

PROCEEDINGS OF SPIE

[SPIDigitalLibrary.org/conference-proceedings-of-spie](https://spiedigitallibrary.org/conference-proceedings-of-spie)

Numerical modeling of the Habex coronagraph

John Krist, Stefan Martin, Gary Kuan, Bertrand Mennesson, Garreth Ruane, et al.

John Krist, Stefan Martin, Gary Kuan, Bertrand Mennesson, Garreth Ruane, Navtej Saini, John Trauger, James Breckinridge, Dimitri Mawet, Philip Stahl, Jeffrey Davis, "Numerical modeling of the Habex coronagraph," Proc. SPIE 11117, Techniques and Instrumentation for Detection of Exoplanets IX, 1111705 (9 September 2019); doi: 10.1117/12.2530462

SPIE.

Event: SPIE Optical Engineering + Applications, 2019, San Diego, California, United States

Numerical modeling of the Habex coronagraph

John Krist^a, Stefan Martin^a, Gary Kuan^a, Bertand Mennesson^a, Garreth Ruane^a, Navtej Saini^a, John Trauger^a, James Breckinridge^b, Dimitri Mawet^b, Philip Stahl^c, Jeffrey Davis^d

^aJet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA;

^bCalifornia Institute of Technology, Pasadena, CA, USA; ^cNASA Marshall Space Flight Center, Huntsville, AL, USA; ^dUniversity of Arizona, Tucson, AZ, USA

ABSTRACT

The Habex study, commissioned by NASA in preparation for the 2020 Decadal Survey, is evaluating a 4 meter space telescope for high contrast imaging and spectral characterization of extrasolar terrestrial planets. Its off-axis configuration, active structural metrology, and low-disturbance pointing control provide an optimal system for coronagraphs. We present predictions of the Habex performance using a charge 6 vortex coronagraph that have been obtained using numerical modeling techniques developed for the WFIRST coronagraph. The models include realistic optical surface and polarization-induced aberrations, pointing jitter, and thermally-induced wavefront variations. Wavefront control using dual deformable mirrors is simulated to create a dark, high-contrast hole around the star. The results show that current technologies can closely approach the Habex performance goals, and with some additional development in key areas (e.g., deformable mirror surface quality, low-polarization coatings, etc.) over the next few years they should reliably meet them.

Keywords: Habex, coronagraph

1. INTRODUCTION

In preparation for the 2020 Astronomy and Astrophysics Decadal Survey, NASA commissioned concept studies of four flagship-level space telescopes covering infrared to X-ray wavelengths. One of these is the Habitable Exoplanet (Habex) observatory¹, a 4 meter-diameter space telescope for finding and characterizing Earth-like planets having visible-light planet/star brightness contrast ratios of $\sim 10^{-10}$. It is optimized for high-contrast coronagraphy from 450 – 1800 nm using an off-axis design that avoids diffraction-producing obscurations. It maintains high optical stability with robust monitoring and control of structural deformations.

Habex observatory and systems

The Habex telescope² uses an off-axis Cassegrain layout with a 4 m diameter monolithic primary mirror (Figure 1). This configuration provides a clear, unobscured wavefront that avoids the additional diffracted light that would be generated in an on-axis layout by the shadows of a secondary mirror and its supports. This is the ideal configuration for coronagraphy, allowing for a larger variety of designs and with higher performance than would be possible on obscured telescopes like Hubble, James Webb, or WFIRST. The telescope assembly consists of the primary, secondary, and tertiary mirrors. The tertiary, located below the secondary and adjacent to the primary, directs the beam into the coronagraph, which lays along the side of the telescope tube (always in shadow from the Sun). The tertiary also feeds adjacent general astrophysics instruments, such as ultraviolet and near-infrared cameras and spectrographs. The near-normal incidence angle of the beam onto the tertiary and up along the side of the barrel avoids the generation of large polarization cross-terms that would reduce coronagraphic performance. If the coronagraph were instead placed behind and along the back of the primary (the WFIRST coronagraph configuration), a $\sim 45^\circ$ fold would be needed that would increase the cross-terms.

In any coronagraphic system optical stability is critical to maintain a static speckle field that the speckles can be reduced using post-processing methods. Habex utilizes laser metrology to sense alignment changes caused by thermal variations,

desorption of water and gases from the truss, and slew-induced stresses between the three telescope optics. These are corrected using actuators on the secondary and tertiary. Vibrations in the system are minimized by using microthrusters instead of reaction wheels, though a fast steering mirror (FSM) is included as part of the low-order wavefront and control system in the coronagraph.

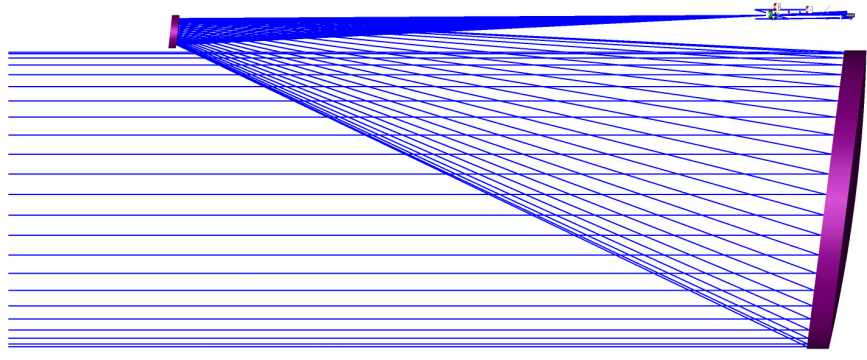


Figure 1. Layout of the Habex telescope. The coronagraph is shown in the upper right, aligned along the side of the system.

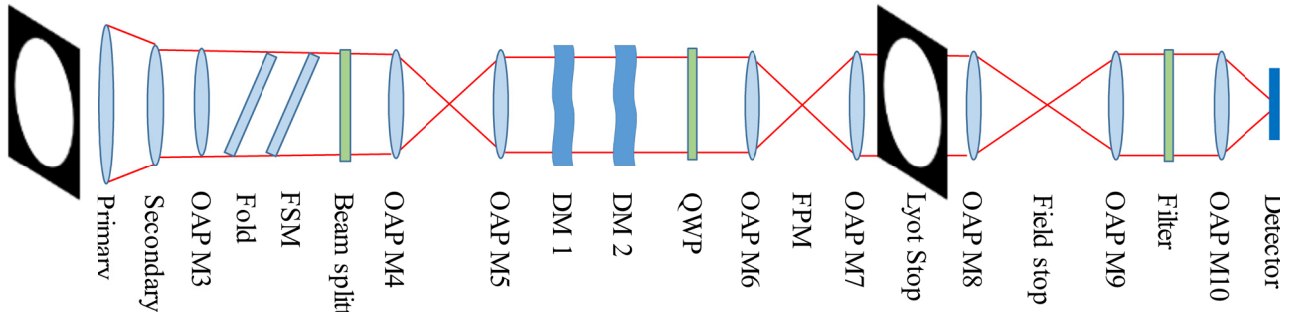


Figure 2. Unfolded layout of the Habex telescope + coronagraph (one polarization channel shown).

Habex coronagraph

In an unobscured system like Habex, a coronagraph suppresses the diffraction of starlight caused by the edge of the primary. The Habex layout (Figure 2) supports Lyot-type coronagraphs with masks at an intermediate focus (focal plane mask or FPM) and at a subsequent pupil image (Lyot stop). Aberrations caused by misalignments and optical surface errors that scatter light into the field are corrected using two 64x64 actuator deformable mirrors (DMs) located in the collimated beam prior to the FPM, one at a pupil and the other some distance downstream. The two mirrors in sequence provide control of both amplitude and phase aberrations. The FSM is located at the entrance of the coronagraph to correct body pointing errors.

The baseline starlight suppression method chosen for Habex is the vortex coronagraph (VC). The VC consists of an azimuthally-phase-modulating FPM that introduces a helical phase delay about the optical axis and a simple clear-aperture Lyot stop. The primary geometric parameter of a VC is its charge, which is the integer multiples of 2π radians of phase change it introduces over a 360° azimuth (the charge must be even-valued). The inner working angle (IWA), which is the angular distance from the star at which the FPM transmission is 50%, becomes larger as the charge increases. A charge 4 VC has an IWA of $1.7 \lambda/D$ while a charge 6 is $2.4 \lambda/D$ (unless otherwise specified, λ is the central wavelength of the bandpass). As the charge increases, the sensitivities to low-order aberrations are reduced. The charge 6 vortex is the default in Habex, providing a compromise between IWA and aberration sensitivity. The vortex FPM can be physically implemented in a variety of ways, either as a physical spiral surface structure (a scalar vortex coronagraph) or as a polarizing coating deposited with an azimuthally-varying polarization axis (a vector vortex coronagraph, VVC). In the VVC there is often a small, opaque spot less than a λ/D in diameter deposited at the center of the FPM to mask any fabrication errors; in Habex, which assumes a VVC, the spot diameter is $0.6 \lambda/D$. The Habex Lyot stop's clear diameter is 95% of the geometric pupil diameter.

When used together with wavefront control (WFC), the VVC must operate in a single polarization to achieve the best possible contrast. The Habex coronagraph uses a beam splitter located just after the FSM to create two orthogonally-polarized beams that feed separate but nearly-identical coronagraphs, each with their own DMs, masks, spectrographs, and detectors. Both coronagraphs have visible-wavelength cameras and spectrographs, while one of them has a selectable mode to observe in the near-infrared. They can be used to simultaneously observe a target in two separate polarizations over two different (visible/IR) or identical (visible) bandpasses. We will only discuss the operation of a single-polarization coronagraph at visible wavelengths here.

The Habex layout provides the option to use other designs besides the VC, as long as they adhere to the Lyot coronagraph layout of an FPM and Lyot stop. Early in the Habex study, for instance, the Hybrid Lyot Coronagraph (HLC) was evaluated, a version of which is the baseline coronagraph design for WFIRST. It had a phase-and-amplitude modulating FPM with an IWA of $2.5 \lambda/D$ and a simple Lyot stop. It used the DMs as part of the diffraction suppression (whether the system is aberrated or not); on WFIRST, the DMs are used to compensate for the telescope obscurations, while on Habex they altered the wavefront to optimize the performance of the FPM. Early modeling analyses indicated that the HLC was much more sensitive to low-order aberration changes compared to the VC, so it was designated as a possible backup design and will not be discussed further here. Other designs such as PIAA (Phase-Induced Amplitude Apodization), shaped pupil, or APLC (Apodized Pupil Lyot Coronagraph) would require additional optics not in the baseline Habex layout.

Photon-counting detectors would be used for a mission like Habex, given the low flux rates from the exoplanets. The modeling results presented here do not include any detector effects.

Modeling the Habex coronagraph

Nothing like Habex has been built before. Both HST and WFIRST are smaller (2.4 meter), on-axis space telescopes, neither of which was optimized for coronagraphs. The planet/star contrast ratios of $\sim 10^{-10}$ (compared to $>10^{-9}$ for WFIRST) require wavefront error stabilities and corrections measured in the single-digit picometers range. Effects such as polarization aberrations, slew-induced vibrations, truss desorption, and others may introduce significant limits on performance that are not present at less demanding contrasts. A primary purpose of the Habex study is to identify those properties (e.g., polarization, DM quality) of such a system that are critical to achieving the required performance and determine the associated technologies (e.g., coatings, DM surface improvement) that need to be advanced over the next decade³. This depends critically on numerical modeling, given that testbed experiments can only evaluate what technologies exist today.

The Habex modeling described here encompasses two categories: (1) diffractive modeling of the combined telescope-and-coronagraphic system defined by static optical aberrations (surface fabrication and polarization aberrations), producing an image of the field of speckles surrounding the star, and (2) time-resolved finite element modeling of the system as it undergoes thermal changes due to solar attitude variations, resulting in optical instabilities that impact the ability to reduce the level of speckles via post-processing techniques.

Nearly all of the modeling presented here utilized code and techniques developed in previous studies and mission proposals, including TPF-C, Exo-C, and the JPL High Contrast Imaging Testbed, but mostly the WFIRST coronagraph⁴. The WFIRST modeling software, including diffractive propagation, wavefront control, and thermal/structural modeling, was used with little modification except for incorporating the Habex layout and coronagraphic masks.

NOTE: The term *contrast* is used in this document to define the relative brightness of a pixel of residual starlight in the coronagraphic field or the relative brightness of a planet. When used to describe a planet's intensity, contrast is the observed flux of the planet divided by the observed flux of the unocculted star. When used to describe field brightness at a given location (x,y), contrast is the intensity of the starlight in pixel (x,y) divided by the peak pixel value if the star were offset to (x,y). By this convention, a field pixel of contrast *c* would have the same brightness as the peak of a planet of contrast *c*.

2. MODEL DEFINITIONS

Diffractive propagation models

Wavefront propagation

The three-dimensional Habex optical layout was implemented in Zemax and then ray traced to provide chief ray propagation distances between the optics and their effective focal lengths. These defined an unfolded version of the system that could be implemented using the freely-available PROPER optical propagation software library^{5,6}, which has versions for IDL, Python, and Matlab. PROPER applies Fresnel and angular spectrum algorithms to propagate the diffractive wavefront from optic to optic, picking up the effects of surface errors along the way. It also has model DMs that can be used for wavefront control. PROPER was used to simulate the two-dimensional coronagraphic field from the primary mirror to the final focus. Broadband models were created by adding monochromatic image generated at wavelengths spanning the camera's bandpass.

Surface errors

As was done for modeling the WFIRST coronagraph⁴, each telescope and coronagraph optic was represented by a map of surface errors (from figuring and polishing) appropriate for the type of optics (e.g., flat or off-axis parabola); see Figure 3. For the smaller optics (neither the primary nor secondary) suitable synthetic error maps were generated based on wavefront error (WFE) measurements of similar optics, namely those used in the Gemini Planet Imager (GPI). Measured surface error maps of multiple GPI flats and off-axis parabolas (OAPs) were decomposed into low-order errors (Zernikes up to 5th-order spherical) and the remaining mid-spatial frequency ones, from which power spectral density (PSD) curves were derived. For each optic type, flat or OAP, these provided a typical range of values for each Zernike and an average 2-D PSD (which turns out to have an $f^{3.5}$ power law, where f is spatial frequency). The synthetic map for each Habex coronagraph optic was then generated by first building a low-order error map by computing a random value for each Zernike within its measured range for that optic's type. This was then added to the mid-frequency map that was generated using the PSD with random phases. Given the orders-of-magnitude deeper contrast of Habex, we assumed default errors of 1/2 of GPI's, still comfortably within the range of modern fabrication capabilities (extreme UV lithography optics are nearly an order of magnitude better and much more expensive). The default errors for the flats were approximately (low/mid/total WFE) 0.5/0.5/0.7 nm RMS and 1.6/0.7/1.7 nm RMS for OAPs.

There was a much larger uncertainty on what errors to assume for the primary and secondary. A high-quality, 4 meter-class, light-weighted, space-worthy, off-axis optical system has never been built, so error parameters had to be derived from other large optics and some educated assumptions. In the end, parameters were derived from measurements of a WFIRST-like set of optics, which despite having only a 2.4 meter on-axis primary, represent the closest match in light-weighting and recent fabrication techniques available to the author. Like the GPI optics, a representative set of Zernike aberrations were derived as well as PSDs, which follow the typical $f^{2.5}$ power law of most large telescope optics. From

these synthetic maps were created. A guess of the low-order gravity release errors in the primary were also added. As with the smaller optics, we opted to assume $\frac{1}{2}$ the error of the measured optics, still keeping within the realm of current fabrication capabilities. Our primary map errors were (low/mid/total WFE) 11.5/5.6/12.6 nm RMS and 4.6/3.5/5.7 nm RMS for the secondary.

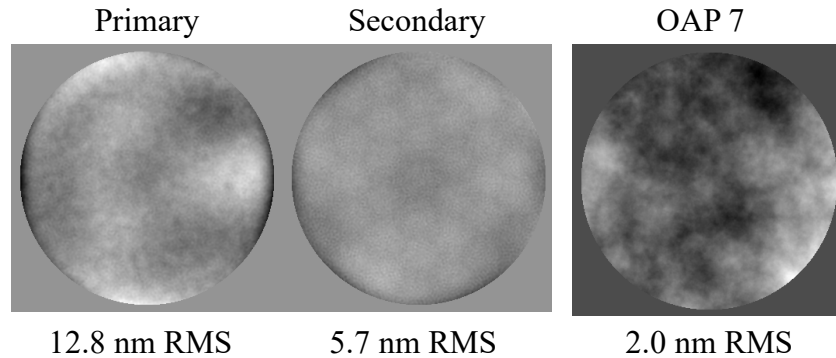


Figure 3. Synthetic optical surface error maps used in the Habex model of the primary, secondary, and a representative off-axis parabola. The RMS wavefront errors of each map are provided. The primary and secondary maps are shown at the same intensity scale. Each map is shown spatially scaled to the illuminated diameter.

Polarization-dependent aberrations

Unpolarized light entering the telescope can be decomposed into incoherent orthogonal components (X_{in}, Y_{in}). These two beams each encounter what appear to be effectively different surfaces when incident on an optic, introducing different aberrations in each polarization. These are dependent on the surface gradient of the optic and the coatings. During reflection or refraction, a small fraction of the light in one polarization may be rotated into the orthogonal direction, creating cross-term components, (X_{in}, Y_{out}) and (Y_{in}, X_{out}), in addition to the “pure”, direct ones, (X_{in}, X_{out}) and (Y_{in}, Y_{out}). Any surface encountered at non-normal incidence will create additional cross-terms, and the greater the gradient across the beam, the larger the amplitude of the cross-term. While the cross-terms have significantly lower amplitudes than the direct ones, they have as large if not larger phase errors, and with significantly different spatial distributions. The direct terms account for the large majority of the light prior to the coronagraph, about 99.99% in Habex.

In a system without a polarizer, the incoherent sum of the four components is observed. Unpolarized wavefront sensing will essentially measure the mean of the two direct terms. Correcting this portion of the wavefront error using the DMs, the remaining uncorrectable WFE in each of the direct terms is half of the difference in WFE between the two plus the cross-terms (which have now been amplified due to the additional WFE introduced into them by correcting the direct terms). If a polarizer is used at the back end, then one direct and one cross-term is removed (e.g., (Y_{in}, Y_{out}) & (X_{in}, Y_{out})). In this case, the remaining direct term, (X_{in}, X_{out}), can be corrected almost fully with the DMs, leaving the cross-term, (Y_{in}, X_{out}), which may dominate in the resulting image. In reality, the polarization WFE is dependent on wavelength and the DMs cannot fully correct errors over a broad bandpass.

At the contrast levels the Habex coronagraph operates, polarization is a significant contributor to WFE and must be included in the simulations. The WFE at the pupil prior to the FPM was computed for each polarization component using Zemax, which included the material properties of the coatings that impact polarization (Figure 4). For verification, a cross-comparison was done of polarization-induced WFE computed using Zemax, Code V, and Polaris-M (an in-house program at U. Arizona) with the same prescription and materials, producing very similar results (the largest differences were actually due to how each program interpolated the tables of material properties). As described elsewhere, input polarizations of $\pm 45^\circ$ were used with 0° (X) and 90° (Y) outputs in order to produce four smooth output wavefronts with similar total intensities that could be well characterized with Zernike polynomials. These can easily be resized as needed.

and can be converted to direct and cross-term components if necessary. Since cross-term phase maps often have singularities, they are difficult to directly resize as needed to arbitrary beam samplings.

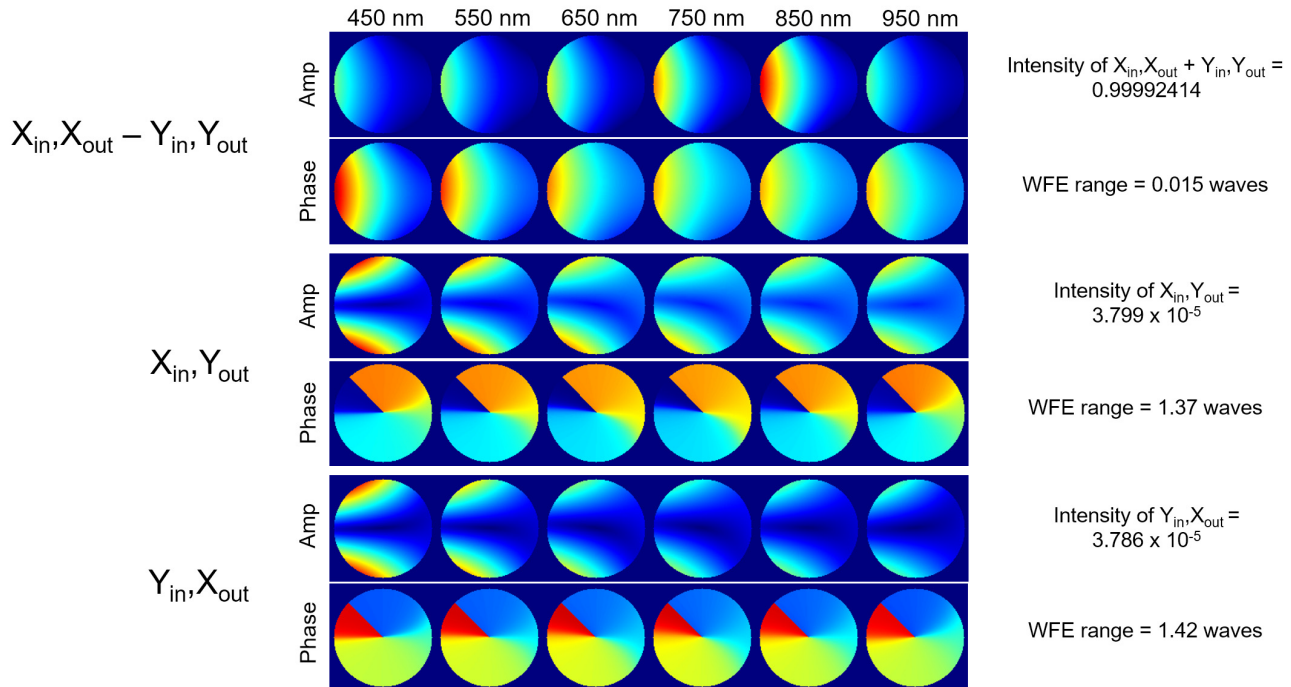


Figure 4. Habex polarization-induced wavefront errors versus wavelength computed using Zemax for the system layout and coatings. The cross-terms (X,Y and Y,X) are shown along with the differences between the direct (X,X and Y,Y) terms.

Vortex coronagraph

The representation of the VC is somewhat more complicated than for other coronagraphs due to the need to sample the vortex phase spiral and occulting mask at high resolution near the center. Failure to do so can lead to orders of magnitude worse performance than expected. The vortex spiral pattern is first generated in a large array at a sampling 7x that of the final representation. It is then 7x7 binned so that the values represent the mean modulation over the area of a pixel, and the result is converted to phase and stored as a complex-value array for repeated use. For this study a perfectly achromatic vortex was assumed. Vector vortex masks designs exist utilizing multiple coatings to allow for sufficient performance over a broad bandpass.

The wavefront was sampled by 1000 pixels across the beam and was propagated to the plane of OAP6 located in the collimated beam space after the DMs and prior to the vortex. That OAP's aberrations are then applied, but not its quadratic powered-optic phase term. The wavefront is then back-propagated to DM1, located at a pupil (called here the *DM1 field*). Up to this step, for efficiency, the propagations use a 2048 x 2048 pixel grid, which would provide a sampling of $1000/2048 = 0.488 \lambda/D$ at a focus. At this stage we need to propagate to focus at somewhat higher sampling ($1000/4096 = 0.244 \lambda/D$) in the outer region ($r > 15 \lambda/D$) and much higher sampling ($0.025 \lambda/D$) inside of it to sufficiently capture the modulation of the PSF by the vortex. To do this we step outside of the PROPER prescription and pad the wavefront to 4096 x 4096 pixels and then propagate it to the focus at the plane of the vortex using an FFT, resulting in a sampling of $0.244 \lambda/D$. This field is then multiplied by a cosine-tapered inverse window of $r = 15 \lambda/D$ to mask the inner region, multiplied by the previously-generated vortex phase function, and propagated back to the pupil with another FFT (what we call the *low-resolution field pupil*). The original DM1 field is separately propagated to focus

at $0.025 \lambda/D$ resolution using a Matrix Fourier Transform (MFT). A cosine-tapered window, reverse of the other, is applied to mask the field outside of $15 \lambda/D$. The vortex phase function is applied as well as a circular $r = 0.3 \lambda/D$ opaque mask. The field is then propagated back to the pupil (the *high-res field pupil*) using an MFT. The low-res and high-res field pupils are then added together and trimmed back to a 2048×2048 pixel grid. The field is then reinserted into the PROPER prescription and propagated again from DM1 to OAP6, where OAP6's power phase term is finally applied. The beam is then propagated to the OAP after the FPM and on to the Lyot stop.

Wavefront control

The wavefront control procedure used in these Habex simulations was the same as those for the WFIRST coronagraph models. The PROPER model of Habex was used to represent the actual system with all of its optics. The beam was propagated from the primary, through the coronagraph, and to the final image plane, with modifications by the DMs along the way. With 62 actuators across the beam, the maximum “good” dark hole size (the outer working angle, OWA) would be $\sim 25 \lambda/D$ over the 20% bandpass, but for practical reasons and to ensure the best contrast, the simulations limited it to $15 \lambda/D$. Prior to any wavefront control, the aberrated system's mean dark hole region contrast was 3×10^{-5} (Figure 5). The initial wavefront control step was to measure the phase error at the Lyot stop plane (a pupil) for the central wavelength without any coronagraphic masks in place. This provided an estimate of the large and low-spatial-frequency errors in the system at the FPM, approximating what would be derived in reality using phase retrieval. DM1, which is at a pupil, was then set to “flatten” these errors, reducing the dark hole contrast to 4×10^{-7} .

Electric Field Conjugation (EFC)⁷ was used to derive DM settings to reduce the dark hole intensity. Inputs to EFC are the measured complex-valued electric field (E-field) values inside the dark hole at wavelengths sampling the science bandpass and a response matrix (the Jacobian) that relates the predicted E-field change for each DM actuator at each sensing wavelength. A least-squares solver was used to determine the DM settings at multiple iterations. In reality, the E-fields would be sensed in sub-band filters by using specific DM patterns (probes) and the Jacobian, but to save time the computed E-fields were used instead at 9 wavelengths spanning the 20% bandpass. The Jacobian was computed using a simplified, “compact” model of the system that did not have all of the optics and used simple Fourier transforms to propagate between pupil and image planes, using Fresnel propagations only between the DMs. Regularization was used in EFC to damp non-linear responses of the system. After running EFC, the dark hole contrast in the system with only optical surface errors and the X polarizer aberrations was 3×10^{-11} ,

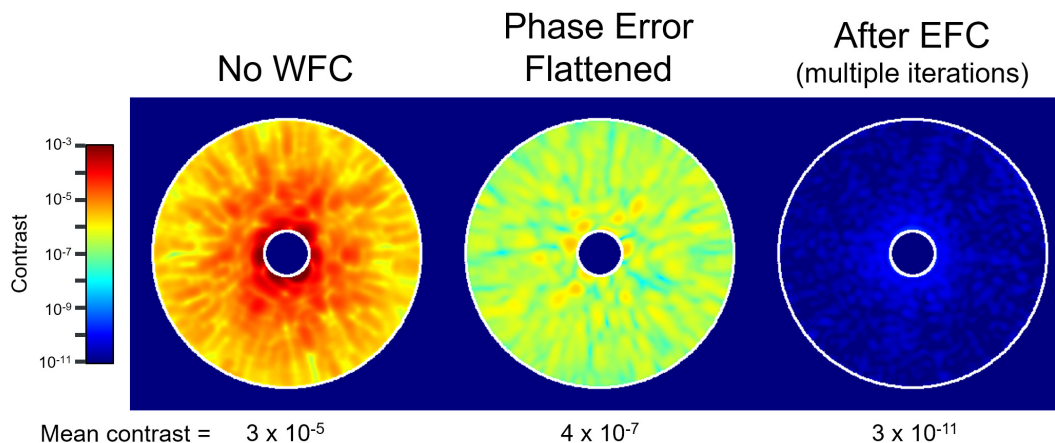


Figure 5. Simulated Habex coronagraph X polarization channel dark hole (450 – 550 nm) for the aberrated system (optical surface errors & polarization) before wavefront control, after flattening, and after EFC. The circles are $r = 2.5$ & $15 \lambda/D$.

Finite-diameter stars and pointing jitter

In a telescope as large as Habex, nearby stars can be partially resolved, potentially degrading contrast as they represent the incoherent summation of pointing offsets defined by the circular disc of the star. Pointing jitter (that portion uncorrected by the FSM) is also a set of incoherently averaged pointing offsets with a Gaussian weighting distribution. Both of these effects can be simulated by generating images at a finely-spaced grid of pointing offsets and then summing them together with the appropriate weights. Due to the low-vibration environment of Habex, pointing jitter is not a major factor, but stellar diameters may be.

STOP models

The diffractive model provides an estimate of the dark hole brightness for an instant of time given specific wavefront errors. The system is not, however, exactly stable, and thermal variations caused by changes in solar attitude will alter the optics, and hence the wavefront, over time. This impacts the stability of the speckle pattern within the dark hole, which limits the ability of post-processing methods such as reference differential (RDI) or angular differential (ADI) imaging to subtract the speckles to reveal any planets or disks. The time variations are predicted using finite-element modeling (structural/thermal/optical performance, or STOP) in which the observatory structure is meshed into a myriad of points with realistic material properties (e.g., CTE).

The primary time variation in Habex is the orientation of the observatory with respect to the Sun, which is defined over time by an observing scenario (OS). The OS begins with an initialization by pointing at some location on the sky for a long (day or two) period. The telescope is then slewed to a reference star, usually much brighter than the science target star, on which the dark hole is dug or tuned and reference images taken to enable RDI. The scope is then slewed to the target star that is observed at multiple roll orientations to enable ADI. The thermal response of each point to the solar angle at each time step is computed (using Thermal Desktop) and its corresponding change in position (using NASTRAN) recorded. The positional offsets of the optics are translated by Sigfit into perturbations of the Code V optical prescription, through which ray tracing computes the low-order wavefront changes. In the Habex study, the WFIRST coronagraph integrated modeling pipeline⁸ was used which automates this sequence.

3. ANALYSES

Polarization

The VVC requires a single polarization to work effectively, so Habex has the necessary polarizing beam splitters and quarter-wave plates to allow this, splitting the two orthogonal polarizations into separate coronagraphs. The X channel is presented here, though the performance should be similar in the Y one. The X channel sees the direct (X_{in} , X_{out}) polarization component along with the cross-term (Y_{in} , X_{out}). The primary aberrations in the direct term are tilt (because the system is off-axis) and astigmatism. Because these vary slowly over the bandpass they are easy to correct with wavefront control, and the limiting direct term contrast is set by the optical surface errors. The cross-term, however, is essentially impossible to correct with wavefront control because any attempt to do so would immediately degrade the direct term solution, which has orders of magnitude more energy. The post-EFC simulations show that the cross-term dominates the dark hole contrast, especially near the IWA, when only optical surface and polarization aberrations are included. The contrast limit due to the system aberrations by themselves is represented by the direct term contrast, which is very low, indicating some margin may be possible for surface errors.

Given that care has already been applied in defining the Habex layout to avoid large angle of incidence optics and symmetrizing the system, there are no easy fixes for reducing the cross-terms. Perhaps some small reductions may be made by large changes to the layout as a whole. Advances in coatings and polarization compensators may also hold hope. This is one technological gap that needs addressing in the next decade.

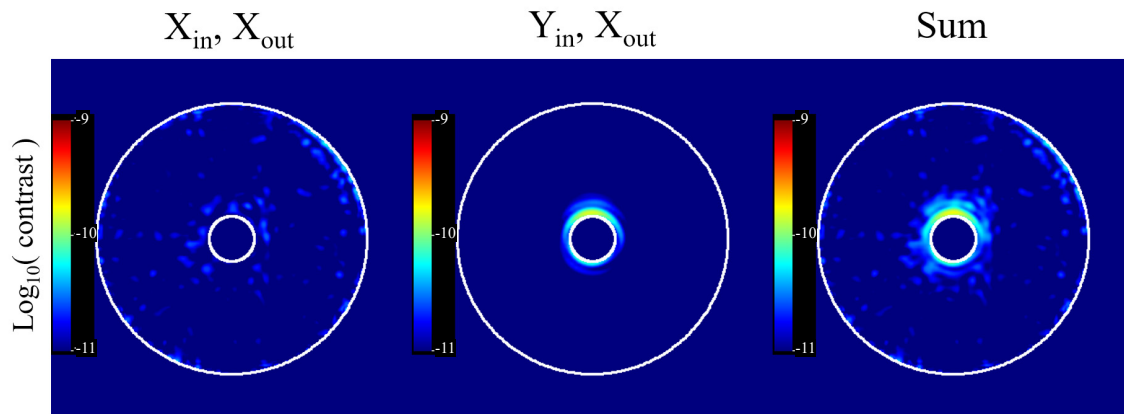


Figure 6. Post-EFC dark hole with polarization and optical surface errors for the X polarization channel (450-550 nm) split into its direct and cross-term components. The direct term is limited by the optical surface errors while the cross-term cannot be corrected by wavefront control. The circles are $r = 2.5$ & $15 \lambda/D$.

Deformable mirror surface errors

There are currently two high-actuator-density DM technologies that are viable candidates for the Habex coronagraph. The first is the electroceramic DM produced by Northrup Grumman (Xinetics). It has an actuator pitch of 1 mm, which is a bit larger than optimal for Habex, which would like to have a large number of actuators cross the pupil (>90) to provide larger fields. One issue with these DMs is that they are thermally sensitive, which would require some type of real-time on-orbit monitoring and control such as an interferometer to keep stable. Their optical surface quality is high, however, because they have a thick facesheet attached that can be polished to the quality of a good flat, with mid-frequency errors of <5 nm RMS. There are no expectations that their surface quality is of concern. These have been used for years in the High Contrast Imaging Testbed at JPL and have provided the highest contrast results so far.

The other technology is the MEMS DM produced by Boston Micromachines (BMC). It is an integrated component with 0.3 – 0.4 mm actuator pitch, allowing for smaller beam sizes for a given actuator/beam sampling. Purportedly thermally stable (but with some recent evidence of instability after shake tests), their drawback for extreme high-contrast coronagraphy is that their integrated facesheet is only 3 microns thick and not polishable, and there is considerable surface structure at the sub-actuator level. These induce wavefront errors that may be an issue.

Figure 7 shows a surface height measurement of a recent BMC DM. Within the region of one actuator the wavefront error (2x surface) varies by ~ 145 nm peak-to-valley, with narrow ridges near the actuator edge and a central depression (where the post pushes and pulls on the facesheet). The RMS wavefront error of ~ 15 nm of the variations is deceptive in terms of coronagraphic impact given that most of the error is concentrated within a small total area.

The grid of replicated actuator patterns generates strong diffraction spots with significant wings in the image plane at spacings equal in λ/D to the number of actuators across the beam (in the case of the Habex simulations, 62). The narrowness of the error structures diffract light from these spots far into the wings. The dark hole region is contaminated with light from these errors. Figure 8 shows the dark hole when errors are present only on DM1 or DM2. DM1 appears to create slightly higher speckles (lower left of Figure 8), but since DM1 is at a pupil, these are phase-error-generated speckles. This pattern smoothly grows in size with increasing wavelength while simultaneously weakening. Such errors are easy to fix with wavefront control, especially using two DMs. The DM2 speckles (lower right of Figure 8) look slightly weaker. However, as wavelength increases the speckles tend to “boil” more than grow in size as a group, a result of surface errors on DM2 generating chromatic amplitude errors. This is not surprising given that DM2 is purposefully located away from a pupil in order for phase modulations (actuator pokes) to propagate into corrective amplitude modulations. It is simply inconvenient that, for the same reason, unwanted errors on DM2 result in chromatic speckles that are difficult to fix with wavefront control.

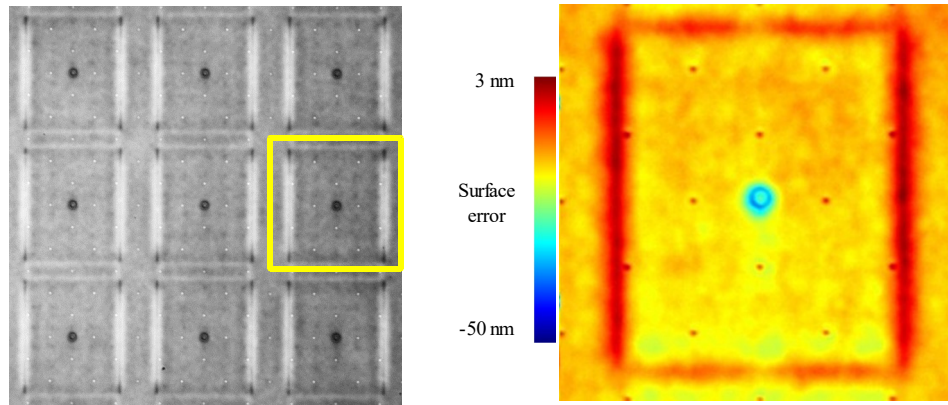


Figure 7. Scan of the surface height variations on a Boston Micromachine's 50x50 actuator MEMS DM. (Left) A 3x3 actuator subregion of the DM. (Right) A single actuator, 0.4 mm on a side. The small dots are actually holes in the facesheet used in the etching process. The central depression is caused by the post that actuates the facesheet.

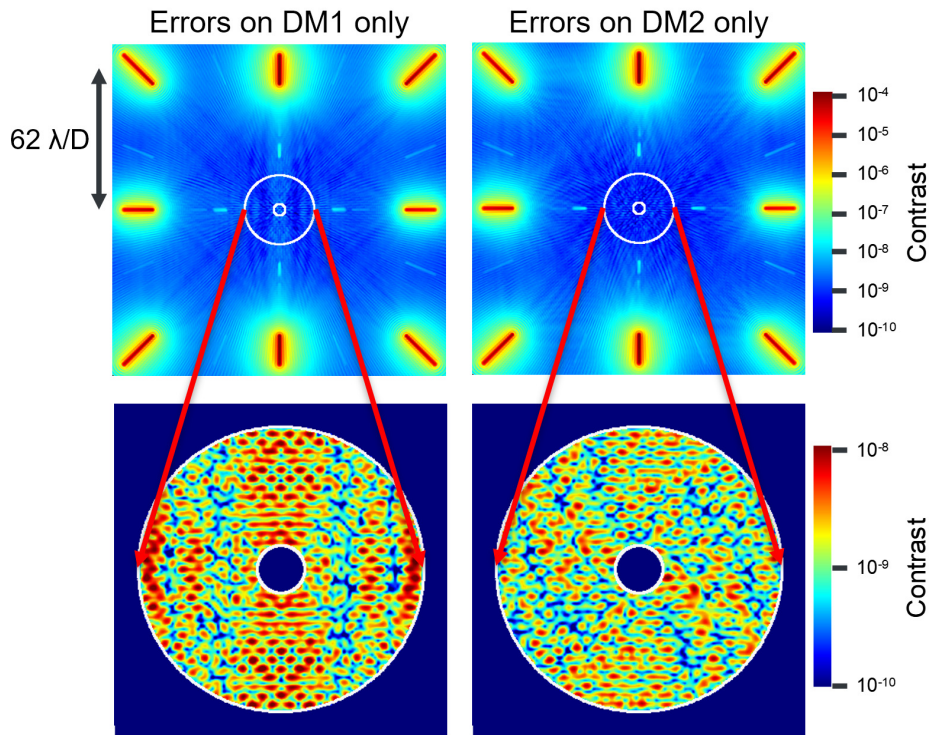


Figure 8. Image plane fields in the Habex coronagraph in the absence of any aberrations except on either (left) DM1 only or (right) DM2 only. In the top row are large-angle views of the final image plane (450 – 550 nm, no WFC). The actuator grid generates spots at $62 \lambda/D$ separations, corresponding to the 62 actuators across the beam. The spots are elongated due to the 20% bandpass. The bottom row shows the light within the dark hole region at a stronger contrast stretch (circles are $r = 2.5$ & $15 \lambda/D$) at 450 nm (no WFC).

Since these surface errors occur on a scale smaller than an actuator, it is impossible for the DMs to directly correct them. Instead, one must rely on the ability of EFC to find a solution where combinations of DM patterns cause interference between speckles in a manner that nulls the light within the dark hole region. To investigate this, EFC was run in a Habex system in which only the BMC surface errors existed. The Jacobian was computed including these errors, so it represents a somewhat unrealistically ideal control case. As shown in Figure 9, when errors exist only on DM1 (but with two DMs used to correct), and thus generate phase-error speckles, the contrast level is $<10^{-12}$. However, when both DMs have surface errors, the post-EFC dark hole mean contrast is above 10^{-10} and even more than 10^{-9} at the IWA, despite repeated attempts to find solutions by adjusting EFC regularization. This demonstrates the difficulty in trying to correct the chromatic amplitude errors from DM2. The impact of the DM2 errors can be reduced by shrinking the DM separation from the default of 0.64 m to 0.32 m. This provides less distance for the phase errors on DM2 to propagate relative to a pupil (DM1) and turn into amplitude errors. This is opposite of the convention of increasing DM separation to gain more authority to control amplitude errors elsewhere in the system.

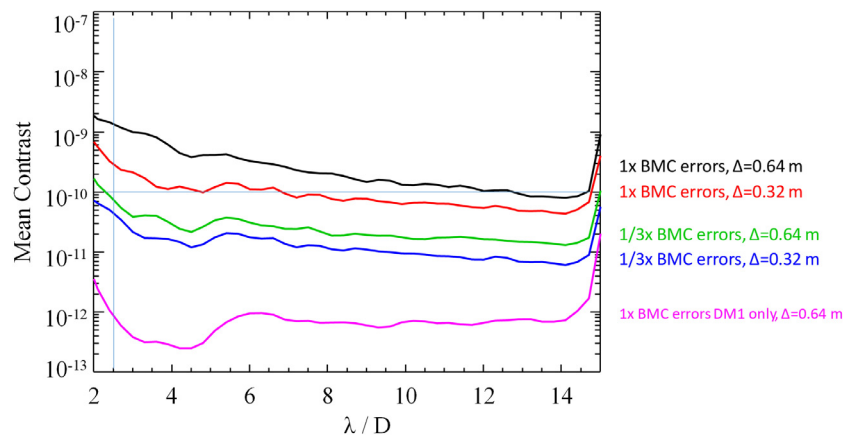


Figure 9. Post-EFC mean radial contrast plots for the Habex coronagraph (450 – 550 nm) in the presence of surface errors on DM1 only (bottom curve) or errors on both DMs. No other errors are present in the system. The separation between DM1 and DM2 (Δ) are specified.

The DM surface errors are problematic by themselves, but in reality they will be used in a system with other aberrations that cannot be fully corrected and will magnify their effect. As shown in Figure 10, when the reduced BMC errors are added with 0.32 m DM separation to the system with polarization and optical surface errors, the speckles increase substantially, especially the polarization cross-term component.

Figure 11 shows the post-EFC dark holes in the presence of certain combinations of errors. The case with current-level BMC DM errors is especially degraded. Figure 12 shows these and more cases as radial contrast curves. With the reduced BMC errors and 0.32 m DM separation along with polarization and optical surface errors, EFC can get the contrast at the IWA to 10^{-10} . Once a 1 mas diameter star is added to the solution (Figure 13), the contrast at the IWA nears 3×10^{-10} , the Habex contrast requirement. It would require 0.2 mas RMS of jitter to reach 2×10^{-10} , an unlikely amount given the low-vibration conditions of Habex.

The mean dark hole contrast, including at the IWA, only reaches below 10^{-10} when the BMC errors are reduced to 1/3 their current level. Because the facesheet is too thin to polish these errors down this much, it will be necessary to modify the fabrication process. BMC has shown examples of non-functional DM surfaces with this much reduction, but it has yet to do so on functional devices. This is another technology development step that is needed to achieve a 10^{-10} contrast mission.

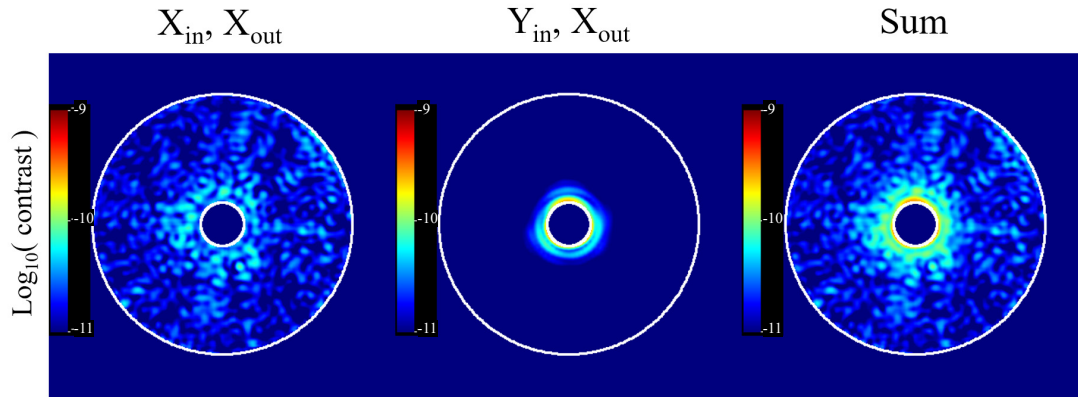


Figure 10. Post-EFC dark hole with polarization, optical surface, and 1/3 BMC DM (0.32 m separation) errors for the X polarization channel (450-550 nm) split into its direct and cross-term components. This is to be compared to Figure 6 that lacks the BMC errors. The circles are $r = 2.5$ & $15 \lambda/D$.

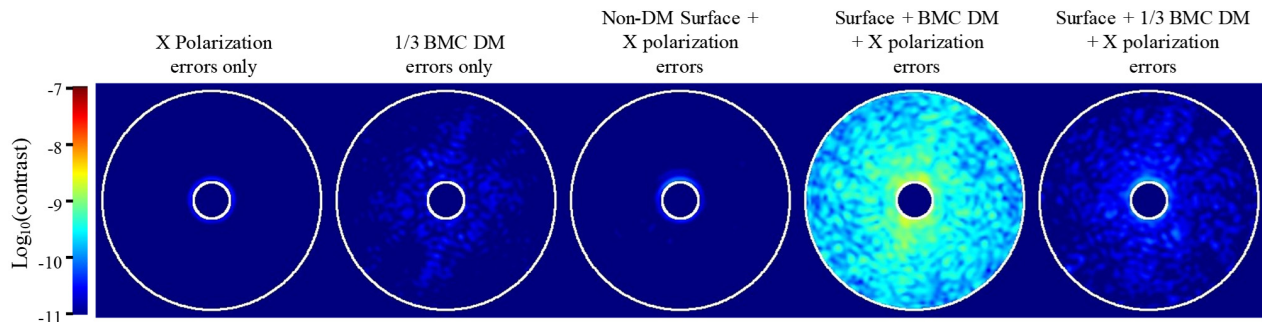


Figure 11. Post-EFC Habex coronagraph (450 – 550 nm) contrast maps for a variety of aberrations present in the system. The circles are $r = 2.5$ & $15 \lambda/D$.

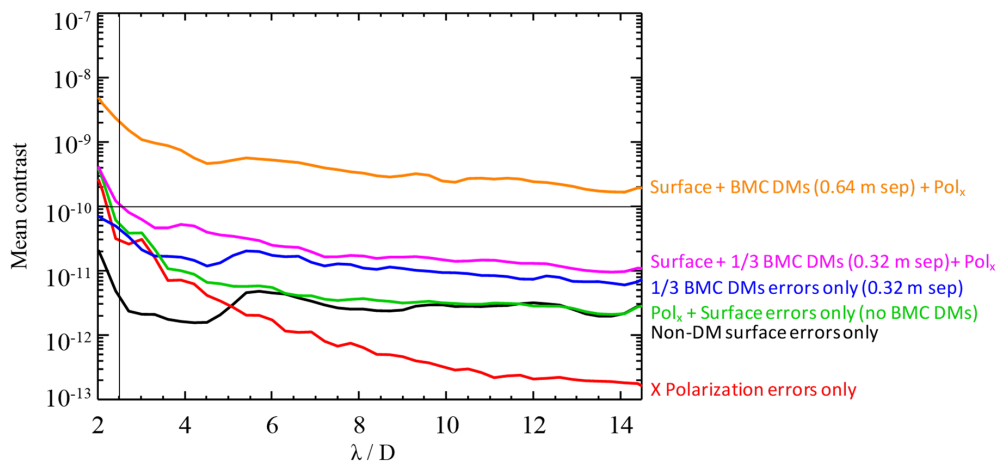


Figure 12. Post-EFC Habex coronagraph (450 – 550 nm) radial contrast curves for a variety of aberrations present in the system.

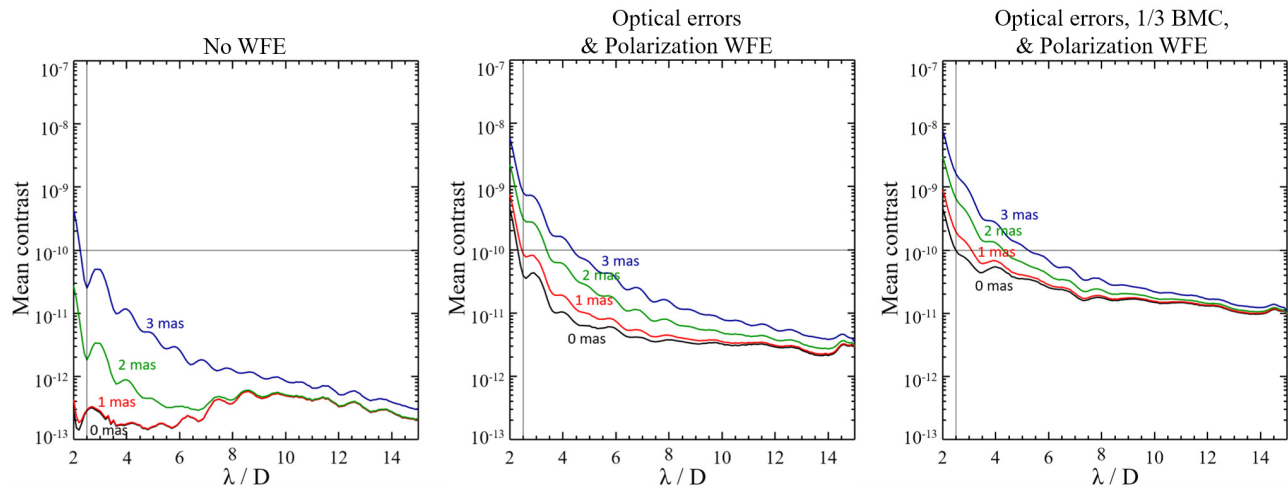


Figure 13. Mean radial contrast curves versus stellar diameter for the Habex coronagraph (450 – 550 nm, X polarization). (Left) With no system aberrations; (middle) with optical surface errors and polarization using an EFC solution; (right) with optical surface, polarization, and BMC DM errors using an EFC solution. A rule of thumb is that x amount of RMS jitter is equal to $4x$ mas stellar diameter in terms of contrast.

STOP modeling

A simple observing scenario was developed as inputs to a Habex STOP model. It begins with the telescope pointed 100° away from the Sun and spending 90 hours there while the model reaches steady state. The scope is then pitched to 110° for 10 hours (a stand-in for reference star observations needed for reference differential imaging) followed by a pitch to 120° for 50 hours (the target star). Without changing pitch, the scope is rolled about the line of sight to the target by $\pm 15^\circ$, spending 50 hours at each roll, representing the type of observations required for angular differential imaging.

The corresponding solar angles and time steps defined the inputs to the WFIRST Integrated Modeling Pipeline⁸ using a Habex finite-element model and prescription. The pipeline automatically runs Thermal Desktop, NASTRAN, SigFit, and Code V in sequence, producing a set of wavefront errors and beam shears (displacement at the FSM) versus time. Such models take days to compute. One factor in the Habex STOP model that is not in WFIRST's is the monitoring of the primary-secondary-tertiary alignment using laser metrology and correction by actuators on the secondary and tertiary.

The variations in the low-order errors versus time are shown in Figure 14. All of the variations are below a few picometers RMS over the span of tens of hours. Note that these do not include sensing and control using the low-order wavefront sensor that is part of the Habex coronagraph but not modeled here. The pupil shift at the FSM is shown in Figure 15. The shift is important because it results in the telescope aberrations shearing against the solution on the DM, which would degrade contrast. These aberration and shear variations are about two orders of magnitude less than those for WFIRST (which is not nearly as well optimized for wavefront stability and thermal isolation as Habex).

To examine the impact these small levels of changes make to the speckle field, the largest change over time for each aberration and shear were added to the post-EFC model (X polarization, optical surface, & 1/3 BMC errors). This represents a contrived worst case (at least to the fidelity of the model). The contrast changes are below 10^{-11} everywhere (Figure 16).

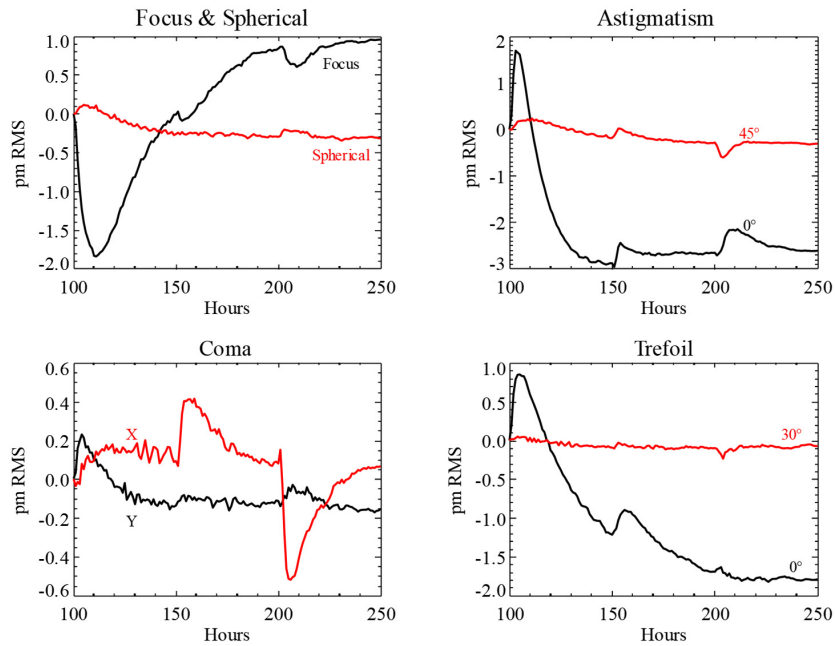


Figure 14. Low-order aberration changes versus time computed by STOP modeling for Habex over the period of the three roll orientations (50 hours each) of 0° , -15° , and $+15^\circ$.

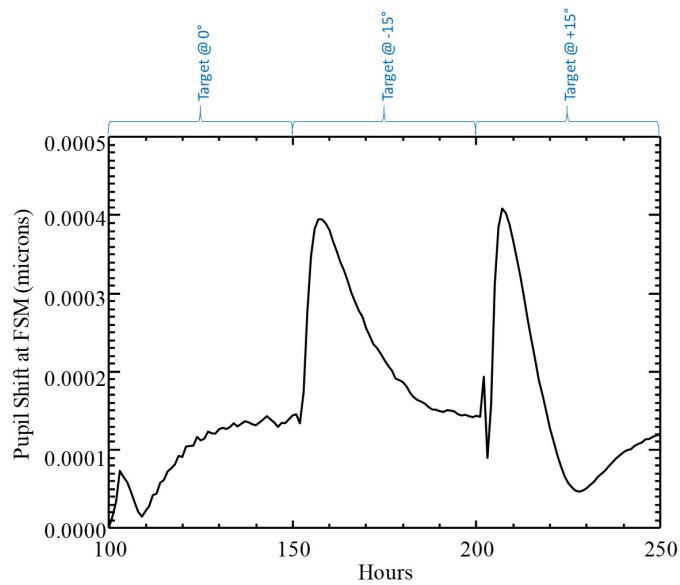


Figure 15. Shift of the pupil relative to the FSM versus time computed by the STOP model for Habex.

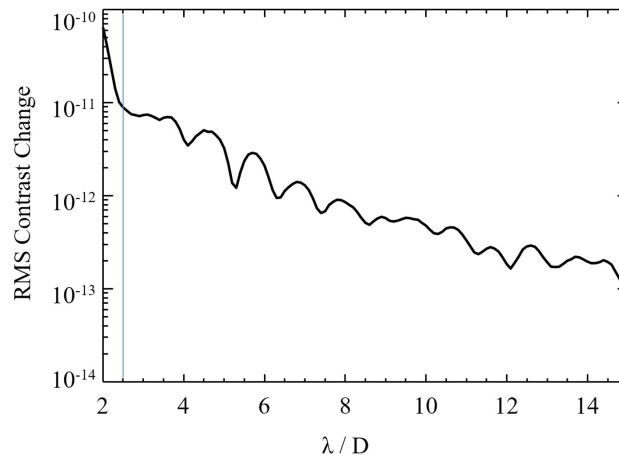


Figure 16. Radial contrast change for the contrived worst-case conditions from STOP modeling versus the default state.

4. SUMMARY

The Habex study was tasked to identify technological advances needed over the next decade to support a mission capable of finding and characterizing Earth-like planets that have contrasts of 10^{-10} . Critical components to achieving such extreme contrasts are optical surface quality (ultrasmooth optics), wavefront control (DMs, algorithms), and wavefront stability (thermal control, actuation, vibration avoidance). These require numerical modeling to evaluate given that testbeds can only judge the state of current technologies.

Using modeling tools developed in prior coronagraphic studies and missions, especially WFIRST, detailed simulations of the Habex telescope and coronagraph were computed with realistic optical surface and polarization aberrations, including the use of wavefront control (EFC). These results highlighted the current state of optical fabrication relative to the desired goals of the Habex telescope, namely

- The optical surface errors from current figuring and polishing methods are likely low enough, though the effects of extreme lightweighting and gravity release need additional study.
- The impact of polarization errors is currently dominated by cross-terms. These might be reduced by carefully specifying layouts, especially reducing optics counts (difficult for coronagraphs given the need to have multiple pupils and focus planes). More work is needed to develop polarization-reducing coatings and compensators.
- The actuator surface errors on Boston Micromachines DMs are currently too high to support 10^{-10} contrasts and need to be reduced by at least a factor of 3. Xinetics DMs do not have these large errors, but have known thermal responses that would require on-orbit monitoring.
- The Habex system has a high level of stability as estimated from STOP models. The use of microthrusters rather than reaction wheels avoids the introduction of pointing jitter, leaving finite stellar diameters as a major incoherent contrast contributor.

Using WFIRST as a technological stepping stone in regards to development, testing, and modeling, the Habex study defines a good path toward advancements that can be achieved in the next decade.

ACKNOWLEDGEMENTS

This work was performed at the Jet Propulsion Laboratory/California Institute of Technology under contract to NASA. The decision to implement WFIRST will not be finalized until NASA's completion of the National Environmental Policy Act (NEPA) process. This document is being made available for information purposes only.

REFERENCES

- [1] Gaudi, S., et al., "The Habitable Exoplanet Observatory," Proc. SPIE, 11115-21 (2019).
- [2] Martin, S., et al., "Habitable Exoplanet Observatory (HabEx) telescope and optical instruments," Proc. SPIE, 11117-4 (2019).
- [3] Morgan, R., et al., "Technology maturity update for the Habitable-zone Exoplanet Imaging Observatory (HabEx) concept," Proc. SPIE, 11115-22 (2019).
- [4] Krist, J., Nemati, B., and Mennesson, B., "Numerical modeling of the proposed WFIRST-AFTA coronagraphs and their predicted performances," JATIS 2(1), 11003 (2016)
- [5] Krist, J., "PROPER: an optical propagation library for IDL," Proc. SPIE 66750P (2007)
- [6] <http://proper-library.sourceforge.net>
- [7] Give'on, A., et al. "Broadband wavefront correction algorithm for high-contrast imaging systems," Proc. SPIE 6691, 6691A (2007).
- [8] Saini, N., Anderson, K., Chang, Z., Gutt, G., and Nemati, B., "IMPipeline: an integrated STOP modeling pipeline for the WFIRST coronagraph," Proc. SPIE 10400, 1040008 (online presentation; 2017).