Assessing the performance limits of internal coronagraphs through end-to-end modeling: a NASA TDEM study

John E. Krist\textsuperscript{a}, Ruslan Belikov\textsuperscript{b}, Laurent Pueyo\textsuperscript{c}, Dimitri P. Mawet\textsuperscript{d}, Dwight Moody\textsuperscript{a},
John T. Trauger\textsuperscript{a}, Stuart B. Shaklan\textsuperscript{a}
\textsuperscript{a}Jet Propulsion Laboratory/California Inst. of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109 (USA)
\textsuperscript{b}NASA Ames Research Center, Moffett Field, CA 94035 (USA)
\textsuperscript{c}Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218 (USA)
\textsuperscript{d}European Southern Observatory, Alonso de Córdova 3107, Vitacura, 763 0355, Santiago, Chile

ABSTRACT

As part of the NASA ROSES Technology Development for Exoplanet Missions (TDEM) program, we are conducting a study of three internal coronagraphs (PIAA, vector vortex, hybrid bandlimited) to understand their behaviors in realistically-aberrated systems with wavefront control (deformable mirrors). This study consists of two milestones: (1) develop wavefront propagation codes appropriate for each coronagraph that are accurate to 1% or better (compared to a reference algorithm) but are also time and memory efficient, and (2) use these codes to determine the wavefront control limits of each architecture. We discuss the results from the study so far, with emphasis on representing the PIAA coronagraph and its wavefront control behavior.

Keywords: Coronagraph

1. INTRODUCTION

1.1 Exoplanet imaging technologies and the need for optical modeling

NASA has established a goal to undertake future missions to image and characterize extrasolar planets around nearby stars as stepping stones to searching for signs of life beyond the solar system. This will not be easy to achieve given that a planet like Earth would appear only $10^{-16}$ as bright as its sun in reflected visible light, and it could only be seen within a fraction of an arcsecond from the star. A large planet like Jupiter would have a brightness contrast of only $10^{-9}$. Even the best telescope will produce enough diffracted and scattered light (i.e. glare) to overwhelm the feeble signals from such planets. Because of these obstacles, there is a strong probability that one or more of these missions will involve the combination of a coronagraph (to suppress diffracted light from telescope obscurations) and wavefront control with deformable mirrors (to suppress scattered light from optical fabrication errors) to allow planet detection and characterization.

There are a wide variety of coronagraphic techniques spanning the feasibility range of “just a theory” to “flight-ready” (or at least their proponents believe so). Which coronagraph is actually most appropriate to fly in terms of performance and technical maturity is a critical question. A complete telescope and coronagraph combination cannot be fully tested prior to launch because a sufficiently stable, large-scale, space-like environment cannot be created on the ground. So, this question can only be answered with accurate computational modeling and evaluations of prototype subsystem hardware under space-like conditions in testbeds.

The modeling required to demonstrate readiness must include:

- A telescope and instrument layout with all of the optical elements that would exist in a real system.
- Realistic representations of aberrations on each optical surface, including phase errors (from figuring and polishing) and amplitude errors (from non-uniform coatings).
The ability to efficiently and accurately propagate an arbitrarily-aberrated wavefront from surface to surface in the system, accounting for the transformation of phase errors into amplitude errors and vice-versa due to propagation effects.

Wavefront control using a representation of a deformable mirror (DM) with realistic influence functions that describe how the DM surface is modified by movement of an actuator.

Wavefront sensing and control algorithms to determine the optimum DM settings to provide a region of suppressed scattered light in which high contrast imaging is possible.

Rigorous, physics-based computational techniques (e.g. Huygens) are capable of accurately modeling a system, but they are extremely slow and can require many hours or days to propagate an aberrated wavefront just once through a system. Because thousands of separate propagations through the entire optical system are required simply for creating the system response matrix used by the wavefront control algorithm, it is impractical to use these methods even with supercomputers (the response matrix often needs to be regenerated as system components change). Efficient but accurate modeling procedures that require only minutes to execute are needed.

Having the means to simulate propagation of a wavefront through a system is not sufficient to accurately predict the contrast limit of a given telescope and coronagraph combination. Each coronagraph responds differently to wavefront aberrations and sets particular limits on the ability to control those errors, especially over a broad wavelength range\(^1,2\). Therefore, the models must be executed within a wavefront control framework\(^1,3\) that senses the simulated electric field at the image plane and then determines the DM actuator settings necessary to minimize the scattered light around the star. Because of nonlinearities in the system, this is an iterative process. Once a “dark hole” around the star has been created, additional analyses can be done to measure the sensitivity of the coronagraph to pointing errors and thermally-induced aberrations, and to examine the morphology and spectral characteristics of the residual background light to develop techniques to differentiate it from any planets.

### 1.2 Goals of this study

This study combines optical modeling of a realistic telescope with wavefront control algorithms to predict the contrast performance limits of coronagraphic techniques that are theoretically capable of suppressing diffraction down to contrast levels of \(10^{-10}\) to within \(2.5 \lambda/D\) radians of the star (\(D\) is telescope diameter). The key to deriving the limits is advancing the current state of optical modeling of each of these coronagraphs using methods that are both accurate and computationally efficient. There are advantages to simulating all of these coronagraphs together using the same basic framework. The performance of each can be compared to the others on even terms, given that the same aberrations and wavefront control methods will be used (an apples-to-apples comparison). Piecing together the results from different studies with varying layouts and modeling methods will not provide the consistency necessary to identify the true capabilities of each technique. Unforeseen and perhaps subtle causes for the poor performance of a coronagraph might be more readily identified by comparing the behaviors of the others, especially how each responds to changes in the wavefront caused by the DM. Any problems can be diagnosed as system-level (if all coronagraphs have poor performance) or particular to just one coronagraph (or its modeling technique).

Progress in the study is measured by the achievement of specified technical milestones, which are the same for each of the three coronagraphic techniques. Milestone #1 is the development of fast but efficient methods for propagating an arbitrarily-aberrated wavefront through the coronagraphic elements. The efficiency is judged by the total amount of time it takes to compute the effect of individually pushing each DM actuator on each DM at each of five sampling wavelengths. Assuming a perfectly symmetrical system and making use of the eight-fold symmetry of the deformable mirror, the milestone’s time limit is 48 hours on a modern workstation. The accuracy metric is the root-mean-square of the difference of the electric fields produced by the efficient method \((E_{\text{eff}})\) and a more rigorous, slower reference method \((E_{\text{ref}})\), as measured within a specified field around the occulted source and expressed in terms of contrast:

\[
\text{accuracy} = \frac{\text{RMS}\left|E_{\text{eff}} - E_{\text{ref}}\right|^2}{\max(PSF)}
\]
Dividing the RMS field difference by the peak value of the unocculted stellar PSF converts the result into contrast. The milestone accuracy requirement is ≤1% for fields with mean contrasts of $10^{-5}$ and $10^{-10}$ (produced by including an appropriately aberrated wavefront).

Milestone #2 is the determination of the design parameters for each coronagraphic type that produce a $10^{-10}$ mean contrast field over $\lambda = 500 – 600$ nm for a realistically aberrated optical system when wavefront control is used (via two DMs in sequence).

A benefit of end-to-end modeling is that it exercises wavefront control algorithms on data having realistic wavefront errors of the sort that will be encountered in flight. The algorithm used here, Electric Field Conjugation\(^3\) (EFC), has been used both in models and in testbeds to determine DM settings that produce a dark hole for high contrast imaging\(^5\).

### 1.3 Procedures

The following steps are followed for each coronagraph:

An efficient and accurate method of representing wavefront modification by the image plane mask (HBLC, VVC) or optical subsystem (PIAA mirrors and apodizer) are established and verified for selected test cases against more rigorous, physics-based algorithms. A telescope/coronagraph layout is specified and implemented as a prescription used by the PROPER propagation library\(^6\) for IDL. The modeling routines that represent the coronagraphic masks and PIAA subsystem are integrated into the prescription in places where the PROPER routines do not suffice.

With the means available to efficiently compute wavefront propagation through the entire optical system, the system response matrix is generated for use by the wavefront control algorithm. For each actuator on each DM and at each of the wavelengths that samples the bandpass of interest, the actuator is pistoned by a small amount and the effect propagated through an aberration-free optical system model. The resulting complex-valued electric field at the image plane is stored in the response matrix.

A dark hole around the simulated point source is created using the EFC wavefront control algorithm. The simulated field at the broadband image plane is measured at a few sample wavelengths, then the DM settings are updated using EFC, and finally the resulting new field is computed. This process continues until convergence. If the contrast limit at convergence is poor, various parameters are adjusted and the process repeated to try to improve the result, including:

- Regularization factor (prevents divergence due to large DM pistons)
- Field point weighting (emphasizes contributions from certain field areas)
- Update of the system response matrix incorporating measured aberrations

If the desired contrast of $10^{-10}$ cannot be achieved, the sensitivity of the coronagraph to aberrations on critical, near-focus optics is investigated by changing the aberrations. The layout might be adjusted, if possible, to avoid those optics. The entire process described above is then be repeated. In parallel, the sensitivity of the coronagraph to these particular aberrations may be investigated to identify its physical basis. The results are compared to those from the other coronagraphs to identify any subtle, unforeseen behavior.

### 1.4 Modeling framework

The propagation of the simulated wavefronts through the telescope and coronagraph optics will accomplished with the PROPER library. PROPER was developed at JPL for the Terrestrial Planet Finder Coronagraph program to provide a common set of tools for the community for simulating optical systems, especially coronagraphs. It is freely available from www.openchannelsoftware.com and is well documented. It is a set of routines that uses standard Fourier-based Angular Spectrum and Fresnel propagation algorithms for surface-to-surface sequential wavefront propagation. The library includes routines for the creation of complex apertures and of phase and amplitude aberrations (specified by Zernike coefficients or power spectrum curves). It also includes a deformable mirror model that includes measured actuator influence functions. To simulate a polychromatic image, multiple monochromatic ones generated at wavelengths sampling the passband are be added together.

For the particular components of the coronagraphs that cannot be implemented using PROPER, modeling routines are developed in IDL (for easy integration with PROPER). For PIAA it is necessary to implement some particular routines in C that can be called by IDL for improved speed (especially to take advantage of multithreading on multiple CPU systems). The wavefront control software, separate from PROPER, implements the EFC algorithm.
1.5 Telescope and coronagraph parameters

We create circular regions around the star that are devoid of scattered light using wavefront control. The inner working angle (IWA) that defines the inner radius of the annulus is set by the 50% transmission radius of the specific coronagraphic being modeled. For this study the IWA is $2.5 \frac{\lambda}{c}/D$ radians at $\lambda_c = 550$ nm, the central wavelength of the $\lambda = 500 – 600$ nm passband.

In this study the actual size of the telescope is not strongly relevant and does not significantly affect instrumental contrast performance as long as angular separations in the sky are measured in units scaled with the telescope diameter (e.g. units of $\lambda/D$ radians). Only when absolute angles (e.g. arcseconds) and sky background sources (e.g. zodiacal light) are involved does the actual size matter. The simulated telescope is an 1.5 m off-axis Gregorian. This provides an unobscured aperture ideal for high contrast imaging and a focus prior to the secondary mirror where a small stop can be placed for stray light reduction. The beam from the secondary mirror is directed past the edge of the primary and a fold mirror directs the beam behind the primary and into the instrument. An off-axis parabola (OAP) creates an image of the entrance pupil onto the first DM. From there the beam travels to a second DM. Up to this point the configuration is the same for all three coronagraphs.

For the HBLC and VVC, the beam is focused by an OAP onto the image plane mask. It is then directed via a fold mirror onto an OAP that collimates the beam and forms a pupil image, where the Lyot stop is located. An OAP or lens then refocuses the light onto the detector. In the PIAA coronagraph the beam from the 2nd DM is directed to a pair of OAPs that reform an image of the pupil onto the 1st PIAA mirror (M1). This apodizes the beam on the 2nd PIAA optic (M2). An OAP then images M2 onto the post-apodizer, followed by another OAP that focuses the beam onto the occultor. The beam is then sent through a reverse set of PIAA optics to correct for the distortion, and then to optics to focus onto the detector.

The specifications for the coronagraphs are:

- **Hybrid Band-Limited Coronagraph (HBLC):** The HBLC uses a circular amplitude and phase transmission pattern at an intermediate focus that provides an inner working angle of $2.5 \frac{\lambda}{c}/D$ radians. An approximately 1-sinc($r$) amplitude transmission profile will be used that has 4th order aberration rejection. As much as is practical, the wavelength-dependent effects of the multilayer dielectric coating will be implemented. A simple circular Lyot stop is appropriately matched to the size of the occultor.

- **Vector Vortex Coronagraph (VVC):** The VVC is implemented with a representation of the effective phase screw with a charge of 4 (4th order aberration rejection). The opaque spot at the center of the screw has a diameter of $0.5 \frac{\lambda}{c}/D$ radians. This provides an inner working angle of $\leq 2.0 \frac{\lambda}{c}/D$. A simple circular Lyot stop will be used with an opening 90% of the pupil diameter.

- **Phase Induced Amplitude Apodization (PIAA):** The PIAA remaps a wavefront to create an apodized beam. Prescriptions for the current best PIAA optics are used, and a binary post-apodizer has been designed; theoretically, this system suppress diffraction to contrast levels of $<10^{-10}$ with an IWA of $2.5 \frac{\lambda}{c}/D$.

In the following sections we will discuss our results for PIAA. Our results for the VVC and HBLC will be discussed in following reports. To date, we have achieve Milestone #1 for the VVC.

2. PIAA PROPAGATION

2.1 Overview

Apodization in a telescope will produce a point spread function with significantly reduced wing intensity, making it a technique of interest for high contrast imaging (Figure 1). Conventional apodizers using transmission-altering masks reduce throughput too much to be acceptable for exoplanet imaging, and they would be difficult to accurately fabricate and operate over broad bandpasses. An alternative solution, Phase-Induced Amplitude Apodization (PIAA), utilizes two optics (lenses or mirrors) that geometrically distort the wavefront to produce an apodized beam. The first optic, M1, remaps the beam onto the second optic, M2, that corrects for the phase errors induced by the remapping to create a “flat” wavefront (Figure 2).
Pure PIAA systems that provide $10^{-10}$ contrast are impractical due to diffraction effects from the edge of M1 and stringent surface tolerances. These limitations can be reduced in a hybrid system using a weak apodizer (post-apodizer) placed near or at a subsequent image of M2. This reduces throughput compared to a pure PIAA arrangement but only by a small amount (the additional apodization occurs only at the edge of the already-PIAA-produced apodized beam). In practice, post-apodizers are implemented using binary transmission masks composed of a series of narrow, opaque rings spaced and sized to provide the desired diffraction reduction. These have been fabricated using lithographic techniques. Greyscale post-apodizers are not feasible due to fabrication tolerances, wavelength-dependent transmission effects, and coating-induced phase modification.

The PIAA apodized beam is brought to focus where a simple circular occulter blocks the core of the point spread function, where nearly all of the star’s light has been concentrated. The remapping of the entrance pupil by the PIAA optics creates a highly distorted, field-dependent PSF, significantly reducing the sharpness (and hence contrast) of any field sources such as exoplanets. To negate this, another pupil image is formed after the occulter, and the beam is sent through another set of PIAA optics but in reverse order, restoring the wavefront to its original mapping and the field PSF to its original sharpness. Afterwards, the beam is focused onto the detector.

2.2 Modeling PIAA

The optical modeling of PIAA is much more difficult than for other coronagraphic techniques because the 1st PIAA optic that creates the apodized beam has a surface with a high rate of curvature change near its outer edge. With conventional Fourier-based propagation methods (e.g., Fresnel, angular spectrum) this change in the phase of the wavefront is too great to be accurately represented by a uniformly spaced grid without aliasing unless the array is unreasonably large (tens of thousands of points in each dimension). These techniques also require a nearly-parabolic or flat wavefront phase and uniform spacing between rays during propagation, which is not the case for PIAA. One can use brute force methods,
such as integration of the Rayleigh-Sommerfeld diffraction equation, but the time required to propagate just one wavefront would be prohibitive (weeks, or even on a large parallel processing cluster, days).

A new, accurate means of propagation has been devised called S-Huygens\(^{12,13}\). This is an approximation to the integration of the Huygens wavelet diffraction integral, but with fewer simplifications than used by the Fourier-based algorithms. It is too slow for practical modeling because any arbitrarily-aberrated input wavefront must be decomposed into numerous (hundreds) of orthogonal terms that are then separately propagated, and the propagation of a single term can take tens of seconds or more. It is, however, useful as a means of verifying other, more efficient methods without resorting to the even more computationally intensive methods like numerical integration of the Rayleigh-Sommerfeld equation.

One efficient method, used in this study, is geometric remapping with propagation of individual component harmonic aberrations, which we call PIAA Angular Spectrum Propagation (PASP)\(^{14,15}\). In this case the aberrated wavefront is decomposed into Fourier components (ripples in X and Y) that are each analytically propagated followed by wavefront geometric distortion mapping via interpolation. To handle diffraction from the edge of M1, which is poorly sampled when using PASP, S-Huygens is used for just that region, and the results merged with those from PASP (the transition region is weighted to smoothly join the two wavefronts).

### 2.3 Binary post-apodizer

As previously noted, a binary post-apodizer (BPA) is used to offload some of the apodization function of the M1 optic to ease fabrication as well as to reduce diffraction effects from the edge of M1. The BPA is a series of concentric rings of varying thicknesses placed at a pupil plane conjugate to and after M2 (but before the occulter). It is designed to work with the PIAA optics to produce a field within a given radius of the star that has a specified maximum contrast level. Note that it does not approximate a grayscale apodizer, which would produce a field with no outer field radius limit; approximating such an apodizer with finite binary structures would be impractical at \(10^{-10}\) contrast levels.

The ring radii and thicknesses are chosen to produce the required contrast at the expense of outer field imaging. The BPA is designed using a method similar to Vanderbei et. al.\(^{16}\) but with some differences to incorporate a minimum ring thickness constraint. The method is essentially a linear optimization on the discrete samples of the radial apodizer profile. Some iterating was required to arrive at the final BPA design used in this study. There were certain requirements: the field created by the unaberrated PIAA optics and the BPA must have a mean contrast of \(<10^{-10}\) inside of 20 \(\lambda/D\) radians over a wavelength range of \(\lambda = 500 - 600\) nm, and the minimum post-apodizer ring thickness must be \(>4\) \(\mu\)m for a 108 mm beam diameter, to ease fabrication in the real world. The outer edge of the PIAA-apodized beam must also be masked and devoid of any rings, since the diffraction effects are greatest there and vary significantly with wavelength. Our final design meets all of these goals.

The thin rings that comprise a BPA cannot be directly represented with the sampling typically used in 2-D wavefront propagation (e.g. ~100-1000 samples across the pupil). Representing the rings using grayscale patterns, where transmission in each pixel is proportional to the pixel area covered by a ring, has been shown to provide insufficient accuracy. One way to implement it with limited sampling is to compute a spatially-filtered version of a high resolution mask, creating a non-physical representation (one with negative transmission values) that produces the same net effect in the image region of concern. To construct it, a very high resolution (tens of thousands of points), one-dimensional array representing the uniform transmission across the wavefront is created, and elements are set to zero where masked by the apodizer rings. This array is then Fourier transformed to produce the frequency spectrum of the apodizer (alternatively, the spectrum can be analytically computed). The spectrum is multiplied by an apodized windowing function to limit the spatial frequencies to those that can be sufficiently represented with the lower resolution of the 2-D wavefront array. The resulting 1-D spectrum array is inverse Fourier transformed, creating the filtered apodizer radial profile that is then interpolated onto a 2-D grid.

To create a filtered version, a 1-D array with a sampling of 240,000 pixels across the radius of M2 was used to represent the BPA and was then Fourier transformed. A tapered windowing function was then applied to this spectrum with an amplitude of 1.0 out to a spatial frequency of 48 cycles/diameter with a cosine-squared taper from 48 – 54 cycles/diameter. After transforming back to apodizer space, the resulting field contained oscillations and excursions with negative real values. This filtered apodizer is shown in Figure 3.

To determine the validity of the filtered apodizer, a 240,000 element, 1-D radial wavefront at M2 was created using the ideal PIAA-induced apodization profile (no diffraction, no errors). This was multiplied by the finite BPA and Hankel
transformed to the PIAA focus. The same ideal wavefront was separately multiplied by the filtered version of the apodizer and transformed to focus. The contrast profiles for these two versions agree well. Note that a perfect grayscale post-apodizer applied to the same wavefront would produce even better contrast with an unlimited outer radius.

The next step was to verify the two-dimensional representation of the filtered BPA. First, a 1-D, 90,000 element, aberrated wavefront at M1 was created. The aberrations (a pattern of radial ripples on M1) were iteratively derived to produce a \(10^{-10}\) contrast field at the PIAA focus at \(\lambda=500\) nm. This was propagated via S-Huygens to produce a 60,000 element aberrated field along the radius of M2. One copy of this field was multiplied by the finite BPA and Hankel transformed to the PIAA focus. The second copy was converted into a 2-D field by downsampling via binning and interpolation, with 1434 samples across M2 and embedded in a 2048 \(\times\) 2048 array. This was multiplied by a similarly sampled version of the filtered apodizer and then Fourier transformed to the PIAA focus.

To determine the contrasts in the final image plane, we continued propagation beyond the PIAA focus. First, the fields at the PIAA focus were multiplied by a central occulter of radius 2.1 \(\lambda/D\) and an outer radius mask of 20 \(\lambda/D\). They were then transformed (Hankel or Fourier) to the reverse PIAA optics, where a mask with a 96% opening diameter was applied (to mask residual diffraction effects from the binary post-apodizer). Then, rather than explicitly propagating between the reverse optics, interpolation was used to reverse the wavefront distortion and PIAA-induced apodization. The fields were then propagated to the final focus and the contrasts computed. The 1-D field has a contrast of \(1.30 \times 10^{-10}\) and the 2-D one is \(1.24 \times 10^{-10}\), agreeing to the accuracy required in this study.

![Figure 3. Binary post-apodizer radial profile. (Left) Finite ring profile; (Right) Filtered version of profile.](image)

### 2.4 Combining PASP and PROPER

We use PASP as described by Krist et al.\(^{17}\) in an earlier comparison with S-Huygens. From the primary mirror up to PIAA M1, the wavefront is propagated using PROPER with a grid size of 2048 \(\times\) 2048 pixels and an entrance pupil diameter of 819 pixels. At M1 the wavefront is Fourier transformed to obtain the spatial frequency spectrum, which contains the amplitudes of the X and Y Fourier components. The forward PIAA propagation is limited to spatial frequencies of <60 cycles/diameter as a compromise between computation time and accuracy. In the forward PIAA the wavefront (ripple) for each Fourier component is analytically propagated from M1 to create a wavefront at M2 with 460 pixels across the beam. This is done for each term, and then the individual wavefronts are added together. Interpolation is used to resample the M2 wavefront to create a beam 1433 pixels across within a 2048 \(\times\) 2048 grid. PROPER is then used from M2 to the entrance of the reverse PIAA, where the wavefront is again transformed into Fourier components, limited to 22 cycles/D (a field stop located at the focus of the forward PIAA system actually limits spatial frequencies to <20 cycles/D). Each component wavefront is analytically propagated to the exit of the reverse PIAA (again, 460 pixels across the beam), and all of them are added together. This wavefront is then interpolated to the original sampling of 819 pixels in a 2048 \(\times\) 2048 array. From there, PROPER completes the propagation to the final focus.
In this study we make one addition to the PASP algorithm described by Krist et al. Because of the limited number of spatial frequencies propagated by PASP for efficiency reasons, a PASP+S-Huygens propagation of an unaberrated wavefront is not equal to that produced by a pure S-Huygens propagation of the entire wavefront. To compensate for this, the interior portion (taper weighted at the edge) of an unaberrated M1 wavefront is propagated to M2 using PASP and added to the S-Huygens-generated M1 edge wavefront at M2 to create a complete M2 wavefront, as is done when propagating an aberrated wavefront. Separately, the full unaberrated M1 wavefront is propagated to M2 using S-Huygens. The difference between these fields produces a correction term that compensates for the PASP limitations for the perfect, unaberrated term of the wavefront (which contains most of the energy in low-aberration systems like those we study here). This function needs to be computed only once per wavelength and stored.

2.5 Verifying the S-Huygens algorithm

To demonstrate the accuracy of the PASP algorithm, an aberrated, two-dimensional wavefront was propagated through the model PIAA coronagraphic system separately using PASP and S-Huygens, and the resulting fields were compared. To validate the use of S-Huygens as the reference, we first compared its results to those generated by numerical integration of the Rayleigh-Sommerfeld diffraction equation. Each method was used to propagate a circularly-symmetric aberrated wavefront (circular phase ripples) from M1 to M2, after which the wavefronts were multiplied by a greyscale post-apodizer and then propagated to focus using a simple Fourier transform. The Rayleigh-Sommerfeld calculation took 31 hours on a 256 processor system with 6.4 billion samples over the surface of M1. S-Huygens took 70 seconds on an 8-core workstation using parallelized code with 45,000 points along the radius of M1 (because a completely symmetrical system was used, only one S-Huygens radial term needed to be calculated). Assuming the Rayleigh-Sommerfeld results were exact, the accuracy of S-Huygens was 0.04%, far better than the milestone requirement. We therefore consider S-Huygens a valid reference algorithm against which to judge other methods.

2.6 Verifying PROPER

Although PROPER uses commonly accepted Fourier-based algorithms, we took the opportunity to compare results from it and S-Huygens. Each method was used to propagate a circularly-aberrated input wavefront through an entire simple bandlimited Lyot coronagraphic system. Based on these results, the accuracy of PROPER is 0.6% when using a grid size of 2048 × 2048 with the beam at the entrance pupil spanning ¼ the diameter of the grid.

2.7 Verifying the PASP algorithm

To verify PASP against S-Huygens, we created an aberrated 2-D wavefront that produces a generally uniform-intensity speckle field in the coronagraphic image plane. The wavefront was propagated through a simple but complete PIAA system composed of both forward and reverse PIAA optics. As discussed, the PIAA beam remapping compresses the wavefront, increasing the spatial frequency of an aberration. Therefore, the error map is filtered to reject spatial frequencies above 48 cycles/diameter to prevent aliasing during beam compression. This is sufficient to include aberrations that directly create speckles inside the dark hole region (those of ~20 cycles/diameter or less) as well as higher spatial frequency ones that create speckles in the dark hole due to frequency folding. The wavefront errors are scaled to produce 10^{-5} and 10^{-10} mean contrast inside the dark hole area. PASP and S-Huygens were used to propagate the wavefront between the two PIAA sets while PROPER was used between all other surfaces.

The accuracies, as computed using the equation in section 1.2, are 0.01% for 10^{-5} contrast and 0.03% for 10^{-10} contrast. These easily meet the 1% requirement by nearly two orders of magnitude, which is satisfyingly surprising given that the algorithms and codes are different.

3. PIAA END-TO-END MODELING

3.1 Overview

With a suitably efficient and accurate propagation method in hand, the general behavior of a PIAA system containing realistic wavefront errors and a wavefront control mechanism (DMs) can be derived. The first step is to generate a response matrix that describes how each DM actuator alters the field at the image plane at each wavelength being used to sample the bandpass (in this study, 5 wavelengths evenly spaced over λ = 500 – 600 nm). This matrix is inverted so that the DM actuator settings required to minimize the energy in the measured field around the star can be determined. In real life, the complex-valued field can be derived by applying multiple known patterns to the DM and solving for the common field, but for speed and simplicity we directly used the field computed by the simulation.
3.2 Generating the DM response matrix

Each actuator on each DM alters the field in the image plane in a unique way. To create the response matrix of the field changes, each DM actuator is pistonned by a small amount (< 1 nm), the field is propagated through the unaberrated system model, and the resulting field change is stored. This is done at each wavelength that is being used to sample the broad bandpass. In our system we have two DMs each with 46 actuators across the pupil (about 1662 active actuators per DM). This means we must individually propagate 1662 actuators × 2 DMs × 5 wavelengths = 16620 times through the system. We can make use of the fact that we have a circularly-symmetric optical system and take advantage of 8-fold symmetry on the DM to reduce the number of propagations to 2077, poking only actuators within a 45° sector on each DM. The fields for the other actuators are then simply obtained by appropriately transposing and rotating the fields computed for the one sector.

We start the sequence by propagating at each wavelength an unaberrated wavefront from the primary mirror to the DM using PROPER and saving the field (this forward part of the system does not change if the DMs are altered). Next, an actuator is poked and the perturbed field is propagated to PIAA M1. A Fourier transform is then used to compute the X,Y spatial frequency spectrum of the field, which is stored to disk with the field spectra for all the other pokes. The spatial frequencies in the spectra are limited to 60 cycles/diameter in the forward PIAA and 22.4 in the reverse (the field stop at the occulter actually limits the spatial frequencies to 20 cycles/D).

Once all the field spectra at M1 for all of the pokes have been computed, the PASP forward-PIAA propagation matrix for each wavelength is generated for each X,Y Fourier component (ripple). This involves analytically propagating and remapping the rippled wavefront corresponding to each component from M1 to M2 and storing the field in the matrix. We again take advantage of perfect symmetry in our model system to reduce the computational load by a factor of 8 by propagating only the ripples corresponding to a 45° sector of the frequency spectrum. We can then use those fields, with appropriate rotates and/or transposes, for the fields corresponding to the other sectors of the spectrum. There are over 70,000 values in the spatial frequency spectrum (which is limited to <60 cycles/D with a sampling of 0.4 cycles/D/element), and each corresponding wavefront (ripple pattern) is 460 × 460 pixels using double precision complex numbers (16 bytes/pixel). Even using the factor of 8 reduction, the propagation matrix is 30 gigabytes.

The M1 field spectrum corresponding to each actuator poke is read in and multiplied by the propagation matrix, producing the wavefront at M2, to which the S-Huygens-generated edge diffraction component of M1 is added, and the result is then stored to disk. Using a propagation matrix and processing all of the DM poke wavefronts in one stage avoids the need to recompute the Fourier term propagations for each poke. The M2 wavefronts are then serially read in and propagated using PROPER from M2 to the filtered version of the post-apodizer, and then on to the occulter and the entrance of the reverse PIAA. There, the spatial frequency spectrum is computed and saved to disk. Another propagation matrix is computed for the reverse PIAA, all the DM poke wavefronts propagated through it using PASP, and then PROPER is used to the final focus. The total computation time for generating the response matrix was 33.8 hours on a dual quad-core Xeon workstation with 48 GB of memory, which meets the milestone requirement of <48 hours. It required the combined use of PROPER and custom, multithreaded C code.

The field changes within the annulus of interest around the star at final focus for all of the actuators and wavelengths are assembled into a single DM response matrix. Regularization constants are also included to prevent solutions with anomalously high actuator strokes. The matrix is then pseudo-inverted using singular value decomposition (SVD). We have found that using the SVD in the optimized math library from the CPU vendor provides an order of magnitude improvement in speed over IDL’s SVD routine.

The generation of the response matrix for other coronagraphs follows a similar procedure as for PIAA, though only PIAA requires the stage-to-stage progression through the system due to the need to optimize the use of PASP.

3.3 The aberrated system

With the inverted response matrix generated, the next step is to simulate an aberrated system with wavefront control. For each optic, synthetic phase and amplitude error maps are generated from power spectral density curves derived from actual optics. For the forward PIAA M1 and M2 mirrors, measured phase error maps (Figure 4) of the current best PIAA optics have been provided by the fabricators, Tinsley, Inc. Prior to use, these maps were filtered to <48 cycles/D and then interpolated using a damped sinc method to match the sampling of the wavefront. Use of these maps allow us to identify aberrations that would not otherwise be accounted for by the synthetic maps, which have isotropic surfaces that lack spatially-correlated structure. The most obvious example of this is the region of narrow rings along the edge of M1,
which have surface departures of more than ±30 nm. The curvature is very high here, making it difficult to polish out errors even with ion beam techniques.

3.4 Creating the dark hole

At each of the five component wavelengths (500, 525, 550, 575, 600 nm) the wavefront is propagated through the system to the final image plane, where the complex-valued electric fields are extracted from within the control annulus. In a real system, these fields would be derived by placing a specified pattern on a DM and measuring the change in intensity in the image plane. This would be repeated for a series of specified DM patterns, and the results used to solve for the electric field. However, for the sake of efficiency we simply use the computed fields here.

The field values within the annulus at all of the wavelengths are stored in a field matrix. Assuming linearity, the field matrix and inverted DM response matrix are then used to solve for the actuator settings that will minimize the energy to create a dark hole around the star. The DMs are set and additional rounds of propagation-sense-control continue until a contrast goal is reached or the contrast converges.

Because we are propagating an entire wavefront at each wavelength in sequence, a more direct progression through the system is used than when the response matrix was computed. As before, the wavefront is propagated from the primary mirror to the DM and then on to PIAA M1 using PROPER. At M1, the Fourier transform of the wavefront is taken and each corresponding Fourier component (ripple) is propagated through the system. This is different from before, where a propagation matrix was constructed for all components. We found that with a single wavefront at sequential wavelengths it is quicker to compute the propagation of each component separately than to read in the entire propagation matrix from disk (which is faster when propagating multiple wavefronts at a single wavelength). The same happens at the reverse PIAA. The total time to propagate once through the system is 5 minutes on the previously detailed workstation (1.7 minutes just from M1 to M2).

3.5 Unaberrated system contrast

The use of a binary post-apodizer creates a contrast floor, even in the absence of any aberrations. Figure 5 (left panel) shows the contrast map of the unaberrated system. The mean contrast in the dark hole is $2.0 \times 10^{-11}$ and the median is $1.0 \times 10^{-11}$. At the inner working angle the contrast is $6 \times 10^{-10}$, and it falls rapidly to $<10^{-11}$ towards the outer working angle. A couple of iterations of EFC is sufficient to reduce the mean contrast to $3.5 \times 10^{-12}$ and the contrast at the IWA to $2 \times 10^{-10}$. Thus, our implementation of PIAA is not a limiting factor for contrast, except that it would be desirable for the IWA contrast to be $<10^{-10}$. The bulk of the residuals are at the longer wavelengths, indicating that making the occulter larger might reduce some of the light.
3.6 Results when only M1 is aberrated

The M1 mirror provides the bulk of the apodization in a PIAA system. The most critical region of the optic is the edge, where the curvature rapidly varies. The measured M1 maps shows a series of concentric rings along the edge, remnants of the figuring and polishing process. The spatial frequency of these in the unfiltered map is ~66 cycles/diameter, which would require a DM with at least $66 \times 2 = 133$ actuators across the pupil to correct. In the map filtered to <48 cycles/D used in our simulations, they are ~37 cycles/D, requiring a DM with 74 actuators/D, still far above our DM with 46 actuators/D (the filtering also reduces their amplitude by about 40%). They therefore can only be weakly corrected by the DM in the simulations. The remaining errors scatter light over the entire remapped pupil and over the entire dark hole. The filtered map has an RMS wavefront error (2× surface) of 6.2 nm with peak amplitudes of ±40 nm.

Figure 6 shows the dark hole region after running EFC in the two DM system where only M1 is aberrated. The mean contrast in the dark hole over the control annulus is $1.7 \times 10^{-9}$ and the median is $7.1 \times 10^{-10}$. To verify that the edge errors are the dominant cause of the high level of light in the dark hole, we zeroed the interior of the map, leaving only the outer 10% of the radius. Running EFC again produced a dark hole almost identical to the full map’s. A pronounced, nearly-horizontal streak running across the dark hole is oriented perpendicular to the highest ridges along the edge of the map.

The next step was to diminish the edge errors by multiplying the map with a cosine taper on the outer 30% radius (3.6 nm RMS wavefront error, ±18 nm peak errors, the highest outer edge errors reduced by ~3×) and running EFC. This significantly improved the contrast in the dark hole, but there were still high values in the streak and around the inner radius (mean = $2.8 \times 10^{-10}$, median = $9.5 \times 10^{-11}$). After that, a cosine-squared taper was used (same full-map statistics as with the cosine taper, but the outer errors reduced by ~10× relative to the non-tapered map), reducing the mean after EFC to $1.5 \times 10^{-10}$ and the median to $4.1 \times 10^{-11}$. There were still small regions of $\sim 10^{-9}$ contrast near the inner radius and along the outer radius in the direction of the streak. Given that these results remained too large, the entire map (cosine-squared tapered) was reduced in amplitude by a factor of two. The end result has a mean contrast of $3.3 \times 10^{-11}$ and a median of $8.0 \times 10^{-12}$. This indicates that the errors in the center of the M1 optic need to be reduce by 2× and those along the outer edge by ~20×. While this may be feasible for the interior, the large reduction in the outer errors may be beyond current polishing techniques.

3.7 Results when only M2 is aberrated

As we did for M1, we ran EFC on a system in which only M2 was aberrated, using the measured error map. The <48 cycles/D filtered map has an RMS wavefront error of 6.8 nm with maximum excursions of ±30 nm. With more gradual curvatures, the errors appear more like those on a conventional high-quality optic, lacking the ridges around the edge that are seen on M1. The M2 error levels, though, are still much higher than can routinely be achieved (<1 nm RMS) on custom-figured mirrors, indicating that some additional time (and, of course, money) would produce a smoother optic.

The EFC result (Figure 7) has a mean contrast of $1.7 \times 10^{-10}$ and median of $4.2 \times 10^{-11}$. There is a prominent $\sim 10^{-9}$ contrast vertical streak crossing a portion of the dark hole. The DM patterns show large strokes on a small number of actuators at the edge of DM #1, an attempt by the algorithm to introduce a compensating wavefront to null the streak.
3.8 Correcting post-PIAA optical errors with pre-PIAA DMs

The preceding results highlight the limitations of using DMs upstream of the PIAA optics (i.e., in undistorted wavefront space) to correct wavefront errors from optics located after PIAA, specifically from M2 to the occulter (i.e., optics in distorted wavefront space). There are some subtle problems that complicate wavefront control in this configuration.

As illustrated in Figure 8, in the forward direction (M1 to M2), the beam is compressed towards the center and stretched along the edge, but when backing through the system (M2 to M1 to the DMs), the opposite happens. A low spatial frequency on M2 is seen by the DM as a higher frequency one along the outer edge. This can push the aberration beyond the spatial frequency control regime of the DMs so that only the central region of the wavefront can be well corrected (lower left image of Figure 8). The remaining errors will scatter light over the entire pupil and inside the dark hole region. A second problem is that when the pre-PIAA DMs are used to correct post-PIAA errors, they effectively introduce what amounts to a phase error at M1, which in turn causes a wavelength-dependent error in the PIAA-induced apodization that cannot be fully corrected.
3.9 Adding a 3rd DM

The previous examples demonstrated that there are inherent difficulties in controlling wavefront errors when the deformable mirrors are in the undistorted wavefront space but critical optics are in the distorted (PIAA apodized) space. A potential solution is to add a deformable mirror after the PIAA optics, so that it can more directly correct aberrations from surfaces between M2 and the occulter, avoiding the remapping by PIAA. This would also offload some of the control responsibility from the pre-PIAA DMs.
We added another DM at the plane of the post-apodizer, with 46 actuators across the opening of the mask (out to the last clear ring). The addition of these new actuators to the DM response function matrix required another stage of propagation. At each wavelength, an unaberrated wavefront was propagated from M1 to M2 with PASP, then to the third DM using PROPER and stored. For each actuator, this wavefront was read in, the actuator poked, and the resulting wavefront propagated to and through the reverse PIAA and on to the final image plane, using a similar procedure as for the prior DMs. The new response functions were added to the DM response matrix, which was then inverted.

We first ran EFC with three DMs in a system with only M1 aberrations and with and without the edge tapering of the M1 map. The resulting dark holes show that the third DM helps reduce the overall contrast level compared to the 2 DM solution, but the streak from the large edge errors remains. With a tapered edge, the three-DM results are only marginally better than those using two DMs. The third DM has little control authority over errors from surfaces prior to PIAA. The beam falling on the DM is highly apodized, so only a small number of actuators can actually have a strong effect.

When the 3-DM EFC is run on a system with only M2 aberrations, the improvement is significant (Figure 10). The mean contrast decreased to 1.1×10^{-11}, down from 1.7×10^{-10} with the 2-DM solution. The 10^{-9} streak has been largely suppressed, and the only significant errors are near the inner radius, which are at the level of those in the unaberrated system (section 3.5).
3.10 Adding all surface errors in a 3-DM system

Using a 3-DM system with the $\frac{1}{2}$ amplitude, cosine-squared tapered, filtered M1 map and the filtered M2 map, plus realistic simulated phase and amplitude aberrations on all other optics, we were able to achieve the dark hole shown in Figure 11. It has a mean contrast of $9.6 \times 10^{-11}$ and median contrast of $3.4 \times 10^{-11}$. This meets the Milestone #2 requirement. Whether this is realistic depends on how much improvement can be made in reducing the M1 errors.

4. CONCLUSIONS

We have succeeded in achieving the milestones for our TDEM study for PIAA. Milestone #1 has also been achieved for the VVC and will be detailed in a future report. At the time of writing, the design of the HBLC was being established.

Our results indicate that without numerical modeling including wavefront control, it is difficult to appreciate how a coronagraphic system will really respond to aberrations. A coronagraph must be evaluated by its performance in an aberrated system, not just by how much it suppresses diffraction in the absence of errors.

The PIAA coronagraph presents unique challenges for wavefront control because optics are located in two different wavefront mapping spaces. Deformable mirrors located before the PIAA remapping occurs are unable to fully correct wavefront errors introduced by optics located after PIAA. Also, any modification of the wavefront by these DMs alters the apodization profile produced by PIAA. A third DM located between the forward PIAA optics and the occulter provides sufficient control to provide the required contrast levels with current post-PIAA optics (namely, M2). The current state of M1 fabrication is not sufficient, however, to produce a $10^{-10}$ contrast field over a 20% bandpass. Errors along the edge of the current M1 must be reduced by ~20× and those in the interior by about 2×. It may be possible to reduce the surface requirements by using a pre-apodizer that offloads some apodization responsibility from M1, at the expense of throughput.

We note that the PIAA configuration used here, with a reverse PIAA and DMs prior to the PIAA optics, is the configuration chosen for proposed space missions. Having the DMs in front allows generation of a larger dark hole than if the DMs were after PIAA. Testbed experiments to date, however, have used a DM after PIAA, which limits the outer control radius and corrects for M1-induced apodization errors after-the-fact. These two configurations have intrinsically different wavefront control behaviors.

We also note that modifications to the classic PIAA configuration studied here have been proposed, including combining PIAA with complex amplitude modulating focal plane masks. Regardless of the permutations, whenever the pupil gets remapped to provide apodization, the wavefront control behavior will have the quirks reported here.
ACKNOWLEDGEMENTS

This work was performed at the Jet Propulsion Laboratory/California Inst. of Tech. and NASA Ames Research Center as part of NASA ROSES 09-TDEM09-0017.

REFERENCES