Coronagraph mask tolerances for exo-Earth detection

Oliver P. Lay, Joseph J. Green, and Daniel J. Hoppe, Stuart B. Shaklan
Jet Propulsion Laboratory, California Institute of Technology, Pasadena California 91109
Email: Oliver.P.Lay@jpl.nasa.gov Tel: (818)354-2521 FAX: (818) 393-4960

ABSTRACT

The direct detection at visible wavelengths of Earth-like planets around nearby stars requires starlight suppression by a factor of $10^{10} - 10^{11}$ at offsets of order 100 mas. It has been shown that perfect suppression is possible in principle, using a combination of a band-limited focal plane coronagraphic mask and a pupil plane Lyot stop. Errors in the transmission amplitude and phase of the mask degrade the performance. These errors can be corrected completely at a given wavelength and polarization using deformable mirrors (DMs) operating in the pupil plane of the system. Both the errors and correction have different chromatic dependences, however, and the DM correction becomes ineffective as the optical bandwidth is increased.

The mask errors can be divided into 2 classes: (1) errors that are uncorrelated with the mask pattern, arising, for example, from the surface roughness of the mask substrate, and (2) errors that are correlated with the mask pattern. We present the results of analysis of random errors and simulate the effects of systematic errors using specific mask designs. In both cases we find that the contrast required by TPF-C imposes very challenging demands on the design and fabrication of the masks. Several potential mitigation approaches are discussed.

Keywords: coronagraph, planet detection, high contrast imaging

1. INTRODUCTION

As part of its Origins theme, NASA is funding the development of a space mission capable of directly detecting and characterizing the starlight reflected from earth-like planets around nearby stars. The Terrestrial Planet Finder Coronagraph[1] (TPF-C) is a high-precision, ultra-stable, large aperture, visible wavelength telescope with a starlight suppression system[2] (SSS) that blocks the starlight while passing the planet light to a filter bank or spectrograph. It is scheduled for launch in the next decade.

The SSS relies on a band-limited coronagraph[3] to remove diffracted starlight before it can reach the image plane. The coronagraph is based on the early designs of Lyot[4], but the mask employed has a band limited Fourier Transform that (ideally) diffracts the starlight into a finite region where it is identically removed by a hard-edged Lyot stop. There are several variations of these masks; most recently, Kuchner, Crepp, and Ge introduced the 8^{th}-order mask[5] with greatly reduced sensitivity to low-order aberrations compared to the earlier 4^{th}-order masks[6,7].

The SSS also requires a pair of high-density deformable mirrors (DMs) to flatten the phase and amplitude of the wave front as it enters the coronagraph. The DMs are capable of perfectly correcting a monochromatic wave front over a range of spatial frequencies up to the Nyquist limit set by the number of actuators. The DMs can also be used to compensate for mask transmission errors. Trauger et al have demonstrated high-contrast imaging approaching the TPF-C requirements in laser light using a band-limited mask and a 32 x 32 actuator DM[8].

A transmission error requirement based on first-order analysis of random mask error, has been defined in the TPF-C error budget[9]. In this paper we study how mask design and the physics of mask transmission over a broad optical band affect the coronagraph performance. The next section derives the relationships connecting the number of leakage photons in the image plane to the transmission errors in the mask. In Section 3 we derive a requirement on phase errors that are uncorrelated with the mask pattern, which includes the surface roughness of the transparent mask substrate that may be present. In Section 4 we study binary masks, performing detailed modeling of the transmission to account for electromagnetic boundary conditions and waveguiding effects. We then simulate how well a pair of DMs can correct for these non-ideal effects over a finite bandwidth, and determine the limiting contrast that can be reached. The results are summarized in Section 5.

2. MASK ERROR ANALYSIS

The electric field incident on the mask at the center wavelength of the band λ_c, uncorrected by the deformable mirrors, is given by

$$E_{uncor}(\mathbf{x}) = E_0(\mathbf{x}) M(\mathbf{x}) * L(\mathbf{x}), \tag{1}$$

where E_0 is the point spread function for the input aperture (a sombrero function for the case of a circular aperture), M is the mask transmission function, and L is the point spread function for the Lyot mask. The * represents convolution and **x** is a vector representing angular offset in the image plane with units of λ_c / D_{in}, where D_{in} is the diameter of the input aperture. The complex mask transmission is a superposition of the ideal mask and the errors:

$$M(\mathbf{x}) = M_0(\mathbf{x}) + \Delta M(\mathbf{x}), \tag{2}$$

and can include both amplitude and phase effects. Substituting into Eq. (1),

$$\begin{aligned} E_{uncor}(\mathbf{x}) &= \left\{ E_0(\mathbf{x}) M_0(\mathbf{x}) + E_0(\mathbf{x}) \Delta M(\mathbf{x}) \right\} * L(\mathbf{x}) \\ &= \left\{ E_0(\mathbf{x}) \Delta M(\mathbf{x}) \right\} * L(\mathbf{x}) \end{aligned}. \tag{3}$$

The term including the ideal mask transmission M_0 is identically zero for a band-limited mask and suitably sized Lyot stop[3]. To generalize to arbitrary wavelength, we simply scale the point spread functions for the input aperture and Lyot stop by the ratio of the wavelengths:

$$E_{uncor}(\lambda, \mathbf{x}) = \left\{ E_0\left(\left[\lambda_c / \lambda \right] \mathbf{x} \right) \Delta M(\lambda, \mathbf{x}) \right\} * L\left(\left[\lambda_c / \lambda \right] \mathbf{x} \right). \tag{4}$$

A system of deformable mirrors (DMs) corrects exactly for the electric field at the center wavelength λ_c:

$$E_{DM}(\lambda_c, \mathbf{x}) = E_{uncor}(\lambda_c, \mathbf{x}) = \left\{ E_0(\mathbf{x}) \Delta M(\lambda_c, \mathbf{x}) \right\} * L(\mathbf{x}) \tag{5}$$

The E-field introduced by the DMs can again be scaled to different wavelengths. Since the chromatic dependence of the DM correction is different for amplitude and phase errors, we must partition the mask errors accordingly:

$$\begin{aligned} E_{DM}(\lambda, \mathbf{x}) &= (\lambda_c / \lambda) E_{DM,\phi}\left(\lambda_c, \left[\lambda_c / \lambda \right] \mathbf{x} \right) + (\lambda_c / \lambda)^2 E_{DM,A}\left(\lambda_c, \left[\lambda_c / \lambda \right] \mathbf{x} \right) \\ &= \left\{ E_0\left(\left[\lambda_c / \lambda \right] \mathbf{x} \right) \begin{bmatrix} (\lambda_c / \lambda) \Delta M_\phi\left(\lambda_c, \left[\lambda_c / \lambda \right] \mathbf{x} \right) + \\ (\lambda_c / \lambda)^2 \Delta M_A\left(\lambda_c, \left[\lambda_c / \lambda \right] \mathbf{x} \right) \end{bmatrix} \right\} * L\left(\left[\lambda_c / \lambda \right] \mathbf{x} \right) \end{aligned}, \tag{6}$$

The 'DM-corrected' electric field in the image plane is now given by

$$\begin{aligned} E_{cor}(\lambda, \mathbf{x}) &= E_{uncor}(\lambda, \mathbf{x}) - E_{DM}(\lambda, \mathbf{x}) \\ &= \left\{ E_0\left(\left[\lambda_c / \lambda \right] \mathbf{x} \right) \begin{bmatrix} \Delta M_\phi(\lambda, \mathbf{x}) - (\lambda_c / \lambda) \Delta M_\phi\left(\lambda_c, \left[\lambda_c / \lambda \right] \mathbf{x} \right) + \\ \Delta M_A(\lambda, \mathbf{x}) - (\lambda_c / \lambda)^2 \Delta M_A\left(\lambda_c, \left[\lambda_c / \lambda \right] \mathbf{x} \right) \end{bmatrix} \right\} * L\left(\left[\lambda_c / \lambda \right] \mathbf{x} \right) \end{aligned}. \tag{7}$$

The number of photons scattered into the image plane by the mask error, as fraction of the number that would have arrived with the mask removed is

$$N = \iint \left| E_{cor}(\lambda, \mathbf{x}) \right|^2 dx_1 dx_1 \Big/ \iint \left| E_0\left(\left[\lambda_c / \lambda \right] \mathbf{x} \right) * L\left(\left[\lambda_c / \lambda \right] \mathbf{x} \right) \right|^2 dx_1 dx_1, \tag{8}$$

where (x_1, x_2) are the components of **x**. This expression, together with Eq. (7), is generally applicable to the problem of mask errors. The specific example given in the next section will highlight the different aspects of these equations. Note

that N does not yet represent the image contrast, which is the number of photons per speckle patch in the image. This will be addressed in the next section.

3. UNCORRELATED SUBSTRATE ERRORS

We now proceed to make assumptions and approximations to derive an estimate for a specific type of mask error: random phase errors in the mask substrate, uncorrelated with the mask pattern.

The input aperture is assumed to be elliptical, with major diameter D_{in} = 8 m, and a minor diameter of 4 m. The input point spread function is proportional to a stretched sombrero function

$$E_0(\mathbf{x}) = k \frac{J_1\left(\pi\left\{x_1^2 + (x_2/2)^2\right\}^{1/2}\right)}{\pi\left\{x_1^2 + (x_2/2)^2\right\}^{1/2}}. \tag{9}$$

The Lyot stop is assumed to be circular, with a diameter D_{lyot} = 4 m

$$L(\mathbf{x}) = k' \frac{J_1\left(\pi\left[D_{lyot}/D_{in}\right]\left\{x_1^2 + x_2^2\right\}^{1/2}\right)}{\pi\left[D_{lyot}/D_{in}\right]\left\{x_1^2 + x_2^2\right\}^{1/2}}. \tag{10}$$

In each case, x_1 and x_2 are in units of λ_c / D_{in}. The center wavelength λ_c = 550 nm. Since we are considering phase errors in the mask

$$\Delta M_\phi(\mathbf{x}) \approx \left|M(\mathbf{x})\right| i\Delta\phi(\mathbf{x}) \approx i\Delta\phi(\mathbf{x}). \tag{11}$$

For simplicity, we have assumed that the mask transmission is approximately unity in the region of the phase error. This assumption will not have a big impact on the final result, and avoids tying the requirement to a specific mask shape. The phase error is represented by a circularly-symmetric Gaussian with peak of α radians at λ_0, centered on $\mathbf{x}_0$, with full-width-to-half-maximum of F:

$$\Delta\phi(\mathbf{x}) = \alpha\left(\frac{\lambda_c}{\lambda}\right)\exp\left\{-\left|\mathbf{x}-\mathbf{x}_0\right|^2 \frac{4\ln 2}{F^2}\right\}. \tag{12}$$

For a bump in the surface with height Δs of a transparent mask substrate material with refractive index n, the peak phase error is given by $\alpha = 2\pi n\Delta s/\lambda_c$. Figure 1 shows the amplitude of the electric field along the x_1-axis (i.e. the high-resolution direction in the image plane corresponding to the long axis of the primary mirror) for the monochromatic case at λ_c = 550 nm. The units of the x_1 axis in the plot are radians, and the Gaussian is centered at an offset of $4\lambda_c/D_{in}$ with F = 0.5 (units: λ_c/D_{in}).

Figure 2 adds a second wavelength, λ = 510 nm. Whereas the mask error is centered near to a minimum of the 550 nm sombrero, it is close to a zero crossing at 510 nm, and the resulting products are quite different for the two cases, both before (Fig. 2b) and after (Fig. 2c) convolution with the Lyot response. Since the DM correction is based on the 550 nm center wavelength, it is wholly inappropriate for 510 nm (Fig. 2c long dashed line).

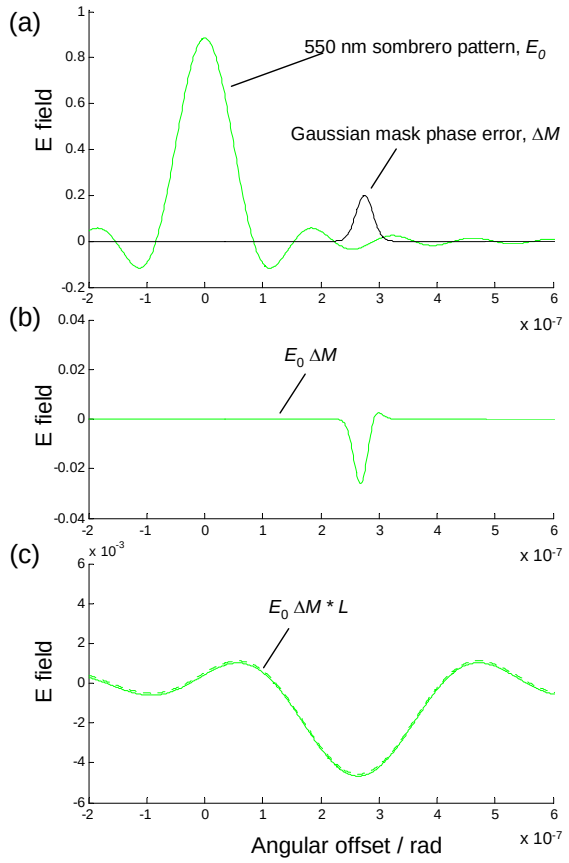

Figure 1: (a) A profile of the incident electric field amplitude at the mask, and a Gaussian error in the mask transmission phase. (b) The product of the two gives the error component of the electric field. (c) Resulting profile of the electric field in the focal plane after passing through the Lyot stop.

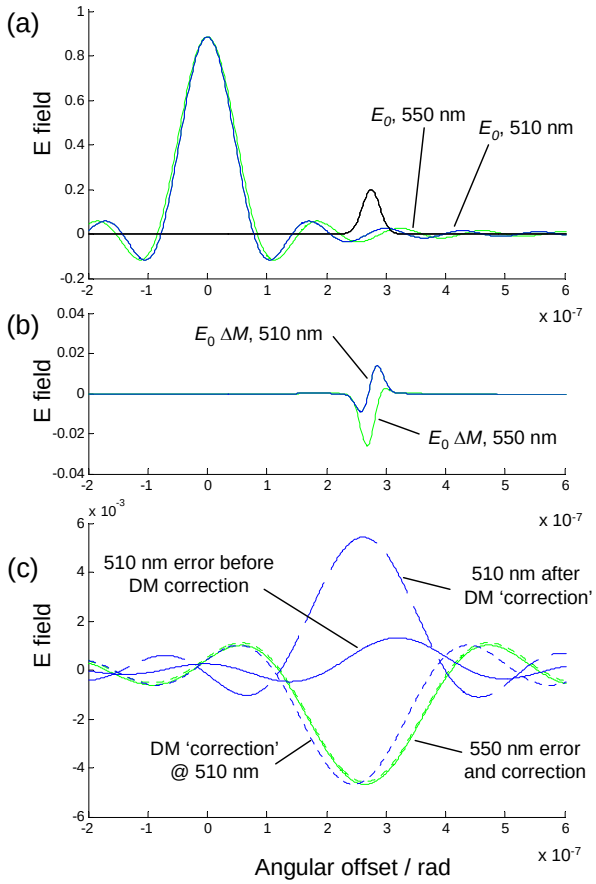

Figure 2: (a) Profiles of the incident electric field at the mask for two different wavelengths. (b) The products are quite different. (c) Deformable mirrors correct for the error in the focal plane at 550 nm, but increase the error at 510 nm.

A simple MatLab calculation was used to determine the number of scattered photons N_1, as a fraction of the number that would be obtained without the mask present, for a single Gaussian bump with $\alpha = 1$, as a function of the angular offset of the bump $|\mathbf{x}_0|$ and the bump width F. The code simulates an optical bandwidth of 500-600 nm by averaging the results obtained for wavelengths of 510, 530, 550, 570 and 590 nm.

For a more meaningful comparison which compensates for the relative size of the bumps, we first normalize the numbers to a common surface rms roughness. To do this, we create a surface profile from a regular grid of Gaussian bumps, each with diameter F, spaced by F, and with random peak amplitude drawn from a normal distribution with rms σ_α. It can be shown that the surface rms roughness is $s_{rms} = 0.75\ \sigma_\alpha\ \lambda_c / 2\pi n$. The expected value of the rms number of photons per bump (averaged over the distribution of possible amplitudes) is given by

$$N_{1,rms} = \sqrt{3}\sigma_\alpha{}^2 N_1\left(\mathbf{x}_0, F\right) = 121.6 n^2 \left(\frac{s_{rms}}{\lambda_c}\right)^2 N_1\left(\mathbf{x}_0, F\right). \tag{13}$$

The square root of 3 comes from the 4th moment of a Gaussian distribution (the variance in the photon rate is proportional to α^4). Since the spacing F of the bumps is in units of λ_c/D_{in}, and the speckle area in the image plane is given by $(\lambda_c/D_{lyot})^2$, then the number of bumps per speckle is $(D_{in}/FD_{lyot})^2$. We can then convert to photons per speckle (relative to the case without the mask) and obtain the contrast:

$$C = 2N_{1,rms}\left(\frac{D_{in}}{FD_{lyot}}\right)^2 = 243.2n^2\left(\frac{s_{rms}}{\lambda_c}\right)^2\left(\frac{D_{in}}{FD_{lyot}}\right)^2 N_1\left(\mathbf{x}_0, F\right). \quad (14)$$

The factor of 2 accounts for the two surfaces that are present on any transmissive substrate. Table 1 shows the contrast obtained for a range of offsets $|\mathbf{x}_0|$ and bump widths F, and assuming a substrate refractive index n = 1.5 (appropriate for Silica). The surface rms has been normalized to 50 pm in each case. Two competing effects underlie the dependence on offset angle: (1) the point spread function from the input aperture is increasing rapidly at small offsets from the star; (2) the relative shift between the point spread functions at two different wavelengths increases linearly with angular offset from the star (see Fig. 2a). In the scenario evaluated here, the former effect is dominant, and the contrast deteriorates close to the star. Table 1 also shows that the scale size of the mask errors is important. First consider the effect of a very narrow Gaussian (small F) in Fig. 2. The product $E_0\Delta M$ (within the curly braces of Eq. (7), and shown in Fig. 2b) is dominated by high spatial frequencies that are blocked by the Lyot stop (i.e. 'smeared' out in the convolution process). A broad Gaussian (high F) spans a similar region of sidelobes for the point spread function at two different wavelengths and the DM correction becomes more effective (the term within the square braces of Eq. (7) is small). It is the intermediate spatial scales – those comparable to the width of the diffraction ring of the input point spread function – that have the largest impact on the contrast.

Table 1: Surface rms required for contrast of 10^{-11}

		Spatial scale of mask error, F (units λ_c / D_{in})				
		0.125	0.25	0.5	1	2
$\|\mathbf{x}_0\|$ (units λ_c/D_{in})	3	13 pm	6 pm	5 pm	7 pm	23 pm
	4	12 pm	7 pm	6 pm	9 pm	27 pm
	5	17 pm	8 pm	7 pm	10 pm	30 pm
	10	28 pm	13 pm	11 pm	15 pm	45 pm

We can also use the 'grid of Gaussian bumps' model to derive a requirement on the power spectrum of the surface roughness, as illustrated in Fig. 3. Each curve on the plot represents the power spectral density (PSD) of substrate surface height for one value of the bump width and spacing F. The cases with broad bumps contain no power at high spatial frequencies and roll off at a spatial frequency proportional to F^{-1}. Each of the 5 spatial scales (i.e. set of Gaussian bumps of a particular size) have been allocated a contrast contribution of 10^{-11}, and Eq. (14) was applied to calculate the corresponding value of s_{rms}, which in turn determines the vertical normalization of each PSD. The sum of the PSDs gives the net requirement on the surface; a surface matching this PSD will give a contrast of 5 x 10^{-11}. The shape is somewhat flexible, since we are free to reallocate to the desired total contrast to the separate spatial scales. There is no requirement on the surface for spatial frequencies higher than 1 cycle per $f\lambda/D_{in}$, since the scattered light is blocked entirely by the Lyot stop.

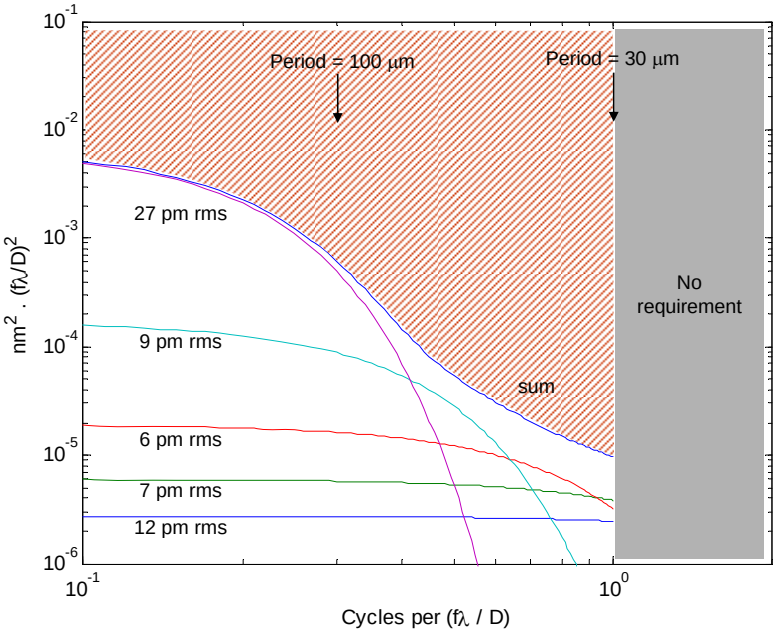

Figure 3: Power spectral densities for Gaussian mask errors with different spatial scales. The surface rms allocated to each curve results in a contrast contribution of 10^{-11}. Spatial frequencies higher than 1 cycle per $f\lambda/D_{in}$ do not impact the contrast. The spatial periods shown at the top apply to the case where f / D_{in} = 60 and λ = 550 nm.

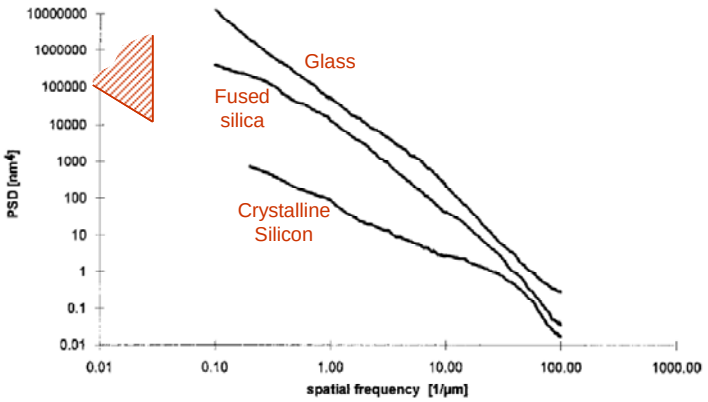

Figure 4: Surface roughness PSDs measured by Duparre and Jakobs for three substrates. The TPF-C mask requirement from Fig. 3 has been scaled to match the units of the plot and overlaid in the upper left corner.

We now compare the derived requirement for the substrate surface rms to values that have been obtained in practice. We assume that the TPF-C optical system has $f / D_{in} = 60$, for which the spatial scales of interest have periods of 33 μm or more (see Fig. 3). These are larger scales than are typically studied. Weis demonstrated superpolishing of sapphire with an rms of 70 pm for scales between 12.5 μm and 500 μm[10]. Figure 4 shows data from Duparre and Jakobs for three substrates[11]. The TPF-C mask requirement has been overlaid, but lies at larger spatial scales than the measured data. Both cases indicate that the TPF-C mask requirement is close to the limit of what has been achieved, but further study will be necessary to establish whether a technology development effort is needed.

4. CORRELATED ERRORS: SIMULATION OF A BINARY MASK

In this section we describe simulations of mask errors that are correlated with the mask pattern, specifically the wave-guiding effects that are obtained in metallic binary masks. The mask design is described, followed by a description of the electromagnetic propagation effects and how they are modeled. We simulate the impact of these errors on the image-plane contrast and show the extent to which they can be corrected over a finite bandwidth.

4.1. Binary mask design

Occulting focal plane masks operate in conjunction with a Lyot stop to eliminate the starlight in a coronagraphic telescope. Band-limited masks make possible complete cancellation of the starlight in the ideal case[12]. Examples are Sin^2 and $1\text{-}\mathrm{Sinc}^2$ masks, or an '8^{th} order' function[5]. They may be implemented in either 'analog' or 'binary' forms. Analog masks are fabricated with a continually varying, graded transmission pattern. Binary masks achieve the same result with a series of transparent windows in an otherwise opaque substrate.

Realization of an arbitrary one-dimensional voltage transmission function, T(x), using a binary (on/off) structure can be considered as a two step process. As a first step the function is sampled in the non-search 'y' direction at a rate sufficient to guarantee that any higher-order diffracted terms generated by this process will strike the opaque parts of the Lyot stop. This requirement specifies the minimum sample spacing in 'y' at $F\#*\lambda$. Next, within each of these y-segments the transmission profile in 'x' must be generated. For the zero-order diffracted wave (in y) at a given x-position the transmission coefficient is the average value (DC value) of the transmission coefficient over the y-period. In the case of a binary mask this is simply the ratio of the opening size to the period. Transmission coefficients for the higher-order y-diffracted waves will also exist, and will in general be different from those of the zero order wave. These waves, and hence their transmission coefficients, are not of concern since they will be intercepted by the Lyot stop. If we denote the period as P_y, and the height of the conductor as 'h(x)', then the gap is given by P_y-h(x). The voltage transmission coefficient for the zero order diffracted wave at a given position 'x' is then $T(x)=[P_y\text{-}h(x)]/P_y$. It should be noted that the total power transmitted by the gap is also given by $=[P_y\text{-}h(x)]/P_y$. The power transmitted by the zero-order wave is $T^2(x)=[P_y\text{-}h(x)]^2/P_y^2$. The remainder of the transmitted power is contained in the higher-order diffracted waves. A large scale view of an example 8^{th} order binary mask is depicted in Fig. 5a, with a more detailed view (not to scale) in Fig. 5b.

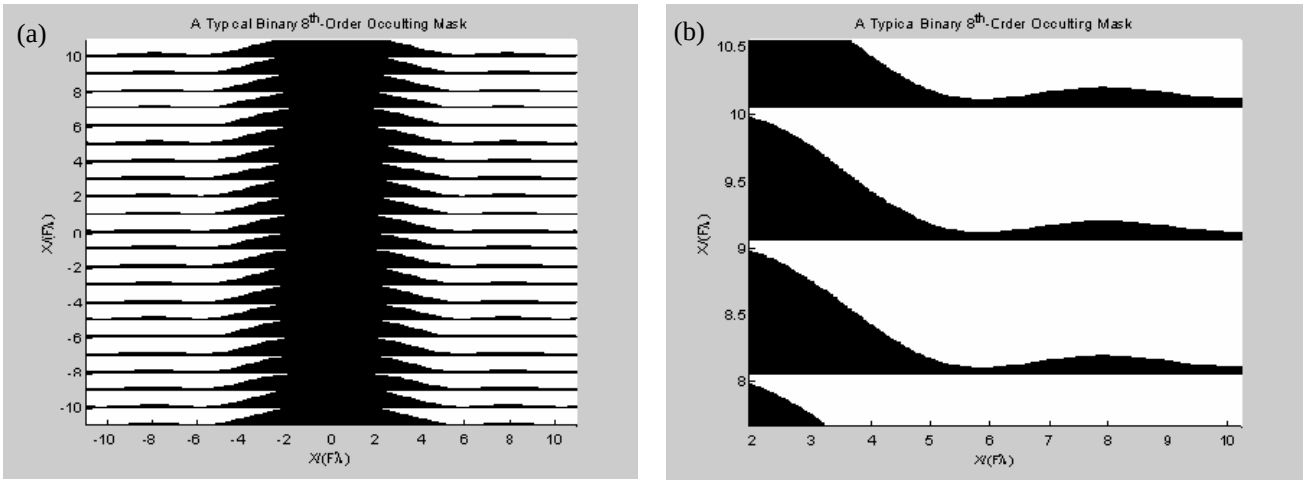

Figure 5: A typical 8th order binary mask, (a), and detailed view (b).

4.2. Vector Electromagnetic Effects

The Fourier optics analysis of the mask, on which the design is based, assumes that the metal portions of the mask are totally opaque. In practice any layer of material of finite thickness, even a metal, will transmit at some level. For aluminum, which is the material that was actually used to create the mask the layer thickness was specified to be 0.200 μm, which guarantees a leakage level of at least 10^{-10} over the (0.5-0.8 μm) wavelength band of interest. This thickness is far from negligible at the wavelengths of operation and has a significant impact on the mask's transmission properties when vector electromagnetic effects and polarization are considered. A second consideration is that no polarization effects are included in the Fourier optics analysis, and it is assumed that the mask is ideal and behaves identically for incident fields of any polarization.

Figure 6 depicts a number of electromagnetic effects that are ignored in the Fourier optics analysis of apertures in thick screens. The most significant effect that occurs as radiation passes through the aperture is that the field must satisfy the boundary conditions on the aperture wall as it passes through. The allowable field configurations that satisfy these boundary conditions are the modes of an infinite waveguide with cross section equal to that of the aperture. As shown in the figure these modes are excited at the input side of the aperture, propagate through the aperture, and then radiate at the aperture's output side. Each mode will experience some reflection at the output side of the aperture, resulting in a backward wave in the aperture. This can lead to resonant effects. The speed of propagation of each of the possible waveguide modes is different, and depends on the wavelength of operation as well as the aperture cross section. Even in the absence of reflection at the waveguide output the difference in propagation constants will cause the phase of the output field to vary across the aperture. The amplitude of the field in the aperture will also be modified, with cross-polarized components added in order to satisfy the boundary conditions on the aperture walls. The excitation coefficients of the modes as well as their propagation through the aperture are both a function of wavelength and the profile and polarization of the incident field, in violation of the Fourier optics approximation. In general small-scale features in the aperture shape are most affected by propagation through the thick aperture.

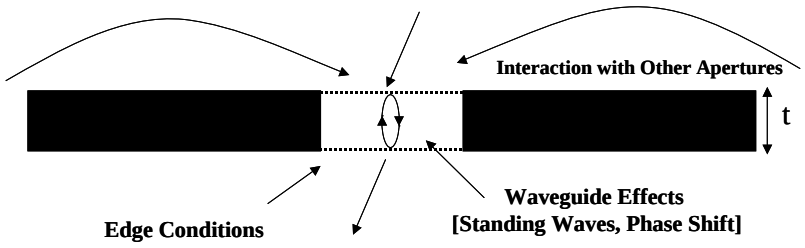

Figure 6. Vector electromagnetic effects in thick conducting screens.

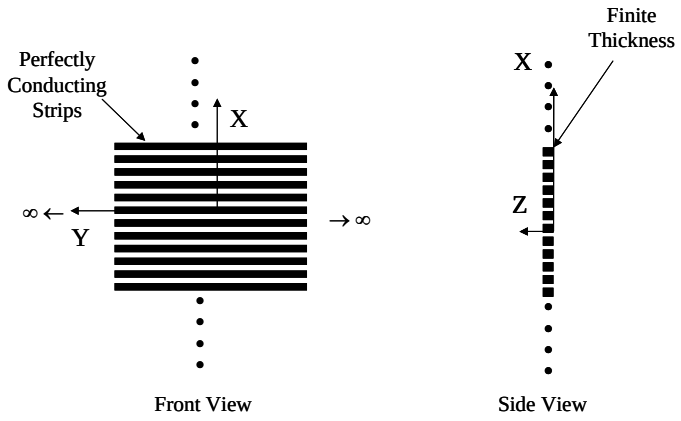

Figure 7. A perfectly conducting, thick, strip grating.

Although the structures making up the binary masks under consideration here are periodic in one dimension their non-periodicity in the 'x' dimension precludes their exact electromagnetic analysis. The problem is simply beyond that of existing numerical methods and computing power. Instead an approximate analysis technique is adopted where the solution to a tractable canonical problem is used to determine local approximation to the transmission coefficient at any point on the mask. This problem is chosen to include the thickness of the metal layer as well as polarization effects. For the binary masks considered here the appropriate canonical problem is the thick periodic grating. In order to simulate the first-order vector electromagnetic effects in a binary mask we compute the x-dependent transmission coefficient of this mask assuming it behaves locally as a thick strip grating. The transmission coefficient is determined for each of the two possible polarizations of incident field at each location on the binary mask by solving the appropriate 'local' strip grating problem as described below. This canonical problem in turn takes into account grating period, local gap dimension, and thickness.

Figure 7 depicts the geometry of the thick strip grating. The grating is periodic in 'y', infinite in 'x' and the perfectly conducting strips have finite thickness in the 'z' direction. In the present implementation of the analysis code the strips exist with an arbitrary number of dielectric layers on each side. The problem is solved by assuming an appropriate set of expansion functions for the vector electromagnetic field in each region and solving for the magnitude of these expansion functions by matching the tangential electric fields over the surfaces in common to the various regions. For the regions behind and in front of the strip grating the field is represented as a series of vector Floquet modes. The modes in the set satisfy the periodicity condition, and the set includes both forward and backward traveling TE_z (Transverse Electric to 'z') and TM_z (Transverse Magnetic to 'z') modes. Between the strips the field is expanded as a set of parallel plate waveguide modes. These modes satisfy the $E_{tan}=0$ condition on the strips and the set includes both forward and backward traveling TE_z and TM_z modes. The periodicity and thickness of the mask are appropriate for the TPF-C application, but are somewhat flexible.

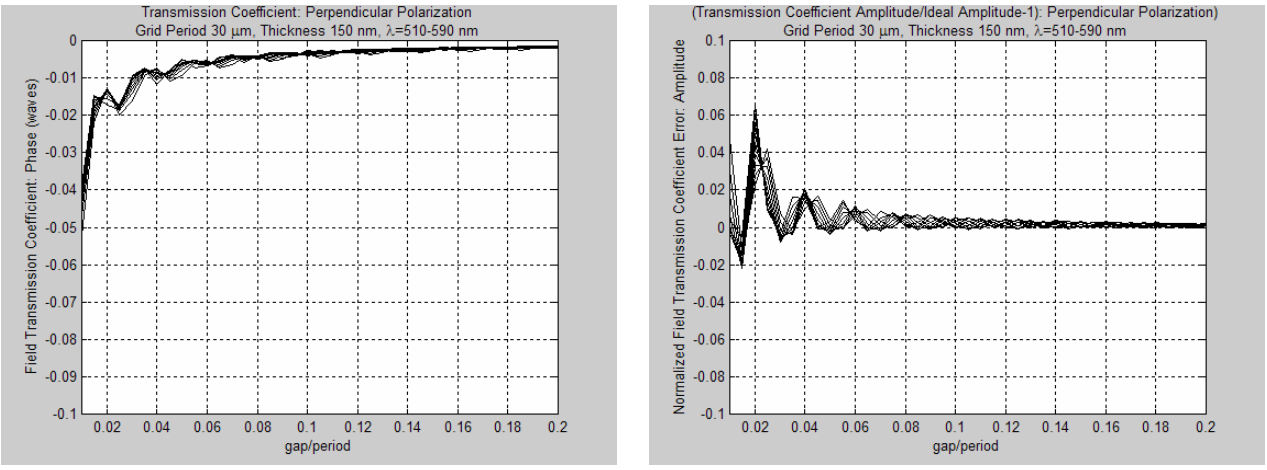

Figure 8: Perturbations in the phase and amplitude transmission of a binary mask introduced by electromagnetic waveguiding effects. The E-field is in the vertical direction in Figs. 5 and 7. Each plot has 9 curves, corresponding to wavelengths of 510, 520,..590 nm.

Figure 8 shows an example of the perturbations in phase and amplitude as a function of the gap size for the perpendicular polarization (E field perpendicular to the strip direction). Each plot has 9 curves spanning the wavelength range 510 to 590 nm in 10 nm increments. The amplitude plot has been normalized to show the amplitude error as a

fraction of the desired value; at a gap/period ratio of 0.04, for example, the actual transmitted amplitude is 1.02 times higher than the desired value. Both amplitude and phase show oscillatory behavior, reflecting resonances in the propagation through the metallic mask structure. As expected, the perturbations are largest for the smallest gap sizes which will be found at offsets close to the star (Fig. 5). The parallel polarization shows similar structure, but the perturbations are ~5 times larger in both amplitude and phase. This is consistent with the more stringent boundary conditions that are imposed by the metallic strips on the parallel polarization than the perpendicular polarization. However, the mask of Fig. 5 is likely to exhibit more adverse behavior for the perpendicular polarization than is suggested from the parallel strip data of Fig. 8, since the edges of the openings are not horizontal and in general will not be perpendicular to the E field. It may be necessary to implement the mask as a 'blocked' structure in which the windows open out in a series of steps as we move away from the center. The simulations of the following sections are based on the parallel strip approximation.

4.3. Controlling waveguide induced binary mask errors

While coronagraphs suppress the diffracted starlight, the discovery space for exosolar planet searches is ultimately defined by the wavefront control (WFC) system. Coherent scattered starlight from the optical aberrations as well as errors induced by coronagraph elements persist over the field of view (FOV). The ultimate ability to suppress this scattered light is dependent upon how well matched the WFC system correction is to the error that need compensation. The correction must both have the right spatial distribution over the pupil as well as match chromatic scaling behavior to guarantee high contrast over a reasonable optical bandwidth. In this section section we examine the limitations to the correctablity of waveguide induced binary mask error over an optical bandwidth.

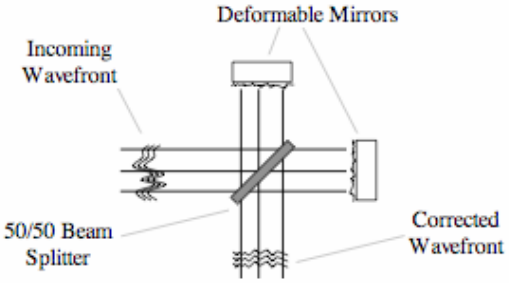

Figure 9: Schematic of the two deformable mirror wavefront control system. The DMs are arranged in a Michelson interferometer configuration to enable control over both amplitude and phase irregularities.

To permit the largest possible discovery space, the baseline WFC system for TPF uses two deformable mirrors (DM) in a Michelson interferometer arrangement. Figure 9 shows a schematic of the baseline configuration. By splitting the beam and applying an independent correction in each arm, both amplitude and phase irregularities in the pupil may be fully addressed[13]. For a given input complex pupil $P_{in}(u,v;\lambda)$, the output beam is described by

$$P_{out}(u,v;\lambda) = P_{in}(u,v;\lambda)\left[\frac{1}{2}\exp\left(j2\pi/\lambda\, opd_1(u,v)\right)+\frac{1}{2}\left(\exp j2\pi/\lambda\, opd_2(u,v)\right)\right], \tag{15}$$

where $opd_1(u,v)$ and $opd_2(u,v)$ are optical path difference (OPD) distributions on each arm of the interferometer. The pupil is corrected by setting the OPD distributions in each arm as

$$\begin{aligned} opd_1(u,v) &= opd_\phi(u,v) + opd_a(u,v) \\ opd_2(u,v) &= opd_\phi(u,v) - opd_a(u,v) \end{aligned}, \tag{16}$$

where

$$\begin{aligned} opd_a(u,v) &= \frac{\lambda}{2\pi}\cos^{-1}\left(\left|P_{in}(u,v;\lambda)^{-1}\right|\right) \\ opd_\phi(u,v) &= \frac{\lambda}{2\pi} arg\left(P_{in}\left(u,v;\lambda\right)\right) \end{aligned}. \tag{17}$$

While this scheme perfectly corrects the complex pupil at one wavelength, the scaling law for the amplitude at phase may not be ideally match to the types of errors that may be encountered. The induced phase scales as $1/\lambda$; a typical path difference correction. This is generally matched to other OPD errors caused by traditional optical aberrations. The amplitude compensation scales as $1/\lambda^2$. As we will see, neither is matched to the chromatic dependence of the binary mask transmission.

4.4. Monochromatic optimization of the wavefront control systen

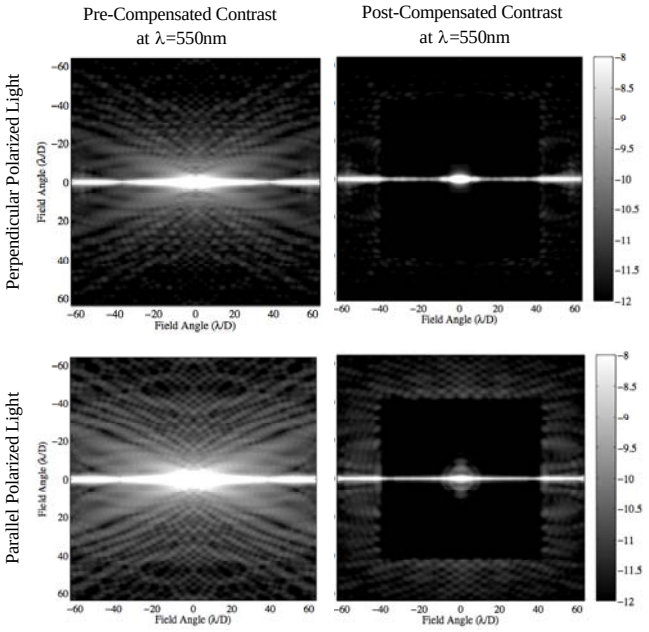

Figure 10: Monochromatic contrast over the field of view is shown at λ=550nm. The top row shows the contrast before (left) and after (right) the compensation of the binary mask error induced by perpendicular polarized light. The bottom row is the same, but for parallel polarized light.

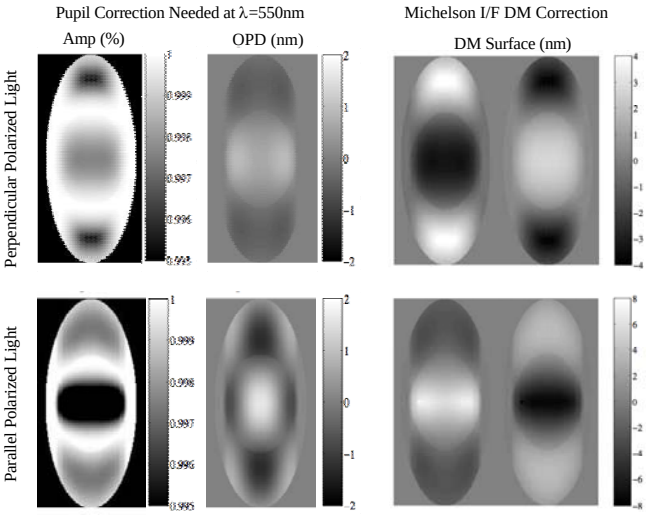

Figure 11: Monochromatic correction required at λ=550nm to achieve high contrast in the field of view is shown The top row shows the pupil compensation needed in the presence of perpendicular polarized light. The amplitude and phase of the pupil-plane compensation are shown on the right while the DM surface shapes needed to achieve that compensation are shown on the left. The bottom row is the same, but for parallel polarized light.

Because waveguide induced errors occur at the occulting mask in the focal plane, we cannot simply minimized the complex pupil errors. Instead we took an approach to minimize the complex field in the focal plane a wavelength of interest. In practice, the complex field is not measured directly and must be estimated through a sequence of intensity measurements in the presence of some known phase diversity[8,14-16]. Although unfair, using the true complex pupil permits to derive ideal control states and study the chromatic nature of the correction.

In Fig. 10, we show contrast over the field of view in the case where the only errors in the system are caused by waveguide effects at the binary occulter. The top row shows the contrast before and after compensation by the WFC system in that case where the incident light was perpendicularly polarized. The mask error compensation is achieved assuming that we have two 96x96 actuator DMs in the WFC system. The bottom row shows the same but for parallel-polarized light. Here, the induced errors are substantially worse but nonetheless correctable monochromatically. In Fig. 11, the required complex-pupil compensation to achieve high contrast is shown along with the DM surface shapes used by the WFC system to affect this control.

4.5. Chromatic limitations of compensation

With the monochromatically optimized control state established, we can now examine the limitations brought fourth in broadband light. In Fig. 12, we show the RMS error of the correction need as a function of wavelength and polarization. These curves are labeled as “Uncompensated” errors. In general terms, the required phase and amplitude compensation increase in magnitude as the wavelength increases. The waveguide induced errors become more severe as the fixed binary mask features appear smaller to the longer wavelength light. Aside from the uncompensated phase and amplitude RMS curves, there are the post-correction phase and amplitude RMS curves. These curves reveal the mismatch between a correction established at λ=550nm and the required compensation needed at other wavelengths. Because we apply a global average zero-path-difference constraint after the control optimization, we do realize perfect amplitude correction even at λ=550nm. Although this compromises the monochromatic optimality of the solution, we feel that this step helps to achromatize the control state.

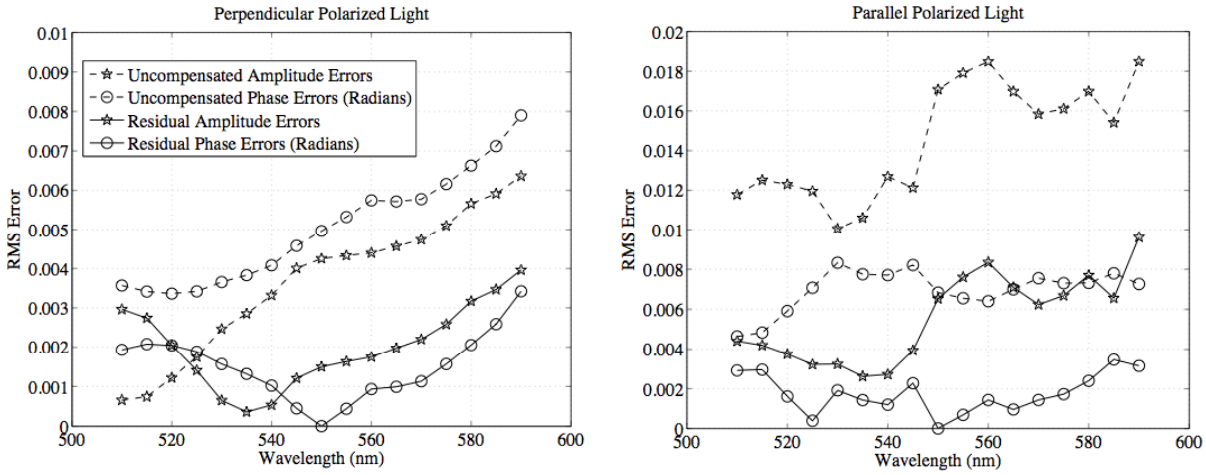

Figure 12: The RMS error in the pupil is shown before and after compensation with the DM pair. The uncompensated error represents the required pupil-domain correction needed to achieve high contrast as a function of wavelength. The residual errors depict how effectively the DM-based corrections compensates the make errors over a bandwidth when the surface shapes at set to correct at λ=550nm. Note the y-axis scale difference between the two plots.

Figure 13 shows how the contrast over the FOV degrades as the optical bandwidth increases. At a 0% optical bandwidth we have our monochromatically optimized dark-hole. As the bandwidth increases the mismatched compensation permits more starlight to leak into the FOV. This is well illustrated by Fig. 14. Here we show the contrast at the 4 λ/D field point degrades with increasing optical bandwidth. These plots reveal that by maintaining perpendicularly polarized light to be incident upon a binary occulter, we may find a suitable mask design to achieve TPF requirements over a significant bandwidth.

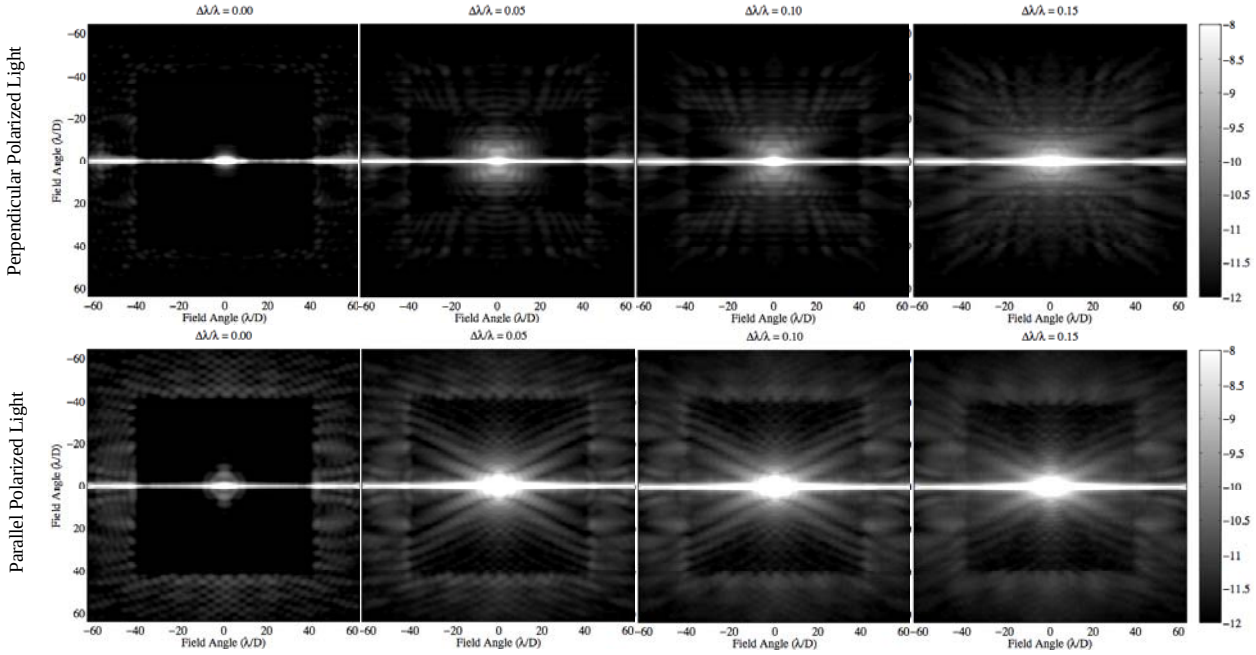

Figure 13: The contrast over the field of view after correction is shown. From left to right, the top row shows the post-correction contrast over 0%, 5%, 10% and 15% bandpasses about λ=550nm in the presence of perpendicular polarized light. The bottom row is the same, but for parallel polarized light.

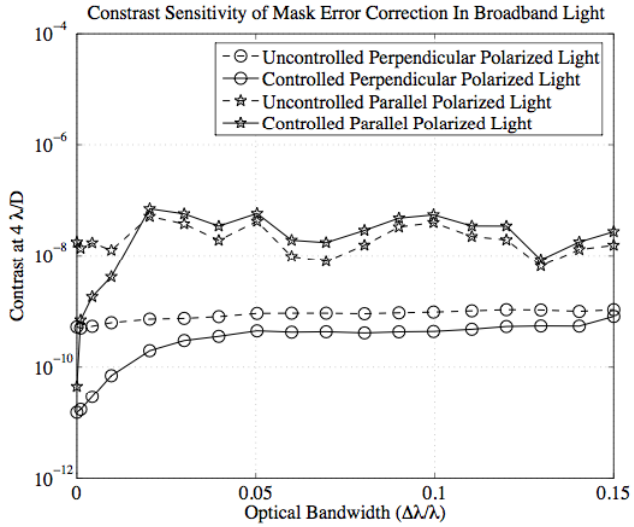

Figure 14: Contrast at 4 λ/D is shown as a function of optical bandwidth. The control state is established at the central wavelength (λ=550nm). At other wavelengths, this correction is less effective and contrast degrades. Further iteration of the wavefront control could be used to improve the contrast for the first two points of each solid curve.

4.6. Aberration sensitivity in the presence of compensated binary mask errors

One last aspect to explore is the impact that mask errors have upon aberration sensitivity. In our previous work we have shown the importance of selecting coronagraph designs that have low aberration sensitivity[6,7,17]. In Fig. 15 we show how the sensitivity of a Lyot coronagraph with an 8th order mask behaves when the occulter possess waveguide-induced mask errors. These sensitivity curves are computed as the absolute change in contrast for the when a particular aberration introduced in the entrance pupil wavefront. Also shown are curves for the ideal 8th order mask.

Clearly, there are significant interactions of the aberration change with the static mask error that result in sensitivity curves. The presence of mask errors largely negates the 8th order mask's insensitivity to low order aberrations. We plot the absolute change in contrast because the stability of the residual stellar light is central to the planet detection problem. Only when the aberration level becomes large do we see that the curve follow the ideal 8th order mask sensitivity curves. While the ideal 8th order masks simulations contained no occulter errors, there is a numerical noise that biases the sensitivities at the 10^{-13} contrast level.

4.7. Mitigation strategies

Figure 14 shows that for a bandwidth of 500 – 600 nm the wave-guiding effects for the binary mask considered here limit the contrast at 4λ/D to ~10^{-9} and >10^{-8} for the perpendicular and parallel polarization states, respectively. The current requirement[9] on contrast for TPF-C is 6 x 10^{-11}, of which mask errors are but one contributor. Meeting this requirement with a binary mask will clearly be a challenge, and we now discuss some of the possible strategies that might be employed.

Decreasing the optical bandwidth. From Fig. 14 we see that the fractional bandwidth must be reduced to ~0.01 before there is a significant improvement in the contrast (the reason we need such a narrow bandwidth is because at 4λ/D the

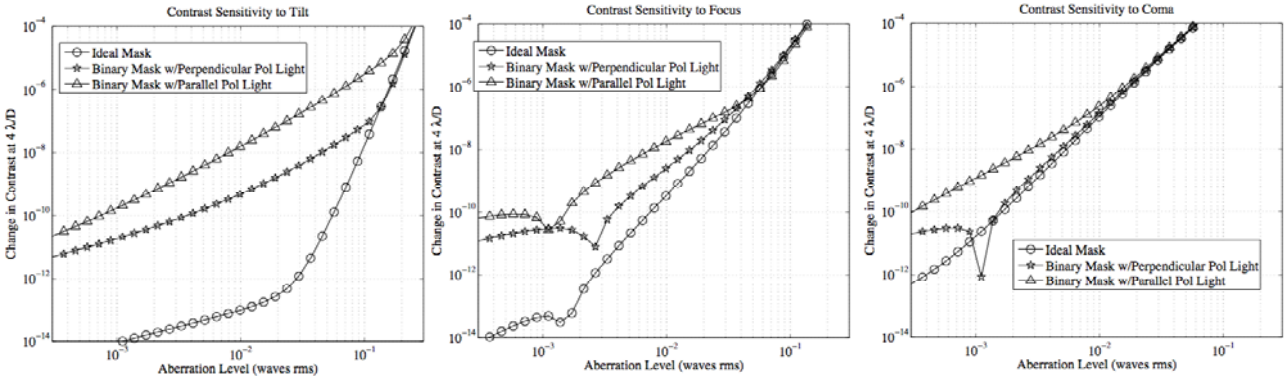

Figure 15: The sensitivity of contrast to low-order aberrations is shown in the presence of compensated binary mask errors and monochromatic light. These sensitivities are computed as the absolute change in contrast to changes in the entrance pupil tilt (left), focus (middle) and coma (right).

openings in the mask are tens of wavelengths across). This would have a severe impact on the instrument sensitivity.

Reducing the mask thickness. Thinner masks will have less impact on the transmission pattern, but will start to leak photons through the nominally opaque regions of the mask. Tapering the thickness profile at each edge might be more effective than an overall reduction in thickness.

Separate masks for each polarization. The parallel polarization is affected by the mask much more than the perpendicular state (Figs. 13 and 14). The performance will be improved if the polarization states are separated and passed through masks tailored individually for each polarization. Simply rotating the mask design of Fig. 7 will not be sufficient, however, since we require that the direction of the strips is aligned with the long (high-resolution) axis of the primary. Each polarization state will require its own wavefront control system.

Increasing the F/#. The physical size of the mask pattern is determined by the F/# of the upstream imaging system. A larger mask means larger openings and reduced wave-guiding effects. The current system has F/# = 60.

Alternative wavefront correction scheme. The amplitude and phase correction from the zero-path Michelson arrangement have specific dependences on wavelength (λ^{-2} and λ^{-1}) which do not match the chromatic dependence of the mask errors. A correction scheme that is better matched to the mask errors will improve the residual contrast. A significant improvement in the correction will be difficult to achieve in practice, however. The waveguide errors in the mask get worse as the wavelength is increased relative to the gap sizes, whereas DM-based systems tend to produce more corrections that diminish as the wavelength is increased. In addition, the chromatic dependence of the mask errors can show many oscillations across the optical bandwidth, and varies according to the offset in the mask.

Rectangular input pupil. A rectangular pupil concentrates the diffracted electric field into sidelobes confined to a 'cross-hair' pattern, with much darker regions in the interspersed quadrants. From Eq. (7) we see that the reduced value of E_0 within the dark quadrants makes us much less sensitive to mask errors (of any kind) in these areas. The penalty is that the useful field of view is reduced by the strip of bright scattered light that will now appear in the image in the region of the high sidelobes.

Each of these should be studied in more detail. There are no doubt other strategies that could also be brought to bear on the problem, and a successful solution may consist of a combination of different approaches.

5. SUMMARY

We have shown how errors in the mask transmission pattern combine with the input point spread function to produce scattered starlight that is not blocked by the Lyot stop. We investigated two specific cases: the impact of surface roughness of a transparent mask substrate and the transmission errors introduced by waveguiding effects in a metallic binary mask.

The surface errors were modeled as a grid of random Gaussian bumps, uncorrelated with the mask pattern. We showed that the coronagraph is most sensitive to mask errors at small offsets from the star that have spatial scales comparable to the width of the diffraction rings of the stellar point spread function incident on the mask. The DM-based compensation system is ineffective at correcting these errors over a bandwidth of 500-600 nm. We find that a surface rms of ~6 pm on spatial scales of ~20 μm contributes ~10^{-11} to the contrast in the image plane at an offset of $4\lambda/D$. Further study is needed to determine whether this requirement can be met by current polishing techniques.

The propagation of light through a metallic binary mask was modeled using electromagnetic simulation software. The need to satisfy the boundary conditions in this 3-dimensional structure results in complex perturbations of the transmitted amplitude and phase as a function of both wavelength and position in the mask. A simulation of the dual-DM Michelson wavefront compensation scheme demonstrated that it was not possible to adequately correct for these errors for optical bandwidths of a few percent or more. A contrast of ~10^{-9} was obtained at $4\lambda/D$ for the perpendicular polarization, and >10^{-8} for the parallel polarization. Several possible mitigating strategies were discussed.

We conclude that the design and fabrication of coronagraphic masks to meet the stringent TPF-C requirements will likely be very challenging and deserves further study.

The work described in this paper was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

REFERENCES

1. V. G. Ford, et al, "The Terrestrial Planet Finder Coronagraph 2005: Overview of System Design Studies and Technology Development," Proc. SPIE Vol. 5905 (2005).
2. Z. Mouroulis. & S. Shaklan, "Optical Design of the Terrestrial Planet Finder Coronagraph Starlight Suppression System," Proc. SPIE Vol. 5874 (2005).
3. M. J. Kuchner, and W. A. Traub, "A Coronagraph with a Band-Limited Mask for Finding Terrestrial Planets," ApJ 570, pp. 900-908 (2002).
4. B. Lyot, "The study of the solar corona and prominences without eclipses", MNRAS, 99, 538 (1939)
5. M.J. Kuchner, J. Crepp, & J. Ge, "Finding Terrestrial Planets using 8th-order Image Masks," ApJ (in press) 2005.
6. S. B. Shaklan, & J. J. Green, "Low-Order Aberration Sensitivity of Eighth-Order Coronagraph Masks," ApJ (in press) 2005.
7. J. J. Green, & S. B. Shaklan, "Optimizing Coronagraph Performance Designs to Minimize Their Contrast Sensitivity to Low-Order Optical Aberrations," Proc. SPIE Vol. 5170, pp. 25-37 (2003).
8. J. T. Trauger, et al, "Coronagraph Contrast Demonstrations with the High Contrast Imaging Testbed," Proc. SPIE Vol. 5487, pp. 1330-1337 (2004).
9. S. B. Shaklan, et al, "The Terrestrial Planet Finder Coronagraph Dynamics Error Budget," Proc. SPIE Vol 5905 (2005).
10. O. Weiss, "Direct contact superpolishing of sapphire", Appl. Opt. 31, 4355 (1992)
11. A. Duparre, & S. Jacobs, "Combination of surface characterization techniques for investigating optical thin-film components", Appl. Opt. 35, 5052 (1996)
12. M.J. Kuchner and D.N. Spergel, "Notch Filter Masks: Practical Image Masks for Planet-Finding Coronagraphs", To Be Published.
13. M. G. Littman, et al, "Phase and Amplitude Control Ability using Spatial Light Modulators and Zero Path Length Difference Michelson Interferometer," Proc. SPIE Int. Soc. Opt. Eng. 4854, 405 (2003).
14. 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 (2003).
15. J. J. Green, et. al., "Post-Coronagraph Wavefront Sensing Using Pupil Imaging and Phase Diversity," Opt. Soc. Am., SRS (Charlotte, NC) June 2005
16. P. Borde and W. A. Traub, "High-contrast Imaging from Space: Speckle Nulling in a Low Aberration Regime," ApJ, Submitted 2005.
17. J. J. Green, et. al., "The sensitivity of shaped pupil coronagraphs to optical aberrations," Proc. SPIE, vol 5487, (Glasgow 2004)