Vortex coronagraphs for the Habitable Exoplanet Imaging Mission concept: theoretical performance and telescope requirements

Garreth Ruane
Dimitri Mawet
Bertrand Mennesson
Jeffrey Jewell
Stuart Shaklan
Vortex coronagraphs for the Habitable Exoplanet Imaging Mission concept: theoretical performance and telescope requirements

Garreth Ruane, a, b Dimitri Mawet, a, b Bertrand Mennesson, b Jeffrey Jewell, b and Stuart Shaklan b

a California Institute of Technology, Department of Astronomy, Pasadena, California, United States
b Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, United States

Abstract. The Habitable Exoplanet Imaging Mission concept requires an optical coronagraph that provides deep starlight suppression over a broad spectral bandwidth, high throughput for point sources at small angular separations, and insensitivity to temporally varying, low-order aberrations. Vortex coronagraphs are a promising solution that performs optimally on off-axis, monolithic telescopes and may also be designed for segmented telescopes with minor losses in performance. We describe the key advantages of vortex coronagraphs on off-axis telescopes such as (1) unwanted diffraction due to aberrations is passively rejected in several low-order Zernike modes relaxing the wavefront stability requirements for imaging Earth-like planets from <10 to >100 pm rms, (2) stars with angular diameters >0.1 λ/D may be sufficiently suppressed, (3) the absolute planet throughput is >10%, even for unfavorable telescope architectures, and (4) broadband solutions (Δλ/λ > 0.1) are readily available for both monolithic and segmented apertures. The latter make use of grayscale apodizers in an upstream pupil plane to provide suppression of diffracted light from amplitude discontinuities in the telescope pupil without inducing additional stroke on the deformable mirrors. We set wavefront stability requirements on the telescope, based on a stellar irradiance threshold set at an angular separation of 3 ± 0.5λ/D from the star, and discuss how some requirements may be relaxed by trading robustness to aberrations for planet throughput.

1 Introduction

The Habitable Exoplanet Imaging Mission (HabEx) concept seeks to directly detect atmospheric biomarkers on Earth-like exoplanets orbiting sun-like stars for the first time.1 Accomplishing this task requires extremely high-contrast imaging over a broad spectral range using an internal coronagraph2 or external starshade.3 Sufficient starlight suppression may be achieved on an ultrastable telescope using an on-board coronagraph instrument (CGI) with high-precision wavefront control and masks specially designed to manage diffraction of unwanted starlight. Each of these critical technologies will be demonstrated in space with the WFIRST CGI at the performance level needed to image gas giant planets in reflected light with a 2.4-m telescope.5 Leveraging the advancements afforded by WFIRST, the HabEx mission concept makes use of a larger (>4 m) telescope whose stability specifications allow for the detection and characterization of Earth-like planets with planet-to-star flux ratios <10−10.

The optimal coronagraph performance for a given mission depends strongly on the telescope design. The possible HabEx architectures currently under study in preparation for the 2020 Astrophysics Decadal Survey are a 4-m monolithic (architecture A) or 6.5-m segmented primary mirror (architecture B). For the purposes of this paper, we assume fully off-axis telescopes in both cases. The unobstructed, circular pupil provided by architecture A is ideal for coronagraph performance. In contrast, the segmented primary mirror of architecture B will introduce a number of additional complications owing to potential amplitude and phase discontinuities in the wavefront.

We present vortex coronagraph5−7 designs for each of the HabEx architecture and their theoretical performance. We demonstrate that the coronagraph may be designed to passively reject unwanted diffraction within the telescope in the presence of temporally varying, low-order aberrations as well as amplitude discontinuities (i.e., gaps between mirror segments). We set wavefront stability requirements on the telescope, including the phasing of the primary mirror segments in the case of architecture B, and discuss how some telescope requirements may be relaxed by trading robustness to aberrations for planet throughput.5

2 Architecture A: 4-m Off-Axis, Unobscured, Monolithic Telescope

The first telescope architecture we analyze is a 4-m off-axis telescope with a monolithic primary mirror. The unobstructed pupil is conducive to highly efficient coronagraph designs, such as the vortex coronagraph, which provide sensitivity to weak planet signals at small angular separations, as demonstrated in the laboratory6 and observations with ground-based telescopes.10−14 Figure 1(a) shows a schematic of a vortex coronagraph with dual deformable mirrors for wavefront control, a focal plane mask, and Lyot stop. The vortex focal plane mask is a transparent optic, which imparts a spiral phase shift of the form exp( ilφ)
on the incident field, where \( l \) is an even nonzero integer known as the “charge” and \( \phi \) is the azimuth angle in the focal plane. Light from an on-axis point source (i.e., the star) that passes through the circular entrance pupil of radius \( a \) is completely diffracted outside of the downstream Lyot stop of radius \( b \), assuming \( b < a \), and one-to-one magnification within the coronagraph. In addition to ideal starlight suppression, the vortex coronagraph provides high throughput for point-like sources at small angular separations from the star [see Fig. 1(b)].

\[
\eta_{p, \text{max}} = \left\{ \begin{array}{ll}
(b/a)^2, & \text{total energy fraction with in } \hat{r} \lambda/D \text{ radius} \\
(b/a)^2 \left[ 1 - J_0(\alpha \hat{r}^2) \right] - J_1(\alpha \hat{r}^2), & \text{otherwise}
\end{array} \right.
\]

where \( J_0(\cdot) \) and \( J_1(\cdot) \) are Bessel functions of the first kind.\(^{15}\) For example, if \( b/a = 0.95 \) and \( \hat{r} = 0.7 \), the theoretical maxima for cases (1) and (2) are 90% and 58%, respectively. The latter value may also be normalized to the same quantity without the coronagraph masks. For example, if \( b/a = 0.95 \), 86% of planet energy remains within \( 0.7 \lambda/D \) of the planet’s position in the image, a value referred to as the relative throughput. In the remainder of this section, we assume a typical value of \( b/a = 0.95 \). In practice, the value of \( b/a \) will be selected based on the desired tolerance to lateral pupil motion and magnification. Definitions (1) and (2) are plotted for various values of \( l \) in Fig. 1(b) for angular separations up to \( 20 \lambda/D \) using numerical beam propagation.

### 2.2 Passive Insensitivity to Low-Order Aberrations

Detecting Earth-like exoplanets in practice will require a coronagraph whose performance is insensitive to wavefront errors owing to mechanical motions in the telescope and differential polarization aberrations, which both manifest as low-order wavefront errors. We describe the phase at the entrance pupil of the coronagraph as a linear combination of Zernike polynomials \( Z_m^l(r/a, \theta) \) defined over a circular pupil of radius \( a \). An isolated phase aberration is written

\[
P(r, \theta) = \exp[i c_m Z_m^l(r/a, \theta)], \quad r \leq a,
\]

where \( i = \sqrt{-1} \) and \( c_m \) is the Zernike coefficient. Assuming small wavefront errors (i.e., \( c_m \ll 1 \) rad rms), the field in the pupil may be approximated to first order via its Taylor series expansion

\[
P(r, \theta) \approx 1 + i c_m Z_m^l(r/a, \theta), \quad r \leq a.
\]

For convenience, we choose to use the set real valued of Zernike polynomials described by

\[
Z_m^l(r/a, \theta) = R_m^l(r/a) q_m(\theta), \quad r \leq a,
\]

where \( R_m^l(r/a) \) are the radial polynomials described in Appendix A and

\[
q_m(\theta) = \left\{ \begin{array}{ll}
\cos(m \theta) & m \geq 0 \\
\sin(m \theta) & m < 0
\end{array} \right.
\]

The field transmitted through a vortex phase element of charge \( l \), owing to an on-axis point source, is given by the product of \( \exp(i l \phi) \) and the optical Fourier transform (FT) of Eq. (3)

\[
F_{nm}^l(\rho, \phi) \approx [f_{00}(\rho, \phi) + i c_m f_{nm}(\rho, \phi)] \exp(i l \phi)
\]

where

<table>
<thead>
<tr>
<th>MMID</th>
<th>Title</th>
<th>Authors</th>
<th>Journal</th>
<th>Volume</th>
<th>Issue</th>
<th>Pages</th>
<th>Year</th>
<th>Month</th>
<th>Date</th>
<th>Terms of Use</th>
<th>Downloaded From</th>
</tr>
</thead>
</table>
where $\rho$ is the radial polar coordinate in the focal plane, $k = 2\pi/\lambda$, $\lambda$ is the wavelength, and $f$ is the focal length, respectively. The field in the subsequent pupil plane (i.e., just before the Lyot stop), $E_{lnm}$, is given by the FT of Eq. (6). The first term, $f_{00}(\rho, \phi)$, is the common Airy pattern, which diffracts completely outside of the Lyot stop for all even nonzero values of $l$. In this case, the Lyot plane field becomes\(^{16}\)

$$E_{l,\text{Airy}}(r, \theta) = \begin{cases} 0 & r \leq a \\ \frac{2}{\pi} R^{|l|-1} \frac{C_{26}}{W_l} e^{i\theta} & r > a \end{cases}$$

More generally, the full Lyot plane field is given by

$$E_{lnm}(r, \theta) \approx E_{l,\text{Airy}}(r, \theta) + i c_{nm} g_{nm}(r, \theta),$$

where

$$g_{nm}(r, \theta) = \frac{k a}{2 f} e^{i\theta} \left\{ \begin{array}{ll} (1)^m e^{im\theta} W_{n+1}^{l+m}(r) + e^{-im\theta} W_{n+1}^{l-m}(r) & m \geq 0 \\ i[(-1)^{m+1} e^{im\theta} W_{n+1}^{l+m}(r) + e^{-im\theta} W_{n+1}^{l-m}(r)] & m < 0 \end{array} \right\}.$$  

Fig. 2 The low-order aberration filtering mechanism of a vortex coronagraph. The top row shows the wavefront at the entrance pupil of the coronagraph (“first pupil” in Fig. 1). The remaining rows show the amplitude distribution just before the Lyot stop (“second pupil” in Fig. 1). A vortex coronagraph is passively insensitive to modes, where the starlight appears outside of a Lyot stop whose radius $b$ is less than the geometric pupil radius $a$ and $|l| > n + |m|$.
and $W_p^q(r)$ is a special case of the Weber–Schafheitlin integral (Appendix B)

$$W_p^q(r) = W_{p,q}^0(r; ka/f, krp/f) = \int_0^\infty J_p(kap/f)J_q(krp/f) dp. \tag{11}$$

The solutions to Eq. (10) are shown in Fig. 2 and listed in Appendix C. In cases where all of the light is located outside of the geometric pupil, the source is extinguished by a Lyot stop if $|l| > n + |m|$. A charge $l$ vortex coronagraph is therefore passively insensitive to the $l^2/4$ Zernike modes rejected by the Lyot stop.

### 2.3 Wavefront Stability Requirements

The wavefront error tolerances of a given coronagraph design depend on the aberration mode and/or spatial frequency content of the error. The coronagraph and telescope must be jointly optimized to passively suppress starlight and provide the stability needed to maintain suppression throughout an observation. We present telescope stability requirements for Earth-like exoplanet imaging with vortex coronagraphs in terms of low-order and mid-to-high spatial frequency aberrations.

#### 2.3.1 Low-order requirements: Zernike aberrations

Figure 3 shows the leaked starlight through the coronagraph (stellar irradiance, averaged over effective angular separations $3 \pm 0.5 \lambda/D$, and normalized to the peak value without the coronagraph masks) as a function of root mean square (RMS) wavefront error. Modes with $n + |m| \geq |l|$ follow a quadratic power law and generate irradiance at the $\sim 10^{-11}$ level for wavefront errors of $\sim 10^{-5}$ waves RMS. However, modes with $n + |m| < |l|$ are blocked at least to first order at the Lyot stop, as described in the previous section. In these cases, the equivalent irradiance level ($\sim 10^{-11}$) corresponds to $\sim 100\times$ of the wavefront error.

We place requirements on the stability of the wavefront by setting a maximum allowable irradiance threshold on the leaked starlight at $3 \pm 0.5 \lambda/D$. This angular coordinate range corresponds to the separations, where a charge eight vortex coronagraph transmits $\sim 50\%$ of the planet light. Here, the threshold is chosen to be $2 \times 10^{-11}$ per Zernike mode (dashed line in Fig. 3) to prevent any single low-order aberration from dominating the error budget. The corresponding wavefront error at $\lambda = 450$ nm, likely the shortest and most challenging wavelength, is shown in Fig. 4 and listed in Table 1. Modes that are passively suppressed by the coronagraph have wavefront requirements $>100$ pm rms, whereas those that transmit tend to require $<10$ pm rms. The minimum charge of the vortex coronagraph may be chosen to preserve robustness to particularly problematic low-order aberrations as well as to relax requirements and reduce the cost of the overall mission. However, increasing the minimum
charge has a significant impact on the scientific yield of the mission, especially since insufficient throughput at small angular separations (i.e., beyond the so-called “inner working angle”) will likely limit the number of detected and characterized Earth-like planets within the mission lifetime.\footnote{18}

The requirements given in Fig. 4 and Table 1 may be scaled to any wavelength by simply multiplying the reported RMS wavefront error by a factor of \(\lambda / \Delta \) for \(\lambda = 450 \text{ nm}\). Although a higher charge (e.g., charges 6 or 8) may be used for the shortest wavelengths to improve robustness, using a lower charge (e.g., charge 4) at longer wavelengths would allow exoplanets detected near the inner working angle of the visible coronagraph to be characterized in the infrared, where the wavefront error requirements are naturally less strict. In that case, the infrared coronagraph would drive requirements in some of the lowest order modes, which would be relaxed by a factor of \(\gtrsim 2\) with respect to high-order requirements driven by the visible coronagraph.

### Table 1

<table>
<thead>
<tr>
<th>Aberration</th>
<th>Noll</th>
<th>Indices</th>
<th>Allowable RMS wavefront error per mode (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(n)</td>
<td>(m)</td>
</tr>
<tr>
<td>Tip-tilt</td>
<td>2,3</td>
<td>1 ± 1</td>
<td>1.1</td>
</tr>
<tr>
<td>Defocus</td>
<td>4</td>
<td>2 0</td>
<td>0.81</td>
</tr>
<tr>
<td>Astigmatism</td>
<td>5,6</td>
<td>2 ± 2</td>
<td>0.007</td>
</tr>
<tr>
<td>Coma</td>
<td>7,8</td>
<td>3 ± 1</td>
<td>0.006</td>
</tr>
<tr>
<td>Trefoil</td>
<td>9,10</td>
<td>3 ± 3</td>
<td>0.007</td>
</tr>
<tr>
<td>Spherical</td>
<td>11</td>
<td>4 0</td>
<td>0.005</td>
</tr>
<tr>
<td>Second astig.</td>
<td>12,13</td>
<td>4 ± 2</td>
<td>0.008</td>
</tr>
<tr>
<td>Quadrafoil</td>
<td>14,15</td>
<td>4 ± 4</td>
<td>0.008</td>
</tr>
<tr>
<td>Second coma</td>
<td>16,17</td>
<td>5 ± 1</td>
<td>0.004</td>
</tr>
<tr>
<td>Second trefoil</td>
<td>18,19</td>
<td>5 ± 3</td>
<td>0.005</td>
</tr>
<tr>
<td>Pentafoil</td>
<td>20,21</td>
<td>5 ± 5</td>
<td>0.005</td>
</tr>
<tr>
<td>Second spherical</td>
<td>22</td>
<td>6 0</td>
<td>0.003</td>
</tr>
<tr>
<td>Third Astig.</td>
<td>23,24</td>
<td>6 ± 2</td>
<td>0.002</td>
</tr>
<tr>
<td>Second quadrafoil</td>
<td>25,26</td>
<td>6 ± 4</td>
<td>0.003</td>
</tr>
<tr>
<td>Hexafoil</td>
<td>27,28</td>
<td>6 ± 6</td>
<td>0.003</td>
</tr>
</tbody>
</table>

### 2.3.2 Mid-to-high spatial frequency requirements

Although the coronagraph design provides degrees of freedom for controlling robustness to low-order aberrations, high throughput coronagraphs are naturally sensitive to mid- and high-spatial frequency aberrations. In fact, any coronagraph that passively suppresses mid-spatial frequency aberrations must also have low throughput for off-axis planets. This is an outcome of the well-known relationship between raw contrast and the RMS wavefront error in Fourier modes.\footnote{19} The pupil field associated with a single spatial frequency is given by

\[
P(r, \theta) = \exp \left[ i2\sqrt{2}\pi \omega \sin \left( \frac{2\pi x}{a} \right) \right], \quad r \leq a, \tag{12}
\]
\[\approx 1 + i2\sqrt{2\pi \omega} \sin \left(\frac{2\pi \xi}{a}\right), \quad r \leq a, \quad (13)\]

where \(r^2 = x^2 + y^2\), \(\xi\) is the spatial frequency in cycles per pupil diameter, and \(\omega\) is the RMS phase error in waves where we have assumed \(\omega \ll 1\). The corresponding field just before the focal plane mask is

\[F(\rho, \phi) = f_{00}(\rho, \phi) + \sqrt{2\pi \omega}[f_{00}(\rho, \phi) - f_{00}(\rho, \phi)], \quad (14)\]

where \(\rho^2 = x^2 + y^2\), \(\rho_0^2 = (x' + \xi \lambda F^\#)^2 + y^2\), \(\rho_f^2 = (x' + \xi \lambda F^\#)^2 + y^2\), and \(F^\# = f/(2a)\). The coronagraph completely rejects the \(f_{00}(\rho, \phi)\) term. Thus, at position \((x', y') = (\xi \lambda F^\#, 0)\) after the coronograph

\[F(\xi \lambda F^\#, 0) = \sqrt{2\pi \rho_0 \omega}[f_{00}(0, 0) - f_{00}(2\xi \lambda F^\#, 0)], \quad (15)\]

where \(\eta_p\) is the coronagraph throughput and \(F^\# = f/(2b)\). Solving for the normalized stellar irradiance, \(\eta_s\), we find

\[\eta_s = \eta_p 2(\pi \omega)^2 \frac{|f_{00}(0, 0) - f_{00}(2\xi \lambda F^\#, 0)|^2}{|f_{00}(0, 0)|^2}. \quad (16)\]

Therefore, for \(\xi \geq 1\), the raw contrast at \((x', y') = (\xi \lambda F^\#, 0)\) is

\[C = \eta_s/\eta_p \approx 2(\pi \omega)^2. \quad (17)\]

For example, a 1 pm RMS midspatial frequency wavefront error described by the vector \(\vec{\xi} = \xi_x \vec{x} + \xi_y \vec{y}\) generates a change in raw contrast of \(\sim 10^{-10}\) at \(\lambda = 450\) nm in the corresponding image plane location \((x', y') = (\xi_x \lambda F^\#, \xi_y \lambda F^\#)\). This implies a stability requirement of \(\sim 1\) pm rms per Fourier mode for midspatial frequency wavefront errors. The stellar irradiance as a function of spatial frequency and charge is shown in Fig. 5.

Rejecting starlight with midspatial frequency phase errors and proportionally reducing the coronagraph throughput at the position of interest degrades performance in the photon-noise-limited regime, where the signal-to-noise ratio (SNR) for planet detection scales is defined as \(\eta_p/\sqrt{\eta_s}\). Provided an optimal coronagraph maximizes the SNR, a coronagraph that is passively robust to spatial frequencies, where \((x', y') = (\xi_x \lambda F^\#, \xi_y \lambda F^\#)\) is in the region of interest (i.e., dark hole), is not desirable.

### 2.4 Sensitivity to Partially Resolved, Extended Sources

The fraction of energy from a point source that leaks through the coronagraph as a function of angular separation, \(\alpha\), may be approximated for small offsets (i.e., \(\alpha \ll \lambda/D\)) through modal decomposition of the source.\(^\text{20}\) The transmitted energy is given by \(T_\alpha = \tau_s (\pi \alpha D/\lambda)^2\), where \(\tau_s\) is a constant [see Fig. 6(a)]. Integrating over an extended, spatially incoherent, stellar source
of angular extent, \( \Theta \), the expression becomes
\[
T_\Theta = \kappa_l (\pi \Theta D / \lambda)^l,
\]
where \( \kappa_l \) is a constant [see Fig. 6(b)]. The theoretical values for \( \tau_l \) and \( \kappa_l \) are given in Table 2. Higher charge vortex coronagraphs are far less sensitive to small tip-tilt errors and sources of finite size. For example, charge six vortex coronagraphs sufficiently suppress light from stars with angular diameters up to \( \sim 0.1 \lambda / D \) or \( \sim 2 \) mas for a 4-m telescope at \( \lambda = 450 \) nm.

An often overlooked potential source of leaked starlight is the presence of an unresolved disk of dust around the star. Figure 6(c) shows the stellar irradiance that appears in the image plane due to scattered light from the debris ring at astrophysical contrast of 1% and angular separation of \( 3 \lambda / D \) from the star as a function of the size of the ring. For example, imaging an Earth-like planet at \( 3 \lambda / D \) around a star with a dust ring of radius \( 0.2 \lambda / D \) requires at least a charge six coronagraph. We note that over the last 10 years, long baseline near infrared interferometric observations\(^{21,22}\) has suggested that \( \sim 10\% \) to \( 20\% \) of nearby main sequence stars have such rings of small hot dust grains concentrating near the sublimation radius, and contributing about 1% of the total solar flux in the near infrared. Assuming that we are looking for a 300 K planet at \( 3 \lambda / D \) separation, 1500 K dust grains would be located 25x closer, i.e., at \( 0.12 \lambda / D \). In addition to being more resilient to low-order wavefront aberrations, higher charge vortex coronagraphs are also less sensitive to astrophysical noise sources in the inner part of the system, such as bright dust rings that may be fairly common.

### Table 2 Coefficients for analytical approximations of transmitted energy from point sources at small angular separations, \( T_\alpha = \tau_\alpha (\pi \alpha D / \lambda)^l \), and extended sources \( T_\Theta = \kappa_\Theta (\pi \Theta D / \lambda)^l \).

<table>
<thead>
<tr>
<th>Charge</th>
<th>( \tau_l )</th>
<th>( \kappa_l )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( l = 2 )</td>
<td>1/8</td>
<td>1/64</td>
</tr>
<tr>
<td>( l = 4 )</td>
<td>1/192</td>
<td>1/9216</td>
</tr>
<tr>
<td>( l = 6 )</td>
<td>1/9216</td>
<td>1/2359296</td>
</tr>
<tr>
<td>( l = 8 )</td>
<td>1/737280</td>
<td>1/943718400</td>
</tr>
</tbody>
</table>

#### 2.5 Effect of Adding an Opaque Spot to the Vortex Mask

Manufacturing processes will limit the minimum size of the central defect in a vortex phase mask. A small opaque occulting spot may be introduced to block the central region, where the phase shift deviates from the ideal vortex pattern [see Fig. 7(a)]. The maximum allowable size of this mask depends on the charge of the vortex. Figure 7(b) shows the stellar irradiance at \( 3 \pm 0.5 \lambda / D \) as a function of the mask diameter for various vortex charges. For charges 6 and 8, the occulting mask can be as large as \( \sim 1.5 \lambda F* \) and \( \sim 1.7 \lambda F* \), respectively, while maintaining sufficient suppression for imaging of Earth-like planets. In each case, the opaque mask does not significantly degrade the planet throughput.

In addition to masking manufacturing errors, there are other potential benefits to introducing a central opaque spot. For instance, the reflection from the spot may be used for integrated low-order wavefront sensing, as recently demonstrated for the WFIRST CGL\(^{23}\) potentially in addition to a reflective Lyot stop sensor.\(^{24,25}\) In the case of a charge 6 vortex coronagraph, \( \sim 80\% \) of the starlight would be available from the reflection off of the opaque mask for fast tip-tilt and low-order wavefront sensing. Combined with the natural insensitivity to low-order aberrations of vortex coronagraphs, this capability will help maintain deep starlight suppression throughout observations and extend the time between calibrations of the wavefront error and reference star images, thereby improving overall observing efficiency.

### 3 Architecture B: 6.5-m Off-Axis, Unobscured, Segmented Telescope

The second potential telescope architecture we study for the HabEx mission concept is a 6.5-m off-axis segmented telescope. This arrangement introduces a few additional complications with respect to the monolithic version. First, a primary mirror with a noncircular outer edge generates diffraction patterns that are difficult to null. To remedy this, we insert a circular sub-aperture in a pupil plane just before the focal plane mask, which provides improved starlight suppression at the cost of throughput (see Fig. 8). Partial segments may also be introduced to form a circular outer edge. Second, the gaps between mirror segments...
must be apodized to prevent unwanted diffraction in the image plane from amplitude discontinuities. In this section, we present a promising vortex coronagraph design for the 6.5-m HabEx concept and address the associated telescope requirements.

3.1 Apodized Vortex Coronagraph Design

Figure 9(a) shows a notional primary mirror with 37 hexagonal segments whose widths are ~0.9 m flat-to-flat. The corresponding pupil masks used in the apodized vortex coronagraph are shown in Figs. 9(b) and 9(c). The apodizer clips the outer edge of the pupil to make it circular and imparts an amplitude-only apodization pattern on the transmitted or reflected field. Most of the starlight is then diffracted by the vortex outside of the Lyot stop. The small amount of starlight that leaks through the Lyot stop (~2%) only contains high-spatial frequencies greater than a specified value of $\xi_{\text{max}} \approx 20$ cycles across the pupil diameter. Thus, in an otherwise perfect optical system, a dark hole appears in the starlight within a $20 \lambda / D$ radius of the star position for all even nonzero values of the vortex charge $l$.

Fig. 7 The influence of an opaque spot at the center of the focal plane mask. (a) Phase shift imposed by a charge 6 vortex mask. Light is blocked within the mask at the center (shown in black). (b) Stellar irradiance, averaged over source positions $3 \pm 0.5 \lambda / D$ and normalized to the peak of the telescope PSF as a function of the mask diameter. Larger masks may be used with higher charges, which alleviate some manufacturing challenges and allow for a reflective low-order wavefront sensor.

Fig. 8 Schematic of an apodized vortex coronagraph. A gray-scale apodizer (see Fig. 9) prevents unwanted diffraction from the noncircular outer edge of the primary and gaps between mirror segments.
We used the auxiliary field optimization method\textsuperscript{26} to calculate the optimal grayscale pattern.\textsuperscript{27,28}

The throughput of the coronagraph with various focal plane vortex masks is shown in Fig. 10. We report both the absolute throughput $\eta_p$ and relative throughput $\eta_p/\eta_{tel}$ within a circular region of interest of radius $\hat{r}\lambda/D$ centered on the planet position, where $\eta_{tel}$ represents the throughput of the telescope with the coronagraph masks removed. After the coronagraph, $\sim 60\%$ of the total energy from an off-axis source remains. Less than 30\% of the total energy appears within $0.7 \lambda/D$ of the planet position, including losses from the apodizer and broadening of the point spread function by the undersized pupil mask and Lyot stop. Approximately 50\% of the planet light remains within $0.7 \lambda/D$ compared with the point spread function with the coronagraph masks removed. Other than a loss in throughput, the apodized version shares most of the same performance characteristics as the conventional vortex coronagraph. Furthermore, a primary mirror that includes partial segments to create a circular outer boundary would allow for drastically improved coronagraph throughput.

3.2 Sensitivity to Low-Order Aberrations and the Angular Size of Stars

Assuming the telescope is off-axis and unobstructed, the leaked stellar irradiance in the presence of low-order aberrations appears identical to the monolithic case, up to a radius of $\xi_{\text{max}}\lambda F^3$ (see Fig. 11). However, to maintain a fixed raw contrast

\begin{figure}
\centering
\includegraphics[width=\textwidth]{fig10}
\caption{The throughput of the apodized vortex coronagraph with charge 4, 6, and 8 focal plane masks. (a) Absolute throughput. The fraction of total planet light that falls within $\hat{r}\lambda/D$ of the planet position, assuming an otherwise perfect optical system. (b) Relative throughput. The fraction of planet light that falls within $\hat{r}\lambda/D$ of the planet position compared with case with the coronagraph masks removed. Throughput losses originate from introducing a semitransparent mask and clipping the outer edge of the pupil to create a circular boundary (see Fig. 9).}
\end{figure}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{fig11}
\caption{The sensitivity of an apodized vortex coronagraph to low-order aberrations on an off-axis, segmented telescope. Log irradiance owing to $\lambda/1000$ rms wavefront error in each mode, normalized to the peak value with the coronagraph masks removed. The dark zone has an angular diameter of $40 \lambda/D$.}
\end{figure}
threshold, the wavefront error requirements presented in Table 1 scale as \( 1/\sqrt{\eta_p} \); i.e., get tougher to guarantee than in the monolithic case. For the sake of brevity, we have not included an updated wavefront error requirement table here.

The stellar leakage due to the angular size of stars is also equivalent to a vortex coronagraph without an apodizer. Figure 12 shows the leaked starlight as a function of stellar angular size. A charge 4 is sufficient to suppress stars \( \lesssim 0.01 \lambda/D \) in diameter. Charges 6 or 8 may be used to maximize SNR (\( \eta_p/\sqrt{\eta_s} \)) in the case of a larger star, such as Alpha Centauri A whose angular diameter is 8.5 mas or \( \sim 0.5 \lambda/D \) in the visible.

### 3.3 Segment Cophasing Requirements

A major challenge for exoplanet imaging with a segmented telescope will be to keep the mirrors coaligned throughout observations. As shown in Fig. 13, small segment motions in piston and tip-tilt cause speckles to appear in the dark hole which may be difficult to calibrate and will likely contribute to both photon and spatial speckle noise. Figures 14(a) and 14(b) show the time average over many realization of the errors shown in Fig. 13 drawn from Gaussian distributions for both piston and tip-tilt with standard deviations of 100 pm and 0.005 \( \lambda/D = 71 \mu \) as rms. When the mirror segments have random piston errors only, the resulting distribution of light resembles the diffraction pattern of a single segment [Fig. 14(a)]. Random segment tip-tilt error tends to spread the leaked starlight to larger separations [Fig. 14(b)]. However, the leaked starlight is well approximated by a similar second-order power law in both cases [Fig. 14(c)], which yields a wavefront error requirement of \( \sim 10 \) pm rms, similar in magnitude to unsuppressed low-order modes. In contrast, if the primary mirror segments undergo a coordinated movement that resembles a low-order Zernike polynomial \( Z_{mn} \), the amount of leaked starlight would be significantly smaller if \( l > n + |m| \) and the tolerance to such a motion would be considerably relaxed.
3.4 Fabrication of Grayscale Apodizing Masks

Achromatic grayscale apodizers have been fabricated using metallic microdots arranged in error-diffused patterns.29,30 Prototype masks produced specifically for the purpose of demonstrating apodized vortex coronagraphs are currently being tested at the High-Contrast Spectroscopy Testbed for Segmented Telescopes (HCST) at Caltech.31 These experiments seek to validate this approach for use on HabEx and prepare for future testing on vacuum testbeds.

4 Conclusions and Outlook

Vortex coronagraphs provide a viable pathway toward imaging Earth-like exoplanets with the HabEx decadal mission concept with a fully off-axis telescope architecture. We have provided an overview of the performance of vortex coronagraphs and wavefront stability requirements. The off-axis design of the HabEx telescope allows for the best possible performance in terms of throughput, inner working angle, and robustness to aberrations. With a vortex coronagraph, the low-order wavefront error requirements for imaging Earth-like planets with HabEx are comparable with those to be demonstrated by the WFIRST- CGI for imaging Jupiters.32 A segmented primary mirror does not fundamentally change the performance characteristics of a vortex coronagraph. However, mirror segments introduce challenging segment cophasing requirements and the need for apodization. In addition to the grayscale pupil mask presented here, alternate apodization approaches are available that shape the pupil amplitude using deformable mirrors.33 A trade study is needed to identify the performance trades between these apodization solutions. In all cases, we find that the throughput of vortex coronagraphs and robust to wavefront errors degrades significantly on centrally obscured telescopes. Several studies are underway to improve performance for such architectures.34

Appendix A: Zernike Polynomials

The Zernike polynomials35 may be written as

\[
Z_m^p(r/a, \theta) = R_m^p(r/a) \left\{ \begin{array}{ll}
\cos(m\theta) & m \geq 0 \\
\sin(m\theta) & m < 0 \end{array} \right., \quad r \leq a,
\]

where \(R_m^p(r/a)\) is the radial Zernike polynomial given by

\[
R_m^p(r/a) = \sum_{k=0}^{\frac{n-m}{2}} \frac{(-1)^k(n-k)!}{k!(\frac{n+m}{2}-k)!(\frac{n-m}{2}-k)!} (r/a)^{n-2k}, \quad r/a \leq 1,
\]

(19)

where \(n - m\) is even. The indices \(n\) and \(m\) are the integers, respectively, known as the degree and azimuthal order. The first few radial polynomials are \(R_0^0 = 1\), \(R_1^0 = r/a\), \(R_2^0 = 2(r/a)^2 - 1\), \(R_2^1 = (r/a)^2\), \(R_3^1 = 3(r/a)^3 - 2(r/a)\), and \(R_3^2 = (r/a)^3\).

Appendix B: Weber–Schafheitlin Integrals

The pupil functions generated by vortex coronagraphs are a subset of solutions of the discontinuous integral of Weber and Schafheitlin,36 which in its conventional form is written

\[
W_{\nu,\mu,\lambda}(\alpha, \beta) = \int_0^{\infty} \frac{J_\nu(\alpha t)J_\mu(\beta t)}{t^\lambda} \, dt,
\]

(20)

where \(\nu\), \(\mu\), \(\lambda\) are the integers and \(\alpha\) and \(\beta\) are the constants. The integral is convergent provided \(\nu + \mu - \lambda \geq 0\) and \(\lambda \geq 0\). If \(0 < \alpha < \beta\)

\[
W_{\nu,\mu,\lambda}(\alpha, \beta) = \frac{\alpha^{\lambda+2} \Gamma(\nu+\lambda+1)}{2 \beta^{\nu+2} \Gamma(\nu+\mu+1) \Gamma(\nu+1)} \times _2F_1 \left( \frac{\nu+\mu-\lambda+1}{2}, \frac{\nu-\mu+\lambda+1}{2}; \nu+1; \frac{\alpha^2}{\beta^2} \right),
\]

(21)

where \(\Gamma()\) is the gamma function and \(_2F_1()\) is a hypergeometric function.37 On the other hand, if \(0 < \beta < \alpha\)

\[
W_{\nu,\mu,\lambda}(\alpha, \beta) = \frac{\beta^{\lambda+2} \Gamma(\nu+\lambda+1)}{2 \alpha^{\nu+2} \Gamma(\nu+\mu+1) \Gamma(\nu+1)} \times _2F_1 \left( \frac{\nu+\mu-\lambda+1}{2}, \frac{-\nu+\mu-\lambda+1}{2}; \nu+1; \frac{\beta^2}{\alpha^2} \right).
\]

(22)
Integrals with the form of Eq. (20), namely a product of Bessel functions, appear in the output function integral in cases, where the input function is circular or may be described by a Zernike polynomial in amplitude.\

\[ g_{200}(r, \theta) = \begin{cases} 2 \left( \frac{\xi}{\eta} \right)^2 - 1 & r \leq a \\ 0 & r > a \end{cases} \]  
\[ g_{202}(r, \theta) = \begin{cases} \left( \frac{\xi}{\eta} \right)^2 e^{i2\theta} & r \leq a \\ 0 & r > a \end{cases} \]  
\[ g_{204}(r, \theta) = \begin{cases} 0 & r \leq a \\ \left( \frac{\xi}{\eta} \right)^4 e^{i4\theta} & r > a \end{cases} \]  
\[ g_{206}(r, \theta) = \begin{cases} 0 & r \leq a \\ \left[ 5 \left( \frac{\xi}{\eta} \right)^6 - 4 \left( \frac{\xi}{\eta} \right)^4 \right] e^{i6\theta} & r > a \end{cases} \]  
\[ g_{208}(r, \theta) = \begin{cases} 0 & r \leq a \\ \left[ 21 \left( \frac{\xi}{\eta} \right)^8 - 30 \left( \frac{\xi}{\eta} \right)^6 + 10 \left( \frac{\xi}{\eta} \right)^4 \right] e^{i8\theta} & r > a \end{cases} \]  

\[ g_{220}(r, \theta) = \begin{cases} \frac{\xi}{\eta} \cos(2\theta) & r \leq a \\ 0 & r > a \end{cases} \]  
\[ g_{222}(r, \theta) = \begin{cases} \frac{1}{2} \left( \frac{\xi}{\eta} \right)^2 - 1 & r \leq a \\ \frac{1}{2} \left( \frac{\xi}{\eta} \right)^4 e^{i4\theta} & r > a \end{cases} \]  
\[ g_{224}(r, \theta) = \begin{cases} \frac{1}{2} \left( \frac{\xi}{\eta} \right)^2 e^{i2\theta} & r \leq a \\ \frac{1}{2} \left[ 5 \left( \frac{\xi}{\eta} \right)^6 - 4 \left( \frac{\xi}{\eta} \right)^4 \right] e^{i6\theta} & r > a \end{cases} \]  
\[ g_{226}(r, \theta) = \begin{cases} 0 & r \leq a \\ \left[ 42 \left( \frac{\xi}{\eta} \right)^10 - 84 \left( \frac{\xi}{\eta} \right)^8 + 105 \left( \frac{\xi}{\eta} \right)^6 - 10 \left( \frac{\xi}{\eta} \right)^4 \right] e^{i10\theta} + \left[ \frac{5}{2} \left( \frac{\xi}{\eta} \right)^6 - 2 \left( \frac{\xi}{\eta} \right)^4 \right] e^{i6\theta} & r > a \end{cases} \]  

\[ g_{310}(r, \theta) = \begin{cases} \frac{3}{2} \left( \frac{\xi}{\eta} \right)^3 - 2 \left( \frac{\xi}{\eta} \right) \cos(\theta) & r \leq a \\ 0 & r > a \end{cases} \]  
\[ g_{312}(r, \theta) = \begin{cases} \frac{1}{2} \left( \frac{\xi}{\eta} \right)^3 e^{i3\theta} + \left[ \frac{5}{2} \left( \frac{\xi}{\eta} \right)^3 - \frac{3}{2} \right] e^{i\theta} & r \leq a \\ 0 & r > a \end{cases} \]  
\[ g_{314}(r, \theta) = \begin{cases} \frac{1}{2} \left( \frac{\xi}{\eta} \right)^3 e^{i3\theta} & r \leq a \\ \frac{1}{2} \left( \frac{\xi}{\eta} \right)^3 e^{i8\theta} & r > a \end{cases} \]
\[ g_{316}(r, \theta) = \begin{cases} 0 & r \leq a \\
 \left[ \frac{3}{2} (\frac{r}{a})^7 - \frac{5}{2} (\frac{r}{a})^5 \right] e^{i7\theta} + \frac{1}{2} (\frac{r}{a})^5 e^{i5\theta} & r > a \end{cases} \]

\[ g_{318}(r, \theta) = \begin{cases} 0 & r \leq a \\
 \left[ 14 \left( \frac{r}{a} \right)^9 - 21 \left( \frac{r}{a} \right)^7 + \frac{15}{2} (\frac{r}{a})^5 \right] e^{9\theta} \\
 + \left[ \frac{3}{2} (\frac{r}{a})^7 - \frac{5}{2} (\frac{r}{a})^5 \right] e^{7\theta} & r > a \end{cases} \]

C6 Spherical \( Z^0_4 \)

\[ g_{400}(r, \theta) = \begin{cases} 6 \left( \frac{r}{a} \right)^4 - 6 \left( \frac{r}{a} \right)^2 + 1 & r \leq a \\
 0 & r > a \end{cases} \]

\[ g_{402}(r, \theta) = \begin{cases} \left[ 4 \left( \frac{r}{a} \right)^4 - 3 \left( \frac{r}{a} \right)^2 \right] e^{2\theta} & r \leq a \\
 0 & r > a \end{cases} \]

\[ g_{404}(r, \theta) = \begin{cases} \left( \frac{r}{a} \right)^4 e^{4\theta} & r \leq a \\
 0 & r > a \end{cases} \]

\[ g_{406}(r, \theta) = \begin{cases} 0 & r \leq a \\
 \left( \frac{r}{a} \right)^6 e^{6\theta} & r > a \end{cases} \]

\[ g_{408}(r, \theta) = \begin{cases} 0 & r \leq a \\
 \left[ 7 \left( \frac{r}{a} \right)^8 - 6 \left( \frac{r}{a} \right)^6 \right] e^{8\theta} & r > a \end{cases} \]

C7 Trefoil \( Z^3_3 \)

\[ g_{330}(r, \theta) = \begin{cases} \left( \frac{r}{a} \right)^3 \cos(3\theta) & r \leq a \\
 0 & r > a \end{cases} \]

\[ g_{332}(r, \theta) = \begin{cases} \left[ 3 \left( \frac{r}{a} \right)^3 - 2 \left( \frac{r}{a} \right) \right] e^{-i\theta} & r \leq a \\
 \left( \frac{r}{a} \right)^5 e^{5\theta} & r > a \end{cases} \]

\[ g_{334}(r, \theta) = \begin{cases} \left[ 3 \left( \frac{r}{a} \right)^3 - 2 \left( \frac{r}{a} \right) \right] e^{i\theta} & r \leq a \\
 \left( \left( \frac{6}{5} \right)^3 - 5 \left( \frac{r}{a} \right)^5 \right] e^{7\theta} & r > a \end{cases} \]

\[ g_{336}(r, \theta) = \begin{cases} \left( \frac{r}{a} \right)^{3\theta} & r \leq a \\
 \left[ 28 \left( \frac{r}{a} \right)^9 - 42 \left( \frac{r}{a} \right)^7 + 15 \left( \frac{r}{a} \right)^5 \right] e^{9\theta} & r > a \end{cases} \]

\[ g_{338}(r, \theta) = \begin{cases} 0 & r \leq a \\
 \left[ 60 \left( \frac{r}{a} \right)^{11} - 126 \left( \frac{r}{a} \right)^9 + 84 \left( \frac{r}{a} \right)^7 - \frac{35}{2} \left( \frac{r}{a} \right)^5 \right] e^{11\theta} \\
 + \frac{1}{2} \left( \frac{r}{a} \right)^5 e^{5\theta} & r > a \end{cases} \]

Acknowledgments

We thank the HabEx Coronagraph Technology Working Group (CTWG) for useful discussions. G. Ruane is supported by an NSF Astronomy and Astrophysics Postdoctoral Fellowship under award AST-1602444. This work was supported by the Exoplanet Exploration Program (ExEP), Jet Propulsion Laboratory, California Institute of Technology, under contract to NASA.

References


Biographies for the authors are not available.