End-to-end coronagraphic modeling including a low-order wavefront sensor

John E. Krist*, John T. Trauger, Stephen C. Unwin, Wesley A. Traub

Jet Propulsion Laboratory, California Institute of Technology,
4800 Oak Grove Dr., Pasadena, CA 91109 (USA)
*E-mail: john.krist@jpl.nasa.gov

ABSTRACT

To evaluate space-based coronagraphic techniques, end-to-end modeling is necessary to simulate realistic fields containing speckles caused by wavefront errors. Real systems will suffer from pointing errors and thermal and motion-induced mechanical stresses that introduce time-variable wavefront aberrations that can reduce the field contrast. A low-order wavefront sensor (LOWFS) is needed to measure these changes at a sufficiently high rate to maintain the contrast level during observations. We implement here a LOWFS and corresponding low-order wavefront control subsystem (LOWFCS) in end-to-end models of a space-based coronagraph. Our goal is to be able to accurately duplicate the effect of the LOWFS+LOWFCS without explicitly evaluating the end-to-end model at numerous time steps.

Keywords: coronagraph, low order wavefront sensor

1. INTRODUCTION

1.1 High contrast imaging of exoplanets

The imaging and characterization of extrasolar, habitable planets requires very high dynamic range observations. As the most obvious example, the brightness ratio (contrast) between the Earth and Sun is 10^{-10} in visible light. At a distance of 10 parsecs (nearby, in astronomical terms), the apparent maximum separation of an Earth twin at 1 astronomical unit would be only 0.1 arcseconds. These circumstances require the use of a space telescope with a means to suppress both the diffracted light (from telescope obscurations) and scattered light (from optical surface errors) that is present in an optical system.

The diffracted light can be reduced using a coronagraph, of which there are many types. Scattered light is suppressed by using one or more deformable mirrors that correct wavefront errors (such as those created by optical polishing errors). The wavefront phase and amplitude are sensed by some means and then the DM actuator pistons that minimize the intensity within the field of interest around the star are computed. To achieve contrasts of 10^{-10}, the optimization algorithm must be more complex than simple phase conjugation that is used for general adaptive optics on the ground, especially to create broadband solutions. The size of the high-contrast region around the star (the *dark hole*) in which scattered light can be controlled by the DMs is limited by the number of actuators that span the beam diameter. Two DMs in series can provide both phase and amplitude control over a circular region centered on the star[1].

1.2 Wavefront stability

It is not enough simply to get the dark hole down to a mean contrast of 10^{-10} with wavefront control and a coronagraph. At that level, any Earth-twin would have the same brightness as, and appear similar to, the average scattered light speckle. This contrast is used as a goal because it means that the planet signal would not be swamped by the shot noise in the background speckles. In order to distinguish the planet from the background, the speckles must be removed, either via spectral filtering (assuming the planet spectrum is known) or by subtracting a reference point spread function. For point sources, the latter may be accomplished by rotating the telescope about the target between exposures. The instrumentally-generated features (the speckles) will remain stationary on the detector while the sky (including the planet) appear to rotate. Various methods can be used to solve for the planet signal based on this diversity[2,3]. However,

this only works if the speckles remain stable to within a fraction of the planet's flux (by convention, ~10% or 10^{-11} contrast).

As we will show via simulation, once a 10^{-10} contrast dark hole has been created, the corrected wavefront phase must be kept stable to within a few picometers (10^{-12} m) RMS (root-mean-square) to maintain the speckle intensity to within 10^{-11} contrast. This is extremely difficult as there are a variety of factors that can cause the wavefront to change over critical timescales (i.e., the science integration time). For instance, the telescope pointing will always be imperfect due to body slew errors and wandering (jitter) caused by vibrations from reaction wheels, gyros, etc. Rapid monitoring of the star's position at the coronagraph's focal plane mask will partially compensate for this, but not completely due to measurement noise, latency errors, and beam walk on non-pupil optics. The optical system will also have thermal variations in the alignments and surface figures caused by attitude changes of the telescope in relation to the Sun. These may come from slewing between targets or by rotating the telescope about the star (for point spread function self-subtraction). There might also be mechanical stresses caused by slewing the spacecraft.

1.3 Low order wavefront sensing and correction

The largest time-dependent wavefront changes one may expect are low order, such as defocus, coma, and astigmatism. These variations must be kept to a few picometers total over the integration time of the science image to maintain the speckle field stability. This requires a means to measure the changes at these levels and compensate for them at intervals more rapid than the integration time. In most coronagraphic systems, the most critical location in the system is the focal plane mask, where all of the light is focused. The wavefront needs to be measured as close to there as possible to capture all of the time-variable components preceding the mask. We require a low order wavefront sensor (LOWFS) that is both capable of measuring wavefront changes at the level of a few picometers at reasonable rates (a reasonable rate is as yet undetermined, as it involves the exact configuration and operating conditions of the telescope, and only approximate estimates can be made using complex finite-element thermal modeling).

1.4 Simulating the effect of a LOWFS

It is useful to have realistically-simulated coronagraphic fields to experiment with planet signal extraction methods and to understand the related statistics. These are computed using end-to-end models of the optical system, with the aberrations from each surface included[4,5]. The wavefront is numerically propagated through the system, picking up aberrations as it encounters each optic. The simulation must also implement deformable mirrors (DMs) for wavefront control and an algorithm to determine the DM actuator pistons that create a dark hole around the central source.

Such simulations are often computed, but in nearly all space-based cases the systems are static. Time-dependent coronagraphic fields are commonly computed for ground-based telescopes, but the wavefront changes are large and the sensing methods are well understood and have already been implemented on real systems.

2. THE MODEL OF THE STATIC SYSTEM

2.1 System definition

In our computational experiments we use the system layout show in Figure 1. It has a realistic design based primarily on the ACCESS study[6]. Light is collected by an off-axis Gregorian telescope and directed via a flat to behind the primary mirror and into the coronagraphic instrument. To provide wavefront control, a 48 × 48 actuator deformable mirror is placed at an entrance pupil image, followed 1.0 meter later by another, similar DM. This dual DM configuration allows both phase and amplitude control within a circular field around the star.

The beam is focused onto a hybrid bandlimited coronagraph (HBLC) mask[7]. This consists of a layered greyscale amplitude transmission pattern and a patterned, phase-modifying coating. Such masks have actually been fabricated using variable-thickness nickel (for amplitude) and cryolite dielectric (for phase) coatings. They have been successfully utilized in the High Contrast Imaging Testbed (HCIT) at the Jet Propulsion Laboratory, where they have reached contrast levels of ~5×10^{-10} in monochromatic light (800 nm) and ~10^{-9} in broadband ($\Delta\lambda/\lambda \approx 10\%$)[8]. In a real system, both coatings would have wavelength-dependent terms that would limit their broadband performance. In our simulations, however, the amplitude and phase modifications are wavelength-independent. The mask has 50% intensity transmission at $r = 3\ \lambda_c/D$ radians (λ_c = 550 nm, D = primary mirror diameter); this corresponds to 28.4 milliarcseconds for a D = 4 m telescope. The pupil is then reimaged onto the Lyot stop, a mask with a circular hole that blocks the

remaining diffracted light at the edge of the beam. At his point, the diffracted light from the star is fully suppressed, leaving only the scattered light. The beam is then refocused onto the detector.

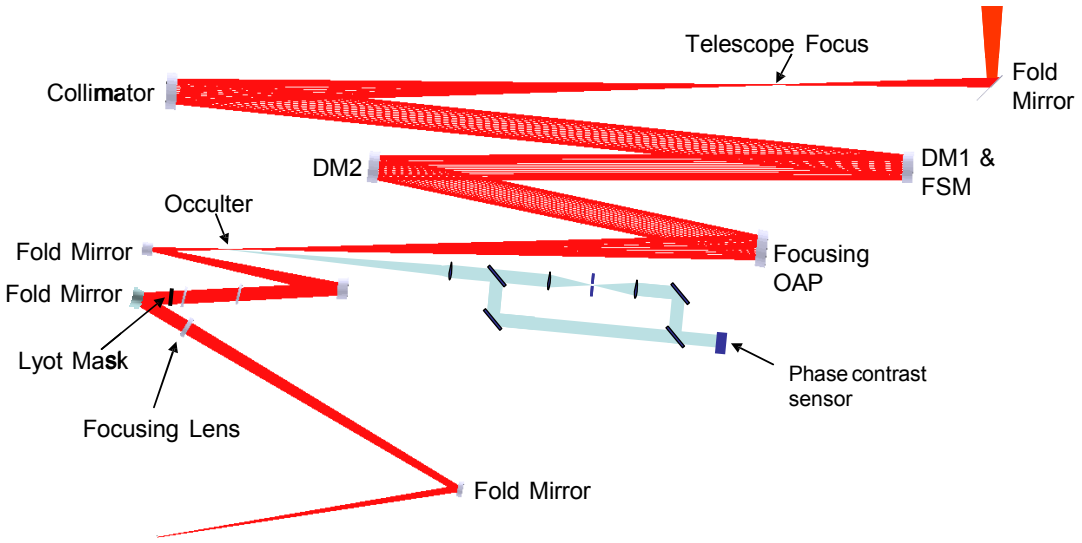

Figure 1. Schematic layout of the back end of the planet imaging telescope, including the Lyot coronagraph and phase contrast sensor, which is fed a portion of the beam reflected by a dichroic coating on the coronagraphic occulter mask.

2.2 Propagation simulation

To compute the broadband ($\Delta\lambda/\lambda \approx 20\%$) performance, the input wavefront is propagated through the entire system at 5 wavelengths spanning the passband and then the intensity images for each are added together. The propagation is done using the PROPER[9] library, which handles both near and far field diffraction. PROPER includes a DM model with actuator surface influence functions derived from a real DM. It can also create synthetic surface error maps (amplitude and phase) from power spectral density specifications. For these simulations, we specified PSDs based on actual optics, with errors included on each optic (amplitude errors only on the primary).

We used computational grid sizes of 2048 × 2048 pixels with a 40% pupil diameter (819.2 pixels). The final image plane sampling was 0.4 λ/D radians at the shortest wavelength (500 nm), slightly better than the Nyquist criterion.

2.3 Wavefront control and generating the dark hole

In a real system, the wavefront errors must be sensed using some calibrated instrument (interferometer) or by modifying the DM with known patterns, measuring the intensity changes at the final image, and deriving the amplitude and phase errors (ref). Because we explicitly compute the field at each wavelength in our models, there is no need to actually simulate sensing it.

Using the computed final-focus wavefronts at each wavelength, the DM actuator pistons that minimize the scattered light in the dark hole are computed using the Electric Field Conjugation (EFC) method[10]. Briefly, EFC relies on knowing how each DM actuator changes the wavefront at the final image plane at each wavelength. This information comprises the DM response matrix, which is computed by poking each actuator and recording its effect. By approximating what in reality is a non-linear system with a linear one, EFC simply solves for the actuator settings that minimize the dark hole field given the measured (computed) wavefronts. This is an iterative process due to the approximation. Note that during the dark hole generation, we assume that the system is static, though in reality it would not be.

The final broadband ($\lambda = 500 - 600$ nm, $\lambda_c = 550$ nm) dark hole shown in Figure 2 has a mean contrast of 10^{-10} over the control region of $r = 2.5 - 18\ \lambda_c/D$ radians (the outer limit is set by the number of DM actuators and the inner by the transmission of the occulter). Our goal now is to keep this field of speckles stable to 10^{-11} contrast when low order aberrations are introduced into the system.

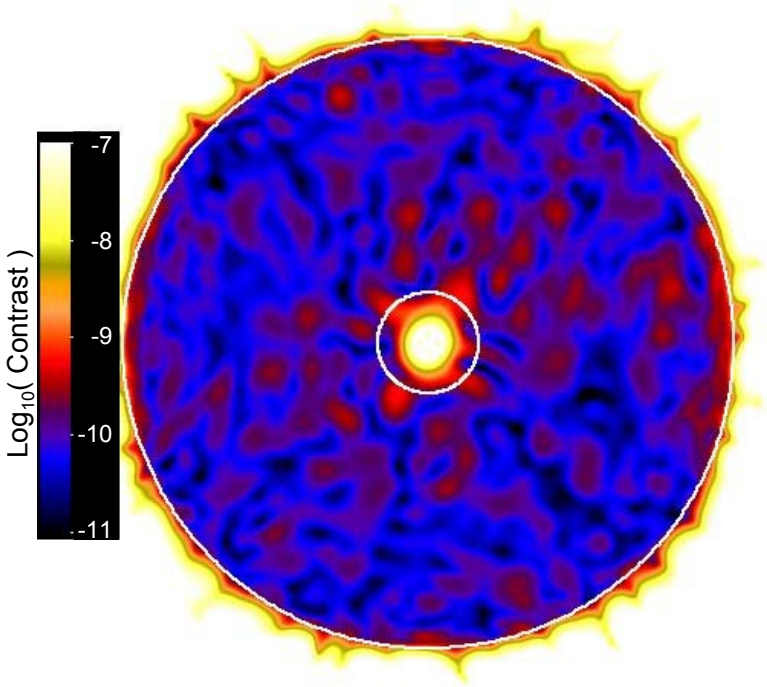

Figure 2. Broadband (λ = 500 – 600 nm) contrast map of the EFC-generated dark hole. The inscribed circles indicate r = 2.5 and 18 λ / D. The color bar indicates $\log_{10}$(contrast). Available in color in the online proceedings.

3. INTRODUCING TIME-DEPENDENT ABERRATIONS

3.1 Computing the wavefront change

The next step is to perturb our up-to-now static system by adding in some low-order aberrations that represent time-dependent variations in the wavefront. To derive realistic aberration estimates one must run finite element modeling software that includes the thermal and mechanical stresses involved with slewing the telescope and changing its attitude with respect to the Sun. To date, very little work on this has been done due to the cost and complexity of such computations. As part of the ACCESS study[6] for a 1.5 m coronagraphic space telescope, just one round of computations was run to derive a wavefront change map for the case where the telescope was rotated by 30º about the boresight and then allowed 12 hours to settle. This map (Figure 3) includes 151 picometers RMS of defocus and 19 pm RMS of other aberrations (mostly coma and astigmatism), for a total RMS wavefront error of 152 pm. We will use this map as a representative example of a wavefront change in the simulations presented here; the exact nature of any wavefront change would depend on the system (size, thermal isolation, active heating, etc.) and the conditions (attitude with respect to the Sun, settling time, etc.).

3.2 Assessing the impact on the dark hole field

Next, we added this wavefront map to the primary mirror in our model system and computed a new broadband field. We did not alter the DM settings or perform any wavefront control to compensate. The result (Figure 4) shows that some portions of the field have increased by nearly magnitude, with some speckles now around the 10^{-9} level. These changes could completely hide any planets if roll subtraction were used. Even when we omit the defocus term and only include the 19 pm RMS of the remaining aberrations there are a significant number of speckles that have changed by more than 5×10^{-11} in contrast (Figure 5). This shows that the wavefront that corresponds to a 10^{-10} contrast dark hole must be kept stable with within a few picometers RMS between science exposures. It is impossible to build a system that can

maintain this stability over the multi-hour timescales of these exposures, so some means of actively measuring and compensating for these changes is necessary.

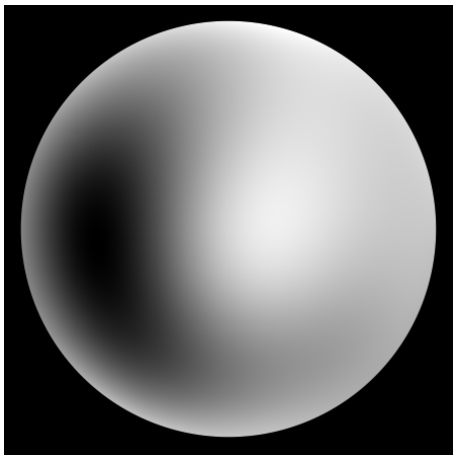

Figure 3. Computed (from the ACCESS[ref] study) thermally-induced wavefront change at a 1.5 m diameter primary mirror after a 30º change in orientation about the target star and after 12 hours of settling time. The change in focus (151 picometers RMS) has been removed to better reveal the remaining aberrations (19 pm RMS).

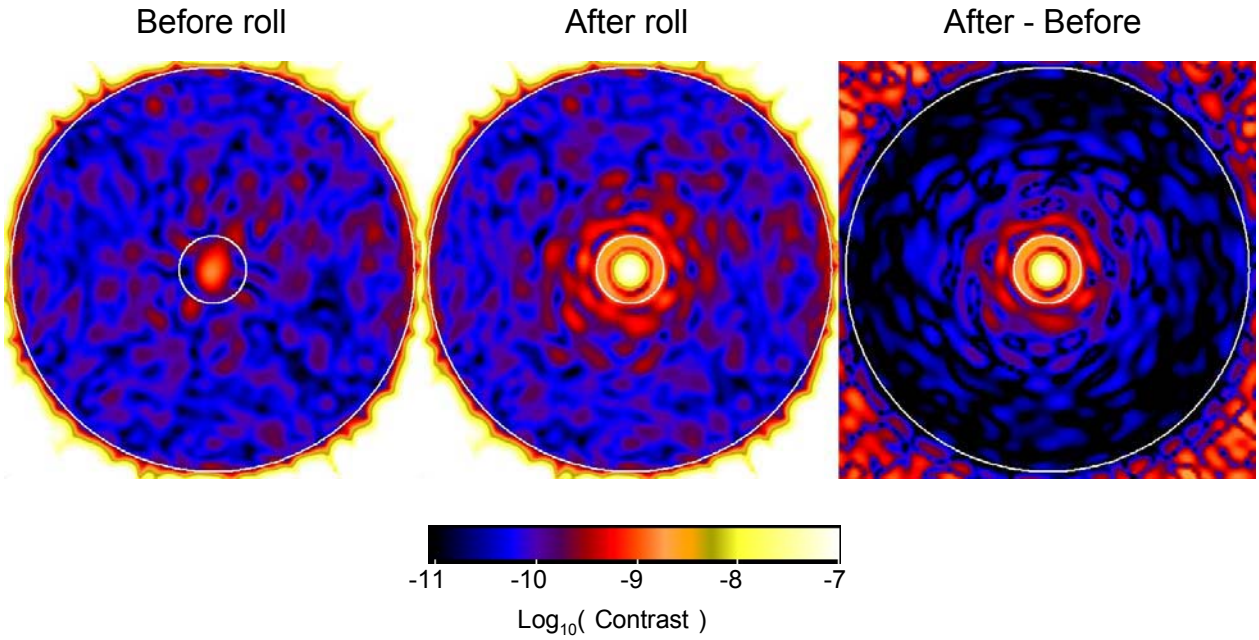

Figure 4. Field contrast before (left) and after (middle) rotating the telescope by 30˚ and waiting 12 hours for settling, without altering the wavefront control solution. The change in contrast is due to the thermal changes in the primary mirror. The right image shows the difference between this and the original dark hole field (a wavefront change of only 152 picometers RMS, 151 pm of which was defocus). Ideally, the entire subtracted region would have speckles no higher than 10^{-11} in contrast to allow easy identification of 10^{-10} contrast planets. Available in color in the online proceedings.

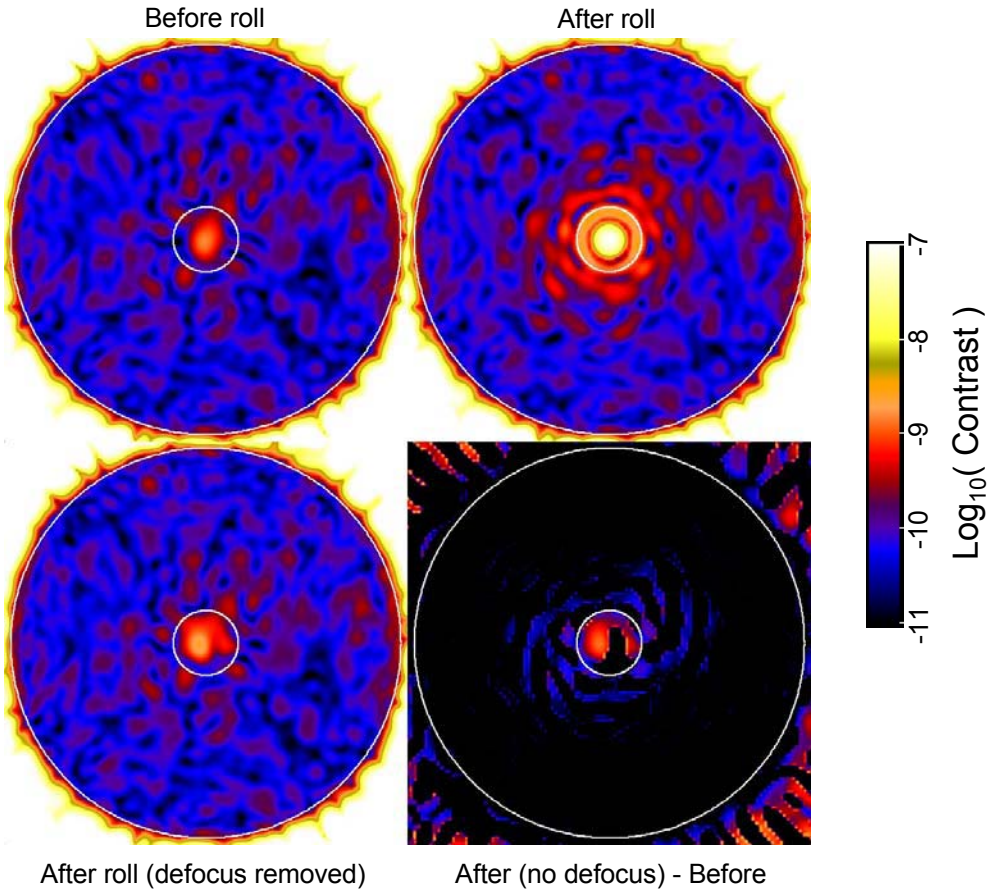

Figure 5. Like Figure 4 above, except the defocus component of the wavefront change has been removed (19 pm RMS of aberrations left). (Left) Field contrast after introducing the wavefront change. (Right) Absolute difference between the dark hole contrast before and after changing the wavefront. This shows that even 19 pm RMS of aberrations are sufficient to introduce changes in the speckle field of almost 10^{-10} contrast (ideally, they would be limited to ~10^{-11}). Available in color in the online proceedings.

4. LOW ORDER WAVEFRONT SENSING AND CONTROL

4.1 Implementation

There are a variety of techniques for sensing low order aberrations, some well tested and implemented on ground-based telescopes, while others are largely theoretical. Ground based adaptive optics typically use Shack-Hartman or curvature sensing schemes (there are also pyramid sensors, which are generally used for pointing corrections). Designed to measure wavefronts that vary by nanometers or microns, these methods are not suited for monitoring changes at the level of a few picometers.

Instead of these traditional techniques, we choose to use a phase contrast sensor[11] fed by the beam reflected off of the coronagraphic focal plane mask. A schematic of this system is shown in Figure 6. We assume that the focal plane mask substrate has a dichroic coating that is designed to transmit a particular wavelength range while reflecting all other

wavelengths. The reflected light can be used for wavefront sensing. This alleviates the need for a separate beam splitter prior to the mask.

The reflected beam is collimated and divided by a beam splitter into reference and sensing beams. The reference beam is focused onto a 0.4 λ_c/D diameter pinhole or fiber that acts as a spatial filter, creating a generally uniform, flat-phase wavefront. It is then collimated and added to the sensing beam via a beam combiner. The path length between the two beams is adjusted to introduce a $\pi/2$ radian offset at the central wavelength (this non-achromatic phase shift does not appear to limit the performance significantly). The combined beam (a pupil image) is then imaged by the LOWFS detector. The addition of the phase shifted reference beam can be viewed as introducing an offset of the imaginary component of the sensing beam's complex-valued electric field so that it now varies both in amplitude and phase, rather than just phase (we are assuming that we are concerned only with changes in the phase errors, as those are expected to have the largest time dependence). The induced amplitude spatial variations are seen as intensity variations superposed on an otherwise uniform pupil image, which can be recorded by a camera and then analytically converted to the equivalent phase errors (Figure 7).

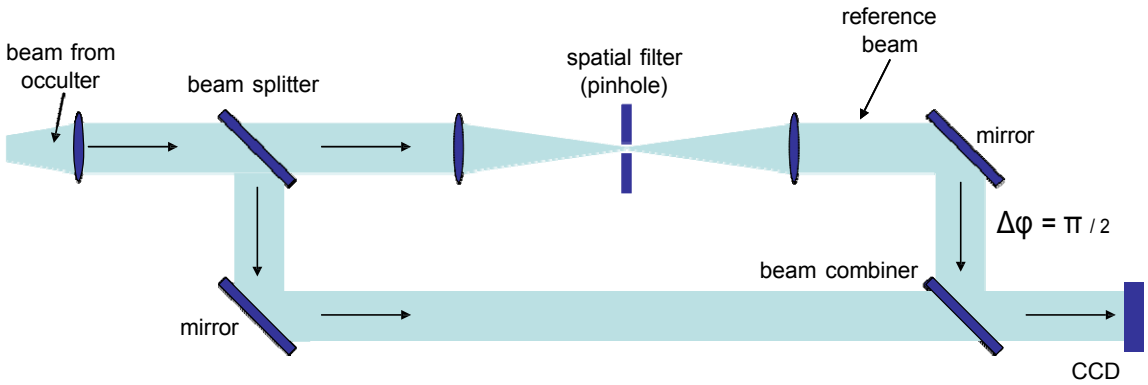

Figure 6. Schematic layout of a phase contrast sensor. The beam reflected from the front surface of the coronagraphic focal plane mask (occulter) enters from the left and is collimated before being split into two beams. The reference beam (top beam) is focused onto a spatial filter (pinhole or fiber) and is then emitted and collimated. It is combined, after a $\pi/2$ shift in phase, with the unfiltered (sensing) beam (lower beam). Available in color in the online proceedings.

We do *not* want to flatten the absolute wavefront but rather just the wavefront *change*. To produce a broadband dark hole within a limited control region, the EFC algorithm intentionally introduces wavefront errors via the DMs that destructively interfere with the errors caused by the optical aberrations. We do not want to disturb this solution by getting rid of these intentional errors. So, rather than measure the absolute wavefront from one LOWFS image, we take two consecutive phase contrast images, difference them, and measure the phase change (lower row of Figure 7). We then add the reverse of this change to the DM to maintain the desired wavefront. This provides the additional benefit that instrumental errors common to both exposures subtract out, including static errors in the LOWFS optics.

The phase difference image is fit with Zernike polynomials. This effectively limits the allowed range of spatial frequencies that are measured on what is an over-constrained dataset. This also allows bad pixels (cosmic rays, hot pixels) to be easily identified and ignored.

4.2 Including the LOWFS in the simulations

To determine the measurement accuracy of the phase contrast system, we compute the wavefronts at the occulter for both the static (before telescope roll) and perturbed (after roll) conditions. Assuming these are the wavefronts reflected by the dichroic on the occulter substrate, we propagate them through two separate optical systems, one representing the reference arm of the LOWFS and the other the sensing arm. The reference beam is focused onto a $r = 0.4\ \lambda_c$/D pinhole that acts as the spatial filter and is then recollimated. The electric fields from both arms (which include the $\pi/2$ phase

shift between them) are then added together and converted to intensity. These steps are done for a number of wavelengths that sample the sensing bandpass, and the results are added together to represent a broadband measurement.

The diameter of the pupil image in the LOWFS, 205 pixels, is optimized via trial and error to produce the most accurate measurements, as is the portion of light diverted into the reference beam (75%). The LOWFS intensity image is then normalized to the computed flux for the assumed stellar type, magnitude, passband, and integration time, accounting for realistic system throughput reductions. Shot noise and read noise (2 e^- per pixel) are then added. The after-before difference is taken and then fit with Zernike polynomials.

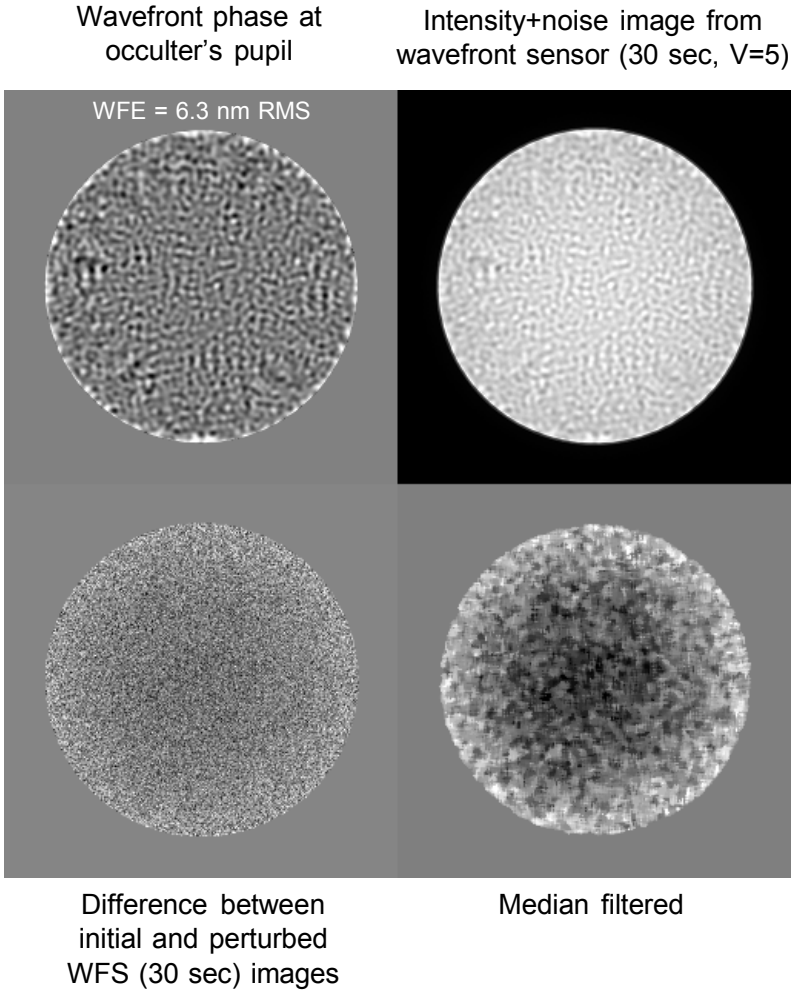

Figure 7. Example measurement with the phase contrast LOWFS. (Upper left) The phase aberrations in the pupil feeding the occulter. (Upper right) The intensity image measured in the LOWFS in a 30 sec exposure on a V=5 star in a 4 m telescope over the V passband (shot & read noise added). (Lower left) The difference between two consecutive LOWFS images, with the roll-induced wavefront errors introduced between the exposures. (Lower right) The difference image filtered to better show the focus change (151 pm RMS).

4.3 Investigating measurement accuracy

To determine how well the phase contrast LOWFS can measure wavefront changes in the presence of noise, we take the same two before-and-after LOWFS images and create separate copies with different normalizations, each representing different LOWFS integration times. Our assumed LOWFS is on a 4 m telescope looking at a V=5 magnitude G2V star

over a λ = 500 – 600 nm passband. We then generate multiple realizations of noisy images, computing the differences, and then fitting Zernikes to them. The derived set of Zernike coefficients for each noise realization can be compared to the known wavefront difference, and this ensemble of these measurements can be used to estimate the mean accuracy and measurement dispersion for each integration time (listed on the left of Figure 8). We can also use these different measurement realizations to compute the corresponding after-roll, low-order-corrected dark holes by adding the negated Zernike polynomials to DM #1. The after-before dark holes then indicate the level of speckle stabilities that can be expected for a given LOWFS integration time (Figure 8).

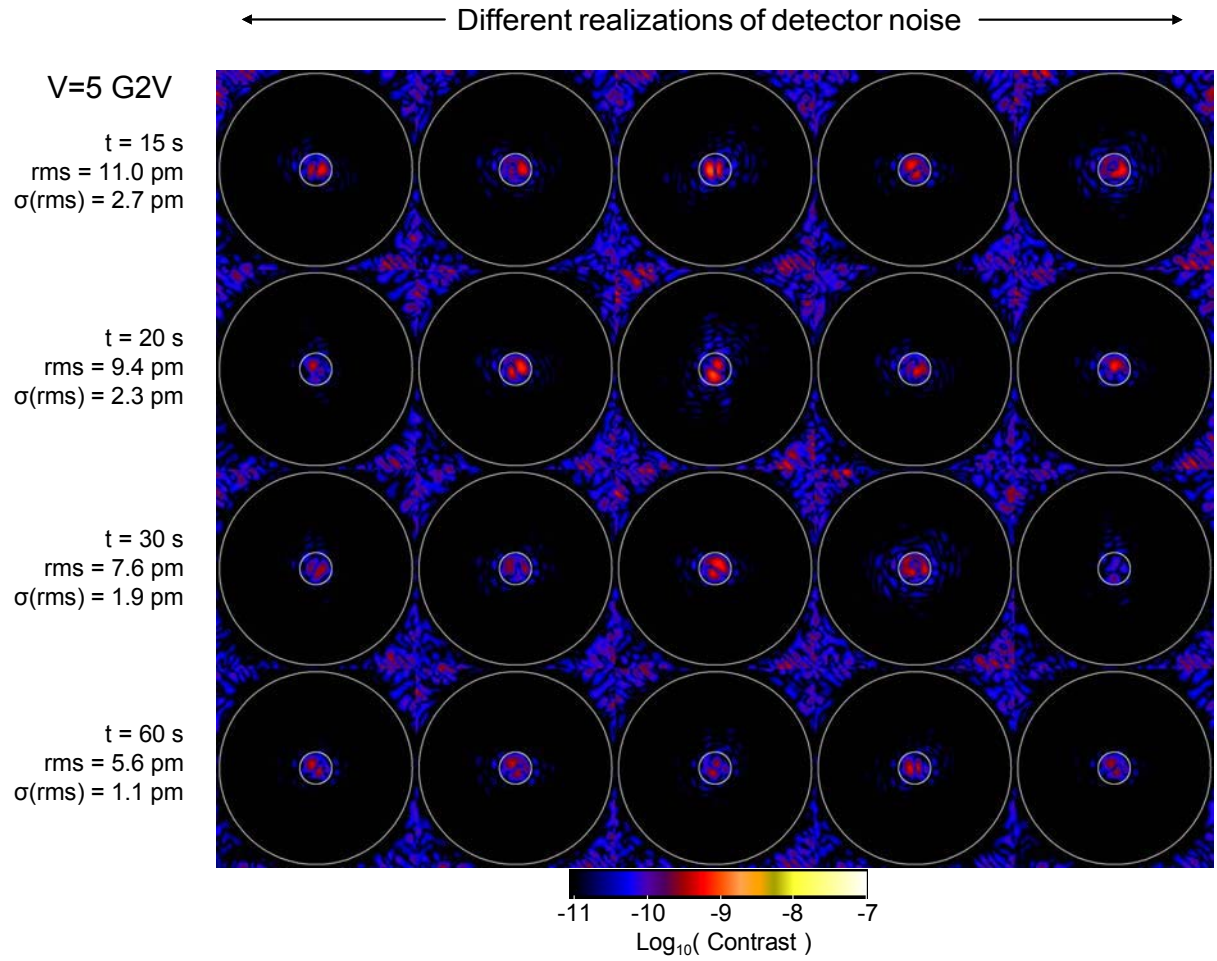

Figure 8. A long-exposure image of the dark hole field around the star is taken by the science camera, and the telescope is then rolled, which introduces the previously-shown aberrations (including defocus). During this time, the phase contrast LOWFS is measuring the wavefront at a specified cadence (shown along the left side). The measured wavefront change is corrected by the DM, then a second science image is taken. The difference of the science camera images for varying assumptions of LOWFS camera integration times (increasing downwards) are shown for different realizations of shot and read noise (along rows). The mean RMS wavefront measurement error is given for each row along with the standard deviation in those measurements. These are for a V=5 star with a 4 m scope in a 500 – 600 nm sensing passband. The region of concern is the annulus between the r = 3 and 18 λ/D radii (inscribed circles); changes inside of 3 λ/D are not of a concern. Ideally, the entire annulus would be black (≤10^{-10} contrast), but these are a tremendous improvement over the uncorrected dark hole (Figure 4). These experiments indicate that a phase contrast sensor can be successfully used to keep speckles stable to ~10^{-11} contrast over 30-60 sec timescales, limited mostly by LOWFS camera noise. Available in color in the online proceedings.

4.4 Caveats

Our simulations cannot be considered exhaustive. We do not include any possible non-common path errors due to time-dependent changes in the LOWFS itself. The LOWFS presented here does not operate at a sufficient rate to also provide rapid pointing adjustment to maintain the position of the star on the occulter. For that, a separate sensing system is needed. Our simulations assume that all of the light from the telescope is going through the system; in reality, the coronagraph may be sensitive to polarization-dependent aberrations, in which case the beam from the telescope must be split into two polarization channels, each with its own DMs, coronagraph, LOWFS, etc. The reduction in throughput due to this is not included in our accuracy-versus-time calculations.

5. CONCLUSIONS

Our modeling indicates that a Zernike phase contrast system could theoretically be used as a low order wavefront sensor to measure wavefront changes with an accuracy of 5 – 8 picometers RMS on a V = 5 star in a 4 m telescope with a cadence of 60 – 120 seconds. This accuracy may generally be scaled proportional to the square root of the exposure time or telescope area. This accuracy is sufficient to ensure that speckles in a 10^{-10} mean contrast, wavefront-corrected coronagraphic field should not vary by more than $1 – 2 \times 10^{-11}$ over a long (>1000 sec) science exposure. This stability would allow the use of roll subtraction to remove the residual speckles and reveal planets of comparable brightnesses. If one assumes that LOWFS measurement accuracy is the limiting factor for ensuring speckle stability, it may be possible to compute different realistic realizations of expected speckle fields without having to explicitly compute the wavefront changes through finite element thermal modeling.

ACKNOWLEDGEMENTS

The research described here was carried out at the Jet Propulsion Laboratory, California Institute of Technology, for the NASA Exoplanet Exploration Program under contract with the National Aeronautics and Space Administration.

REFERENCES

[1] Shaklan, S., & Green, J., "Reflectivity and optical surface height requirements in a broadband coronagraph: 1. Contrast floor due to controllable spatial frequencies", *Applied Optics*, 45, 5143 (2006).

[2] Lafreniere, D., Marois, C., Doyon, R., Nadeau, D., Artigau, E., "A new algorithm for point-spread function subtraction in high-contrast imaging: A demonstration with angular differential imaging", *Astrophysical Journal*, 660, 770 (2007).

[3] Krist, J., Stapelfeldt, K., Bryden, G., et al., "HST and Spitzer Observations of the HD 207129 Debris Ring", *Astronomical Journal*, 140, 1051 (2010).

[4] Krist, J., Belikov, R., Pueyo, L, Mawet, D., Moody, D., Trauger, J., Shaklan, S., "Assessing the performance limits of internal coronagraphs through end-to-end modeling: a NASA TDEM study", *Proc. SPIE*, 8151, 81510E (2011).

[5] Krist, J., et al., "End-to-end simulations of different coronagraphic techniques", *Proc. SPIE*, 7440, 744016 (2009).

[6] Trauger, J., Moody, D., Gordon, B., Krist, J., Mawet, D., "A hybrid Lyot coronagraph for the direct imaging and spectroscopy of exoplanets systems: recent results and prospects", *Proc. SPIE*, 8151, 81510G (2011).

[7] Moody, D., Gordon, B., Trauger, J., "Design and demonstration of hybrid Lyot coronagraph masks for improved spectral bandwidth and throughput", *Proc. SPIE*, 7010, 70103P (2008).

[8] Trauger, J., et al., "ACCESS – A concept study for the direct imaging and spectroscopy of exoplanetary systems", *Proc. SPIE*, 7731, 773128 (2010).

[9] Krist, J., "PROPER: an optical propagation library for IDL", Proc. SPIE, 6675, 66750P (2007).

[10] Give'on, A., Kern, B., Shaklan, S., Moody, D., Pueyo, L., "Broadband wavefront correction algorithm for high-contrast imaging systems", *Proc. SPIE*, 6691, 66910A (2007).

[11] Bloemhof, E., and Wallace, J., "Phase contrast wavefront sensing for adaptive optics", *Proc. SPIE*, 5553, 159 (2004).