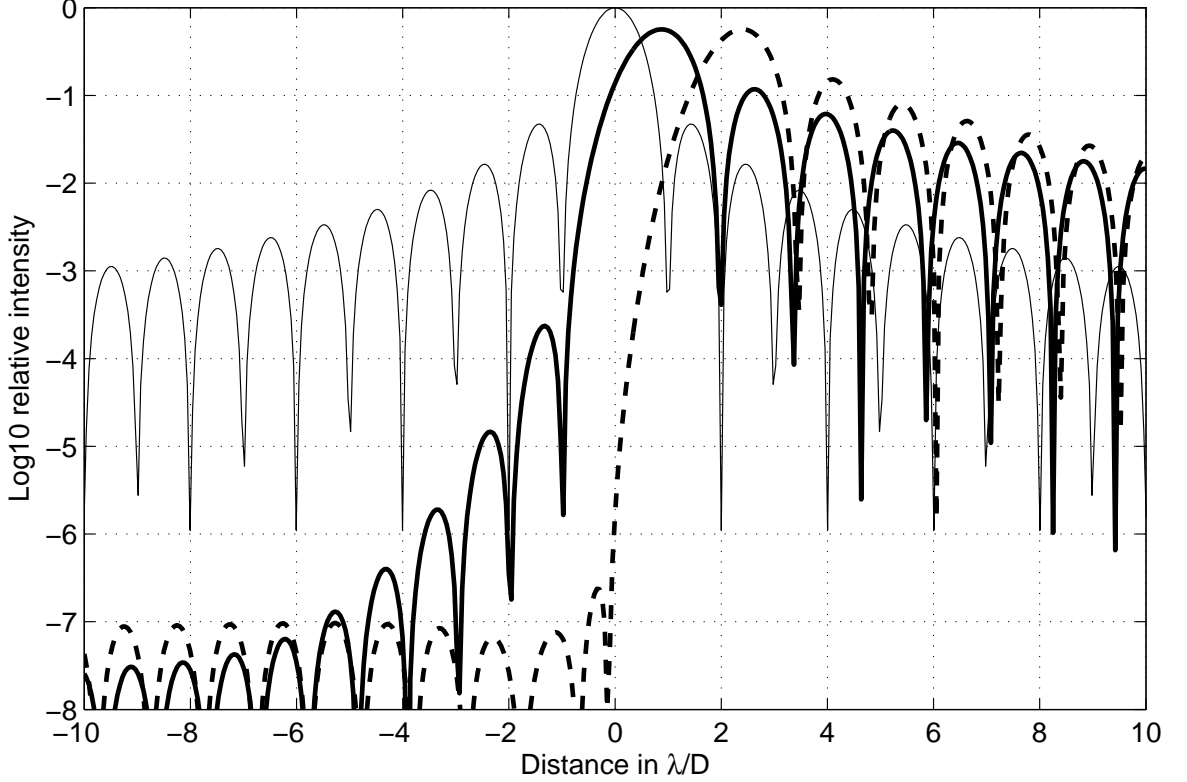e annulling results due to $\phi_2(x) = a \cdot \ln \frac{(1+\varepsilon)+2x/D}{(1+\varepsilon)-2x/D}$ with $a = 3$ and $\varepsilon = 0.001$. The thinner solid line is, as a reference, the diffraction intensity without phase modulation. One can see a sharper reduction in employing ϕ_2 than that in employing ϕ_1.

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 3.4 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 λ/D and the 20th λ/D. The corresponding curves are shown in Figure 3.5, labelled (a) and (c). The curve (a) corresponds to ϕ_1 while (c) corresponds to ϕ_2. It can be seen that the reduced intensity level is strongly
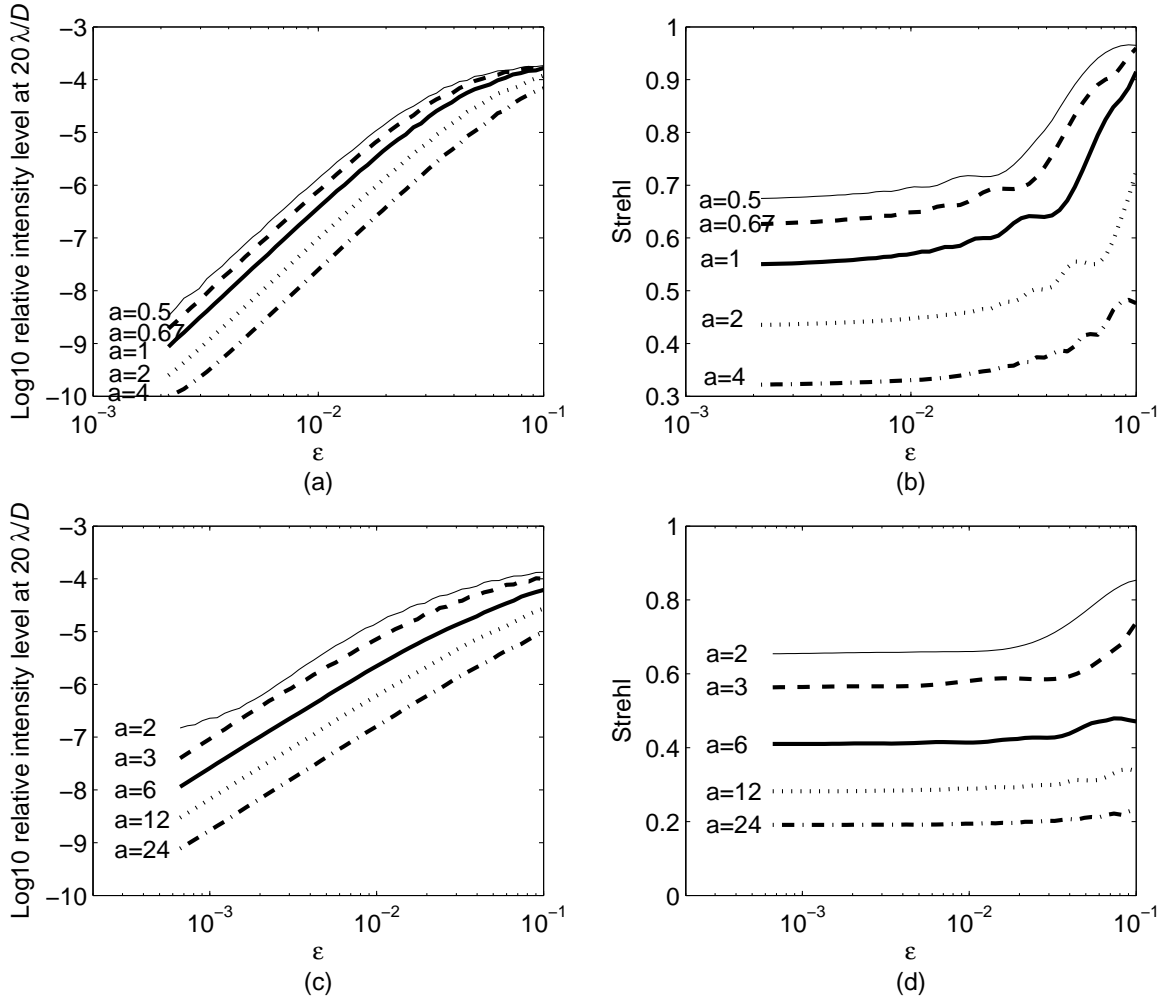
33

Figure 3.5: 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. ε at different values of $a = 0.5, 0.67,$ 1, 2 and 4 while (b) shows the relation of the Strehl ratio vs. ε at different values of $a = 0.5, 0.67, 1, 2$ and 4. Panel (c) and (d) are for the cases with phase modulation $\phi_2(x) = a \cdot \ln \frac{(1+\varepsilon)+2x/D}{(1+\varepsilon)-2x/D}$, in which (c) shows the relation of logarithmic (base 10) relative intensity level around $20\lambda/D$ vs. ε at different values of $a = 2, 3, 6,$ 12 and 24 while (d) shows the relation of the Strehl ratio vs. ε at different values of $a = 2, 3, 6, 12$ and 24.

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. In Figure 3.5, the corresponding curves are labelled as (b) and (d), where (b) is associated with ϕ_1 and (d) with ϕ_2.

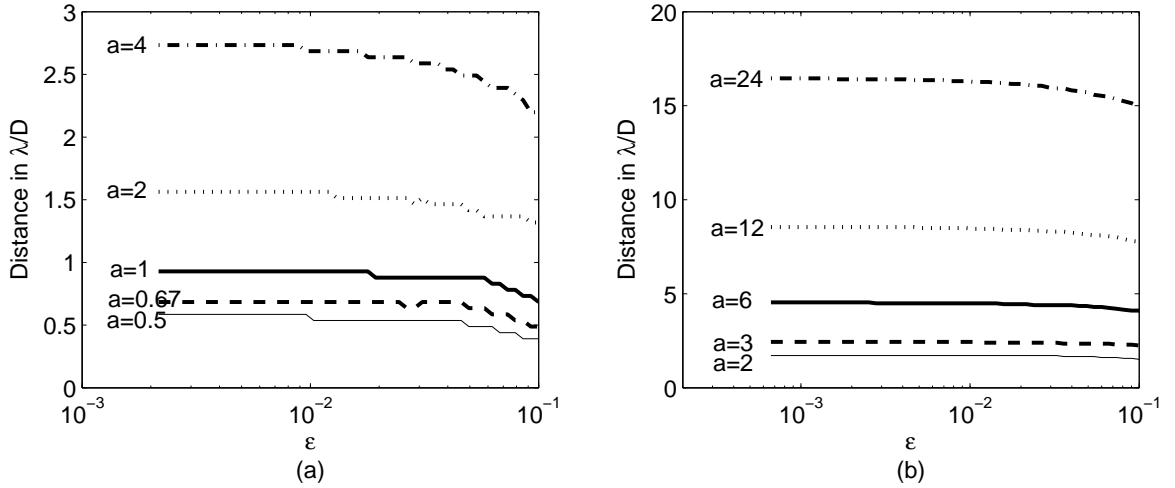Figure 3.6 shows relations between the shifted distance of the peak and the function

34

Figure 3.6: 1-D relation between the shifted peak and the parameters. Panel (a): the shifted distance of the peak in units of λ/D vs. ε 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 λ/D vs. ε at different values of a =2, 3, 6, 12 and 24 with phase modulation $\phi_2(x) = a \cdot \ln\frac{(1+\varepsilon)+2x/D}{(1+\varepsilon)-2x/D}$. The step-like shape in the curves is due to finite number of pixel points within each λ/D interval. In this calculation, there are 20.56 pixels within each λ/D interval.

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 the above Figure 3.5 and Figure 3.6 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 3.5 and equation 3.6 have another common qualitative features: a substantial and rapid rate of change near the edge of the pupil, as shown in Figure 3.1. 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 $N \geqslant \frac{D}{\pi} \cdot \frac{d\phi(x)}{dx}|_{max}$. Figure 3.7 shows the relationship between the estimated minimum number of sampled pixels N_{min} and the parameter ε for the phase
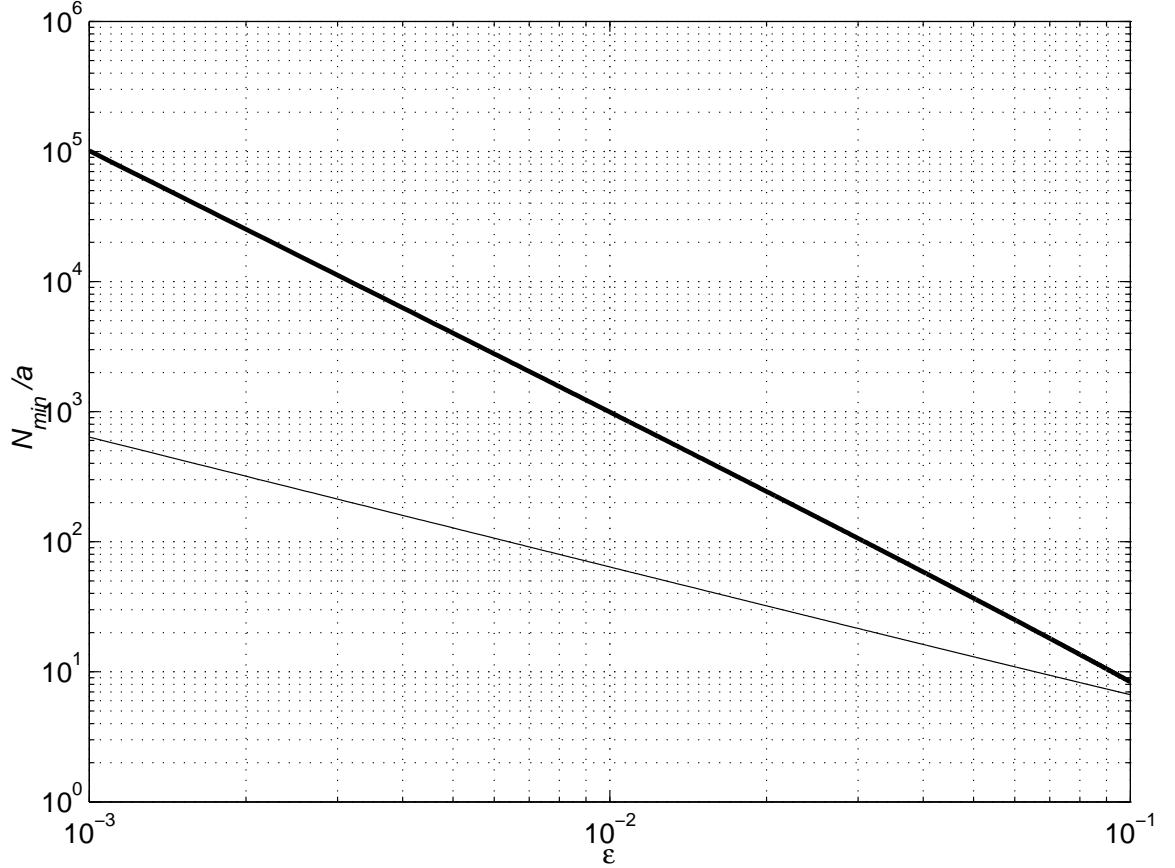
Figure 3.7: 1-D relationship between parameter ε and N_{min}/a, where N_{min} is the estimated minimum number of pixels sampled over the pupil D and a is the parameter in the phase functions. The thicker solid line is for the case with phase modulation $\phi_1(x) = a \cdot \tan[(0.5 - \varepsilon) \cdot 2\pi x/D]$. The thinner solid line is for the case with phase modulation $\phi_2(x) = a \cdot \ln \frac{(1+\varepsilon)+2x/D}{(1+\varepsilon)-2x/D}$.

functions. For smaller ε, 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 ε 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).

3.6 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 interest in the field imaging of extra-solar planet imaging [60, 57]. 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 $I(\eta, \xi) = I(\eta) Sinc(\xi)$, where $I(\eta)$ is the 1-D diffraction field intensity determined by equation 3.4 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 3.4.

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 3.8 where we show logarithmic intensity image produced by the phase function $\phi_1(x) + \phi_1(y)$. The panel (b) of Figure 3.8 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 3.8 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.
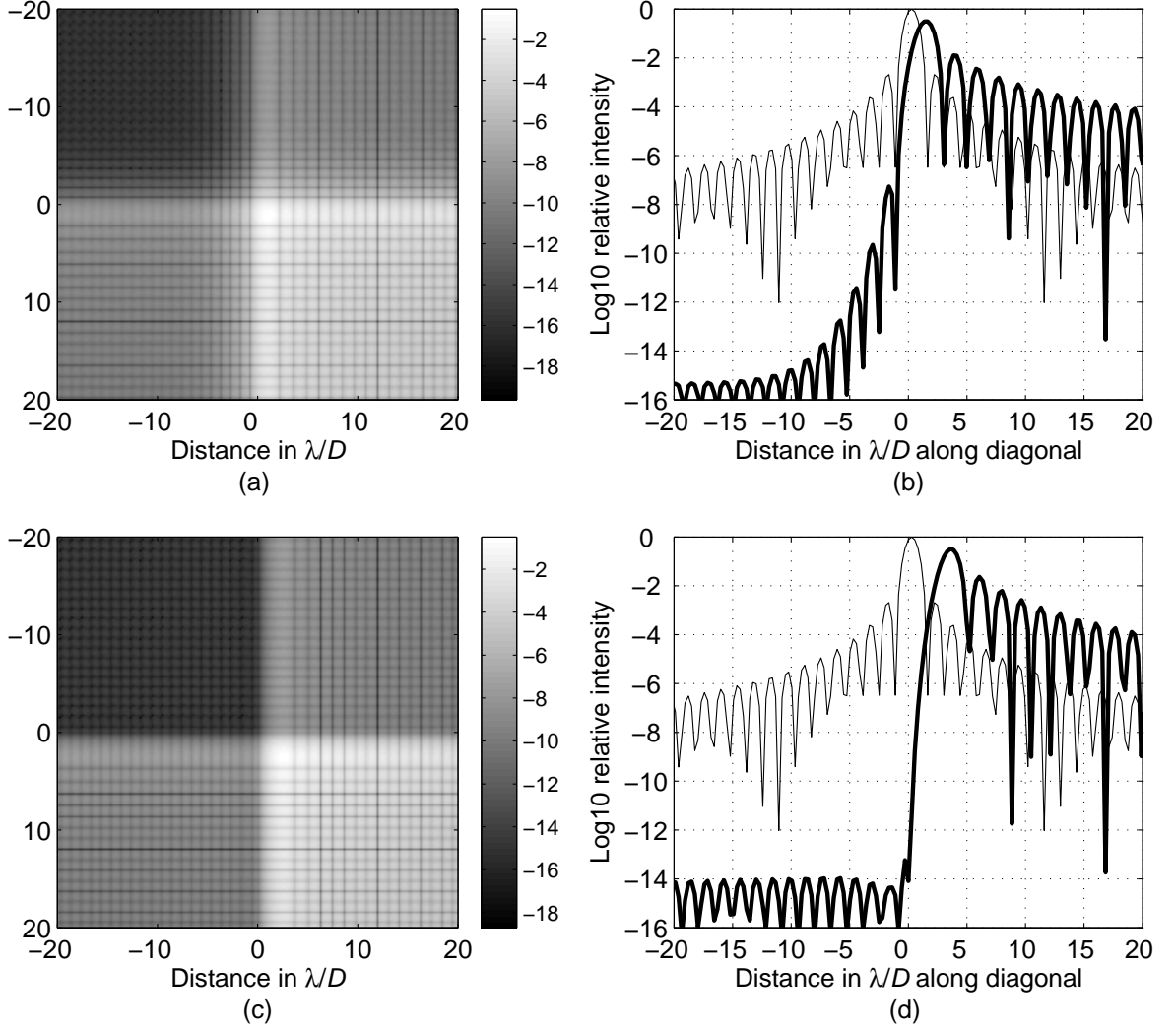
Figure 3.8: Light reduction effect on one quadrant of focal plane when phase function is applied along x and y directions for the square pupil. Panel (a): Logarithmic (base 10) relative intensity image when phase $\phi(x,y) = a \cdot \tan[(0.5 - \varepsilon) \cdot 2\pi x/D] + a \cdot \tan[(0.5 - \varepsilon) \cdot 2\pi y/D]$ 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 $\phi(x,y) = a \cdot \ln[\frac{(1+\varepsilon)+2x/D}{(1+\varepsilon)-2x/D} \cdot \frac{(1+\varepsilon)+2y/D}{(1+\varepsilon)-2y/D}]$ 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.

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

38

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 λ_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 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

$$\phi(x, \lambda) = \frac{\lambda_0}{\lambda} \frac{G(\lambda)}{G(\lambda_0)} \phi(x) \tag{3.7}$$

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 3.5 and 3.6. 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 λ. Therefore, based on parameter relations of Figure 3.5, 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 3.9 and still keep a small inner working angle and low reduction level.
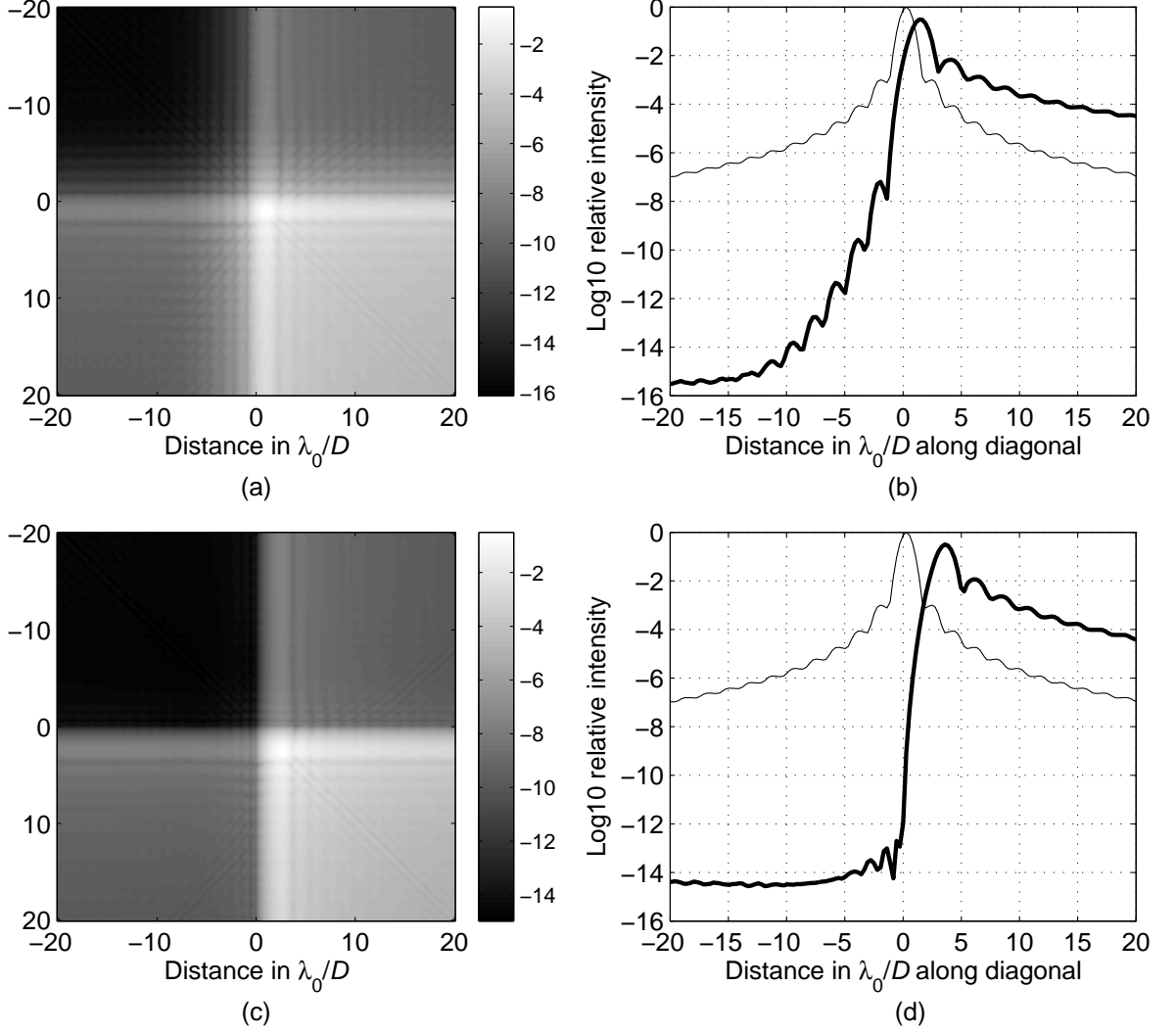
Figure 3.9: Broad bandwidth light reduction effect on one quadrant of focal plane. The simulation is based on a rectangular spectrum distribution with total bandwidth of $60\%\lambda_0$. Panel (a): Logarithmic (base 10) relative intensity image when phase $\phi(x,y) = a \cdot \tan[(0.5 - \varepsilon) \cdot 2\pi x/D] + a \cdot \tan[(0.5 - \varepsilon) \cdot 2\pi y/D]$ 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 $\phi(x,y) = a \cdot \ln[\frac{(1+\varepsilon)+2x/D}{(1+\varepsilon)-2x/D} \cdot \frac{(1+\varepsilon)+2y/D}{(1+\varepsilon)-2y/D}]$ 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 reduction level of 10^{-12} with a 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.

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

40

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

$$E(\eta, \xi) \approx \mathcal{F}\{T(x, y) \cdot e^{i\phi(x,y)}[1 + i\delta\phi(x, y)]\} \tag{3.8}$$

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. Then, the light intensity in the dark region is given by

$$\delta I(\eta, \xi) \approx |E_{ideal}(\eta, \xi) \otimes \mathcal{F}\{\delta\phi(x, y)\}|^2. \tag{3.9}$$

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):

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \delta I(\eta, \xi)d\xi d\eta \approx \int_{-D/2}^{D/2}\int_{-D/2}^{-D/2} |\delta\phi(x, y)|^2 dxdy. \tag{3.10}$$

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 [47, 48] 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 $\phi(x, y) = a \cdot \ln[\frac{(1+\varepsilon)+2x/D}{(1+\varepsilon)-2x/D} \cdot \frac{(1+\varepsilon)+2y/D}{(1+\varepsilon)-2y/D}]$ ($a = 3$ 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 3.10), 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 3.5 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 [21]. This results in phase errors within the critical spatial requency region below $15.5 \times 10^{-4} rad\ rms$ or below $2.5 \times 10^{-4}\ \lambda\ rms$. This is feasible with current technology as reviewed in the Introduction.
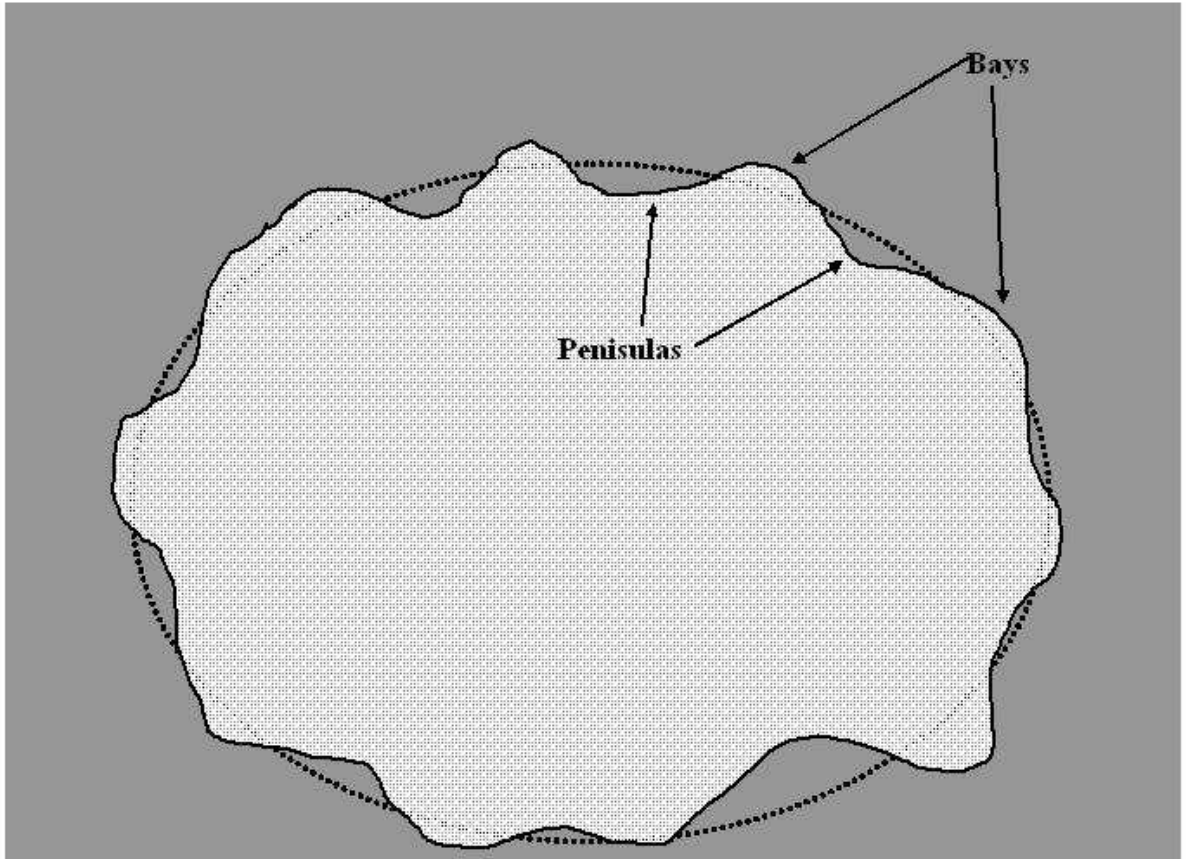
Figure 3.10: Illustration of the actual pupil which is formed by the sum of the ideal pupil and the "peninsulas" and "bays" around the ideal pupil boundary. The ideal pupil is the region encircled by the dotted lines, the actual pupil is the white region encircled by the solid lines.

Let us next address the precision requirements for the pupil shape. The rough edge (pupil boundary) is illustrated in Figure 3.10 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

$$n \sim L \cdot l^3/A^2 \tag{3.11}$$

and, based on equation 3.11, 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.

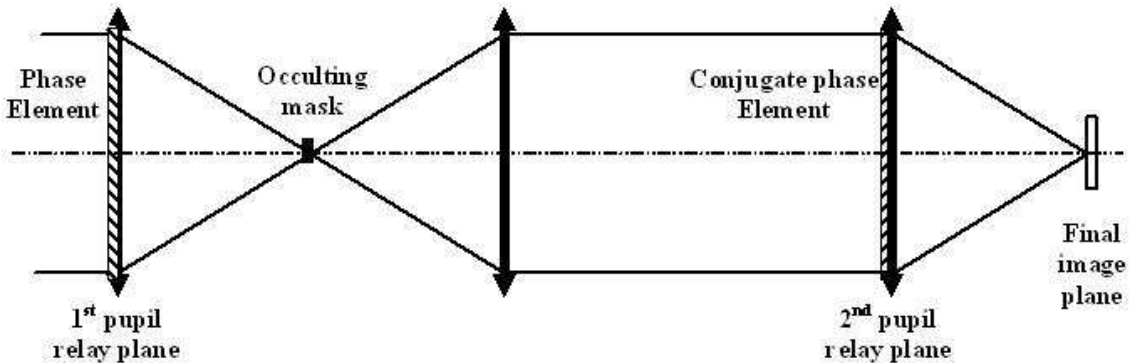**Diffraction-limited Planet Imaging by Combining Coronagraph and a Conjugate Phase Plate**

Figure 3.11: Schematic layout for restoring the diffraction limited images of the extrasolar planets.

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
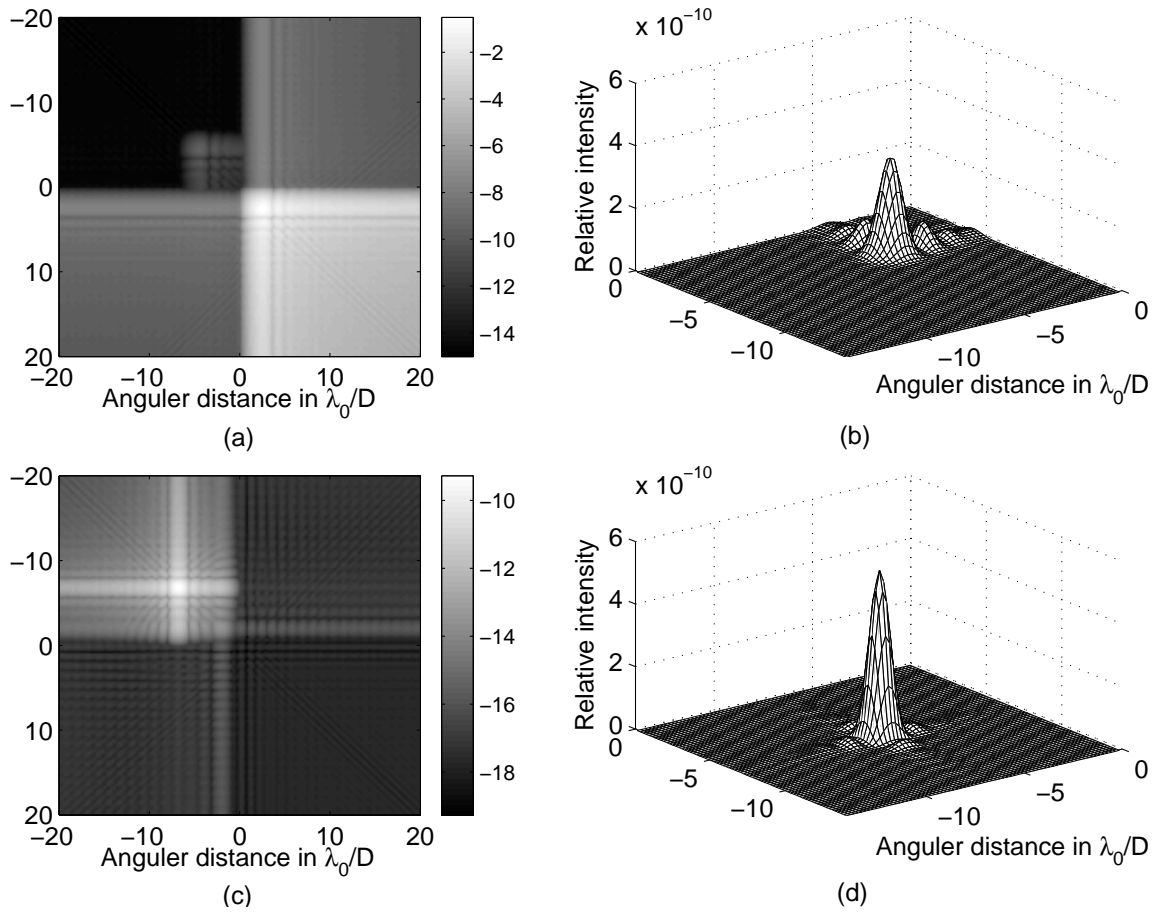
Figure 3.12: Comparison of the images of the planet before and after combination use of an occulting mask and a conjugate phase element. The occulting mask covers the light in the 1st, 3rd and 4th quadrants. The pair of conjugate phase is based on $\phi(x,y) = a \cdot \ln\left[\frac{(1+\varepsilon)+2x/D}{(1+\varepsilon)-2x/D} \cdot \frac{(1+\varepsilon)+2y/D}{(1+\varepsilon)-2y/D}\right]$ 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.

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 3.9. 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?

44

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 3.11, 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 3.12 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.

3.7 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 3.1 corresponds to the 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 [64], Fluid Jet Polishing [65], Ion Beam Milling [66], Wet-Etch Figuring [67], 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 [47, 48, 68], 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., [43]. 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.

Chapter 4

Regional phase apodization via the Gerchberg-Saxton Algorithm

4.1 Introduction

In resolving two adjacent point light sources that are far away in the presence of the phase aberrations, adaptive optics or active optics can help obtain high contrast, diffraction limited images of them. However if the brightness of one of the source is extremely faint compared with the other, even in the absent of phase aberration, the faint resource is still hard to detect because diffraction side lobes from the adjacent brighter source may cover the signal from the faint one and the shot noise from diffraction local side lobes of the brighter source can be larger than the signal of the fainter source within a tolerable integration time. This problem may exist in the observation of the superfine structure in spectroscopy or sonar or radar detections, but may be best exemplified in the imaging of extra-solar planets which have appeal public interest in recent years [7, 16, 29].

The extra-solar terrestrial planets may be 10^{10} or even more fainter in the visible light than the parent star and their separation angle with the parent star could be as close as 0.25 arcsecond which corresponds to a few diffraction ring of their parent star if it is intended to be observed by a moderate sized telescope like Hubble Space Telescope. To detect such a faint signal as the extra-solar planet in the vicinity of a bright star within a reasonable time, the diffraction side lobes must be reduced to about the brightness level of the fainter planets,

at least in the region of interest(the region where the planet might locate).

In diffraction side lobe reduction, many approaches such as coronagraph, shaped binary pupil and pupil amplitude apodization methods have been developed and intensively studied [49, 51, 52, 50, 54, 61, 7, 37, 58, 56, 57, 60, 59]. In the coronagraph approach, the signal of the brighter light source is screened off by the combination use of a mask on the image plane and a stop on the pupil relay plane while in shaped binary pupil and the pupil apodization approach, the side lobe reduction is obtained by adjusting the shape of the pupil or the transmission rate over the pupil to re-design the PSF(Point Spread Function). In spite of these approaches, we found that the deep reduction of diffraction side lobes with broad bandwidth tolerance can also be obtained if the phase over a square pupil is modulated by a series of specific odd functions [69]. Side lobe reduction by phase modulation has many apparent advantages over the other approaches: firstly, there is no loss of light energy in the phase modulation and the "distorted" PSF due to the phase modulation can be compensated by applying a reversal phase modulation, therefore the diffraction limited image of the extra-solar planets and a shorter integration time can be obtained; secondly, realization of side lobe reduction can be potentially controlled by active segmented mirror or deformable mirror, therefore the phase error compensation and the diffraction side lobe reduction can potentially integrated into one system.

Here we raise another question: are there any other phase functions for which the sufficiently deep side lobe reduction can be generated in the region of interest very close to the star,say, a few λ/D? We found that Gerchberg-Saxton algorithm [41], if it is modified, can be used to find the phase functions for the side lobe reduction. The Gerchberg-Saxton's algorithm has widely been used to solve the phase retrieval problems in optical and electronic metrology[70, 71, 72] and synthetic problems [73, 74, 75, 76, 77, 78] . In this paper we will modify the method to solve the problem of how to find the phase function over the pupil that can help to yield a deep reduction region which can satisfy the requirements in the example case of extra-solar planet imaging.

4.2 Modified Gerchberg-Saxton algorithm

As stated in the introduction, the Gerchberg-Saxton algorithm is widely used in the phase retrieval problem and is based on the two or more known(or measured) intensity constraints on both of the pupil plane and the focal (image)plane. During each iteration of forward and backward Fourier Transform between the pupil plane and focal plane, the light amplitudes on the pupil plane and the focal plane are replaced by the square root of the corresponding measured intensities while the phase parts are maintained. Assume the intensity measurements on the pupil plane and focal plane are I_p and I_f respectively, then on the nth iteration, the light field on the focal plane, denoted as E_f^n, and the light field on the pupil plane, E_p^n, are constructed by the following way:

$$E_f^n = \mathcal{F}\{\frac{E_p^{n-1}}{|E_p^{n-1}|} \cdot \sqrt{I_p}\} \tag{4.1}$$

$$E_p^n = \mathcal{F}^{-1}\{\frac{E_f^n}{|E_f^n|} \cdot \sqrt{I_f}\} \tag{4.2}$$

where the superscript n denotes the nth iteration, $\mathcal{F}$ and $\mathcal{F}^{-1}$ denotes the Fourier Transform or Inverse Fourier Transform operation respectively.

In the phase retrieval problem, the constraints on the intensities can impose the iterations to converge to a phase function that satisfies the constraint conditions even though the solutions might not be unique. In the problem of diffraction side lobe reduction by phase-only modulation over the pupil,however, we require different constraint conditions: over the pupil, the amplitude be simply a binary function in which the values are 1 within the pupil opening and are 0 outside the opening, while over the focal plane, the values are imposed to 0 within the region of interest and are left unconstrained outside the region of interest. Based on this modification, during each iteration, the light field on both of the pupil plane and focal plane can be constructed as the following:

$$E_f^n = \mathcal{F}\{\frac{E_p^{n-1}}{|E_p^{n-1}|} \cdot P\} \tag{4.3}$$

$$E_p^n = \mathcal{F}^{-1}\{E_f^n \cdot G\} \tag{4.4}$$

where P is the binary pupil function, G is also a binary function with 0 values in the region of interest and 1 values outside the region of interest.

Because the modified constraint condition for the area outside the region of interest is loose, the iteration result for the phase and PSF is not unique and is strongly dependent on the initial condition. For obtaining a PSF that is not changed much compared with the ideal PSF except in the dark region of interest, the initial phase over the pupil should be set to zero or another very small random phase values.
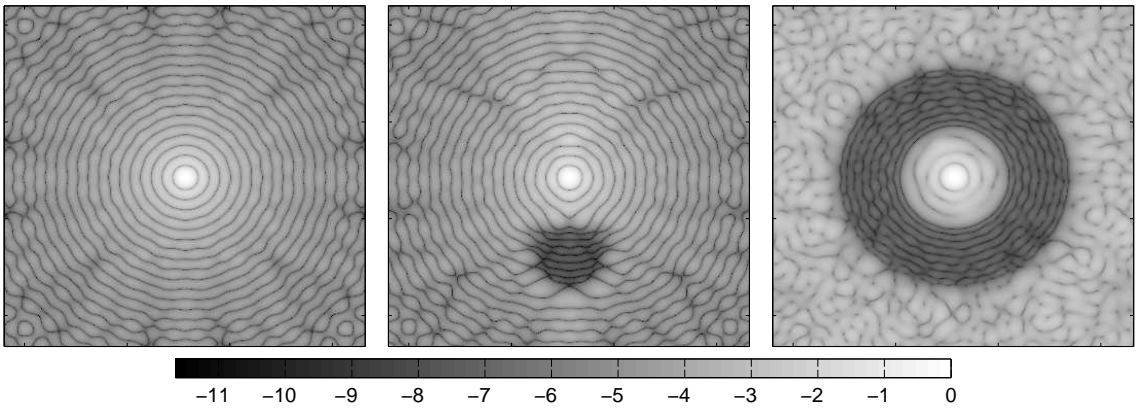
4.3 Examples

Figure 4.1: Modified Gerchberg-Saxton algorithm iteration results for generating different dark region of interest on focal plane for round pupil cases. Left: the logarithm (base 10) image of the PSF without phase modulation. Center: the logarithm (base 10) image of the PSF in which a round dark region with a radius of $3\lambda/D$ and centered at $8\lambda/D$ is generated. Right: the logarithm (base 10) image of the PSF in which a ring-like dark region with inner radius of $5\lambda/D$ and outer radius of $11\lambda/D$ is generated. The image intensities are normalized to the peak power of their corresponding PSFs and share the same gray scale.

Based on the modified Gerchberg-Saxton algorithm described above, we first examined the iteration results for the round pupil case when two different constraint functions G in formula 4.4 are used. The first constraint function we used is a round disk-like binary function in which we expect a round dark region centered at $8\lambda/D$ with radius of $3\lambda/D$ can be generated, here D is the diameter of the round pupil. The second constraint function we used is a ring-like binary function and we expect to generate a PSF with a ring-like dark region which has inner radius of $5\lambda/D$ and outer radius $11\lambda/D$. To implement the Modified Gerchberg-Saxton algorithm for these two cases, 51 pixels were sampled over the diameter of a round pupil and a grid of 512×512 was used in calculation. As shown in Figure 4.1 and in Figure 4.2, both
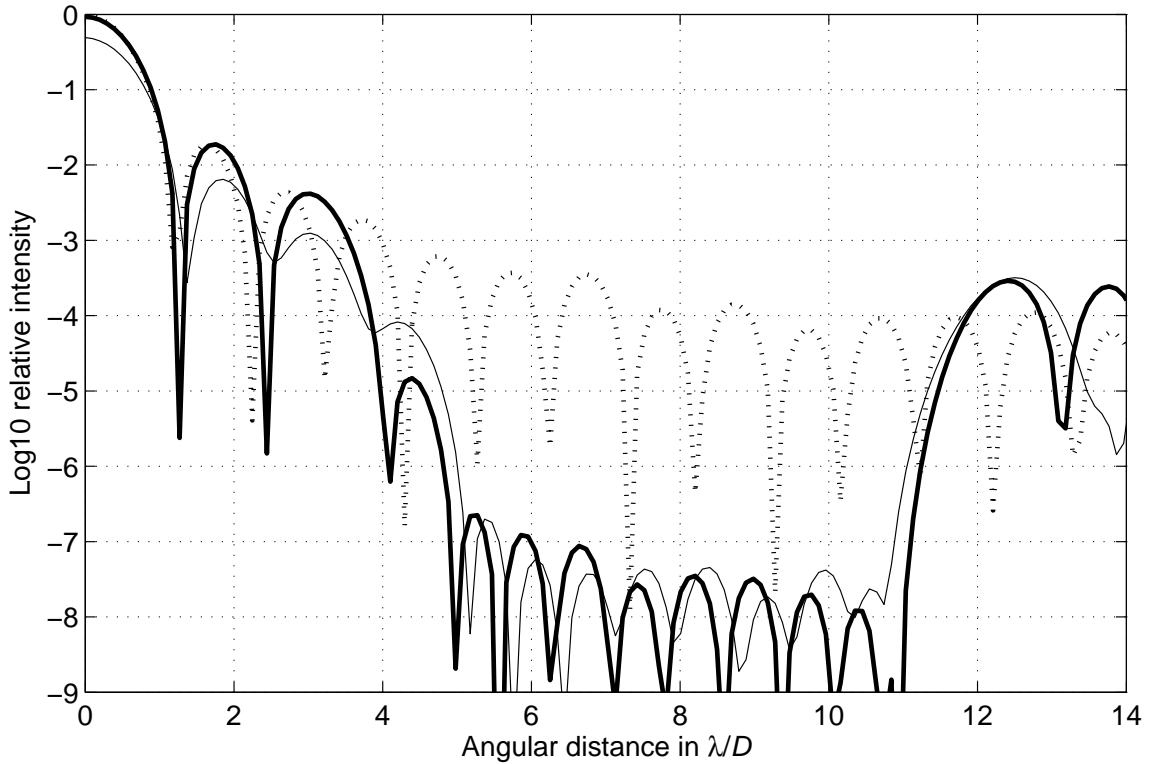
Figure 4.2: Slice plots of the logarithm(base 10)relative intensity distribution along y axis. The intensity is normalized to the peak intensity without the phase modulation. The dotted curve represents the intensity without the phase modulation. The thicker solid curve represents the intensity distribution described by the center image in Figure 4.1. The thinner solid curve represents the intensity distribution described by the right image in Figure 4.1.

of the expected dark regions have been generated and great portion of the light energy in the dark regions have been moved outside.

It should be noted that reduction level shown in Figure 4.2 is about 10^{-7} which is not yet sufficient for the purpose of extra-solar planet detection and the results were obtained after $10,000$ iteration loops. To achieve deeper reduction, more iterations or more sampling points across the pupil are probably needed. But, expanding the array in the FFT will slow down the speed and our experience shows that the iteration converges very slowly after a few hundred loops and the remaining loops (about 9,000 loops) are used to gain only about 1 more order of reduction in the above two calculations. In finding the phase for the purpose of extra-solar planet detection, directly applying the Modified Gerchberg-Saxton algorithm in 2-D computation would be very expensive. However, 2-D square pupil allows for separating the 2-D phase function into the sum of two 1-D phase functions as a consequence of which

the 2-D PSF will be the product of two 1-D PSFs , therefore the problem of optimizing a 2-D regional apodization is altered to a 1-D computation which is rapidly computed.
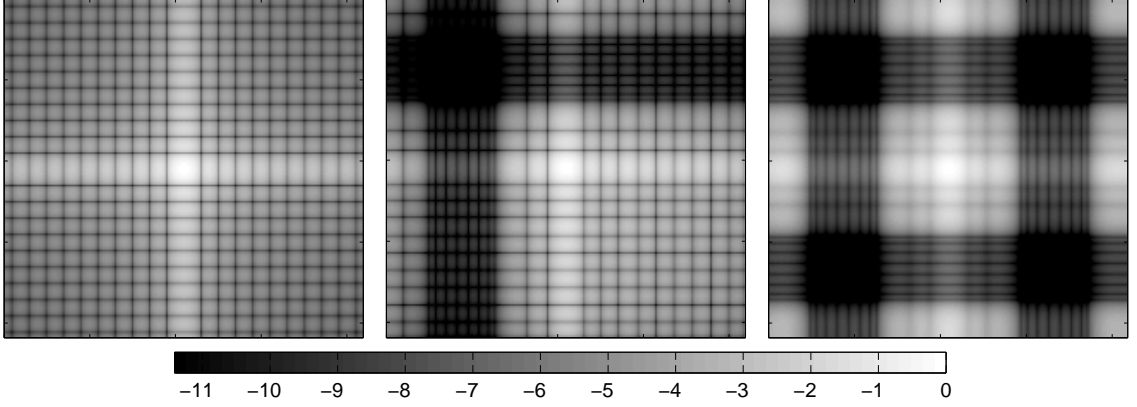

Figure 4.3: Modified Gerchberg-Saxton algorithm iteration results for generating different dark region of interest on focal plane for square pupil cases. Left: the logarithm (base 10) image of the PSF without phase modulation. Center: the logarithm (base 10) image of the PSF in which a square dark region with a side of $4\lambda/D$ and with a distance of $5.6\lambda/D$ to the core is generated in the fourth quadrant. Right: the logarithm (base 10) image of the PSF in which four square dark regions (same size as the one in the center figure) are generated in the all four quadrants. The image intensities are normalized to the peak power of their corresponding PSFs and share the same gray scale.

Based on the idea of separating the 2-D phase function into the sum of two 1-D phase functions, we then examined the PSFs constructed from two different constraint functions G during the 1-D Modified Gerchberg-Saxton iteration and the results are shown in Figure 4.3 and Figure 4.4. The first 1-D constraint function is a binary function in which the values in the region from $4\lambda/D$ to $8\lambda/D$ on one half of the axis are set to 0 while other region are set to 1, where D is the side of the square pupil. The constructed 2-D PSF based on this constraint function has one square shape cross area which experiences twice the 1-D reductions and reaches the level of about 10^{-13}, as shown in the Center panel of Figure 4.3 and in the thicker solid line in Figure 4.4. The second 1-D constraint function has two zero value regions on the axis , each of which has the same distance to the center and the same size as the one in the first constraint condition. As shown in the Right panel of Figure 4.3 and in the thinner solid line in Figure 4.4, the constructed 2-D PSF based on this constraint function has four square shape cross areas each of which experiences twice the 1-D reductions and reaches the level of 10^{-13}. In these calculations, the side of the square pupil was sampled 51 pixels and
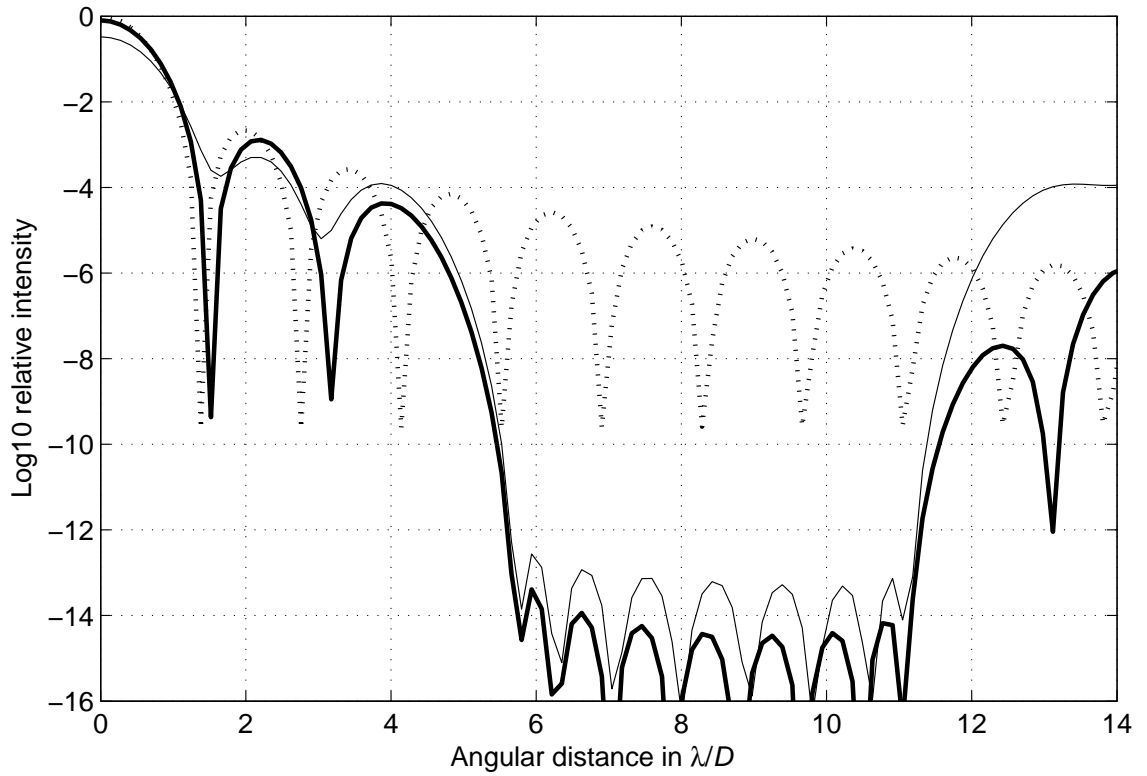
Figure 4.4: Slice plots of the logarithm(base 10)relative intensity distribution along the diagonal. The intensity is normalized to the peak intensity without the phase modulation. The dotted curve represents the intensity without the phase modulation. The thicker solid curve represents the intensity distribution described by the center image in Figure 4.3. The thinner solid curve represents the intensity distribution described by the right image in Figure 4.3.

the calculation was based on a 512×512 grid.

4.4 Application to imaging extra-solar planets by pupil phase modulation on segmented active mirror

In the above examples, it can be seen that the Modified Gerchberg-Saxton algorithm does help to optimize regional side lobe reduction and especially works well for the square pupil because it allows for variable separation. In the square pupil examples, the contrast level of below 10^{-11} can be obtained which is very suitable for extra-solar planet imaging. However, in imaging extra-solar planets through a monolithic telescope like HST (Hubble Space Telescope) in the visible light, stringent requirements such as smaller inner working distance and broader bandwidth tolerance will be needed at the same time.
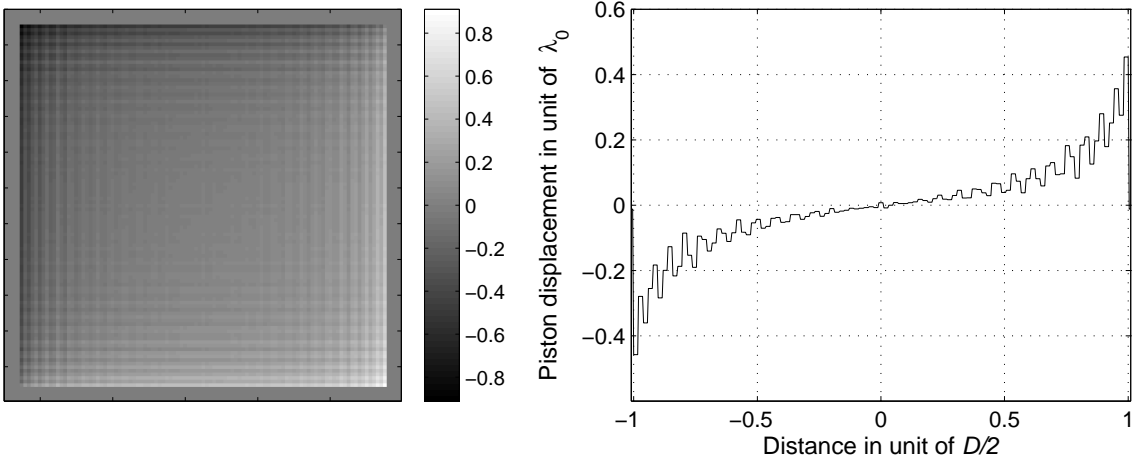

Figure 4.5: The actuator displacement constructed from the 1-D unwrapped phase in which the constraint function G has zero values from $1\lambda_0/D$ to $24\lambda_0/D$. Left panel shows the image of the segmented mirror array (101×101) displacement in units of λ_0. Right panel shows the slice plot of the displacement along the x-axis.

To investigate how small the inner working distance could be achieved by this method, we choose the constraint function that starts the 0 values at $1\lambda/D$, meanwhile for achieving broader bandwidth tolerance, the phase obtained directly from the Modified Gerchberg-Saxton iteration will be unwrapped and forced to be zero-mean. Figure 4.5 and Figure 4.6 show the results obtained through this simulation strategy. Figure 4.5 shows the actuator displacement map of the active mirror, assuming the pupil phase modulation is realized by a 101×101 segmented but gapless active mirror. The displacement value of each of the actuators is assigned based on the unwrapped phase obtained from the 1-D iterations in which

the constraint function on image plane has zero values from $1\lambda_0/D$ to $24\lambda_0/D$. Figure 4.6 shows the broad bandwidth ($20\%\lambda_0$) logarithm (base 10) PSF which is the result of the phase modulation by the segmented active mirror described in Figure 4.5. We can see that this broad bandwidth PSF, which has a shifted peak of 0.45 (normalized to the peak power without phase modulations), possesses a square dark region with contrast level of 10^{-10} and with an inner working distance $2.8\lambda_0/D$ which is sufficiently good for extra-solar planet imaging in a moderate space telescope.
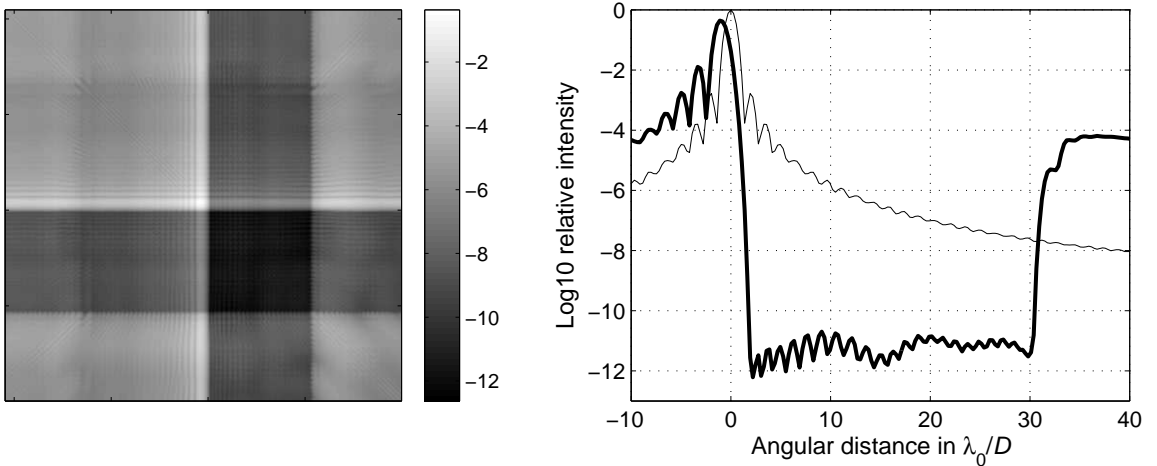

Figure 4.6: The broadband ($20\%\lambda_0$) logarithm (base 10) PSF with the phase modulation shown in Figure 4.5. Left panel: the logarithm relative image of the PSF. Right panel: the thicker solid line represents the slice plot of logarithm relative intensity distribution along the diagonal of the image in the Left panel. The thinner solid line represent the case without phase modulation. The intensity is normalized to the peak power without any phase modulation.

4.5 Conclusion

In previous sections, we modified the Gerchberg-Saxton algorithm and applied it, as a synthetic tool, to the problem of the regional side lobe reduction. Our calculations show that this Modified Gerchberg-Saxton algorithm can help find the phase functions that optimize the regional side lobe reduction efficiently. By applying this method to the problem of imaging extra-solar planet by a square pupil, we showed numerically that pupil phase modulation on a 101×101 segmented mirror can yield a dark region that is sufficiently good for direct imaging of extra-solar planets.

It should be pointed out that the performance of regional side lobe reduction by pupil phase modulation(or called pupil phase apodization) is related to the number of descretized phase elements on the pupil(or the size of the segmented mirror array). From the interference point of view, the PSF of the phase modulated pupil is effectively the interfering results of the focal image of these pupil phase elements. The dark region is generated when the Fourier Transforms of these elements interfer destructively. The more of these elements we have, the more chances we will have in arranging them to have a deeper or wider side lobe reduction.

Chapter 5

Summary and possible future work

In this thesis, by reviewing and analyzing the difficulties in direct imaging of extra-solar planets in visible light and current coronagraph-type techniques, we proposed a new approach for direct imaging of extra-solar planets in visible light—pupil phase apodization.

Through our analysis and the numerical calculations on examples of some specific continuous phase functions in Chapter 3 as well as the discrete phase functions obtained from the Modified Gerchberg-Saxton synthetic algorithm in Chapter 4, we demonstrated that the pupil phase apodization approach is achromatic and feasible.

It should be emphasized once again here that the phase modulation could be realized by a high density deformable mirrors and thus has the potential of integrating the phase error correction and sidelobe reduction into only one high density active mirror, therefore simplifying the whole imaging system and relaxing the need for a separate coronagraph system or making the density modulating mask or complicated binary mask.

Future work could include the experimental examination of the 1D pupil phase apodization effect in a phase plate or reflecting mirror. As to the realization of phase modulation by the high density deformable mirror, there is still much more work to be done. For example, we will have to find the signals to control the stroke of the actual deformable mirror array. In our simulations here, for simplicity, we used segmented and gapless active mirrors in which there is no correlations between adjacent actuators. However, correlations exist in real deformable mirrors. Therefore, how and how well can the actual deformable mirror be fitted into these functions still needs to be investigated. Furthermore, a whole system performance simulation

study could be another direction for future work which may include considering the phase sensing, actuator controlling and noise, etc.

Finally, it should be pointed out that phase-apodization should have many applications in other fields, such as radar detection and communication. For example, in the case of airborne or spaceborne radar detection of rain over the tropical sea area, the radar cross-section of the sea surface could be tens of times or even hundreds of times higher than the radar cross-section of the rain, the un-apodized side-lobes could pick the sea surface signal and the receiving antenna could misinterpret the sea signal as the rain signal. So reduction of the radar side-lobes on the sea surface side is needed. More research regarding this phenomena needs to be done.

Appendix A

MATLAB code

THIS CODE IS FOR SIMULATING THE BROADBAND PERFORMANCE OF THE PUPIL PHASE MODULATION
                                TECHNIQUES

```
%This code is written by Weidong Yang, Physics Department of Michigan
%Technological University, 31. July. 2003


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%The following part is for inputting data from keyboard
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

 chosen_option=0 ;%a tag for checking the correct option is chosen or not , 0->no; 1->yes
 while( chosen_option==0 )
 disp('PLEASE CHOOSE ONE OF THE PHASE DELAY FUNCTIONS : ')
 disp('            1. a*asin(x/r); suggested a be around 18' )
 disp('            2. tan(a*pi*x/r), with |a| <0.5 but close to 0.5; suggested a=0.490')
 disp('            3. a*(x/r)^b, with b odd int; suggested a ~ 30, b ~ 29 ')
 disp('            4. a*ln{(b+x/r)/(b-x/r)}, with b >1 ; suggested a ~ 3, b ~ 1.001')
 disp('            5. a*sinh(b*x/r); suggested a ~ 1.7, suggested b ~ 6.0')
 disp('                                        ')
 phase_option=input('YOUR PHASE DELAY OPTION IS :');
 if ( phase_option==1 | phase_option==2 | phase_option==3 |
    phase_option==4 | phase_option==5 );
    chosen_option=1;
```

```matlab
end;
end;


parameter_a=input('PLEASE INPUT A PARAMETER, a=');
if (phase_option==3 | phase_option==4 | phase_option==5)
    parameter_b=input('PLEASE INPUT B PARAMETER, b=');
end
disp('')
bandwidth=input('HALF BANDWIDTH IN DECIMALS= ');
disp('')
obs_range=input('HALF SIZE OF THE OBSERVATIONAL WINDOW IN LAMBDA/D =');
disp('');
position_mask=input('POSITION OF mask in unit of lambda/D=');
disp('');
n_planets=input('How many planets (including the star numbered 1 in following input)?=');
disp('');


contrast_ratio=input('how many order difference the planet is dimmer than star?=');
disp('');


planet_cordn=zeros(2,n_planets);
planets_count=1;
while(planets_count <= n_planets);
    disp(planets_count);
    planet_cordn(1,planets_count)=input('input_x_coordinate in units of lamda/D=');
    disp(planets_count);
    planet_cordn(2,planets_count)=input('input_y_coordinate in units of lamda/D=');
    planets_count=planets_count+1;
end;



%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%The following part is for calculation
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```matlab
n=(2^7)*512;  % define the total pixel numbers
n2=n/2;
rr_pupil0=(2^7)*50.0; %define the radius of the pupil in pixels
n_cut=fix(obs_range*n2/rr_pupil0); %convert the observational range to pixels
a=ones(1,n);
unit_vector=ones(n,1);
x=[-n2:1:n2-1];% x_axis
x_axis_diag=[-n_cut:1:n_cut]*1.414/(n/(2*rr_pupil0));
      %define the diagonal line in unit of lambda/D
x_axis_bound=[-n_cut:1:n_cut]/(n/(2*rr_pupil0));
      %define the x-axis in unit of lambda/D


mask1D=(x <= position_mask*n2/rr_pupil0);


sum_ideal_inten=zeros(2*n_cut+1,2*n_cut+1);
sum_mix_inten=zeros(2*n_cut+1,2*n_cut+1);
sum_mix_inten_f2=zeros(2*n_cut+1,2*n_cut+1);


for pln_count=1:n_planets;
 for kkk=-10:10 ; %loop for the 41 sub wavelength components
   ratio_lamda=1/(1+kkk*bandwidth/20); %define the ratio of the lambda change
   rr_pupil=ratio_lamda*rr_pupil0; %define the virtual pupil size
   pupil=a.*(abs(x) <= rr_pupil); %prepare a virtual pupil
   if (pln_count == 1)
      focal_ideal=(fftshift(fft(fftshift(pupil))));
            %the complete aberration free PSF on the 1D focal
      focal_ideal_x=focal_ideal(n2-n_cut:n2+n_cut);
            %choose the data only within the observational range
      focal_ideal2D=(focal_ideal_x')*focal_ideal_x;
            %calculate the 2-D light field of aberration free PSF
      intensity_focal=(abs(focal_ideal2D)).^2;
            %the 2D intenisty of the aberration free PSF
      clear focal_ideal; clear focal_ideal2D; clear focal_ideal_x;
      sum_ideal_inten=sum_ideal_inten+intensity_focal;
            %accumulate the light intensity on different wavelength
      clear intensity_focal;
   end
```

```
if (phase_option==1)
   phs_func=(parameter_a*asin(x./rr_pupil)-0)*ratio_lamda;
   phs_func_x=(parameter_a*asin(x./rr_pupil)-x*planet_cordn(1,pln_count)
            *pi/rr_pupil)*ratio_lamda;
   phs_func_y=(parameter_a*asin(x./rr_pupil)-x*planet_cordn(2,pln_count)
            *pi/rr_pupil)*ratio_lamda;
end if (phase_option==2)
   phs_func=(tan(parameter_a*pi*x./(rr_pupil))-0)*ratio_lamda;
   phs_func_x=(tan(parameter_a*pi*x./(rr_pupil))-x*planet_cordn(1,pln_count)
            *pi/rr_pupil)*ratio_lamda;
   phs_func_y=(tan(parameter_a*pi*x./(rr_pupil))-x*planet_cordn(2,pln_count)
            *pi/rr_pupil)*ratio_lamda;
end if (phase_option==3)
   phs_func=(parameter_a*((x./rr_pupil).^parameter_b)-0)*ratio_lamda;
   phs_func_x=(parameter_a*((x./rr_pupil).^parameter_b)-x*planet_cordn(1,pln_count)
            *pi/rr_pupil)*ratio_lamda;
   phs_func_y=(parameter_a*((x./rr_pupil).^parameter_b)-x*planet_cordn(2,pln_count)
            *pi/rr_pupil)*ratio_lamda;
end if (phase_option==4)
   temp001=((parameter_b+x/rr_pupil)./(parameter_b-x/rr_pupil+0.000000013))
            .*(abs(x) <= rr_pupil)+(1-pupil);
   phs_func=(parameter_a*log(temp001)-0)*ratio_lamda;
   phs_func_x=(parameter_a*log(temp001)-x*planet_cordn(1,pln_count)*pi/rr_pupil)
            *ratio_lamda;
   phs_func_y=(parameter_a*log(temp001)-x*planet_cordn(2,pln_count)*pi/rr_pupil)
            *ratio_lamda;
end if (phase_option==5)
   phs_func=(parameter_a*sinh(parameter_b*x/rr_pupil)-0)*ratio_lamda;
   phs_func_x=(parameter_a*sinh(parameter_b*x/rr_pupil)-x*planet_cordn(1,pln_count)
            *pi/rr_pupil)*ratio_lamda;
   phs_func_y=(parameter_a*sinh(parameter_b*x/rr_pupil)-x*planet_cordn(2,pln_count)
            *pi/rr_pupil)*ratio_lamda;
end


pupil_temp1_x=pupil.*(cos(phs_func_x)+i*sin(phs_func_x));
        %pupil function with phase aberrations along x
```

```
pupil_temp1_y=pupil.*(cos(phs_func_y)+i*sin(phs_func_y));
        %pupil function with phase aberrations along y


focal_x=(fftshift(fft(fftshift(pupil_temp1_x))));
        %the 1D PSF on the f1 with phase aberrations along x
focal_y=(fftshift(fft(fftshift(pupil_temp1_y))));
        %the 1D PSF on the f1 with phase aberrations along y


p2_x=fftshift(ifft(fftshift(focal_x.*mask1D)));
p2_y=fftshift(ifft(fftshift(focal_y.*mask1D)));


p2_temp1_x=pupil.*(cos(-phs_func)+i*sin(-phs_func));
        % pupil function with conj phase aberrations along x
p2_temp1_y=pupil.*(cos(-phs_func)+i*sin(-phs_func));
        % pupil function with conj phase aberrations along y


f2_x=(fftshift(fft(fftshift(p2_x.*p2_temp1_x))));
        %the 1D PSF on the f2 with phase aberrations along x
f2_y=(fftshift(fft(fftshift(p2_y.*p2_temp1_y))));
        %the 1D PSF on the f2 with phase aberrations along y



focal_x_tmp=focal_x(n2-n_cut:n2+n_cut);
        %get the 1D PSF within the observational range along x
focal_y_tmp=focal_y(n2-n_cut:n2+n_cut);
        %get the 1D PSF within the observational range along y


focal2D=(focal_y_tmp')*focal_x_tmp; %get the 2-D light field

clear pupil_temp1_x;clear pupil_temp1_y;
inten_temp1=(abs(focal2D)).^2; %get the 2-D intensity
clear focal_x;clear focal_y; clear focal_x_tmp;clear focal_y_tmp;clear focal2D;

if (pln_count == 1)
   sum_mix_inten=sum_mix_inten+inten_temp1;
     %accumulate the intensity on f1 for different wavelength
end if(pln_count > 1)
```

```matlab
        sum_mix_inten=sum_mix_inten+inten_temp1*10^(-contrast_ratio);
            %accumulate the intensity on f1 for different wavelength
      end;
      clear inten_temp1;


      f2_x_tmp=f2_x(n2-n_cut:n2+n_cut);% 1D PSF within the observational range on f2
      f2_y_tmp=f2_y(n2-n_cut:n2+n_cut);% 1D PSF within the observational range on f2
      f2_2D=(f2_y_tmp')*f2_x_tmp;% get the 2-D light field on f2


      clear p2_temp1_x;clear p2_temp1_y;
      inten_temp_f2=(abs(f2_2D)).^2;%get the 2-D intensity
      clear f2_x;clear f2_y; clear f2_x_tmp;clear f2_y_tmp; clear f2_2D;


    if (pln_count == 1)
       sum_mix_inten_f2=sum_mix_inten_f2+inten_temp_f2;
          %accumulate the intensity on f2 for different wavelength
    end if(pln_count > 1)
       sum_mix_inten_f2=sum_mix_inten_f2+inten_temp_f2*10^(-contrast_ratio);
          %accumulate the intensity on f2 for different wavelength
    end;
    clear inten_temp_f2;


 end;
end;


sum_ideal=(sum_ideal_inten./(21*1))/(max(max(sum_ideal_inten./(21*1))));
sum_mix=(sum_mix_inten./(21*1))/(max(max(sum_ideal_inten./(21*1))));
sum_mix_f2=(sum_mix_inten_f2./(21*1))/(max(max(sum_ideal_inten./(21*1))));


 clear sum_ideal_inten; clear sum_mix_inten;


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%The following part is for plotting
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
 figure;plot(x_axis_diag,diag(log10(sum_ideal)),x_axis_diag,diag(log10(sum_mix)));
```

```
ylim([-18 0]); title('2-D intensity along diagonal');
xlabel('Angular distance in \lambda/D'); ylabel('Log10 relative intensity');


figure;plot(x_axis_diag,diag(log10(sum_ideal)),x_axis_diag,diag(log10(sum_mix_f2)));
ylim([-18 0]); title('2-D intensity along diagonal');
xlabel('Angular distance in \lambda/D'); ylabel('Log10 relative intensity');


figure;colormap(hot);imagesc(x_axis_bound,x_axis_bound,log10(sum_mix));colorbar;
title('2-D intensity distribution');
xlabel('Angular distance in \lambda/D'); ylabel('Angular distance in \lambda/D');


figure;colormap(hot);imagesc(x_axis_bound,x_axis_bound,log10(sum_mix_f2));
colorbar;


figure;colormap(hot);
imagesc(x_axis_bound(1:n_cut),x_axis_bound(1:n_cut),sum_mix_f2(1:n_cut,1:n_cut));
colorbar;
figure;colormap(hot);
imagesc(x_axis_bound(1:n_cut),x_axis_bound(1:n_cut),sum_mix(1:n_cut,1:n_cut));
colorbar;
figure;colormap(hot);
mesh(x_axis_bound(1:n_cut),x_axis_bound(1:n_cut),sum_mix_f2(1:n_cut,1:n_cut));
colorbar;
figure;colormap(hot);
mesh(x_axis_bound(1:n_cut),x_axis_bound(1:n_cut),sum_mix(1:n_cut,1:n_cut));
colorbar;

%this is the end of the code
%This code is written by Weidong Yang, Physics Department of Michigan
%Technological University, 31. July. 2003
```

Bibliography

[1] D. W. Latham, R. P. Stefanik, T. Mazeh, M. Mayor, and G. Burki. The unseen companion of hd114762 - a probable brown dwarf. *Nature*, 339:38, 1989.

[2] A. Wolszczan and D. A. Frail. A planetary system around the millisecond pulsar psr1257 + 12. *Nature*, 355:145, 1992.

[3] M. Mayor and D. Queloz. Jupiter-mass companion to a solar-type star. *Nature*, 378:355, 1995.

[4] G.W. Marcy and R.P. Butler. A planetary companion to 70 virginis. *Astrophys. J. Lett.*, 464:Ll47, 1996.

[5] J. Schneider. The study of extrasolar planets: methods of detection, first discoveries and future perspectives. *C. R. Acad. Sci. Paris t.*, 327:621, 1999.

[6] M. A. C. Perryman. Extra-solar planets. *Rep. Prog. Phys.*, 63:1209–1272, 2000.

[7] J. R. P. Angel, A. Y. S. Cheng, and N. J. Woolf. A space telescope for infrared spectroscopy of earth-like planets. *Nature*, 322:341–343, 1986.

[8] L. Arnold, S. Gillet, O. Lardiere, P. Riaud, and J. Schneider. A test for the search for life on extrasolar planets - looking for the terrestrial vegetation signature in the earthshine spectrum. *Astron. and Astrophys.*, 392:231–237, 2002.

[9] D. J. Des Marais et al. Remote sensing of planetary properties and biosignatures on extrasolar terrestrial planets. *Astrobiology*, 2:153–181, 2002.

[10] C.A. Beichman, N.J. Woolf, and C.A. Lindensmith. *The Terrestrial Planet Finder: A NASA Origins Program to Search for Habitable Planets*. JPL Publication, USA, 1999.

[11] L.I. Caroff and D.J. Des Marais. *Pale Blue Dot 2 Workshop: Habitable and Inhabited Worlds Beyond Our Solar System.* NASA Publication, USA, 2000.

[12] R. N. Bracewell. Detecting nonsolar planets by spinning infrared interferometer. *Nature*, 274:780–781, 1978.

[13] R. N. Bracewell and R. H. MacPhie. Searching for nonsolar planets. *ICARUS*, 38:136, 1979.

[14] C. E. KenKnight. Methods of detecting extrasolar planets. i, imaging. *ICARUS*, 30:422–433, 1977.

[15] D. W. Davies. Direct imaging of planetary systems around nearby stars. *ICARUS*, 42:145, 1980.

[16] J. R. P. Angel. Ground-based imaging of extrasolar planets using adaptive optics. *Nature*, 368:203–207, 1994.

[17] 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.

[18] A. E. Siegman. *Laser.* University Science Books, Sausalito, CA, 1986.

[19] C. Ftaclas. Personal communications.

[20] J. W. Goodman. *Statistical Optics.* Wiley, New York, 1985.

[21] M. Born and E. Wolf. *Principles of Optics.* Cambridge University Press, United Kingdom, 1999.

[22] J. M. Elson, J. P. Rahn, and J. M. Bennet. Relationship of the total integrated scattering from multilayer-coated optics to angle of incidence, polarization, correlation length, and roughness cross-correlation properties. *Appl. Opt.*, 22:3207–3219, 1983.

[23] J. M. Elson. Theory of light scattering from a rough surface with an inhomogeneous dielectric permittivity. *Phys. Rev. B*, 30:5460–5480, 1984.

[24] C. Ftaclas, A. L. Nonnenmacher, A. Gruszczak, R. J. Terrile, and S. H. Pravdo. Direct detection of extra-solar planets by imaging: definition of the background. *proc. SPIE*, 2198:1324–1335, 1994.

[25] P. Glenn. space telescope performance prediction using the optical surface analysis code(osac). *Opt. Eng.*, 25:1026–1033, 1986.

[26] V. I. Tatarski. *Wave Propagation in a Turbulent Medium*. McGraw-Hill, New York, 1961.

[27] R. E. Hufnagel and N. R. Stanley. Modulation transfer function associated with image transmission through turbulent media. *J. Opt. Soc. Am.*, 54:52–61, 1964.

[28] D. L. Fried. Optical resolution through a randomly inhomogeneous medium for very long and very short exposures. *J. Opt. Soc. Am.*, 56:1372–1379, 1966.

[29] J. W. Hardy. *Adaptive Optics for Astronomical Telescopes*. Oxford University Press, New York, 1998.

[30] M. C. Roggemann and B. Welsh. *Imaging Through Turbulence*. CRC Press, Boca Raton, 1996.

[31] D. G. Sandler, S.Stahl, and J.R.P. Angel. Adaptive optics system for direct imaging of extra-solaer planets from the ground. *Proc. SPIE*, 2534:378–385, 1995.

[32] A. Labeyrie. Images of exo-planets obtainable from dark speckles in adaptive telescopes. *Astron. and Astrophys.*, 298:544–548, 1995.

[33] R. V. Shack and B. C. Platt. Production and use of a lenticular hartmannscreen. *J. opt. Soc. Am.*, 61:656, 1971.

[34] F. Roddier. Curvature sensing and compensation: a new concept in adaptive optic. *Appl. Optics*, 27:1223–1225, 1988.

[35] F. Roddier, C. Roddier, and N. Roddier. Curvature sensing: a new wavefront sensing method. *Proc. SPIE*, 976:203–209, 1988.

[36] R. A. Gonsalves. Phase retreval and diversity in adaptive optics. *Opt. Eng.*, 21:829–832, 1982.

[37] R. Gonsalves and P. Nisenson. Calculation of optimized apodizers for a terrestrial planet finder coronagraphic telescope. *PASP*, 115:706–711, 2003.

[38] R. G. Lyon, P. E. Miller, and grusczak. Hubble space telescope phase retrieval: a parameter estimation. *Proc. SPIE*, 1567:317–326, 1991.

[39] R. G. Lyon, J. e. Dorband, and J. M. Hollis. Hubble space telescope faint object camera calculated point-spread functions. *Appl. Optics*, 36:1752–1765, 1997.

[40] D. J. Lee, M. C. Roggemann, B. M. Welsh, and E. R. Crosby. Evaluation of least-squares phase-diversity technique for space telescope wave-front sensing. *Appl. Optics*, 36:9186–9197, 1997.

[41] R. H. Gerchberg and H. O. Saxton. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik*, 35:237–246, 1972.

[42] D. L. Misell. A method for the solution of the phase problem in the electronic microscopy. *J. Phys. D*, 6:L6–L9, 1973.

[43] J. J. Green, D. C. Redding, S. B. Shaklan, and S. A. Basinger. Extreme wave front sensing accuracy for the eclipse coronagraphic space telescope. *Proc. SPIE Int. Soc. Opt. Eng.*, 4860:266–276, 2003.

[44] S. B. Shaklan, D. Moody, and J. J. Green. Residual wave front phase estimation in the reimaged lyot plane for the eclipse coronagraphic telescope. *Proc. SPIE Int. Soc. Opt. Eng.*, 4860:229–239, 2003.

[45] R. Hudgin. Wave-front compensation error due to finite corrector-element size. *J. Opt. Soc. Am.*, 67:393–395, 1977.

[46] F. Malbet, W. Yu, and M. Shao. High-dynamic-range imaging using a deformable mirror for space coronagraphy. *PASP*, 107:386–398, 1995.

[47] J. Trauger et al. The eclipse mission: a direct imaging survey of nearby planetary systems. *Proc. SPIE Int. Soc. Opt. Eng.*, 4854:116–128, 2003.

[48] J. Trauger et al. Performance of a precision high-density deformable mirror for extremely high contrast imaging astronomy from space. *Proc. SPIE Int. Soc. Opt. Eng.*, 4854:1–8, 2003.

[49] B. Lyot. A study of the solar corona and prominences without eclipses. *MNRAS*, 99:580–596, 1939.

[50] M. J. Kuchner and W. Traub. A coronagraph with a band-limited mask for finding terrestrial planets. *Astrophys. J.*, 570:900–908, 2002.

[51] F. Roddier and C. Roddier. Stellar coronagraph with phase mask. *PASP*, 109:815, 1997.

[52] D. Rouan, P. Riaud, A. Boccaletti, Y. Clenet, and A. Labeyrie. The four-quadrant phase-mask coronagraph. i. principle. *PASP*, 112:1479–1486, 2000.

[53] R. A. Mayo, V. A. Spector, and C. F Lillie. Pointing and jitter control for the eclipse mission. *Proc. SPIE Int. Soc. Opt. Eng.*, 4854:544, 2003.

[54] M. J. Kuchner and D. N. Spergel. Notch-filter masks: Practical image masks for planet-finding coronagraphs. *Astrophys. J.*, 594:617–626, 2003.

[55] D. W. Wilson, P. D. Maker, J. T. Trauger, and T. B. Hull. Eclipse apodization: realization of occulting spots and lyot masks. *Proc. SPIE Int. Soc. Opt. Eng.*, 4860:361, 2003.

[56] P. Jacquinot and B. Roizen-Dossier. Apodisation. *Prog. Optics*, 3:29–186, 1964.

[57] N. J. Kasdin, R. J. Vanderbei, D. N. Spergel, and M. G. Littman. Extrasolar planet finding via optimal apodized-pupil and shaped-pupil coronagraphs. *Astrophys. J.*, 582:1147–1161, 2003.

[58] O. Guyon. Phase-induced amplitude apodization of telescope pupils for extrasolar terrestrial planet imaging. *Astron. and Astrophys.*, 404:379–387, 2003.

[59] R. J. Vanderbei, D. N. Spergel, and N. J. Kasdin. Spiderweb masks for high-contrast imaging. *Astrophys. J.*, 590:593–603, 2003.

[60] P. Nisenson and C. Papaliolios. Detection of earth-like planets using apodized telescopes. *Astrophys. J.*, 548:L201–L205, 2001.

[61] C. Aime, R. Soummer, and A. Ferrari. Total coronagraphic extinction of rectangular apertures using linear prolate apodizations. *Astron. and Astrophys.*, 389:334–344, 2002.

[62] W. Yang and A. B. Kostinski. Phase-modulated pupil for achromatic imaging of faint companions. *Phys. Lett. A*, 320:5–8, 2003.

[63] J. W. Goodman. *Introduction to Fourier Optics*. McGraw-Hill, New York, 1996.

[64] Y. Mori, K. Yamauchi, and K. Endo. Elastic emission machining. *Precision Eng.*, 9:123–128, 1987.

[65] O. W. Fähnle, H. van Brug, and H. J. Frankena. Fluid jet polishing of optical surfaces. *Appl. Opt.*, 37:6771, 1998.

[66] T.W. Drueding, S.C. Fawcett, S.R. Wilson, and T.G. Bifano. Ion beam figuring of small optical components. *Opt. Eng.*, 34:3565–3571, 1995.

[67] M. C. Rushford et al. Wet-etch figuring for precision optical contouring. *Appl. Opt.*, 42:5706–5713, 2003.

[68] T. Hull et al. Eclipse telescope design factors. *Proc. SPIE Int. Soc. Opt. Eng.*, 4860:277–287, 2003.

[69] W. Yang and A. B. Kostinski. One-sided achromatic phase apodization for imaging of extra-solar planets. *Astrophys. J.*, in press:to appear in April, 2004.

[70] J. R. Fienup. Phase retrieval algorithms: a comparison. *Appl. Opt.*, 21:2758–2769, 1982.

[71] J. M. Zuo, I. Vartanyants, M. Gao, R. Zhang, and L. A. Nagahara. Atomic resolution imaging of a carbon nanotube from diffraction intensities. *Science*, 300:1419–1421, 2003.

[72] K. A. Nugent, D. Paganin, and T. E. Gureyev. A phase odyssey. *Physics Today*, 54:27–32, 2001.

[73] J. R. Fienup. Iterative method applied to image reconstruction and to computer generated holograms. *Opt. Eng.*, 19:297–305, 1980.

[74] J.R. Fienup. Reconstruction and synthesis application of an iterative algorithm. *Proc. SPIE*, 373:147–160, 1981.

[75] G. F. Schils and D. W. Sweeney. Iterative technique for the symthesis of optical-correlation filters. *JOSA A*, 3:1433, 1986.

[76] K. M. Nashold and B. E. A. Saleh. Image construction through diffraction-limited high-contrast imaging systems: an iterative approach. *JOSA A*, 2:635, 1985.

[77] F. Wyrowski and O. Bryngdahl. Iterative fourier-transfor algorithm applied to computer holography. *JOSA A*, 5:1058, 1988.

[78] M. T. Eismann, A. M. Tai, and J. N. Cederquist. Iterative design of a holographic beamformer. *Appl. Opt.*, 28:2641–2650, 1989.