

Method for deriving optical telescope performance specifications for Earth-detecting coronagraphs

Bijan Nemati,^{a,*} H. Philip Stahl,^{b,*} Mark T. Stahl,^b Garreth J. Ruane,^c
and Leah J. Sheldon^a

^aUniversity of Alabama in Huntsville, Huntsville, Alabama, United States

^bMarshall Space Flight Center, Huntsville, Alabama, United States

^cCalifornia Institute of Technology, Pasadena, California, United States

Abstract. Direct detection and characterization of extrasolar planets has become possible with powerful new coronagraphs on ground-based telescopes. Space telescopes with active optics and coronagraphs will expand the frontier to imaging Earth-sized planets in the habitable zones of nearby Sun-like stars. Currently, NASA is studying potential space missions to detect and characterize such planets, which are dimmer than their host stars by a factor of 10^{10} . One approach is to use a star-shade occulter. Another is to use an internal coronagraph. The advantages of a coronagraph are its greater targeting versatility and higher technology readiness, but one disadvantage is its need for an ultrastable wavefront when operated open-loop. Achieving this requires a system-engineering approach, which specifies and designs the telescope and coronagraph as an integrated system. We describe a systems engineering process for deriving a wavefront stability error budget for any potential telescope/coronagraph combination. The first step is to calculate a given coronagraph's basic performance metrics, such as contrast. The second step is to calculate the sensitivity of that coronagraph's performance to its telescope's wavefront stability. The utility of the method is demonstrated by intercomparing the ability of several monolithic and segmented telescope and coronagraph combinations to detect an exo-Earth at 10 pc. © The Authors. Published by SPIE under a Creative Commons Attribution 4.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JATIS.6.3.039002](https://doi.org/10.1117/1.JATIS.6.3.039002)]

Keywords: coronagraph; exoplanet; imaging; modeling; space telescope.

Paper 20013 received Feb. 4, 2020; accepted for publication Jul. 31, 2020; published online Aug. 18, 2020.

1 Introduction

“Are we alone in the Universe?” is one of the most compelling science questions of our generation.^{1–3} Per the 2010 *New Worlds, New Horizons in Astronomy and Astrophysics* decadal report:⁴ “one of the fastest growing and most exciting fields in astrophysics is the study of planets beyond our solar system. The ultimate goal is to image rocky planets that lie in the habitable zone of nearby stars.” The survey recommended, as its highest priority, medium-scale activity, such as a “New Worlds Technology Development Program” to “lay the technical and scientific foundations for a future space imaging and spectroscopy mission.” And, per the National Research Council report *NASA Space Technology Roadmaps and Priorities*,⁵ the second-highest technical challenge for NASA regarding expanding our understanding of Earth and the universe, in which we live is to “develop a new generation of astronomical telescopes that enable discovery of habitable planets, facilitate advances in solar physics, and enable the study of faint structures around bright objects by developing high-contrast imaging and spectroscopic technologies to provide unprecedented sensitivity, field of view, and spectroscopy of faint objects.”

Directly imaging and characterizing Earth-like, habitable-zone planets require the ability to suppress the host star's light by many orders of magnitude. This can be done with either an external star shade or an internal coronagraph. Performing exoplanet science with an internal coronagraph requires an ultraprecise, ultrastable optical telescope. Wavefront errors can cause

*Address all correspondence to Bijan Nemati, E-mail: bijan.nemati@uah.edu; H. Philip Stahl, E-mail: h.philip.stahl@nasa.gov

stellar light to leak through the coronagraph and introduce noise.^{6–8} Sources of these errors can be rigid-body misalignments between the optical components, mounting error, low-order, and mid-spatial frequency figure errors of the optical components themselves. For example, a lateral misalignment between the primary and secondary mirror introduces coma into the wavefront. If an error is static, it is possible to correct it via wavefront sensing and control and deformable mirrors (DMs)—limited by the DM actuator number, range, and spatial frequency. Or, its effect (i.e., speckle noise) can be removed via calibration and subtraction. For either approach, static errors should not be too great. But the most important error sources are dynamic. They arise from changing conditions (like the thermal loads) on the telescope or coronagraph assembly. Dynamic errors are difficult to correct for a number of reasons. Sensing these errors requires long integration times because the photon rates are very low with the starlight suppressed. Also sensing many of the most important error modes requires interruption of the science integration time. As we shall see later, the total observation time for direct imaging is already many tens of hours when all the aspects of the measurement are taken into account, and spectroscopy takes even more time (by an order of magnitude), making time a scarce resource. Real-time, concurrent sensing of the dynamic errors is conceivable, but so far many of these approaches suffer from noncommon path errors. These challenges may be overcome in the future, but, at this time, the lowest risk approach is to assume that the system is operated in open-loop during the science integration. Thus the telescope system must be designed to minimize dynamic errors. The problem is how to specify an ultrastable telescope.

To achieve robust open-loop control, insensitive to dynamic wavefront error, the telescope and coronagraph must be designed as an integrated system. Engineering specifications must be defined that will produce an on-orbit telescope performance that enables exo-Earth science. Stahl et al.^{9,10} used science-driven systems engineering to develop specifications for aperture, primary mirror surface error, aperture segmentation, and wavefront stability for candidate telescopes. One conclusion of this work was the “poetic” specification that the telescope needs to be stable to 10 picometers per 10 min. In reality, the specification is more complicated. The control system’s stability duration depends on factors such as the target star’s brightness, telescope’s aperture diameter, and coronagraph’s core throughput.¹¹ And the tolerable amplitude depends on the coronagraph’s sensitivity to that error, as well as the error mode’s spatial and temporal characteristics.^{11–18} References 11–18 each calculated candidate coronagraph’s contrast leakage as a function of wavefront error mode. References 11–15 used numerical simulations to calculate contrast leakage for Seidel aberrations and segmented aperture piston and tip/tilt error. Lebouilleux et al.¹⁶ developed an analytical method for calculating segmented aperture piston and tip/tilt error. Ruane et al.¹⁷ calculated contrast leakage as a function of Zernike polynomials, sinusoidal spatial frequencies, and segment piston and tip/tilt errors. And Coyle et al.¹⁸ developed a power spectral density (PSD)-based description. Each of these papers yielded essentially the same result for the same boundary conditions. This paper significantly extends this previous work¹⁵ to present a new systems-engineering process for deriving a telescope’s wavefront stability error budget from the sensitivity of its coronagraph’s performance to wavefront stability and provides specific implementation examples.

Section 2 outlines the parameters that go into creating such a wavefront stability error budget. Section 3 reviews the basics of coronagraphy and defines the coronagraph attributes that most directly affect their performance in planet detection: core throughput, raw contrast, and stability of raw contrast. Section 4 provides a detailed description of the error budget approach, including the analytical model that governs it and creates the error budget for exo-Earth detection. Section 5 provides an in-depth description of the coronagraph diffraction modeling approach used to derive the error budget sensitivities. Section 6 applies the method to five representative architectures: two vector-vortex and a hybrid Lyot coronagraph (HLC), all with a 4-m off-axis monolithic unobscured telescope; a vector-vortex charge-6 coronagraph with a 6-m off-axis hexagonal segmented aperture unobscured telescope; and an apodized pupil Lyot coronagraph (APLC) with a 6-m on-axis hexagonal segment telescope. Note that the wavefront stability error budget examples in Sec. 6 are to detect an exo-Earth at 10 pc (i.e., at a separation of 100 mas from its host star). Also, note that, while we study specific cases, the purpose of this paper is to present a process for generating a wavefront stability error budget. And the examples in Sec. 6 may or may not represent the current state of the art. Finally, Appendix A contains the detailed

mathematics for creating the contrast error budget. The analytical methodology presented in Appendix A was developed and is currently in use in the coronagraph instrument to be flown on the Nancy Grace Roman Space Telescope (hereafter referred to as “Roman”). The Roman coronagraph instrument is currently in development and will demonstrate the technologies needed for the Earth-detecting coronagraphs we are addressing in this paper. Appendix B describes a method for modeling polychromatic diffraction.

2 Science Drives Systems Performance

Direct imaging of exoplanets requires coronagraph/telescope systems capable of rejecting the light from the host star and enabling imaging of its companions. Planets can be directly detected using either reflected sunlight (which peaks in the visible band for sun-like stars) or the planet’s own blackbody radiation (which peaks in the infrared). Although the latter offers a number of advantages in terms of improved flux ratio and better mitigation of atmospheric effects for ground-based telescopes, Jupiter-class or smaller planets are still too dim for ground-based instruments to image. Space-based coronagraphs, however, are not subject to atmospheric effects and can, in principle, detect far dimmer companions.¹⁹

A special goal for future missions (beyond Roman) is to image an Earth-like planet in the habitable zone of a nearby sun-like star. Viewed from a distance of 10 pc, this planet would have an angular separation α (Fig. 1) of 100 mas (0.1 arc sec) at maximum separation. The flux ratio of the planet’s reflected light relative to its host star’s direct light can be estimated if we have a model of the albedo and phase function. Traub and Oppenheimer²⁰ gave a simple expression to estimate the flux ratio of a planet based on its size, location, albedo, and phase function:

$$\xi = A_g \phi(b) r_p^2 a^{-2}, \quad (1)$$

where A_g is the geometric albedo, $\phi(b)$ is the geometric phase function, b is the phase angle, r_p is the planet radius, and a is the distance from the planet to the star. This is illustrated in Fig. 1.

Using the Lambertian sphere approximation, the phase function is given by

$$\phi(b) = \frac{1}{\pi} [\sin b + (\pi - b) \cos b]. \quad (2)$$

At quadrature phase (i.e., “half-moon”), $\phi(\pi/2) = 1/\pi$. Assuming this planet has a geometric albedo of 0.37 for this planet, its flux ratio is 2.1×10^{-10} , or 210 ppt (parts-per-trillion). By comparison, the flux ratio for an exo-Jupiter has a flux ratio of 1.5 ppb (parts-per-billion).²¹ Figure 2 illustrates the challenge of directly detecting a companion relative to its host star by plotting the point spread functions (PSFs) for Jupiter and Earth analogues surrounding an exo-Sun located 10 pc away. The Jupiter analogue, at its angular separation, is dimmer than the scattered starlight by a factor of about 8000 while the Earth, closer in and smaller, is dimmer by a factor of 5,500,000 at its separation. Although the Roman coronagraph is being designed to detect exo-Jupiters, achieving the level of starlight suppression required to detect an exo-Earth with a flux ratio of 210 ppt is beyond the range of all current telescopes and coronagraphs.

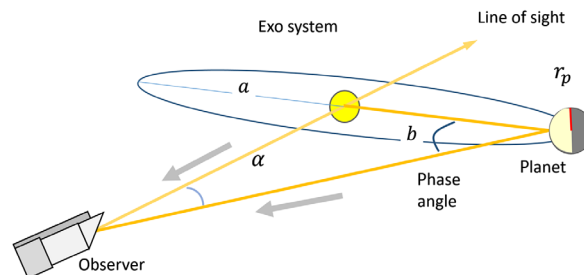


Fig. 1 The flux ratio of an exoplanet, seen at separation angle α , depends on its radius, orbital semimajor axis, and the phase angle.

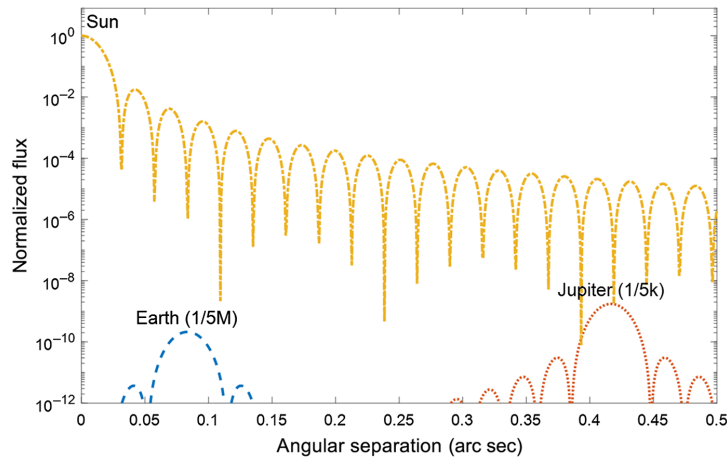


Fig. 2 Looking at the solar system 10 pc at 550 nm wavelength using a perfect, unobscured telescope with a 4-m primary mirror. The Earth peak is suppressed by a factor of over 5 million relative to the light from the Sun, whereas Jupiter is suppressed by a factor of a few thousand.

Historically, general astrophysics missions (from Hubble to Webb) have assumed that the observatory-level error budget can be bifurcated between the telescope and science instruments. But this is not the case for direct imaging of exoplanets, especially exo-Earths. To achieve this science, the telescope and coronagraph must be designed as an integrated system with an integrated performance error budget. The first step is to flow down an error budget from well-defined science objectives. Once the error budget has been conceived, the derivation of performance requirements in their native units additionally requires knowledge of the sensitivity of the performance to each given error source. With these two ingredients in place, tolerances can be derived that gauge the relative advantages of different telescope-coronagraph approaches.

Figure 3 gives an overview of the methodology. We propose for creating the error budget, the sensitivities, and the tolerances. Two models are used together to derive the error budget and the tolerances starting from some specific observing scenario assumptions. At the top level, there is an analytic performance model, which calculates the expected signal, noise, and signal-to-noise ratio (SNR) for a given observing scenario, along with the integration time required to obtain

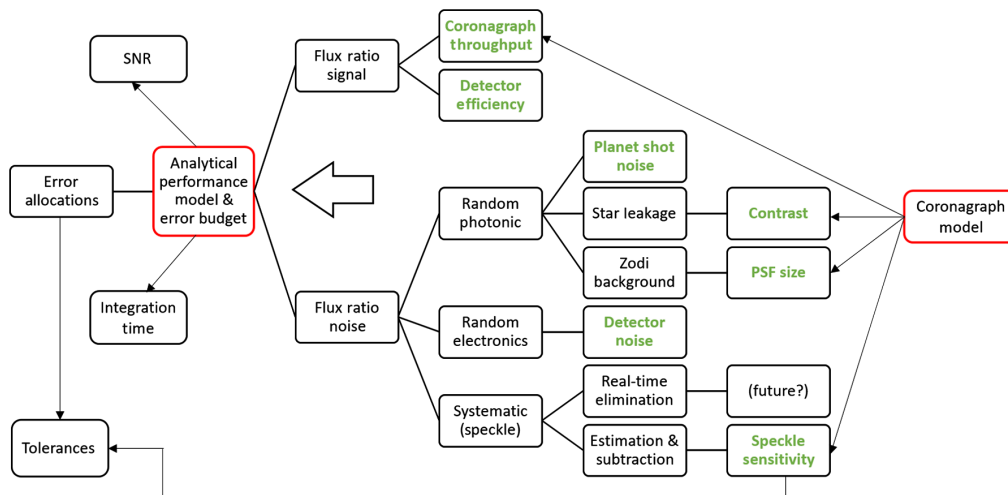


Fig. 3 Deriving optimal tolerances require an integrated approach between the error budget and the coronagraph performance model. The coronagraph model is based on specific design parameters including telescope architecture and coronagraph design. The two models (the coronagraph model and the analytical model) are highlighted with red borders, whereas the inputs to the analytical model are highlighted in bold green font.

a given SNR. The analytical performance model also provides the error budget. The second major component is the diffraction model of the optical system, labeled “coronagraph model.” This is a much larger and more computationally intensive model, where an incoming wavefront’s propagation through the coronagraph is simulated all the way to the focal plane. The two models together provide a comprehensive set of products, including the error budget allocations and tolerances to various modes of wavefront instability. This methodology draws from the Roman coronagraph error budget approach.²²

3 Coronagraphs and Their Key Attributes

Coronagraphs offer a compact, on-demand, ready apparatus for suppressing starlight and detecting planet images and spectra. The advantage of coronagraphs relative to external occulters is that, when built into a space telescope, a coronagraph offers the advantage of access to a large field of regard on the sky at any given time. The main components of a basic coronagraph are illustrated in Fig. 4. Only one DM is shown, but many designs use two DMs to control both the phase and amplitude of the incoming light before entering the masks. The first mask, when present, usually shapes (“apodizes”) the amplitude profile. This mask is often referred to as the shaped pupil mask or apodizer mask. From this pupil, the light is focused onto a focal plane mask (FPM), which modifies the central part of the starlight image-plane electromagnetic field. The planet light, which comes in at a slight angle, misses this mask in part and proceeds less altered. After recollimation, a third, so-called “Lyot” mask removes the largest portion of the remaining on-axis light. After this final alteration, the beam is focused onto the image plane of a detector, creating what is usually referred to as a “dark hole,” a zone where starlight has been strongly suppressed.

The key attributes of any coronagraph needed to derive a direct-imaging error budget are core throughput (Sec. 3.1), inner and outer working angles (OWAs) (Sec. 3.2), and raw contrast (Sec. 3.3).^{7,8,17,23} Radially, the inner working angle (IWA) of the dark hole is set by the loss of core throughput and increase of starlight leakage. The OWA, beyond which starlight suppression is not provided by the coronagraph, is usually set by the number of DM actuators. The IWA will be defined more formally in Sec. 3.2. Note that while these parameters are helpful in describing the shape of the dark hole, as will be discussed in Secs. 3.4 and 6, discriminating among the different coronagraph approaches requires a more holistic systems-engineering approach.

Arguably, the most important attribute for a telescope to be used with a coronagraph is its collecting aperture geometry. The ideal telescope aperture, from the standpoint of starlight suppression, is an unobscured circle (i.e., an off-axis telescope with a monolithic primary). An unobscured circle has well-defined diffraction properties and is easier to control. For a telescope with a circular collector aperture of diameter D , diffraction causes the broadening of the image to a “PSF” whose full-width at half-max (FWHM) scales with λ/D , where λ is the mean wavelength of the detection band. A smaller D implies a larger IWA, and hence a smaller maximum distance

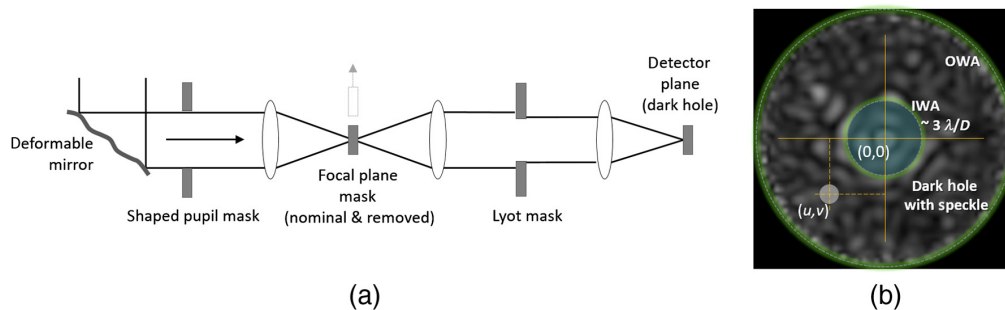


Fig. 4 Typical coronagraph setup. Incoming light from the left is shaped in phase by a DM (or in phase and amplitude by two separated DMs), then sent through a succession of masks. The result in the final focal plane is a dark hole where (on-axis) starlight is strongly suppressed relative to the (off-axis) planet light.

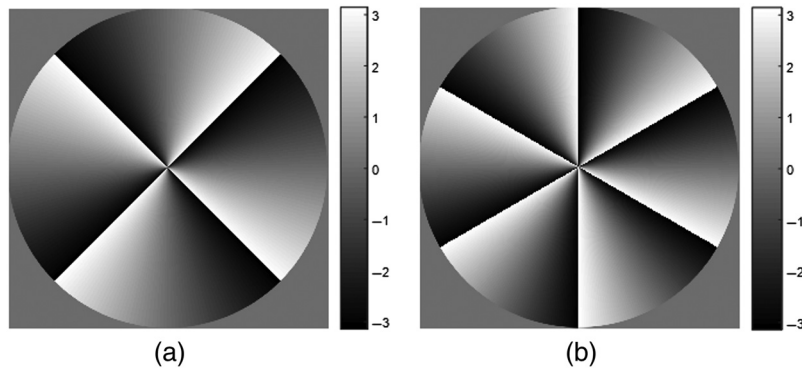


Fig. 5 The FPM phase maps for the two vector-vortex coronagraph cases presented: the (a) charge-4 and (b) charge-6. The gray-scale color bars indicate phase in radians. In the vector-vortex case, the initial wavefront is assumed to be flat. There is no pupil apodization nor FPM amplitude variation. The Lyot mask is a simple circle whose diameter is 90% of the pupil diameter.

out to which planets in circumstellar habitable zones can be directly imaged. Telescopes with larger diameters reach farther, but larger diameters may require segmenting the primary and going on-axis (with a central obscuration for a secondary mirror). For a space telescope, the limit on the primary mirror diameter usually comes from the launch vehicle fairing size and mass constraints. An off-axis configuration is preferred because a central obscuration and its associated struts significantly degrade coronagraph performance: diffraction from pupil discontinuities can be suppressed but at a high cost to the coronagraph throughput and IWA.

To illustrate these points, this paper considers three telescope cases: (1) an unobscured, off-axis telescope with a 4-m monolithic primary mirror, (2) an obscured, on-axis, 6-m segmented aperture telescope, and (3) an off-axis 6-m segmented aperture telescope. To achieve this for the segmented aperture telescopes, we are imposing an arbitrary circular aperture onto the primary mirror. Also for the segmented cases, for simplicity, we assume no gaps. Case 1 is the current baseline for the HabEx telescope. Case 2 is similar to JWST, whereas case 3 is similar to the LUVOIR alternative design. For case 1, we studied three different coronagraphs: a vector-vortex coronagraph²⁴ with charge-4 (VVC-X4) and charge-6 (VVC-X6) variations (Fig. 5), and an HLC (Fig. 6).²⁵ For case 2, we used an APLC originally designed for the ATLAST study (Fig. 7).²⁶ And, for case 3, we studied one coronagraph, the VVC-6.¹⁷ Note that the HLC and segmented APLC designs may not be current and do not necessarily represent their current best performance.

We now proceed to define the basic coronagraph attributes that will be needed in this study.

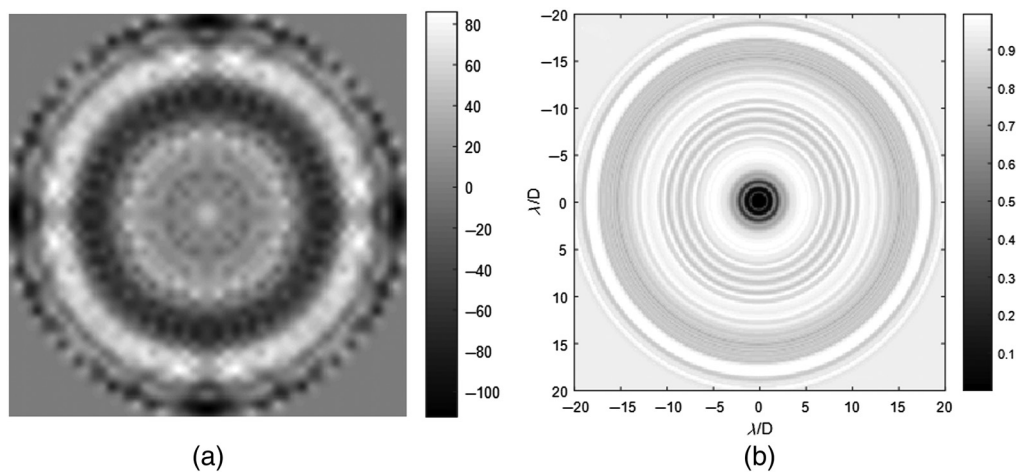


Fig. 6 The HLC design used here features predefined shapes for two DMs and an azimuthally symmetric FPM. (a) The wavefront specification for the first DM is shown, with the color scale in units of nanometers. (b) The transmission of the FPM is shown.

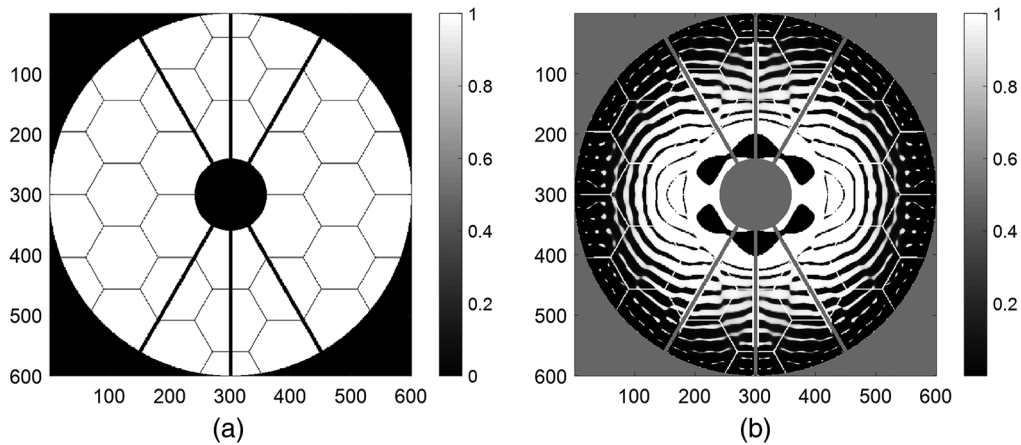


Fig. 7 (a) For the on-axis, obscured, segmented-primary telescope case, the pupil transmission appears, showing the obscurations from the secondary mirror assembly, the struts, and the mirror intersegment gaps. (b) The transmission map for the corresponding (APLC) shaped pupil mask is shown.²⁶

3.1 Core Throughput

A key attribute of a coronagraph is its core throughput. The planet PSF's “core” is defined as the area circumscribed by its half-max contour. Core throughput is the fraction of the planet light collected by the telescope primary mirror that ends up inside the core region (see Fig. 8). The photometric SNR is influenced most strongly by the high-signal part of the PSF, and the core is a good representation of that domain. The solid angle subtended by the core Ω_{core} depends on the aperture. For an unobscured circular collector (primary mirror) and no coronagraph, the PSF core is a circle of diameter $\sim 1 \lambda/D$.

Core throughput includes two effects: the loss of light due to partial or complete obscuration by the masks (particularly the FPM) and the spread of the PSF beyond the core boundary. Both of these are diffractive effects. In searching for Earth-like planets, to get a sufficient sample of Sun-like stars, one must search as large a volume of space as possible. This, in turn, drives the need for good performance at smaller working angles. The IWA is one metric that is sometimes used for this purpose.

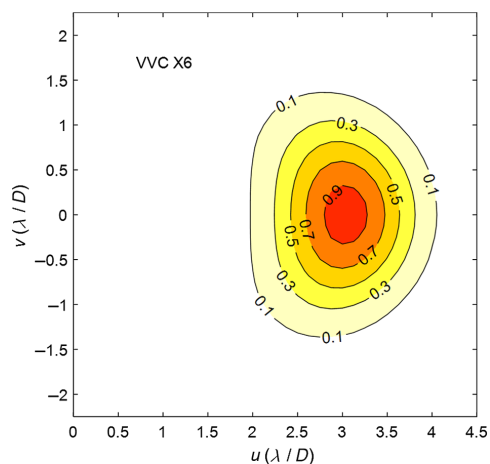


Fig. 8 The PSF for a vector-vortex coronagraph charge-6 (VVC-X6), when the source is $3 \lambda/D$ from the nominal LOS. The core area is enclosed by the half-max (0.5) contour. This core area corresponds to some solid angle Ω_{core} on the sky. The PSF is asymmetric because of the proximity of the source location to the IWA of the coronagraph, which is $2.3 \lambda/D$ in this case.

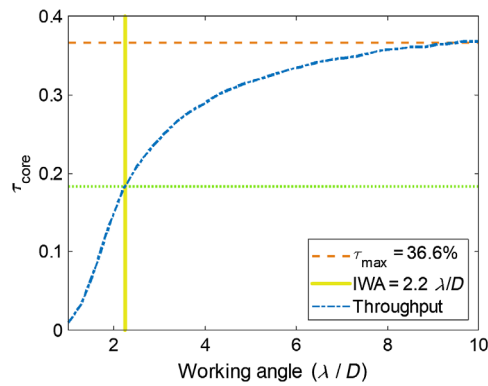


Fig. 9 Example of azimuthally averaged core throughput. Shown is the vector-vortex coronagraph, charge-4 case. The maximum throughput is about 38%. This is indicated by the orange, dashed horizontal line. Another such line, in green, indicates the half-max level, and the IWA is indicated by the intersection point of the throughput curve and the half-max line. The IWA is seen to be $1.6 \lambda/D$.

3.2 Inner and Outer Working Angles

The IWA is defined as the angular separation from the line of sight (LOS) below which the azimuthally averaged core throughput falls below 1/2 of its maximum value within the dark hole. As an example, Fig. 9 shows the azimuthally averaged core throughput for the vector-vortex coronagraph, charge-4 case. The maximum throughput is seen to be 38%. The IWA for this coronagraph is a very favorable $1.6 \lambda/D$, but it has good throughput ($>5\%$) all the way down to $1 \lambda/D$.

The existence of an OWA can be inherent to the coronagraph architecture (for example, if the FPM is designed with an annular opening), but, even when it is not, it is practically limited by the number of actuators in the DM. This is because the DM must always be used for high contrast, at least to compensate for the optics imperfections.

3.3 Raw Contrast

The coronagraph attribute to compare with the planet flux ratio is its raw or initial contrast. The qualifier raw is used to distinguish it from the residual contrast after differential imaging and other postprocessing. Unless otherwise indicated, we will henceforth use the unqualified form to refer to raw contrast.

Contrast is the measure of the effectiveness of the coronagraph in suppressing starlight near the planet. Considering an angular location (u, v) within the field of view, the raw contrast is the ratio of the starlight scatter throughput to that point, over the planet throughput at that point.

For a star located at the nominal LOS of the instrument, angular coordinates $(0, 0)$, the throughput to (u, v) is given by the fraction of the incident light from the light source that ends up within some region of interest Ω_r , centered at (u, v) . By “the incident light,” we refer to the total power incident on the usable, unobscured portion of the collecting aperture, which is the telescope primary mirror. We label this throughput as $\tau(u, v)$: a quantity evaluated at (u, v) , with the source at $(0, 0)$. The reference region of interest Ω_r can be thought of equivalently as a two-dimensional area on the image plane or a solid angle on the sky. In hardware, the width of Ω_r is typically chosen to be that of a detector pixel. In modeling, it is chosen to be that of a modeling pixel. The exact choice is not critical as long as it is small compared to λ/D . For a planet located at (u, v) , the throughput is simply the fraction of the incident light that is detected within Ω_r . We label this as $\tau_{pk}(u, v)$: i.e., the throughput into a reference region centered at (u, v) , with the source located also at (u, v) . Contrast at (u, v) is simply the ratio of the two throughputs:

$$C(u, v) \equiv \frac{\tau(u, v)}{\tau_{pk}(u, v)}, \quad (3)$$

This way of defining contrast makes the correspondence between flux ratio (a planet attribute) and contrast (an instrument attribute) more direct and free from hidden, uncommon throughput factors: the numerator and denominator both are evaluated at the same location, namely (u, v) .

Evaluation of contrast is inherently cumbersome: to map out the contrast one needs to evaluate the denominator, which means placing a test source (whether in a hardware test or in computer modeling) at a grid of (u, v) points within the relevant part of the field of view. In a testbed, this means the incoming beam is tilted with a mirror so that it now appears to come from (u, v) , while in modeling the incoming wavefront phase is given a tilt of (u, v) . By comparison, the numerator is obtained from a single image, whether in hardware or in modeling. Because of the time-consuming nature of measuring contrast, a simplified approximation is often computed, called normalized intensity (NI). Its definition is very similar to contrast:

$$NI(u, v) \equiv \frac{\tau(u, v)}{\tau_{nf}(u, v)}. \tag{4}$$

The numerator is the same, but the denominator is now also from a single image, from a source located at $(0, 0)$, only now with no FPM (hence the subscript *nf*). In hardware, the FPM is temporarily removed (as illustrated by the dashed rectangle near the FPM in Fig. 4) and in computer models, the matrix representing the FPM is replaced with a matrix of 1 s (which leaves the field unchanged after multiplication).

Henceforth, the term working angle (designated by α) will be used to refer to the separation angle between a point of interest (such as a planet location) and the LOS in λ/D units: $\alpha = \sqrt{u^2 + v^2}/(\lambda/D)$. Also a subscript of α (such as in C_α defined above) shall imply an azimuthally averaged quantity where the remaining dependence is on the radial distance α in the dark hole. For a radial band of width $\delta\alpha$ centered on working angle α , the azimuthally averaged contrast is given by

$$C_\alpha = \frac{1}{2\pi\alpha \cdot \delta\alpha} \int_{\alpha-\frac{1}{2}\delta\alpha}^{\alpha+\frac{1}{2}\delta\alpha} \int_0^{2\pi} C(\alpha', \phi) d\alpha' d\phi, \tag{5}$$

where $\phi = \tan^{-1}(v/u)$ is the azimuthal coordinate corresponding to (u, v) . A similar definition applies to NI_α . (In what follows, plots of quantities versus α will always implicitly mean azimuthally averaged quantities, and the subscript α will be dropped in those cases.) Figure 10 shows the ratio C_α/NI_α for a number of coronagraph cases. Note how the ratio (and hence the

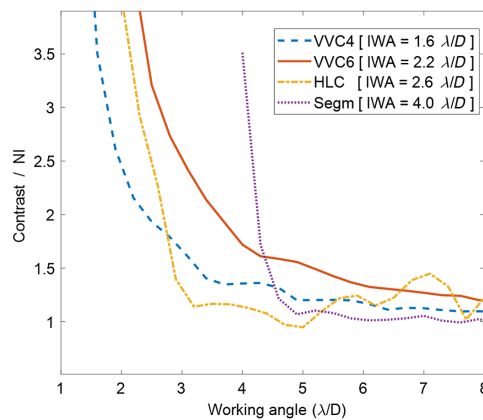


Fig. 10 The importance of using contrast over NI is shown here by plotting the azimuthally averaged synthetic contrast C_α over NI_α for the different coronagraph cases as a function of working angle. The legend indicates the IWA for each case. Near the IWA, the ratio C_α/NI_α is seen to be quite significant.

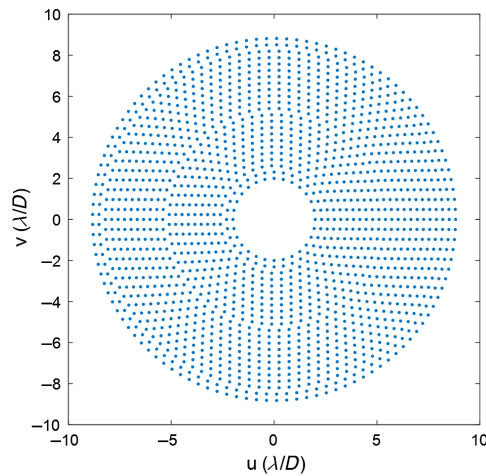


Fig. 11 Example of a grid of pointing offsets (u, v) for evaluating contrast and core throughput.

difference between C and NI) is quite substantial as the working angle approaches the IWA for every coronagraph.

The appeal of using (NI) instead of contrast C is that it replaces hundreds of propagations (each with a different incoming wavefront) with a single operation (in hardware) or propagation (in software). For example, as shown in Fig. 11, to evaluate the contrast from 2 to 9 λ/D in radial steps of 0.3 λ/D , a total of 1703 separate “pointings” (u_i, v_i) are needed, if 2 azimuthal samples per λ/D are used. The disadvantage of NI is that it tends to underestimate the contrast at small working angles. This is because contrast accounts for throughput loss for imaging off-axis sources—i.e., planet throughput (see core throughput below)—while NI does not (due to the FPM being removed). The difference between C and NI diminishes as the source working angle α becomes larger than the IWA by a few λ/D . Conveniently, it is just as NI becomes a good approximation that calculating C becomes most cumbersome. The number of pointings (u, v) is small at the most important, smaller working angles α , while it grows as $2\pi\alpha$ as α is increased. This fortuitous condition can be exploited by defining a “synthetic contrast, C_{syn} ,” which equals contrast near the IWA, equals NI near the OWA, and transitions at some intermediate working angle α_t between the two, given by

$$C_{\text{syn}} = t \cdot C_{\alpha} + (1 - t) \cdot \text{NI}_{\alpha}, \quad (6)$$

where $0 \leq t(\alpha) \leq 1$ is a transition function. A workable choice for t is $t(\alpha) = 1 / \{1 + \exp[(\alpha - \alpha_t) / \alpha_s]\}$, where α_t and α_s are the free parameters determining the point of transition and the sharpness of the transition, respectively. This makes it possible to avoid generating all the pointings out to the OWA, but instead only out to some intermediate working angle.

Figure 12 shows contrast, NI and synthetic contrast for the VVC-6 case. The IWA for this coronagraph is 2.3 λ/D , and, as can be seen from the plot in Fig. 10, NI underestimates C by a factor of ~ 4 near the IWA. The discrepancy differs for other coronagraph cases, but this case illustrates how important the distinction is between C and NI. In the literature, sometimes the distinction between C and NI is not made, and the quantity called “contrast” is, in fact, NI. But leakage should be measured in C , and NI does not approximate C well enough near the all-important IWA. The IWA is usually the region of greatest interest for exoplanet searches.

For all cases of interest $C \ll 1$ and as a result, it can be shown that the contrast C is proportional to the square of the wavefront error w (i.e., $C \propto w^2$). The proportionality factor depends on the wavefront error mode and the coronagraph design (masks and DM configuration).

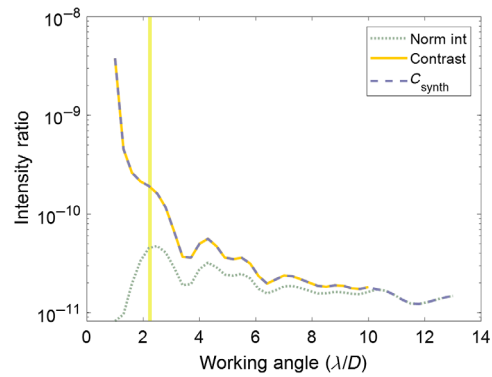


Fig. 12 Azimuthally averaged values of contrast (solid), NI (dotted), and synthetic contrast (dashed) for the vector-vortex coronagraph, charge-6 case. The transition working angle is chosen near $9.5 \lambda/D$.

3.4 Comparing Core Throughput and Raw Contrast among Architectures

So far, we have shown plots of core throughput and contrast in the more common λ/D units. However, comparisons in these units are inconvenient. For example, at 550 nm, λ/D is ~ 28 mas, for $D = 4$ m and ~ 19 mas for $D = 6$ m. Hence, the exo-Earth at 100 mas separation would be located at $\sim 3.6 \lambda/D$ for a 4-m telescope and $\sim 5.3 \lambda/D$ for a 6-m telescope.^{27,28} To facilitate meaningful comparison of candidate architectures, it is better to plot core throughput in scientifically relevant angular separation units. Figure 13 plots core throughput versus angular separation for four cases: three different coronagraphs (VVC-4, VVC-6, and HLC) on a 4-m off-axis monolithic aperture telescope and an APL coronagraph on a 6-m on-axis segmented aperture telescope. Note that Fig. 13 excludes throughput losses other than core throughput. Losses, such as those from reflection off mirrors or transmission through filters, are bookkept in the photometric calculations. Two key takeaways from Fig. 13 are the following: (1) different coronagraphs on the same telescope have significantly different core throughputs and (2) the central obscuration of an on-axis telescope greatly reduces core throughput.

It is similarly helpful to plot raw contrast versus angular separation. But the value of the raw contrast achieved depends on wavefront control performance. In keeping with the modeling approach outlined in Sec. 4, we use a per-design specified wavefront plus an additional post-wavefront control surface error. There is sufficient difference between the raw contrast of the

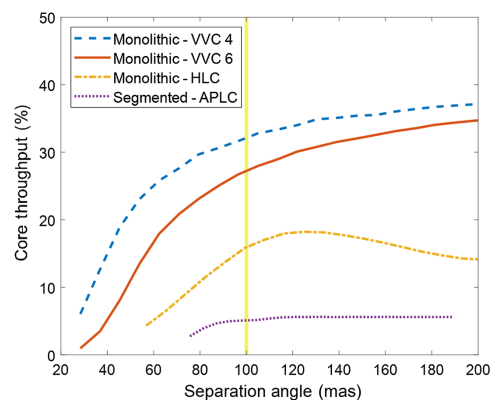


Fig. 13 Comparison of core throughput versus separation angle for five cases (three coronagraphs with a 4-m off-axis monolithic and two 6-m segmented cases). The 6-m off-axis case with a VVC6 has the same core throughput profile as the 4-m off-axis monolith with VVC4. The separation for an exo-Earth at 10 pc (100 mas) is indicated with the vertical line. Note that while the 6-m (on-axis) segmented aperture has $\sim 2\times$ more collecting area than the 4-m aperture, its “comparative” throughput at 100 mas is only $\sim 5\%$.

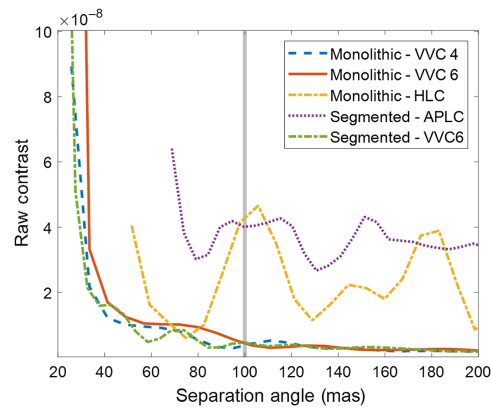


Fig. 14 Comparison of raw contrast versus separation angle for four architectures. Raw contrast depends on wavefront control, but here we instead assumed a postcontrol residual surface error of 120 pm rms. Notice that the vortex coronagraphs, particularly charge-4, have much smaller IWAs and a reach that is many times that of the obscured segmented case.

different architectures to reliably distinguish the two vector-vortex cases as a group from the HLC case and the APLC case. Figure 14 shows the raw contrast versus angular separation for each case with an assumed postcontrol residual surface error of 120 pm rms (picometers root-mean-square). Again, the key takeaways are that different coronagraphs perform differently and that a central obscuration significantly degrades performance. Also the off-axis unobscured monolithic aperture with a vector-vortex coronagraph has the smallest IWA.

4 Formulating the Error Budget for Exo-Earth Detection

The first step in forming an error budget is to choose a representative science target and a corresponding observing scenario. Tied closely to that is also an error metric which will form the exchangeable “currency” of the error budget, allowing trades of allocations to different error sources. Any real mission, of course, will involve many targets of various kinds with different scientific interests. But a challenging objective can be taken as the enveloping case for the purposes of setting requirements. For this study, we chose the detection of an Earth-like planet in the habitable zone of a nearby Sun-like star.

The second step is to choose an error metric that forms the exchangeable currency of the error budget—allowing trades of allocations to different error sources. Typically, this metric for space telescopes is rms wavefront error. But, for the case of exoplanet direct imaging and photometry, the more suitable metric is the noise in measuring the planet flux ratio. To directly image an exo-Earth, its PSF needs to be clearly discernable against the background arising from the residual starlight halo. As discussed in Sec. 1, the flux ratio for an exo-Earth at 10 pc is 210 ppt. If we require an SNR of 7, then the combined noise from all sources (including the residual starlight speckle) must contribute no more than $\sim 210/7 = 30$ ppt in noise. Also important is where this exo-Earth’s PSF is located relative to the diffractive fundamentals of the instrument. If we assume that the telescope has a primary mirror diameter of $D = 4$ m and is operating in the visible band (e.g., $\lambda = 550$ nm), this level of starlight suppression must be achieved at $3.5 \lambda/D$ from the nominal LOS (100 mas separation). For targets that are closer to us or for planets orbiting farther from their host stars, the requirement has to be met farther out in separation, which is easier to achieve.

To reiterate, our error budget will be based on the noise accompanying the planet flux ratio measurement and must roll up to 30 ppt total. It includes fundamental (inevitable) effects, such as the photon noise associated with the detection, as well as potentially improvable imperfections in the telescope and coronagraph. In addition to the planet, target specification must include some assumptions about the exosolar system, e.g., that the host is a sun-like G star with an absolute magnitude of 4.8 (like our sun). We also assume the exo-zodi is $3\times$ solar in optical depth.

This somewhat conservative choice is motivated by the current lack of knowledge about dust characteristics around the nearby stars.

4.1 Flux Ratio Noise as the Error Budget Metric

As described in Appendix A, the planet (electron) count rate is given by

$$r_{\text{pl}} = F_{\lambda} \Delta\lambda \xi_{\text{pl}} A \tau_{\text{pl}} \eta, \quad (7)$$

where F_{λ} is the spectral flux, $\Delta\lambda$ is the filter bandwidth, ξ_{pl} is the planet flux ratio, A is the collecting area, τ_{pl} is the throughput for the planet light, and η is the detector quantum efficiency. The signal count after integrating over some time t is given by

$$S = r_{\text{pl}} t. \quad (8)$$

The photometric quantity of interest is the planet flux ratio ξ_{pl} , which is proportional to S :

$$\xi_{\text{pl}} = (F_{\lambda} \Delta\lambda A \tau_{\text{pl}} \eta t)^{-1} \cdot S, \quad (9)$$

We define κ as the “flux ratio factor:”

$$\kappa \equiv (F_{\lambda} \Delta\lambda A \tau_{\text{pl}} \eta t)^{-1}. \quad (10)$$

Note that the flux ratio factor depends on observing scenario parameters, such as the optical band and the total integration time. Since ξ_{pl} is the flux ratio, the noise in this quantity can be written as $\delta\xi_{\text{pl}}$.

Although the signal S consists of photoelectrons at the detector over some integration time t , the noise comes from a variety of sources. We enumerate these as: (1) shot noise in the planet signal (σ_{pl}), (2) shot noise in the underlying speckle (σ_{sp}), (3) shot noise in the underlying zodiacal dust background (local + exo) (σ_{zo}), (4) detector noise (σ_{det}), and finally (5) the residual speckle instability error $\sigma_{\Delta I}$. The total variance is given by

$$\sigma_{\text{tot}}^2 = \sigma_{\text{pl}}^2 + \sigma_{\text{sp}}^2 + \sigma_{\text{zo}}^2 + \sigma_{\text{det}}^2 + \sigma_{\Delta I}^2. \quad (11)$$

The first four of these contribute random noise to the signal, and their variance increases only linearly with time. The last source has a variance that often grows faster, typically as t^2 .

If σ_{tot} is the total noise associated with the signal S , then the noise in measuring the planet flux ratio is given by

$$\delta\xi_{\text{pl}} = \kappa \cdot \sigma_{\text{tot}}. \quad (12)$$

This quantity, which we simply refer to as flux ratio noise, is the error budget metric. This is the metric used by the Roman coronagraph. Just as the different contributors to σ_{tot} add up in quadrature, by linearity so do the corresponding contributors to $\delta\xi_{\text{pl}}$. If σ_i is the i 'th contributor to the noise in S , then its contribution to the flux ratio noise is simply:

$$\delta\xi_i = \kappa \cdot \sigma_i. \quad (13)$$

The noise terms σ_i are in units of electron counts, and, when multiplied by κ , they become noise-equivalent flux ratio, or flux ratio noise. The error budget boxes are thus the set of $\delta\xi_i$.

4.2 Strawman Observing Scenario for the Error Budget

Having identified the science objective and the error metric, the next step is to form a representative observing scenario. This is also called a “strawman” scenario in the sense that it may not correspond to any actual observation in perfect detail, but contains enough of the aspects of the expected observations to fill the convenient role of a single operating concept for quantitative analysis.

For our target case, the scenario includes a certain period of staring, pointed at the host or target star. The coronagraph operates to suppress unwanted starlight and any system errors. Some filtering is employed in ultrahigh-precision measurements to assist in driving down the errors. For example, if the errors were strictly random (such as photon noise or detector noise), simply extending the observation duration t would reduce the relative error at a rate of $1/\sqrt{t}$. This by itself would call for a long integration time in the observing scenario. However, in a real instrument, unsensed drift errors, possibly from thermal sources, begin to dominate as integration times are lengthened. The effect of these drift errors is variation in the starlight residual speckle (i.e., contrast instability). Some new technical innovations are currently being developed for sensing speckle instability in real time,²⁹ but as of this writing it remains to be seen whether they will be feasible. At present, the best understood method of mitigating speckle instability is via chopping, where a measurement is taken of a reference star and subtracted from the target star measurement. In the context of coronagraphs, this is called differential imaging. Another form of differential imaging, called angular differential imaging, is based on observing the same star at different roll angles. For either method, if the speckle subtraction is perfect, the residual image has no speckles. But it will still have the shot noise of the subtracted speckle patterns—which can be reduced via longer integration time.

Many types of differential imaging have been employed in the various ground-based coronagraphs currently in operation. One of the more common techniques, and one currently based by the Roman coronagraph, is called reference differential imaging (RDI).³⁰ It calls for the instrument to point to a “reference” star to generate the dark hole and point back to the reference star every few hours to recalibrate the dark hole (Fig. 15). The final image is the sum of separate reference-subtracted images, in each of which a reference-star image is subtracted from a target-star image (see also Sec. 1 and Appendix A). For the purpose of this paper, we adopt an RDI observing scenario.

What differential imaging makes possible is a relaxation of the requirements on direct starlight suppression, if the speckle pattern in the reference image is sufficiently close to that of the target image. Typically, the goal is to achieve an order of magnitude improvement in the suppression using differential imaging. This benefit comes at the cost of a requirement on the stability of the optical system—a requirement that is often the most challenging in a coronagraph, and hence one of the most important tolerances to determine. It is the purpose of this paper to develop a process for determining these tolerances. Note that the stability error budget (Sec. 6) applies to the telescope from the end of the reference integration period to the end of the target integration period. Telescope stability sets the chop period. For a telescope with no instability, there is no need to return to the reference star.

The foregoing discussion, however, should not be interpreted as implying that raw contrast is not important. Understanding the interplay between the existing residual starlight and its change arising from optical instability is important to the analysis that follows. At the end of the wavefront control procedure that gives the coronagraph its final level of starlight suppression, some residual optical error remains, leading to a “leakage” field $E(u, v)$ in the image plane (where u, v are image plane coordinates). For this discussion, it is adequate to think of this as a complex scalar function of position. The speckle intensity pattern is then simply given by $|E(u, v)|^2$.

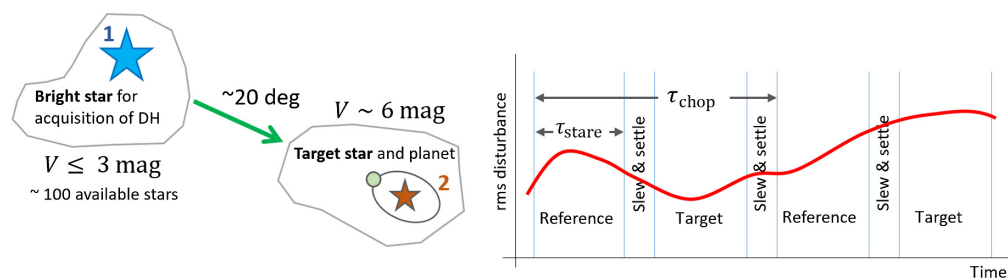


Fig. 15 A Roman-like observing scenario for coronagraph applications. A bright reference star serves both as an efficient object for dark-hole creation and to provide a reference speckle pattern for differential imaging.

When a disturbance or drift error occurs in the optomechanical configuration of the telescope or coronagraph, this field changes by a small amount. We can equivalently think of this as a perturbation field $\Delta E(u, v, t)$ being coherently added to the “initial” or “static” part of the field $E(u, v)$. The coherent mixing of the original field E and perturbation field ΔE creates the modified speckle pattern, the intensity of which is given by

$$|E + \Delta E|^2 = |E|^2 + |\Delta E|^2 + 2\mathcal{R}\{E^* \Delta E\}. \tag{14}$$

The contrast at any moment in time is this mixed quantity. Note that the mixing term $2\mathcal{R}\{E^* \Delta E\}$ is not positive-definite like the first two terms: it can be positive or negative. For example, consider a perturbation field ΔE that is small and oscillates just in amplitude, sinusoidally, as illustrated in Fig. 16. Given the amplitude of the multiplication factor, the mixing term drives the temporal observing strategy. Thus if the measurement integration period is sufficiently long, the mixing term will average to zero and the only impact to contrast is the average perturbation modulus. Similarly, if the integration period is much shorter than the mixing term’s period, then the mixing term will appear as a slow drift and its impact can be mitigated by averaging multiple independent (i.e., uncorrelated) measurements or more frequent RDI operations. The problem that arises is when the integration period duration is close to the mixing term’s period.

Viewed in terms of filters, observing scenarios can be designed to reduce the effect of the instability terms (usually dominated by the mixing term) by a judicious choice of integration times and chopping. Integration is a low-pass filter, and chopping, which is temporal differentiation, is a high-pass filter. Their combined application can produce a limited band filter to minimize the impact of disturbances.

Another implication of the mixing term is that the amplitude of raw contrast or initial contrast is important in determining speckle amplitudes, so that the assumption of initial contrast cannot be decoupled from setting requirements on optical stability. We will revisit this in Sec. 5 when we discuss the modeling of the modal instability errors.

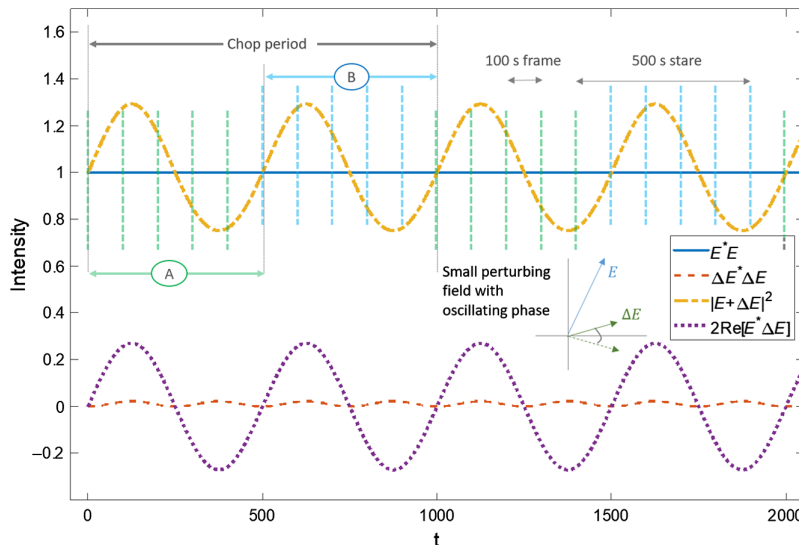


Fig. 16 Example case of mixing fields, with nominal chop (RDI switch) segments, highlighting individual image frames and single-stare integrations. The perturbing field ΔE , if coherent with the larger existing field E , is amplified when mixing with this field. If the perturbation (ΔE) has multiple oscillations over a long integration time, the mixing term ($2\mathcal{R}\{E^* \Delta E\}$) will average to zero, leaving only the perturbation term ($|\Delta E|^2$). If the perturbation term is slow, the mixing term has a large effect. The perturbation is defined as the change in the field between the reference and measurement.¹⁴

4.3 Converting Contrast Instability to Flux Ratio Noise

Earlier, we derived an expression for calculating the flux ratio noise that arises from a source of photometric noise. Having introduced some of the considerations with regard to speckle instability, we now derive the corresponding relationship between speckle instability and flux ratio noise.

In each RDI differential image, there are two stars involved: one at the target star and one at a reference star. There are, correspondingly, two speckle patterns. We can label the two-dimensional contrast maps in the dark hole as $C_{\text{tar}}(u, v)$ and $C_{\text{ref}}(u, v)$, for the target and reference stars, respectively. We will call their difference the residual contrast map $\Delta C(u, v)$:

$$\Delta C(u, v) = C_{\text{tar}}(u, v) - C_{\text{ref}}(u, v). \quad (15)$$

The spatial nonuniformity of ΔC within the dark hole causes confusion noise in the differential image. This is quantified by the spatial standard deviation (SSD), on the λ/D scale of ΔC :

$$\sigma_{\Delta C} = \text{SSD}[\Delta C(u, v)]. \quad (16)$$

Further reduction of the residual speckle through postprocessing has been shown to be possible in certain circumstances.³¹ Without going through the various possible postprocessing algorithms, we simply summarize their impact by assuming a further “postprocessing factor” f_{pp} , a number between 0 and 1 that, when multiplied by $\sigma_{\Delta C}$, gives the final residual contrast. Conversion of this quantity to the differential imaging flux ratio noise $\delta\xi_{\Delta I}$ is derived in detail in Appendix A. Here we merely quote the result:

$$\delta\xi_{\Delta I} = \kappa_c \cdot f_{\text{pp}} \cdot \sigma_{\Delta C}, \quad (17)$$

where κ_c is the flux ratio noise factor for contrast instability and can be derived using a diffraction model of the coronagraph. In Appendix A, it is shown that κ_c is given by

$$\kappa_c = \frac{\tau_{\text{pk}} n_{\text{core}}}{\tau_{\text{core}}}. \quad (18)$$

Recall, from Eq. (3), that τ_{pk} is the throughput to a pixel in the dark hole. Thus κ_c can be thought of as the ratio of the throughput per pixel at the peak of the PSF, over its average within the core.

Since the peak of the PSF is centered within the region of interest in this case, $\tau(u, v)$ is also referred to as the peak throughput (τ_{pk}). The numerator contains the peak throughput τ_{pk} and n_{core} the number of diffraction modeling pixels in the image plane covering the PSF core. Though not exactly 1, κ_c is usually not far from 1. For the vector-vortex charge-6 coronagraph, near the IWA $\kappa_c = 1.12$.

4.4 Integration Time Needed to Achieve SNR

The observing scenario, in which the error budget is based, includes the all-important integration time. Time is a key parameter for at least two reasons. First, different types of errors have different time dependencies: random errors can be reduced by integrating longer, while (systematic) drift errors (such as thermal errors or DM actuator drift) grow with time. Second, for a space mission, integration time is a scarce resource, and target selection and the science objectives cannot be decided without counting this cost. Thus the integration time chosen for the observing scenario must be realistic from the standpoint of the random and drift errors. Success in direct detection of a planet can be parameterized in terms of achieved SNR in a given amount of time, or conversely the time required to achieve a desired SNR. This time defines the maximum desired duration for the telescope’s wavefront stability. If sufficient stability within this duration cannot be achieved, then the dark hole will need to be recalibrated. In this section, we develop an analytical expression for the time required to achieve the desired SNR and use the result to calculate the random noise part of the error budget.

To begin, SNR is simply the ratio of the signal S to the noise N :

$$\text{SNR} = S/N. \quad (19)$$

There are two kinds of SNR that could apply to direct imaging, depending on the goal of the observation. If we are merely interested in detection, the SNR requirement guards against a false positive. In the limiting case of a noise-free background, a single excess or signal event gives an SNR of infinity and absolute certainty of detection. This, of course, is never the case, but it serves to emphasize that for detection only the background noise matters:

$$\text{SNR}_{\text{det}} = S/\sigma_B, \quad (20)$$

where σ_B is the background noise. Setting the SNR threshold in this case starts from choosing the false positive rate that is considered tolerable. For example, if the background follows Gaussian statistics, the false positive probability is given by

$$P_{\text{FP}} = \frac{1}{\sqrt{2\pi}\sigma} \int_z^{\infty} e^{-\frac{s^2}{2\sigma^2}} ds = \frac{1}{2} \text{erfc}\left(\frac{\text{SNR}_{\text{det}}}{\sqrt{2}}\right). \quad (21)$$

The function erfc is the complementary error function and $\sigma = \sigma_B$. For example, if the dark hole extends from an IWA of $2 \lambda/D$ to an OWA of $12 \lambda/D$ and if a planet signal falls on a core area of roughly $(\lambda/D)^2$, there are, in each direct image, about 110 planet-signal-sized core areas which have the potential to create a false positive. If we produce 200 such images over the course of the mission, 22,000 core areas could give false positives. Suppose we require the probability of a false positive to be $<1\%$ over the duration of the mission and assume Gaussian statistics apply to the background noise. Then the probability of having the background (in a single core area in a dark hole out of the 22,000 over the mission) look deceptively like a signal needs to be $<4.5 \times 10^{-7}$. Using Eq. (21), we see that if we choose $\text{SNR}_{\text{det}} = 5$, then $P_{\text{FP}} = 2.9 \times 10^{-7}$, which is low enough to meet our requirement. Therefore, for direct imaging, an SNR of 5 is adequate to limit the per mission false positive probability to $<1\%$, assuming Gaussian background noise statistics.

Two important caveats are in order at this point. The first is that the background, which is the result of an RDI differential image taken from images that contained speckles, may not be strictly Gaussian, which means Eq. (21) may be too optimistic.³² The second is that any analytical calculation that estimates S gives the expected value of S and not its distribution. That is, the SNR that we calculate represents (all else being equal) the average SNR from an ensemble of visits. In a typical direct-imaging case, tens of hours must be spent integrating to receive a significant and unambiguous signal. The assessment (partway through a planet-discovery observation) of whether or not a planet has been detected at some location in the dark hole depends on the significance of the excess. Even for the case of a true planet existing in the dark hole, when an integration time suggested by Eq. (21) has elapsed, in half of the cases the SNR will be less than the chosen threshold (because S will be less), and in half of the cases it will be more. To have a high probability that the lack of significance is due not to statistics but from the absence of a true planet, a larger SNR target than suggested by Eq. (21) will be necessary.

In most cases, detection is not the only goal; photometry (i.e., the calculation of the flux ratio of the planet) is also desired. For photometry, the uncertainty in the flux ratio measurement is an important factor. As such, the noise N must now also include the noise associated with the signal. Since the signal S follows Poisson statistics, with the special feature that the mean (S) is equal to the variance (σ^2), we can then say that the Poisson noise (commonly called shot noise) associated with the signal is simply \sqrt{S} . Then N will be the quadrature sum of the two noise contributions, and we get the following for the photometric SNR:

$$\text{SNR}_{\text{ph}} = \frac{S}{\sqrt{S + \sigma_B^2}}. \quad (22)$$

In the case of an exo-Earth concept mission, the coronagraph will be used for both direct imaging and spectroscopy. In both modes, the goal is not only discovery but photometry. We can think of a spectrum as a series of photometric measurements at consecutive spectral bins. Hence, we will hereafter only consider the photometric SNR.

We can re-express this SNR equation in terms of σ_{tot} as simply $\text{SNR} = S/\sigma_{\text{tot}}$. Furthermore, using Eq. (7), we can break up σ_{tot} into a random part (shown below) and a systematic part [which is just $\sigma_{\Delta I}$ in Eq. (7)]. The random variance is given by

$$\sigma_{\text{md}}^2 = \sigma_{\text{pl}}^2 + \sigma_{\text{sp}}^2 + \sigma_{\text{zo}}^2 + \sigma_{\text{det}}^2. \quad (23)$$

We now define a random variance rate r_n as

$$r_n = \sigma_{\text{md}}^2/t. \quad (24)$$

For the differential imaging error $\sigma_{\Delta I}$, which we expect to grow linearly with time, we define, instead of a variance rate, a standard deviation rate:

$$r_{\Delta I} = \sigma_{\Delta I}/t. \quad (25)$$

In Appendix A, we show that $r_{\Delta I}$ is also given by

$$r_{\Delta I} = (f_{\text{pp}} \cdot f_{\Delta C}) \cdot r_{\text{sp}} = f_{\Delta I} \cdot r_{\text{sp}}, \quad (26)$$

where $f_{\Delta C} = \sigma_{\Delta C}/\bar{C}$, per Eq. (60), is the dimensionless measure of the effectiveness of differential imaging. The factor $f_{\Delta I}$ is the differential imaging suppression effectiveness from both contrast stability and postprocessing. When r_{sp} is the speckle rate, $r_{\Delta I}$ can be thought of as the residual speckle rate. With these terms replacing the variances, the photometric SNR equation becomes

$$\text{SNR} = \frac{r_{\text{pl}}t}{\sqrt{r_n t + r_{\Delta I}^2 t^2}}. \quad (27)$$

Inverting this equation gives the time to reach the desired SNR:

$$t = \frac{\text{SNR}^2 r_n}{r_{\text{pl}}^2 - \text{SNR}^2 r_{\Delta I}^2}. \quad (28)$$

When used to describe the SNR and time to SNR in a differential (RDI) image, the planet signal count rate r_{pl} [see Eq. (38)] and random variance rate r_n [see Eq. (46)] must also include contributions from the reference star observation. Detailed expressions for these are derived in Appendix A.

In Eq. (28), the subtraction in the denominator causes a divergence in the dependence of the integration time on SNR. If the speckle subtraction is not effective (i.e., $r_{\Delta I}$ is too large) or the required SNR is too high relative to r_{pl} , the available count rate from the planet, the denominator, can vanish or become negative, indicating no solution. Thus it is useful to define, for a given observation, the critical SNR, SNR_{crit} , where the denominator goes to zero. This is the infinite-time limit of the maximum achievable SNR:

$$\text{SNR}_{\text{crit}} = \frac{r_{\text{pl}}}{r_{\Delta I}}. \quad (29)$$

A higher SNR is not achievable in any amount of time. Only a brighter planet with a higher r_{pl} can achieve a higher SNR.

The integration time needed to achieve a desired SNR is architecture-dependent. To calculate the integration time for a specific architecture, we can use the steps in Sec. 2 to compute the peak throughput (τ_{pk}), core throughput, PSF size on the sky, contrast, and the number of core modeling pixels. As outlined in Appendix A, the noise, planet, and speckle rates are all calculable based on the observing scenario assumptions for these quantities.

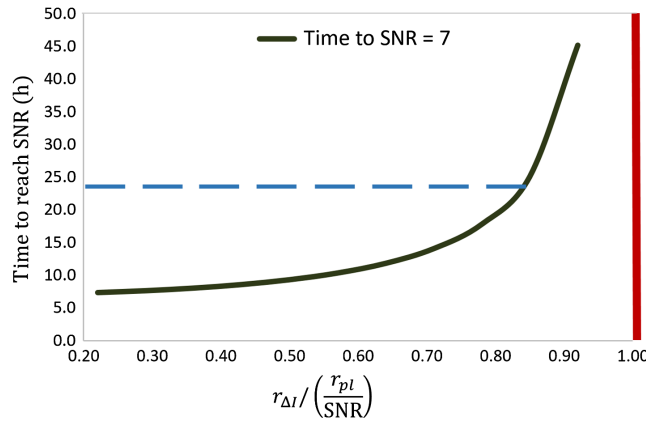


Fig. 17 The calculated time to reach SNR as the residual speckle $r_{\Delta I}$ is increased. In the horizontal axis, $r_{\Delta I}$ is normalized by r_{pl}/SNR . The rapid increase of needed integration time as $r_{\Delta I}$ is increased indicates that there are diminishing returns beyond about 25 h of integration time. We hence choose 25 h as the nominal integration time for the observing scenario.

As an example, Fig. 17 plots the time needed to reach an SNR of 7 for a 210-ppt flux ratio exo-Earth. The horizontal axis is $r_{\Delta I}$, normalized to r_{pl}/SNR . Per Eq. (28), when this quantity reaches unity, the time to SNR becomes infinite. Larger values of this quantity imply more relaxed requirements on speckle stability and postprocessing effectiveness, but, beyond some point, consuming further integration time to allow more relaxed requirements on speckle suppression has no value. We, therefore, choose an integration time of 25 h for the observing scenario.

4.5 Error Budget at the Top Level

With the target, observing scenario, and error metric all defined, it is now possible to create an error budget. For a 210-ppt exo-Earth target, desired to be observed with $SNR = 7$, the total error from all sources combined must be <30 ppt. Using equations derived in Appendix A for the photometric error contributions, Fig. 18 plots the total random noise and its constituents as a function of integration time. Also plotted in Fig. 18 is the maximum allowable systematic error from residual speckle, i.e., the quadrature difference between the total allowable error ξ_{tot} (30 ppt) and the total random error ξ_{rnd} :

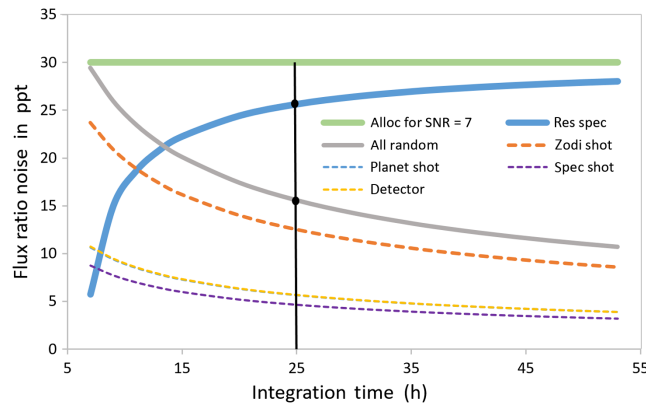


Fig. 18 The relative contributions to the flux ratio noise from the main noise sources versus integration time. The selected integration time of 25 h is highlighted. Coincidentally, the planet shot noise trace is hidden beneath the detector trace.

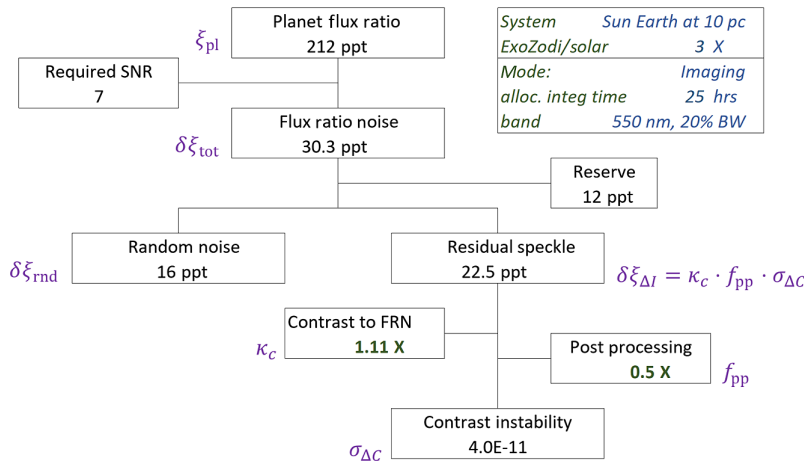


Fig. 19 Top-level error budget for direct imaging of exo-Earth at 10 pc.

$$\delta \xi_{\Delta I} = \sqrt{\delta \xi_{\text{tot}}^2 - \delta \xi_{\text{rnd}}^2} \tag{30}$$

For the strawman observing scenario, at an integration time of 25 h, Fig. 18 shows that the total expected random error is about 15.6 ppt and the total allowed residual speckle error is about 25.6 ppt.

A top-level error budget based on these numbers is shown in Fig. 19. The requirement of SNR = 7 on the exo-Earth means a maximum allowable flux ratio noise of 30 ppt for the target system of the observing scenario. As shown in Fig. 18, the random error, for a 4-m telescope with a VVC-6, is estimated to be just under 16 ppt, and we use this number for the allocation to random noise. The remainder, in a quadrature sense, goes to residual speckle and reserve. Here we choose 22.5 ppt for the residual speckle error $\delta \xi_{\Delta I}$, which leaves 12 ppt for reserve. Obviously, some degree of freedom exists at this point, but our 12 ppt of reserve amounts to a modest 16% reserve, in a quadrature sense, relative to the total allowed flux ratio noise of 30 ppt. Note that reducing the reserve to zero would only modestly increase the residual speckle allocation to 26 ppt. The last step is to calculate via Eq. (17) the allowed final residual contrast instability after postprocessing. Assuming a residual speckle allocation of 22.5 ppt and a post-processing suppression factor of 0.5, the allowed residual contrast instability is 40 ppt.

The remainder of this paper focuses on how this 40 ppt contrast instability is suballocated to modal instability errors in their native units. The starting point for that process is the coronagraph diffraction model that yields the sensitivities to the various modal errors.

5 Modeling Contrast Stability and its Sensitivity to Modal Errors

The previous section discussed how changes in the optical system can be thought of as producing a perturbation field ΔE that is coherently added to some existing field $E(u, v)$. Equation (14) showed that the intensity of the combined field has a contribution from the mixing term $2\mathcal{R}\{E^* \Delta E\}$. This mixing term is usually the dominant instability term since it involves an amplification of the perturbation field by the existing field. The consequence is that, in understanding the effects of optical instability on the speckles, one must also be cognizant of the existing initial speckles, specifically the field that gives rise to them. Hence, assumptions about the initial contrast are needed to formulate a meaningful answer regarding sensitivity of the speckles to optical instability.

The purpose of this section is to describe a modeling approach that leads to an optical telescope error budget based on contrast stability. The key step is RDI. But first, we start with a discussion on wavefront control and the generation of the initial dark hole. This lays the foundation for the rest of the modeling approach.

5.1 How Wavefront Control is Used to Generate the Dark Hole

All high-contrast coronagraphs depend on wavefront control to create the deep null in the dark hole. One widely used technique is the so-called electric field conjugation (EFC) method.²⁸ EFC is an iterative process of measuring then cancelling the electric field in the dark hole. In each iteration, the existing dark-hole electric field is measured through the application of specific surface changes at the DM. These deliberate perturbations, called “pokes,” cause a relatively large amount of additional coherent light (ΔE) to enter the dark hole, in a specific pattern, and mix with the existing unknown field. Equation (14) applies here too, but in this context E represents the existing small field leading to the raw contrast, whereas ΔE is the intentional, large additional field arising from the probe. The coherent mixing term amplifies the faint initial dark-hole field, giving rise to a now conveniently bright intensity pattern. Ideally, ΔE is generated to be flat over the domain of interest. A forward propagation model then guides the determination of the DM shape needed to produce the desired amplifying coherent field at the dark hole. Pairwise, complementary real and imaginary amplification fields in the dark hole, generated via the appropriate poke patterns on the DM, allow the estimation of the dark-hole field.³³ Once the (preamplification) dark-hole field is estimated, the DM is restored to its unpoked state. In the next step, the new DM shape, that would cancel the measured dark-hole field, is estimated. The estimation process uses a stored library of sensitivities obtained using the forward diffraction model. These are sensitivities of the real and imaginary parts of the electric field at each dark-hole pixel to a perturbation in each DM actuator. A pseudoinverse solution, with constraints added to limit actuator strokes in a single iteration, is used to compute the shape. Errors in knowledge of the optical configuration, the DM actuators’ response to commands, the detector noise, as well as the first-order approximation implicit in the EFC method, all contribute to error in the estimated shape. Therefore, it is necessary to repeat the correction process many times to achieve the final contrast—i.e., “dig” the dark hole. Whether the desired contrast is attained in an acceptable number of iterations, or not at all, depends to a large measure on the accuracy of the system model, the DM and the detector noise.

DM actuator commands cannot perfectly compensate for the imperfections of the DM surface. To achieve a 10^{-10} level of raw contrast, the DM must produce a wavefront that deviates by no more than order of (10 pm) from ideal. But real DMs never reach anywhere near this level of perfection. If a DM was commanded to a flat surface, the residual surface error would never be better than a few nanometers rms. This apparently insurmountable obstacle can be circumvented because many more solutions than the assumed ideal wavefront exist. A gap of nearly three orders of magnitude between the flatness requirement and what can be achieved is closed using a wavefront control algorithm that searches for a local “optimum” solution near the current surface. It makes minute, subnanometer adjustments to the DM’s current shape, in the vicinity of the surface shape the DM has already reached (with its nanometers of error). Put another way, the final iterations of wavefront control do not necessarily improve the root-mean-square surface error, but they do move light from the inside to the outside of the dark hole. Some coronagraph designs, like the hybrid Lyot, require a specific nonflat initial wavefront error to be applied by the DM. In this case, a perfect stellar wavefront enters the telescope, the instrument optics add error to this flat wavefront, and the DM further adds a design-specified pattern to the wavefront. The DM’s shape must both compensate for the wavefront error of the optics and add the needed wavefront shape required by the coronagraph. For other coronagraph designs, the coronagraph needs a flat wavefront. In these cases, the DM merely compensates for the wavefront error incurred from reflections and transmissions through the optics.

5.2 Wavefront Error Approximation

With these considerations in mind, the approach employed in this paper involves generating the initial raw contrast by the addition of a small wavefront error (a few pm rms) to the coronagraph’s per-design ideal wavefront. For vector-vortex and apodized pupil coronagraphs, the ideal wavefront is flat, while for the hybrid Lyot it is a specific nonflat shape. This approach is summarized in Fig. 20.

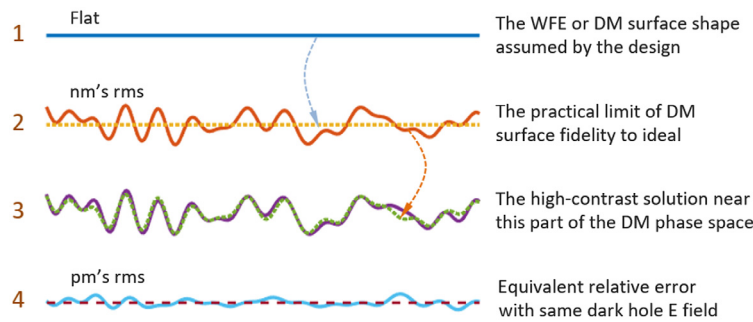


Fig. 20 Logical steps to the guiding assumption on wavefront error simulation. In this example, it is assumed that the ideal wavefront in this coronagraph design is a flat wavefront. This would be true for the typical apodized pupil or vector-vortex case.

As Fig. 20 illustrates, the coronagraph masks are designed for an ideal (e.g., flat) incoming wavefront. Real DMs, as noted above, can only reach a desired wavefront with the fidelity of a few nanometers rms. This is because their phase sheets have deformations at spatial frequencies above the actuator spacing and because the actuators themselves can produce residual phase-sheet deformations. Obviously, nanometers of error are far from ideal for reaching the required level of contrast. To overcome this limitation and dig the dark hole, wavefront control makes small, pm-level adjustments to the DM to find a high-contrast solution near the best-effort attained shape. Finally, from a modeling perspective, the important point (illustrated as step 4 in Fig. 20) is that the (actuator height) phase-space distance between the optimal local solution and the EFC-settled solution can just as well be applied to a simple-to-simulate, ideal surface (e.g., the ideal flat surface for a vector-vortex coronagraph).

For the purposes of this paper, the DM error spectrum is assumed to be flat for all spatial frequencies out to the edge of its sampling. For example, a DM with 64 actuators across its surface is assumed to have a surface error power spectrum that stays flat out to the edge of its Nyquist frequency of 32 cycles per aperture then rises abruptly beyond this range. An azimuthally integrated PSD distribution and the resulting surface shape for such a DM is modeled in Fig. 21. Note that this is not meant to represent a real DM surface but a DM surface error relative to the final wavefront-control computed surface. This error (in both a real system and here in the model) is the dominant contribution to the raw contrast.

The procedure outlined is relatively straightforward to implement, but it does not predict the initial contrast that can be expected. Instead, we assume that a certain level of contrast has been

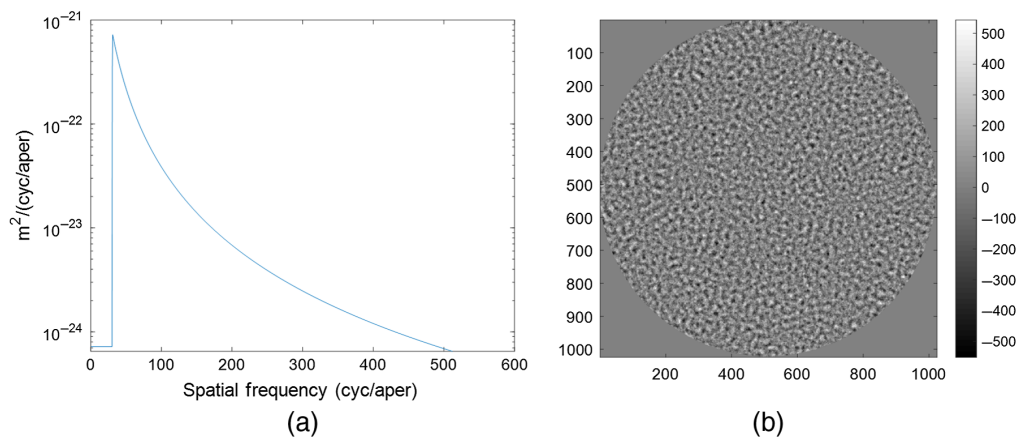


Fig. 21 Wavefront error implementation following the approach outlined in this section. (a) The assumed azimuthally integrated PSD, which is flat and small out to 32 cycles per aperture (where a 64×64 DM is expected to have strong control authority) and rises abruptly afterward, falling off with spatial frequency as $f^{-2.5}$. (b) The resulting surface shape.

achieved via wavefront control, and then we evaluate sensitivities to various modes of wavefront error in the presence of an initial field consistent with the assumed contrast. The Roman mission technology effort has demonstrated contrast approaching 10^{-9} in testbeds.³⁴ It is reasonable to expect that a future telescope, customized for a specific coronagraph, could achieve another order of magnitude better performance in contrast. Therefore, we assume a range of initial contrast values in the 10^{-10} vicinity and compute a sensitivity curve for each wavefront error mode.

5.3 Contrast Instability and Residual Speckle

After all measures to create the deepest null have been taken, the last step in improving signal detection is to calibrate and subtract the remaining speckle pattern. We assume RDI is the method used. For the reference image to remain accurate, the speckle pattern needs to remain stable. When speckles are perfectly stable, RDI yields a residual that only has random (photon and detector) noise. But if the speckles have instability, RDI yields a residual with “structure.” As we noted already, the residual speckle standard deviation $\sigma_{\Delta C}$ depends on raw contrast. But, in modeling, setting the raw contrast to a precise value is not convenient. It is easier to start with an assumed input field and propagate that field forward to arrive at a contrast map. Repeating this exercise with increasing amplitudes of effective post-EFC surface error (Fig. 21) produces a set of dark holes with increasing amounts of initial raw contrast. The coronagraph error budget in general will call for a specific initial raw contrast. That particular value might not exactly correspond to any of the raw contrast values we have created, but, as long as our range straddles the desired contrast, we can interpolate our results to the called-for initial raw contrast.

To get the sensitivity of the contrast to specific wavefront perturbation modes, we start with some initial field and add the perturbing field. The initial field creates the initial raw contrast in the dark hole, and the perturbed field causes the changed contrast. Subtracting the perturbation-added speckle map from the original speckle map gives the residual speckle map. We are interested in the structure of the residual contrast map as quantified by its SSD within the dark hole. Since the contrast, and residual contrast, change most strongly with angular separation (i.e., radius), it is usual to divide the dark hole into radial slices and compute the standard deviation within each radial slice. The sensitivity is the ratio of this quantity over strength of the perturbation:

$$S_i = \frac{\sigma_{\Delta C}}{\Delta w_i}, \quad (31)$$

where Δw_i is the rms applied mode- i wavefront perturbation. For this paper, we chose wavefront perturbations from a list likely to be produced by telescope-structure rigid-body motions or optical component deformations (see Table 1). Other perturbation modes could be Zernike polynomials, pair-wise combinations of segments, or just sinusoidal spatial frequencies.

Note that monolithic and segmented aperture telescopes have different perturbation modes. For example, regardless of whether the telescope has a monolithic or segmented primary mirror, physical displacements of the secondary mirror relative to the primary mirror produce global tip/tilt, defocus, astigmatism, coma, and spherical errors. But only a segmented aperture mirror has segment-to-segment piston and tip/tilt error, and only a segmented aperture mirror is affected by backplane bending because the individual mirror segments are mounted to the backplane.

Table 1 List of error modes considered.

Study case		Error modes considered					
Monolithic	Tip/tilt	Defocus	Astigmatism	Coma	Trefoil	Spherical	Sec. trefoil
Segmented global	Bend	Power (Seidel)	Spherical (Seidel)	Coma (Seidel)	Coma (Zernike)	Trefoil (Zernike)	Sec. trefoil (Zernike)
Segmented segments	Piston	Tip/tilt	Power (Seidel)	Astigmatism (Zernike)	Trefoil (Zernike)	Sec. trefoil (Zernike)	

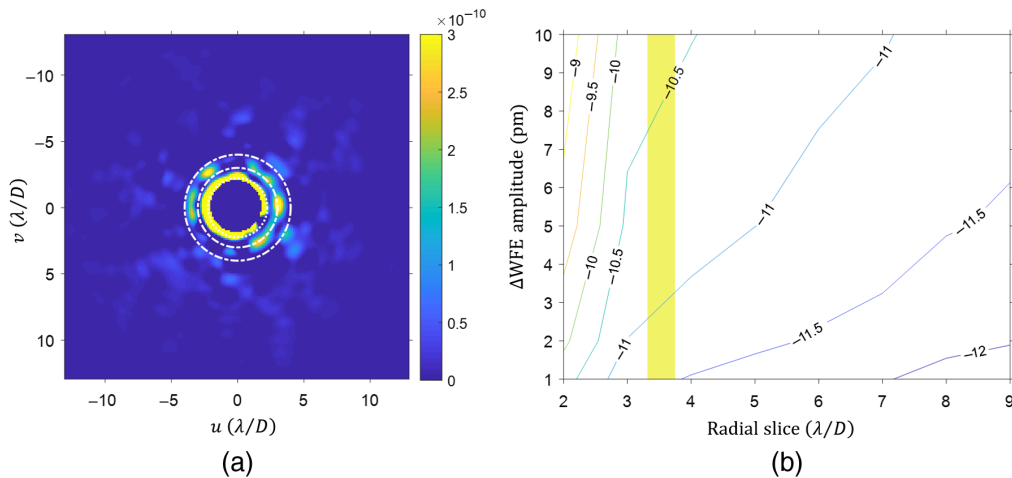


Fig. 22 (a) Residual speckle and (b) contours of $\sigma_{\Delta C}$ for different levels of perturbation wavefront error amplitude and radial slice. (a) A $1 \lambda/D$ -wide annular region at this separation is highlighted by a pair of concentric circles. (b) The contour values are of $\log_{10} \sigma_{\Delta C}$. These plots correspond to the case of trefoil mode on vector-vortex charge-6 coronagraph. The vertical yellow strip is centered at the separation of our nominal target (exo-Earth at 10 pc) observed at a waveband centered at 550 nm.

For the global modes, we assume that the backplane has the shape corresponding to the mode considered (such as a global bend). We then compute the piston and tip/tilt of a tangent plane centered at each segment's two-dimensional center and apply the piston to the segment at its center. We also apply the local tip/tilt arising from the global deformation. For the azimuthally symmetric cases, such as defocus, we create only one instance, while for the rest, we randomly create a number of instances with different azimuthal orientations (clocking). We then average the results for the set.

Figure 22 shows an example of this procedure for the case of a 4-m off-axis unobscured monolithic aperture telescope with a vector-vortex charge-6 coronagraph. The left image shows the residual speckle map ΔC when the wavefront is changed by 10 pm of trefoil. The residual contrast map can be divided into annular zones. For each zone, the azimuthal standard deviation ($\sigma_{\Delta C}$) of the residual speckle map is computed. The contour plot in Fig. 22 shows how $\sigma_{\Delta C}$ varies with angular distance in λ/D and perturbation amplitude. If we set a requirement that the residual speckle contribution from trefoil shall not exceed 10^{-11} , then the region excluded from observation is that to the left of the -11 contour line (shaded). And, if the goal is to observe an exo-Earth located at $3.6 \lambda/D$ (yellow vertical zone), then this graph shows that the trefoil change between reference and target star observation times must be less than ~ 3 pm rms. Repeating this process for each perturbation mode fully populates the telescope/coronagraph sensitivity table.

6 Linking the Error Budget to Requirements on Wavefront Stability

Because it is cumbersome to show a separate plot for each perturbation mode, Fig. 23–Fig. 27 show summary graphs for each of the cases studied. Each graphic plots the residual speckle error ($\sigma_{\Delta C}$) caused by 5 pm rms of each perturbation mode computed over a $1 \lambda/D$ -wide annular zone near 100 mas (location of a 1 AU exoplanet at 10 pc) as a function of initial raw contrast. The central wavelength in all cases is assumed to be 550 nm. As the raw contrast of the planet increases, so too does the sensitivity of the contrast to wavefront change and hence the residual speckle error. The vertical line at 10^{-10} raw contrast, which appears in all but one of the plots, indicates the target goal needed to observe an exo-Earth.

Figures 23–27 also show an error budget for each case. These error budgets were constructed using a modified allocation method based upon the coronagraph's sensitivities to each error mode. The sensitivity of each coronagraph to a given error mode is given by Eq. (31), where

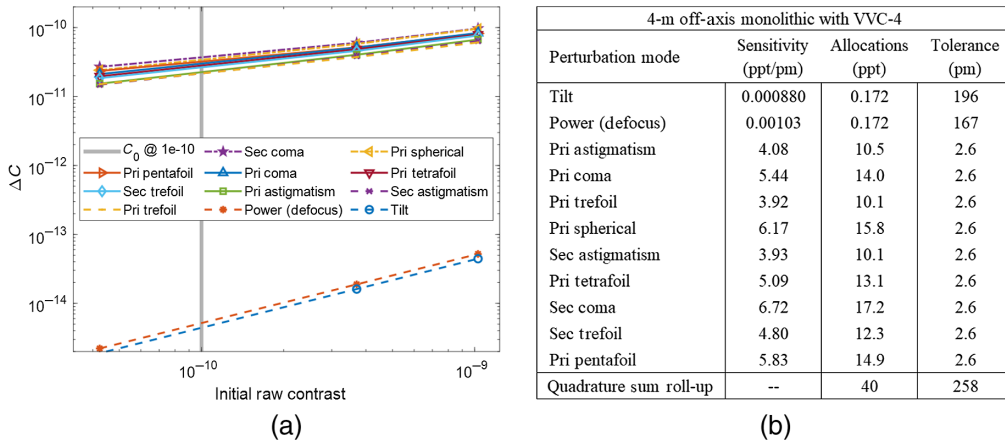


Fig. 23 Case 4-m off-axis monolithic telescope with a vector-vortex charge-4 coronagraph: (a) residual speckle error caused by 5 pm rms of a number of perturbation modes versus initial raw contrast. Sensitivities are taken from the crossing at C_0 of 10^{-10} . (b) The sensitivities and allocations following the algorithm described here. The tolerances are in pm rms, and they are ratios of the allocations to the sensitivities. These are evaluated at the separation of our nominal target (exo-Earth at 10 pc) observed at a waveband centered at 550 nm.

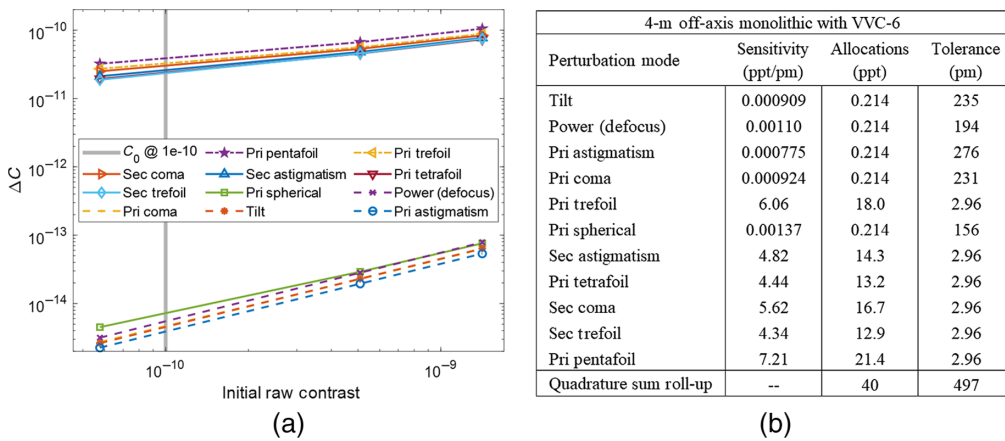


Fig. 24 Case 4-m off-axis monolithic telescope with a vector-vortex charge-6 coronagraph: (a) residual speckle error caused by 5 pm rms of a number of perturbation modes versus initial raw contrast. Sensitivities are taken from the crossing at C_0 of 10^{-10} . As in the previous figure, (b) lists the sensitivities, allocations, and resulting tolerances in pm rms. These are evaluated at the separation of our nominal target (exo-Earth at 10 pc) observed at a waveband centered at 550 nm.

the numerator is read for each mode at the assumed initial contrast in each plot, and the denominator is the perturbation amplitude used to obtain the plots (5 pm rms in all cases). Note that this error budget is only to detect an exo-Earth at a separation of 100 mas from its host star. If the planet is a different size or located at a different angular separation, a different wavefront stability allocation will be required. This is because different Zernike modes introduce speckle errors into different locations in the dark hole, and because different-sized planets have different contrast SNRs.

To simplify the terminology in the allocation process, we rename numerator and denominator of Eq. (31) as

$$S_i = \frac{\sigma_{\Delta C}}{\Delta w_i} \equiv \frac{\partial e}{\partial x_i}. \tag{32}$$

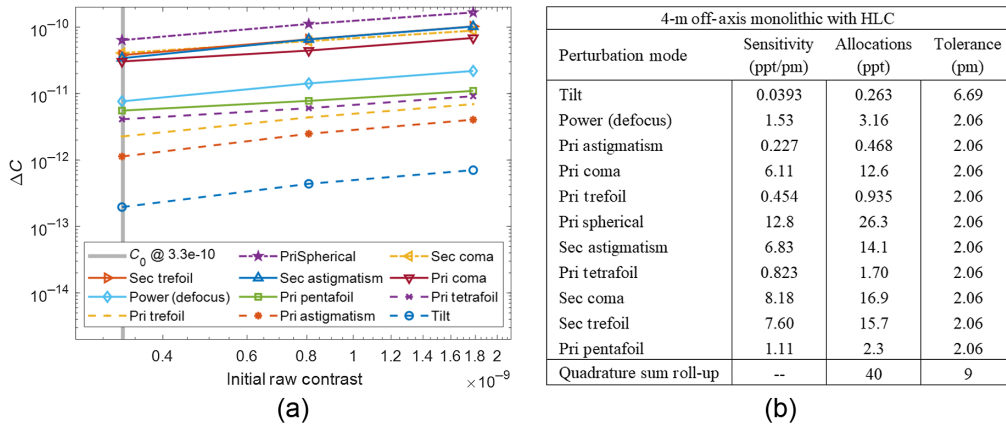


Fig. 25 Case 4-m off-axis monolithic telescope with a HLC: (a) residual speckle error caused by 5 pm rms of a number of perturbation modes versus initial raw contrast. Sensitivities are taken from crossings at C_0 of 3.3×10^{-10} (left edge) of the plot. As in the previous figures, (b) lists the sensitivities, allocations, and resulting tolerances in pm rms. These are evaluated at the separation of our nominal target (exo-Earth at 10 pc) observed at a waveband centered at 550 nm.

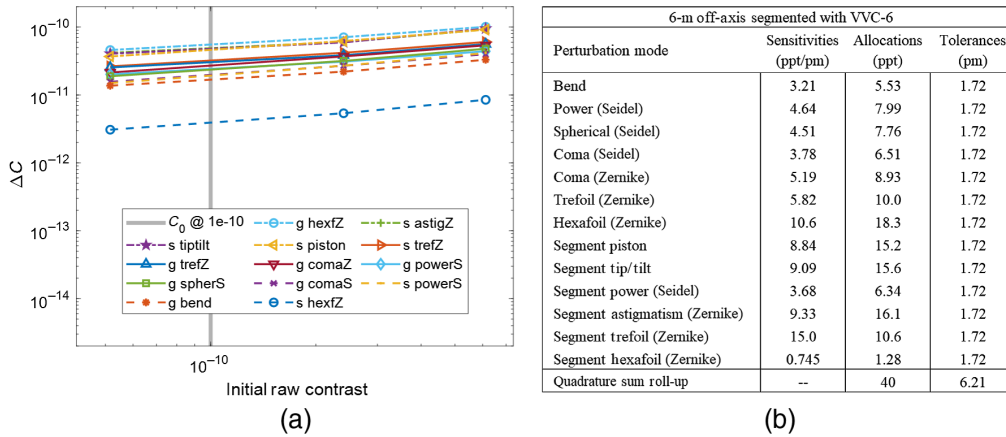


Fig. 26 Case 6-m off-axis segmented telescope with a vector-vortex charge-6 coronagraph: (a) residual speckle error caused by 5 pm rms of a number of perturbation modes versus initial raw contrast. Sensitivities are taken from crossings at C_0 of 10^{-10} . As in the previous figures, (b) lists the sensitivities, allocations, and resulting tolerances in pm rms. These are evaluated at the separation of our nominal target (exo-Earth at 10 pc) observed at a waveband centered at 550 nm.

That is, the error ϵ is the SSD of the residual speckle $\sigma_{\Delta C}$ and the “perturbation” x_i is the rms amplitude of the wavefront change in mode i , namely Δw_i . The modes are selected among those that are most likely to be present in a significant amount (see Table 1).

A very simple approach for creating an error budget allocation could be to allocate each mode the same amount of contrast. This would result in a wavefront stability error budget whose allocations are proportional to the sensitivities. The problem is that an allocation proportional to sensitivities makes the easy modes unnecessarily difficult without providing relief to the other modes. This is obvious just looking at the VVC charge-4 case as an example, we see from Fig. 23 that the sensitivities can be very different for different modes. Therefore, we adopt a weight approach that allocates contrast using a modified version of the sensitivity for each mode, one that has a bound at the low end of the sensitivity set at a reasonably small fraction of the largest sensitivity:

$$s_i = \max\left(\frac{\partial \epsilon}{\partial x_i}, 1\% \cdot \max\left\{\frac{\partial \epsilon}{\partial x_i}\right\}\right). \tag{33}$$

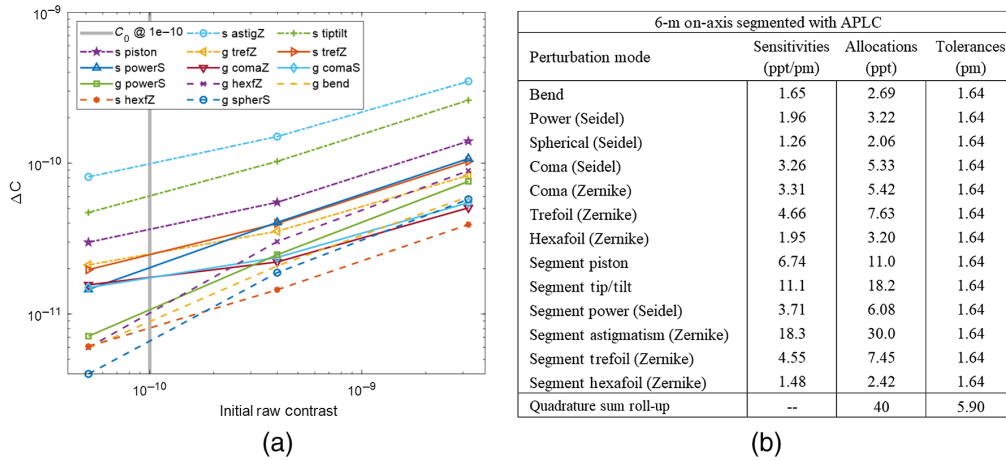


Fig. 27 Case 6-m on-axis segmented telescope with an APLC: (a) residual speckle error caused by 5 pm rms of a number of perturbation modes versus initial raw contrast. Sensitivities are taken from crossings at C_0 of v . As in the previous figures, (b) lists the sensitivities, allocations, and resulting tolerances in pm rms. These are evaluated at the separation of our nominal target (exo-Earth at 10 pc) observed at a waveband centered at 550 nm.

Thus the weight for each mode is set either at the sensitivity to that mode or at one percent of the maximum sensitivity among all the applicable modes, depending on whichever is greater. It is implied that the sensitivities in question are to be evaluated for a given initial raw contrast of C_0 . With these weights, we assign allocations, using a formula that ensures the quadrature sum equals the total allowable error:

$$\epsilon_i = \epsilon_{\text{tot}} \cdot \left(w_i / \sqrt{\sum_i w_i^2} \right). \tag{34}$$

Using these suballocations, the tolerances are given by

$$\delta x_i = \epsilon_i / \left(\frac{\partial \epsilon}{\partial x_i} \right). \tag{35}$$

The quadrature sum of the allocations thus made is manifestly ϵ_{tot} , which we have set to 40 ppt. This algorithm is followed in obtaining the tolerances in the tables that appear to the right of each of the following five figures. Finally, these rigorously derived analytical results are similar to and consistent with our previously published numerical simulation results.^{11–15} Also note that the allocation between modes can be further refined on a case-by-case basis depending upon which modes are mostly likely to occur in a given telescope implementation.

In the vector-vortex cases, there are clearly two populations (of modes): those to which the coronagraph is highly sensitive and those to which it is insensitive. The difference in sensitivity is over three orders of magnitude. This difference allows the weighted allocation method to assign looser tolerances to those modes to which the coronagraph is insensitive and assign as much error as possible to the modes to which the coronagraph is most sensitive.

In the HLC case, it was not possible to reach a contrast as low as 10^{-10} , so the results are quoted for the minimum contrast available. Again, note that the HLC and segmented APLC designs may not be current and do not necessarily represent their current best performance.

The next two cases are for telescopes with a segmented primary mirror. As discussed in Sec. 3, these are both assumed to have a 6-m diameter circular aperture with “zero” gaps. The off-axis 37-segmented case assumes a vector-vortex charge-6 coronagraph (VVC-6). And the on-axis 36-segmented case assumes an APLC. A key takeaway is that a monolithic mirror telescope performs better than a segmented aperture telescope—even if they are both off-axis unobscured and use the same coronagraph. The reason is segment co-phasing. The hexagonal segmentation pattern places speckle noise produced by segment co-phasing error (caused by

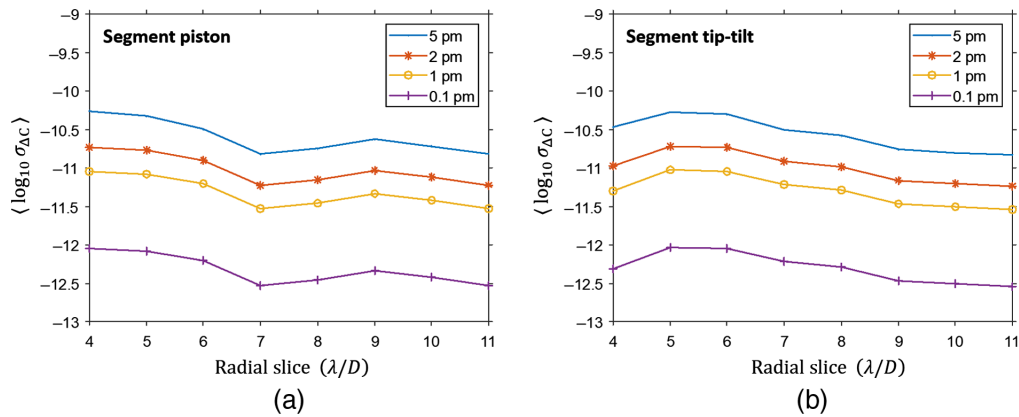


Fig. 28 The challenge of segment stability illustrated by (a) considering segment piston and (b) segment (tip-tilt). For the on-axis segmented case with APLC, the plots are of $\sigma_{\Delta C}$ versus radial slice, in λ/D , for various amounts of error in pm rms. The horizontal line at 2×10^{-11} indicates 20 ppt of allocation. The vertical dashed line is at $5.3 \lambda/D$, where our example exo-Earth would be located.

piston and tip/tilt) into the dark hole near the IWA. And because the segments are essentially rigid body objects, any global bending of the backplane causes co-phasing errors at the segment boundaries.

We can see that the segmented cases end up with very tight tolerances on stability of segments relative to each other. To see this in another way, consider just segment piston and segment tip-tilt. Figure 28 plots the log of $\sigma_{\Delta C}$ versus radial slice, in λ/D , for various amounts of error in pm rms for the 36-segmented on-axis telescope with the APLC coronagraph. For our example case of an exo-Earth at 10 pc, the vertical dashed line represents the location of the signal. The horizontal line at 2×10^{-11} in each plot corresponds to an allocation of 20 ppt. From the allocation table in Fig. 27, we see that 20 ppt is already higher than we were able to allocate, and yet it only allows for 2 pm rms of intersegment tip-tilt error. There are two take-aways from Fig. 28. First, the bump in the contrast leakage near $5 \lambda/D$ is caused by the aperture segmentation pattern. And, second, performing exo-Earth coronagraphy using a segmented aperture primary requires segment-to-segment co-phasing at <5 pm rms.

7 Summary and Conclusions

Exoplanet imaging with an internal coronagraph requires an ultraprecise, ultrastable optical telescope. Since the telescope's specifications depend on the type of internal coronagraph implemented, the telescope and coronagraph must be specified and designed as an integrated system. This paper develops a generic system-engineering approach for deriving a telescope wavefront stability error budget that can be applied to any telescope/coronagraph combination. The method has two inputs: (1) a science mission specification in terms of a target, desired SNR, and an observing scenario and (2) coronagraph raw contrast sensitivity to specific telescope wavefront error modes. To implement the method, we first calculate the coronagraph's performance metrics of core throughput, raw contrast, and stability of raw contrast as a function of angular separation from line-of-sight. Next, we use calculated throughputs and PSF core sizes to determine the random error contributions expected as a function of exposure time per some observing scenario for a specific fiducial target (e.g., an exo-Earth at 10 pc). The desired SNR and the flux ratio of the target give the allowable total error, which is then broken down into random and systematic allocations. Finally, a weighting method is used to allocate the error between different wavefront stability modes based on the coronagraph's residual raw contrast sensitivity to each mode. Thus errors with the smallest impact are given a larger tolerance. The utility of the method is demonstrated for the ability of specific cases (vector-vortex and hybrid Lyot in combination with an off-axis monolithic and segmented telescope and an APLC with an on-axis hexagonal segment telescope) to detect an exo-Earth at 100 mas from its host star. Note that the HLC and

segmented APLC designs may not be current and do not necessarily represent their current best performance.

The cases studied manifested several findings. Different coronagraphs on the same telescope have significantly different core throughputs. A 6-m off-axis telescope with a VVC-6 has nearly the same core throughput as a 4-m off-axis telescope with a VVC-4. The same coronagraph on a monolithic aperture telescope performs better than on a segmented aperture telescope. Coronagraphs perform better when paired with an off-axis unobscured telescope than with an on-axis obscured telescope. Co-phasing errors between segments can place significant contrast noise near the dark-hole IWA. And performing exo-Earth coronagraphy with a segmented aperture telescope requires segment-to-segment co-phasing better than 2 pm rms.

Additionally, several findings were identified while developing the error budget. For the assumed design reference mission of imaging with an SNR of 7 and exo-Earth at 10-pc with 3 exo-zodi of dust, 25 h is the optimum integration exposure time. Initial contrast must be included when setting requirements on optical stability because of how it mixes with the perturbation field. This mixing can be minimized by taking measures to reduce raw contrast and choosing observation integration times and chopping intervals to target the modes of error for which the product of expected amplitude and contrast sensitivity is the highest. The effect of the mixing term and its minimization is important for understanding the error budget. For a harmonic perturbation whose period is much shorter ($\sim 100\times$) than the measurement integration period, the error budget does not care about the amplitude of its periodic motion but, rather, the stability of the average of that motion. For example, in Fig. 26, it is not necessary for the segment tip/tilt motion to have an amplitude of <1.7 pm rms but rather for its rms to be stable to <1.7 pm. Also the ability of the telescope to achieve the derived error budget determines the chopping interval period. The chopping interval period is determined by the shortest length of time before the telescope exceeds the error budget specifications. A detailed temporal optimization process is beyond the scope of this paper because it requires structural thermal optical performance (STOP) analysis of a specific telescope implementation to determine the period of specific perturbation modes.

Also the final step of the error budgeting process—allocating contrast instability based on which errors are mostly likely to occur—is beyond the scope of this paper. With this step, errors that are least likely to occur are given a small tolerance. Combining likelihood of occurrence with sensitivity to the error results in an error budget where all error sources have equal margins against performance. This step is not illustrated in this paper because it is telescope-dependent and requires detailed STOP modeling of a specific design to implement.

8 Appendix A: The Roman Analytical Framework for Direct Detection SNR

In this section, we describe an analytical framework, developed for the Roman coronagraph, which can be used to estimate the SNR of a planet image in a coronagraph. We then extend it to apply also to the case of spectroscopy with an integral field spectrograph (IFS). The context of this derivation is an RDI scenario with a bright reference star.

Consider a region of interest where a planet signal is centered. Assume the signal is simply the excess of light above the background within the core area of the planet PSF. The background includes unrejected starlight “speckle” zodiacal light and detector dark current. The dominant background is the starlight speckle, and differential imaging is the way we assume it is being subtracted.

After background subtraction, the signal is extracted. Signal extraction will typically be some sort of matched filter for optimized performance.³⁵ However, for our purposes, we assume a more simplified approach, namely aperture photometry: a core region of the planet PSF is defined (see Sec. 3.1), and both the signal and background in that core region are estimated. This assumes that the bulk of the contribution to the signal will come from the core region, and it is here that the signal and noise will be counted. Although this is likely to miss some signal relative to matched filtering, we can derive a conservative estimate of the performance by making the simple aperture photometry assumptions.

8.1 Differential Image

We would like to know the contribution of the background to the noise in the differential image. We assume there is a bright (typically three magnitudes brighter) reference star available, and the telescope slews periodically to this star and obtains reference speckles for subtraction. The total time for obtaining the measurement, excluding the slew and settling times, will be $t_{\text{tot}} = t + t_r$, where t is the integration time on the target star, and t_r is the integration time on the reference star.

We can think of the target star image as a two-dimensional map (counts per unit area), which we can write as $I_{\text{tar}}(u, v)$. Dropping the explicit notation of the coordinates from now on, for the reference star, we have a corresponding quantity I_{ref} . The differential image is obtained by subtracting the scaled reference image from the target image

$$\Delta I = I_{\text{tar}} - \beta I_{\text{ref}}. \quad (36)$$

The normalization factor for the reference star β is given by

$$\beta = \frac{F_{\lambda}^t t}{F_{\lambda}^r t_r}, \quad (37)$$

where F_{λ}^t and F_{λ}^r are the spectral flux of the target and the reference stars, respectively. The ideal reference star is much brighter than the target star while being identical in (angular) stellar diameter and spectral type with the target star. The ideal star would also have, at a given epoch, the same solar angle (LOS to the star versus direction of the Sun) as the target star, to keep the thermal loads the same on the space telescope. Finally, it would not be very far in absolute angle, so that the slews are not time expensive. The 100 brightest stars in the sky are roughly 3 mag or brighter, while the typical target for an exo-Earth host is roughly 3 mag less bright. For the representative case of a 3-mag brightness difference and $t_r = 0.2 t$, $\beta = 0.32$. The scale factor β applies to all the error sources since it enters at the level of the subtraction of the two images.

In the absence of noise and any planets, and if the reference star speckle pattern is identical to the target star case, then $\Delta I(x, y)$ should be zero everywhere.

The sources of random noise are:

1. planet shot noise,
2. speckle shot noise,
3. zodi shot noise (both from local zodi and exo-zodi),
4. detector noise [will assume an electron multiplication CCD (EMCCD)].

The first three error sources are photonic in origin and require a discussion of throughput. We have neglected stray light in this list, assuming good rejection of this background. If it is included, the treatment might begin as being most similar to local zodi (smooth incoherent background), for example in the calculation of the RDI noise enhancement factor discussed below.

The main source of nonrandom error is the residual speckle after differential imaging. This, which we will label as systematic error, is much more difficult to model analytically, but we will treat this with a simple parameterization further below.

8.2 Throughput

The throughput can be expressed as a product of factors.

- τ_{obs} : Loss due to occultations of a perfect circular pupil (e.g., from the secondary mirror, struts, or segment gaps).
- τ_{ref} : Reflections and transmissions through the system, from the primary mirror all the way to the final focusing element.
- τ_{occ} : *occulter transmission*. The effect of the three coronagraph masks, namely the pupil, the focal plane, and the Lyot masks. This is the fraction of the light incident on the pupil mask that arrives anywhere in the final image plane. This depends strongly on the working angle of the source.

- τ_{PSF} : *PSF core fraction*. The fraction of the light in the PSF that ends up within the core region of the PSF. This depends strongly on the PSF shape and hence on the working angle.
- τ_{fil} : *filter transmission*. The average transmission within the nominal bandwidth of the filter.
- τ_{pol} : *polarizer transmission*. The fraction of randomly polarized light that goes through the polarizer—typically close to 50%.

In Sec. 2, we defined core throughput. In terms of these factors, core throughput is given by

$$\tau_{core} = \tau_{occ} \cdot \tau_{PSF}. \quad (38)$$

The three photonic error sources have various factors as follows.

1. τ_{pl} : *planet throughput*. Everything applies: $\tau_{pl} = \tau_{obs} \cdot \tau_{ref} \cdot \tau_{occ} \cdot \tau_{PSF} \cdot \tau_{fil} \cdot \tau_{pol}$.
2. τ_{sp} : *speckle throughput*. Coronagraph effects captured elsewhere: $\tau_{sp} = \tau_{obs} \cdot \tau_{ref} \cdot \tau_{fil} \cdot \tau_{pol}$.
3. τ_{zo} : *zodi throughput*. Uniform extended source has no τ_{PSF} loss: $\tau_{zo} = \tau_{obs} \cdot \tau_{ref} \cdot \tau_{occ} \cdot \tau_{fil} \cdot \tau_{pol}$.

These considerations imply

$$\tau_{pl} = \tau_{sp} \cdot \tau_{core}. \quad (39)$$

8.3 Planet Signal

The signal, which appears only when observing the target star, has an expected value of

$$S = r_{pl}t, \quad (40)$$

where t is the time spent integrating on the target star, and r_{pl} is the planet signal count rate:

$$r_{pl} = F_{\lambda} \Delta\lambda \xi_{pl} A \tau_{pl} \eta, \quad (41)$$

where F_{λ} is the spectral flux, $\Delta\lambda$ is the filter bandwidth, ξ_{pl} is the planet flux ratio, A is the collecting area, and η is the detector quantum efficiency. τ_{pl} is the throughput for the planet light.

8.4 Total Noise Variance

The total noise variance can be written as

$$\sigma_{tot}^2 = \sigma_{md}^2 + \sigma_{\Delta I}^2, \quad (42)$$

where the first term σ_{md}^2 includes all the shot noise and detector noise contributions to the subtracted image. The second term $\sigma_{\Delta I}^2$ arises from any lumpiness (which can be confused for a planet signal) in the residual speckle after differential imaging.

8.5 Random Noise Variance

We distinguish the contributions to σ_{md}^2 from the target and reference star observations:

$$\sigma_{md}^2 = \sigma_{tar}^2 + \sigma_{ref}^2. \quad (43)$$

For the target star variance, we have

$$\sigma_{tar}^2 = \sigma_{pl}^2 + \sigma_{sp}^2 + \sigma_{zo}^2 + \sigma_{det}^2 = r_{pl}t + r_{sp}t + r_{zo}t + r_{det}t. \quad (44)$$

The first term σ_{pl}^2 is the noise variance from the planet, which, per Poisson statistics, equals the signal S . We will explore each of the other terms in more detail the following sections. As before, t is the target star integration time.

8.6 Speckle Noise Variance

The second term of Eq. (44) σ_{sp}^2 is the variance from the shot noise in the starlight halo (i.e., the mean speckle background). The first step is to calculate the mean speckle rate r_{sp} . We express this, following the discussion in the main narrative that led to Eq. (3), in terms of the fraction of the starlight that ends up in the region of interest. We assume that the region of interest corresponds to the PSF core region, with size Ω_{core} . Per Eq. (3), the fraction of light diffracting into Ω_{core} is given by

$$C\tau_{\text{pk}}\left(\frac{\Omega_{\text{core}}}{\Omega_r}\right) = C\tau_{\text{pk}}n_{\text{core}}. \quad (45)$$

The count rate (electrons per second at the detector) from the speckle background is thus given by

$$r_{\text{sp}} = F_\lambda \Delta\lambda C\tau_{\text{pk}}n_{\text{core}}A\tau_{\text{sp}}\eta. \quad (46)$$

8.7 Zodi Noise Variance

The third contribution is the noise from zodiacal background $\sigma_{\text{zo}}^2 = r_{\text{zo}}t$. There are contributions both from our own solar system (“local-zodi,” lz) and the exo-system (“exo-zodi,” ez). Since these are extended sources of light, they are expressed in terms of spectral flux per unit solid angle. Thus the zodi count rate is given by

$$r_{\text{zo}} = \left(\frac{\partial F_\lambda^{\text{lz}}}{\partial \Omega} + \frac{\partial F_\lambda^{\text{ez}}}{\partial \Omega}\right) \Delta\lambda \Delta\Omega A\tau_{\text{zo}}\eta. \quad (47)$$

Multiplying each of these by $\Delta\lambda$ gives the associated surface brightness. In the visible band, the local zodi surface brightness is uniform within the dark hole at about 23 mg/as².

Actually, the zodi variation over the dark hole may not be uniform for the exosystem. Here for simplicity, we have assumed it is uniform with the same effective throughput τ_{zo} as local zodi. This is a conservative assumption, as it allows more throughput for this very significant background. At present, statistics on exosystem dust densities is very poor, so a conservative estimate is reasonable.

8.8 Detector Noise Variance

The detector used in the Roman coronagraph is an EMCCD, with very low effective read noise.³⁶ The HabEx concept mission has also baselined a similar detector. The EMCCD detector noise variance is given by

$$\sigma_{\text{det}}^2 = r_{\text{det}}t = i_d m_{\text{pix}}t + q_{\text{CIC}} m_{\text{pix}} \frac{t}{t_{\text{fr}}} + m_{\text{pix}} \frac{t}{t_{\text{fr}}} \left(\frac{\sigma_{\text{rd}}}{G_{\text{EM}}}\right)^2, \quad (48)$$

where i_d is the detector dark current in $e^-/\text{pix}/s$, q_{CIC} is the EMCCD clock-induced charge in $e^-/\text{pix}/\text{fr}$, t_{fr} is the frame duration, σ_{rd} is the read noise, G_{EM} is the electron multiplication gain, and m_{pix} is the number of detector pixels onto which the light of interest (namely the planet core light) falls. Note that m_{pix} is analogous to n_{core} , except it applies to the plate scale of the detector pixels rather than the modeling pixels. Note also that t/t_{fr} represents the number of frames read over the whole integration time t spent on the target star.

8.9 Additional Noise from Differential Imaging

So far, we have only considered the random noise from target star observations. But the differential image is also affected by the noise from the reference image, albeit smaller because the reference star is brighter. Here, we account for the additional noise coming from the reference star.

We will first examine the speckle background shot noise within the region of interest—the core region centered on a putative planet. Since the intensity is proportional to detected (electron) counts, we can restate Eq. (36) as $\Delta n(u, v) = n_{\text{tar}} - \beta n_{\text{ref}}$. Since we are not including the planet, we should have $\Delta n \approx 0$, so that the normalized reference counts are equal to the target counts, or $\beta n_{\text{ref}} = n_{\text{tar}}$. We write the noise in the differential image as $\sigma_{\text{tot}}^2 = \sigma_{\text{tar}}^2 + \sigma_{\text{ref}}^2$. From Poisson statistics, $\sigma_{\text{tar}} = \sqrt{n_{\text{tar}}}$. The reference star contribution, because of the normalization factor, is $\sigma_{\text{ref}} = \beta \sqrt{n_{\text{ref}}}$. But, since $\beta n_{\text{ref}} = n_{\text{tar}}$, the net result for the additional noise from the reference star to the speckle (sp) noise is

$$\sigma_{\text{ref}}^{\text{sp}} = \sqrt{\beta} \sigma_{\text{tar}}^{\text{sp}}. \quad (49)$$

The reference star contributions to the zodi background shot noise must distinguish between the local and exo-zodi backgrounds. The local zodi brightness remains unchanged between the two observations, so the only difference lies in how much time was spent integrating on the target (t) and the reference (t_r). The reference counts from local zodi are smaller than the target counts by a factor of t_r/t , so that the shot noise is smaller by $\sqrt{t_r/t}$. When the normalization factor from differential imaging is included, the net effect for local zodi (lz) shot noise is

$$\sigma_{\text{ref}}^{\text{lz}} = \beta \sqrt{t_r/t} \cdot \sigma_{\text{tar}}^{\text{lz}}. \quad (50)$$

For exo-zodi, the densities of the disks may be different for the target and reference stars, and, of course, the stars have different brightness. If we assume that the densities are the same, then the difference becomes due only to the relative brightness of the two stars. In this case, the factor is identical to the factor for speckles, along the same lines of reasoning, so that the net effect for exo-zodi (ez) is

$$\sigma_{\text{ref}}^{\text{ez}} = \sqrt{\beta} \cdot \sigma_{\text{tar}}^{\text{ez}}. \quad (51)$$

The detector dark current counts from the reference star are related to those from the target star according to $n_{\text{ref}} = (i_d m_{\text{pix}}) t_r = (t_r/t) n_{\text{tar}}$. When the normalization factor from differential imaging is included, the net effect is that $\sigma_{\text{ref}} = \beta \sqrt{t_r/t} \cdot \sigma_{\text{tar}}$ for detector dark current, just like the speckle background.

The detector clock-induced charge contribution from the reference star is related to the contribution from the target star in the same way as the dark current, by simple inspection of the form of these two error sources in Eq. (48). Thus when the normalization factor from differential imaging is included, the net effect is that $\sigma_{\text{ref}} = \beta \sqrt{t_r/t} \cdot \sigma_{\text{tar}}$ for detector clock-induced charge.

Since the two detector terms are related the same way, so is their quadrature sum. Thus, overall, the detector (det) part of the reference noise is related to the target noise according to:

$$\sigma_{\text{ref}}^{\text{det}} = \beta \sqrt{t_r/t} \cdot \sigma_{\text{tar}}^{\text{det}}. \quad (52)$$

Read noise is not a factor since, in an EMCCD, the read noise is effectively zero. However, had read noise been important, the ratio would be the same as CIC and dark current.

The factors are summarized in Table 2 for the five error categories that apply. There are two ways one can bookkeep the effect of differential imaging: either add up all the reference star contributions in quadrature into one total number, to be added in quadrature itself with the target star contribution, or apply an error enhancement factor to each of the error sources to account for the differencing. The bottom row in Table 2 gives the variance enhancement factors k_x for each error source. For the speckle, for example, this table is to be read to give $k_{\text{sp}} = 1 + \beta$, etc.

Table 2 Reference star error relative to the target star error contribution for each of five categories that apply to HabEx operation under an RDI observing scenario.

	Speckle (k_{sp})	Local Zodi (k_{lz})	Exo-Zodi (k_{ez})	Detector (k_{det})
$\sigma_{\text{ref}}/\sigma_{\text{tar}}$	$\sqrt{\beta}$	$\beta\sqrt{t_r/t}$	$\sqrt{\beta}$	$\beta\sqrt{t_r/t}$
$k_{\mathbf{x}} = \sigma_{\text{tot}}^2/\sigma_{\text{tar}}^2$	$1 + \beta$	$1 + \beta^2 t_r/t$	$1 + \beta$	$1 + \beta^2 t_r/t$

We can now define, following Eq. (44), a total random noise variance rate r_n , which includes both contributions:

$$r_n = r_{\text{pl}} + k_{\text{sp}} r_{\text{sp}} + k_{\text{lz}} r_{\text{lz}} + k_{\text{ez}} r_{\text{ez}} + k_{\text{det}} r_{\text{det}}, \quad (53)$$

where we have broken up r_{zo} into $r_{\text{lz}} + r_{\text{ez}}$ for the respective local and exosystem contributions to zodi noise.

8.10 Modifications to Accommodate Spectroscopy

We have so far been assuming a direct imaging scenario. The HabEx coronagraph will also be using an IFS for spectroscopy. To extend the results to spectroscopy, we need to make an abstraction.

We can think of spectroscopy as photometry in the separate spectral elements of a spectrum. For a concrete example, the HabEx concept mission plans to do spectroscopy with a spectral resolution of $R_{\text{IFS}} = \lambda/\delta\lambda = 140$, where $\delta\lambda$ is the spectral width of a single spectral element.³⁷ A filter of bandwidth $\text{BW} = \Delta\lambda/\lambda$ is used in each coronagraph observation. The instantaneous bandwidth of HabEx is planned to be 20%.

The SNR of interest for spectroscopy is for a single spectral element. That is, we have the light in the PSF core now subdivided spectrally into N_{spec} spectral elements:

$$N_{\text{spec}} = \frac{\Delta\lambda}{\delta\lambda} = R_{\text{IFS}} \cdot \text{BW}. \quad (54)$$

Given a resolution of 140 and bandwidth of 20% in the IFS, this means the fraction of the core light that goes into the SNR region of interest is $1/N_{\text{spec}} = 1/28$. The abstraction that helps here is to define a parameter that gives the fraction of the light in the core that ends up contributing to the signal and noise:

$$f_{\text{SR}} = \begin{cases} 1/N_{\text{spec}} & \text{IFS} \\ 1 & \text{imaging} \end{cases}. \quad (55)$$

In the case of direct imaging, where all of the planet PSF core light went into what was considered signal, so that $f_{\text{SR}} = 1$, this parameter was not useful. In the spectroscopy case, it is a significant factor, $1/N_{\text{spec}}$. The introduction of f_{SR} is the only modification we need to make to the equations we have derived for direct imaging count rates. The count rates for the planet in Eq. (38), the speckle in Eq. (43), and the zodi in Eq. (44) each now get multiplied by f_{SR} . For the direct imaging case, $f_{\text{SR}} = 1$, so that the expression remains the same as before. But, for spectroscopy, it is now reduced by a factor of N_{spec} :

$$\begin{aligned} r_{\text{pl}} &= f_{\text{SR}} F_{\lambda} \Delta\lambda \xi_{\text{pl}} A \tau_{\text{pl}} \eta \\ r_{\text{sp}} &= f_{\text{SR}} F_{\lambda} \Delta\lambda C \tau_{\text{pk}} n_{\text{core}} A \tau_{\text{sp}} \eta \\ r_{\text{zo}} &= f_{\text{SR}} \left(\frac{\partial F_{\lambda}^{\text{lz}}}{\partial \Omega} + \frac{\partial F_{\lambda}^{\text{ez}}}{\partial \Omega} \right) \Delta\lambda \Delta\Omega A \tau_{\text{zo}} \eta. \end{aligned} \quad (56)$$

The spectroscopy case has less signal, by f_{SR} , and more noise, because of the larger number of pixels involved—this is the quantity m_{pix} in Eq. (48). For direct imaging, we typically desire a

detector plate scale of $<0.5 \lambda/D$ per pixel. For a design choice of $0.4 \lambda/D$, this means $m_{\text{pix}} \sim 5$. In the spectroscopy case, m_{pix} is a much larger number: the PSF core is sampled by approximately four lenslets in the IFS, and each of these produces a “microspectrum” wherein one spectral element falls onto a 2×2 pixel area on the IFS detector. The greater number of pixels in turn means a larger detector noise contribution. Thus the total number of pixels for measuring one spectral element becomes $\sim 4 \times 4 = 16$. So, insofar as detector noise is concerned, the form of the equation has not changed, but the typical value m_{pix} for the IFS case is approximately greater than the direct imaging case by a factor of more than three.

8.11 Nonrandom, Systematic Background

We now turn to the systematic contribution to the total noise variance $\sigma_{\Delta I}^2$, the second term in Eq. (42). This is the noise from the nonuniformity of the residual speckle in the dark hole, so that a planet can be confused for one of these, or vice versa. Focusing only on the speckle background, the differential image is given by

$$\Delta I = r_{\text{sp}}^{\text{tar}} t - \beta r_{\text{sp}}^{\text{ref}} t_r, \quad (57)$$

where $r_{\text{sp}}^{\text{tar}}$ and $r_{\text{sp}}^{\text{ref}}$ are the speckle rate during the target and reference observations, respectively, and t_r is the integration time while staring at the reference star. The normalization factor β , given by Eq. (37), scales the reference speckles for the subtraction.

The measure of nonuniformity is the SSD of the differential image excluding known planets. Since all the diffractive effects are band-limited to the λ/D scale, no features finer than scale (such as detector noise, which produces effects on the pixel scale) can be optical. Hence, in computing the SSD, it is acceptable to low-pass filter the image (e.g., average the pixels) down to the λ/D scale before computing the SSD. The domain of interest is where a signal could be the entire dark hole, or more typically a radial slice within the dark hole:

$$\sigma_{\Delta I} = \text{SSD}(\Delta I). \quad (58)$$

Here ΔI is a two-dimensional speckle map and $\sigma_{\Delta I}$ is a standard deviation, a scalar number. We can now relate $\sigma_{\Delta I}$ to differential contrast, which is a design/engineering quantity. Using Eq. (56) for the speckle rates, we obtain

$$\Delta I = f_{\text{SR}} \Delta \lambda \tau_{\text{pk}} n_{\text{core}} A \tau_{\text{sp}} \eta \cdot (F_{\lambda} t C_{\text{tar}} - \beta F_{\lambda}^{\text{ref}} t_r C_{\text{ref}}), \quad (59)$$

where F_{λ}^{ref} is the reference star spectral flux. Substituting Eq. (34) for β and factoring out $F_{\lambda} t$, this becomes

$$\Delta I = f_{\text{SR}} F_{\lambda} \Delta \lambda \tau_{\text{pk}} n_{\text{core}} A \tau_{\text{sp}} \eta t \cdot (C_{\text{tar}} - C_{\text{ref}}). \quad (60)$$

Note that $C_{\text{tar}}(u, v)$ and $C_{\text{ref}}(u, v)$ are two-dimensional contrast maps in the dark hole. We will call their difference the residual contrast map:

$$\Delta C(u, v) = C_{\text{tar}}(u, v) - C_{\text{ref}}(u, v). \quad (61)$$

The SSD of this is

$$\sigma_{\Delta C} = \text{SSD}[\Delta C(u, v)]. \quad (62)$$

Using Eqs. (51) and (55) to get the standard deviation of Eq. (60), we now have

$$\sigma_{\Delta I} = f_{\text{SR}} F_{\lambda} \Delta \lambda \tau_{\text{pk}} n_{\text{core}} A \tau_{\text{sp}} \eta t \cdot \sigma_{\Delta C}. \quad (63)$$

This result relates contrast residual standard deviation $\sigma_{\Delta C}$ to the residual image standard deviation $\sigma_{\Delta I}$. Contrast residual more directly relates to modeling and tests, while the image residual is how the speckle morphology and its change from the reference to the target observations affects the residual image speckle background. Note that $\sigma_{\Delta C}$ captures exclusively the

optically induced effect. It is not meant to include shot noise or detector noise in the speckle map that would be incurred from camera images.

8.12 Derivation of the Relation between Contrast Stability and Flux Ratio Noise

The systematic noise $\sigma_{\Delta I}$ is defined as the variance, on the core-area scale, of the postsubtraction residual (e.g., RDI). The core area on the sky is roughly λ/D in width. The temporal characteristics of $\sigma_{\Delta I}$ depend on the type of perturbation. For oscillatory variations in the speckle, such as might arise from dynamic modes of the telescope, the time constant would probably be shorter than the integration time $\tau \ll t$ and thus $\sigma_{\Delta I}$ would decrease rapidly with increasing integration time. However, for long-time constant error sources, such as actuator drift or the slowest thermal variations, the time constant would be large relative to the integration time $\tau \gg t$ so that the variance $\sigma_{\Delta I}^2$ would grow at the same rate as the signal (i.e., $\sigma_{\Delta I}^2 \propto t^2$). Since this mode of temporal dependence dominates over the other, we will use a temporal parameterization that only includes this mode. So, we define a postsubtraction improvement factor relative to the mean speckle rate r_{sp} :

$$f_{\Delta I} \equiv \frac{\sigma_{\Delta I}}{r_{\text{sp}} t}. \quad (64)$$

Perfect subtraction would give $f_{\Delta I} = 0$. In general, $f_{\Delta I}$ is a strong function of the speckle stability, and also a function of the effectiveness of any postprocessing that might attempt to improve the estimate of the background speckle pattern. We can express this as

$$f_{\Delta I} = f_{\Delta C} \cdot f_{\text{pp}}, \quad (65)$$

where the first factor $f_{\Delta C}$ is the improvement that comes from speckle stability, and the second factor f_{pp} comes from further improvement by postprocessing.

In error budgeting, $f_{\Delta C}$ is usually informed by separate integrated modeling results where the effects of various thermal and jitter errors are propagated to the final image. Similarly, the choice of a good factor to use for f_{pp} is informed by postprocessing of model-generated speckle histories.

Thus for the systematic contribution $\sigma_{\Delta C}$, we have, by the definition above, the explicit integration-time dependence:

$$\sigma_{\Delta I} = f_{\Delta I} r_{\text{sp}} t. \quad (66)$$

Without postprocessing, $f_{\text{pp}} = 1$ and $f_{\Delta I} = f_{\Delta C}$. Thus we can use Eq. (57) to define $f_{\Delta C}$, replace $\sigma_{\Delta I}$ and r_{sp} with their respective definitions from Eqs. (49) and (56), perform some cancellations, and arrive at a precise meaning for $f_{\Delta C}$:

$$f_{\Delta C} = \frac{\sigma_{\Delta C}}{C}, \quad (67)$$

where $\sigma_{\Delta C}$ is the SSD (within the dark hole) of the post-RDI differential contrast map, and is the contrast averaged over the same region over which the standard deviation is defined.

It is useful to define a differential image error rate $r_{\Delta I}$, which is the expected rate at which the differential image SSD grows:

$$r_{\Delta I} \equiv f_{\Delta I} r_{\text{sp}}. \quad (68)$$

This expression summarizes all the relations:

$$\sigma_{\Delta I} = r_{\Delta I} t = f_{\Delta I} r_{\text{sp}} t = f_{\text{pp}} f_{\Delta C} r_{\text{sp}} t = f_{\text{pp}} \frac{\sigma_{\Delta C}}{C} r_{\text{sp}} t. \quad (69)$$

We are interested in converting $\sigma_{\Delta I}$ to the equivalent residual-speckle flux ratio noise $\delta \xi_{\Delta I}$. Starting with flux ratio Eq. (13), we have

$$\delta\xi_{\Delta I} = \kappa \cdot \sigma_{\Delta I} = \kappa \cdot f_{pp} f_{\Delta C} r_{sp} t. \quad (70)$$

Rearranging the terms and using the relationship between the image residual standard deviation $\sigma_{\Delta I}$ and the contrast residual standard deviation $\sigma_{\Delta C}$ from Eq. (63), we have

$$\delta\xi_{\Delta I} = \left(\frac{\kappa r_{sp} t}{C} \right) \cdot f_{pp} \cdot \sigma_{\Delta C}. \quad (71)$$

In the quantity in parentheses, many of the factors in r_{sp} are cancelled by the corresponding inverses in κ , per Eqs. (9) and (11). Using also the definition in Eq. (39) for τ_{sp}/τ_{pl} , we have

$$\frac{\kappa r_{sp} t}{C} = \frac{\tau_{pk} n_{core}}{\tau_{core}} \equiv \kappa_c. \quad (72)$$

Finally, using this definition for κ_c , we arrive at a helpful expression for $\delta\xi_{\Delta I}$:

$$\delta\xi_{\Delta I} = \kappa_c \cdot f_{pp} \cdot \sigma_{\Delta C}. \quad (73)$$

The significance of this result is that it allows conversion of contrast instability, a coronagraph engineering measure, into the science-relevant planet flux ratio noise.

8.13 RDI Scenario SNR Equation

Putting all that we have now defined together, we rewrite the expression for the photometric SNR, R , given by Eq. (22), to make the dependence on integration time t explicit:

$$R = \frac{r_{pl} t}{\sqrt{r_n t + f_{\Delta I}^2 r_{sp}^2 t^2}}. \quad (74)$$

This is an expression for the SNR achievable after integrating on the target star for a total time t , and implicitly, a time t_r on the reference star under an RDI observing scenario [see Sec. 8.7 for error contribution factors k_x in Eq. (46) and Table 2].

8.14 Time to Reach SNR

We can ask how long we have to integrate on the target star before we reach a desired SNR, R , given the error rates, including the enhancement factors. Solving Eq. (74) for the integration time t , we have

$$t_R = \frac{R^2 r_n}{r_{pl}^2 - R^2 f_{\Delta I}^2 r_{sp}^2}. \quad (75)$$

In terms of the differential image error rate, $r_{\Delta I}$ from Eq. (61), this can be written as

$$t_R = \frac{R^2 r_n}{r_{pl}^2 - R^2 r_{\Delta I}^2}. \quad (76)$$

9 Appendix B: Polychromatic Diffraction Modeling

Diffraction effects are always wavelength dependent, so coronagraphs are always designed to work within a limited spectral band. Coronagraphs that work over a wider band are more desirable: in imaging they provide more signal photons, and in spectroscopy they yield a wider spectrum. Modeling finite-band propagation and diffraction of light is done by breaking the optical band into a number of subbands, propagating each subband, and adding the intensities incoherently in the final image. Care must be taken that the chromatic change of sampling in the focal plane is appropriately accounted for.

The ideal coronagraph produces a null across a wide spectral bandwidth. Diffraction propagation, on the other hand, is done assuming a single wavelength. To extend the model to cover a finite wavelength band, we divide the band into a number of subbands, each of which is treated monochromatically. Since the subbands are incoherent with respect to each other, the intensity of the full band is sum of the intensities from the subbands. It is convenient to choose an odd number of subbands so that the central wavelength of the full band corresponds to the central wavelength of the central subband.

In polychromatic diffraction modeling, care must be taken to ensure the plate scale is the same across all subbands. The plate scale can be written as

$$p = \frac{\delta\theta}{\delta x}, \quad (77)$$

where $\delta\theta$ is some angular extent on the sky, and δx is the corresponding spatial extent at the focal plane. In a simple imaging system (e.g., a single lens) with focal length f , the plate scale is simply $p = 1/f$. In a space telescope, the initial collected beam, which has the same diameter D as the telescope primary mirror, is compressed by a factor m and recollimated before being introduced to the coronagraph instrument. Within the coronagraph, usually no further compression is applied, so that at every pupil the same compression factor m applies. At the last pupil, after the Lyot stop, a final imaging lens of focal length f_d creates the detected image and sets the plate scale. In this case, the plate scale is given by

$$p = \frac{1}{mf_d} \equiv \frac{1}{f_e}, \quad (78)$$

where f_e is the effective focal length that a hypothetical single lens of diameter D would have, with the same plate scale p . This is useful when modeling the details of the detector imaging system is not of immediate interest.

For full reconstruction of the speckle pattern, the coronagraph must be designed with adequately fine sampling of the speckle. One can set the detector plate scale to attain a desired sampling s_0 at some reference wavelength λ_0 . For a detector with pixel size x_d and a system effective focal length f_e , the sampling at wavelength λ is defined as

$$s = \frac{x_d}{f_e\lambda/D}, \quad (79)$$

where D is the collecting aperture diameter. Sampling is thus the fraction of $1/\lambda/D$ captured by the width of a single detector pixel x_d . A typical choice is to set $s_0 = 1/2$ for Nyquist sampling and choose the mid-band wavelength as λ_0 . So the plate scale, set by f_e , is constant across the waveband, whereas the sampling varies with wavelength. Samplings at all other wavelengths are thus given by

$$s(\lambda) = \frac{s_0\lambda_0}{\lambda}. \quad (80)$$

Frequently, a fast Fourier transform (FFT) is used to propagate the complex field from a pupil to the image plane. The location of the results within the FFT output matrix is described in detail elsewhere.³⁸ For our purposes, it suffices to note that in a direct FFT of a matrix representing the complex pupil field, the output elements will represent Fourier components (spatial frequencies) in steps of 1 cycle per linear dimension of the matrix being transformed (the pupil matrix). To get a desired sampling s of spatial frequencies at the output, the pupil matrix needs to be zero-padded, that is, embedded within a zero matrix of greater dimension than the pupil matrix. If the pupil diameter is D , and the corresponding pupil field matrix has dimensions $n_D \times n_D$, then this matrix is embedded within a larger matrix of zeros, size $N \times N$, such that the center of the pupil is at $(N/2 + 1, N/2 + 1)$. For a given pupil array size n_D (per dimension), and desired sampling s , the pupil needs to be embedded within a zero matrix of size:

$$N = \frac{n_D}{s}. \quad (81)$$

Note that this result is independent of wavelength. Zero-padding a matrix before doing a discrete Fourier transform changes the sampling to n_D/N regardless of the assumed wavelength.

The task is then to define the subbands and set the appropriate paddings N_i within each subband to produce the correcting sampling for that subband. Since most FFT algorithms compute faster for even-sized arrays than odd-sized ones, we impose an additional constraint that N_i be even. Starting with a desired starting dimension for the (unpadded) pupil array (n_D), we use the following procedure to set the paddings and wavelengths of the subbands:

1. Choose the central wavelength λ_0 , the wavelength bandwidth $\Delta\lambda$, and the number of subbands n_{sub} . We assume n_{sub} is an odd number to ensure that λ_0 will also be the center wavelength of the central waveband. Then the edges of the full band will be at $\lambda_0 \pm \Delta\lambda/2$. We will call the lower edge λ_m .
2. Set the padding N_0 for the central subband and sampling s_0 close to the desired sampling, such that together they satisfy Eq. (81): $N_0 s_0 = n_D$.
3. For each subband i , define an interim wavelength $\lambda_i' = \lambda_m + (\Delta\lambda/n_{\text{sub}}) \cdot (i - 1/2)$. The set of λ_i' would be the centers of equal-sized subbands, but they do not necessarily satisfy the constraints.
4. Set the subband paddings: $N_i = 2[n_D/(2s')]$, where $[\dots]$ is the nearest-integer (“round”) operator and $s' = s_0 \lambda_0 / \lambda_i'$.
5. Set the subband center wavelengths: $\lambda_i = N_i \cdot \frac{s_0 \lambda_0}{n_D}$.

This algorithm produces a set of wavebands that are symmetric in position and width about the central wavelength λ_c . It gives subband samplings precisely consistent with the chosen central sampling s_0 and achieves this with FFT-efficient, even-sized paddings. Note that the subbands, while symmetric about the center, are not necessarily of equal width. When adding the subband contributions to the total intensity, the appropriate weighting, based on the subband widths, must be applied. One useful test to see that a polychromatic model is working correctly is to introduce a point source that is located at some angular offset from the LOS and ensure that the PSF peak location is the same for all the subbands.

In the Fourier propagation package called PROPER,³⁹ developed at the Jet Propulsion Laboratory and widely used for coronagraph modeling, consistent sampling across the wave band is not produced by adjusting the padding. Rather, PROPER provides a function that applies a user-specified magnification, applied to the field, using a choice of interpolation methods. This provides an additional degree of freedom; zero-padding is not tied to the detector plate scale and sampling. If the pupil matrix linear dimension is n_D and the array size is N , then the sampling of the Fourier transform is $s_F = n_D/N$ in Eq. (79), regardless of the wavelength. The actual system’s image plane sampling, on the other hand, is wavelength dependent and is given by Eq. (80). The magnification to be applied by PROPER, therefore, needs to be

$$M(\lambda) = \frac{s_F}{s(\lambda)} = \left(\frac{n_D/N}{s_0} \right) \cdot \frac{\lambda}{\lambda_0}. \quad (82)$$

Acknowledgments

This work was supported by the HabEx study office at the Jet Propulsion Laboratory, under the oversight of the Exoplanet Exploration Program (ExEP), for NASA. The analytical framework presented in Appendix A was developed by the Roman coronagraph project. The authors thank Keith Warfield for his support of this work and Brian Kern for discussions on the modeling approach. The authors would also like to thank Nicholas Reich, technical editor through ASRC Federal’s communications contract with NASA MSFC.

References

1. L. J. Henderson, “The fitness of the environment: an inquiry into the biological significance of the properties of matter,” *Am. Nat.* **47**(554), 105–115 (1913).

2. G. Gonzalez and J. W. Richards, *The Privileged Planet: How Our Place in the Cosmos is Designed for Discovery*, Regnery Publishing, Washington, D.C. (2004).
3. P. D. Ward and D. Brownlee, *Rare Earth: Why Complex Life is Uncommon in the Universe*, Springer, New York (2000).
4. The National Academies of Sciences, Engineering, and Medicine, *New Worlds, New Horizons in Astronomy and Astrophysics*, The National Academies Press, Washington, D.C. (2010).
5. NASA, *NASA Space Technology Roadmaps and Priorities: Restoring NASA's Technological Edge and Paving the Way for a New Era in Space*, p. 469, The National Academies Press, Washington, D.C. (2012).
6. J. E. Krist et al., "End-to-end coronagraphic modeling including a low-order wavefront sensor," *Proc. SPIE* **8442**, 844253 (2012).
7. J. J. Green and S. B. Shaklan, "Optimizing coronagraph designs to minimize their contrast sensitivity to low-order optical aberrations," *Proc. SPIE* **5170**, 25 (2003).
8. S. B. Shaklan et al., "Stability error budget for an aggressive coronagraph on a 3.8 m telescope," *Proc. SPIE* **8151**, 815109 (2011).
9. H. P. Stahl, M. Postman, and W. S. Smith, "Engineering specifications for large aperture UVO space telescopes derived from science requirements," *Proc. SPIE* **8860**, 886006 (2013).
10. H. P. Stahl et al., "AMTD: update of engineering specifications derived from science requirements for future UVOIR space telescopes," *Proc. SPIE* **9143**, 91431T (2014).
11. M. T. Stahl, S. B. Shaklan, and H. P. Stahl, "Preliminary analysis of effect of random segment errors on coronagraph performance," *Proc. SPIE* **9605**, 96050P (2015).
12. H. P. Stahl, "Advanced mirror technology development (AMTD) project: overview and year four accomplishments," *Proc. SPIE* **9912**, 99120S (2016).
13. M. T. Stahl, H. P. Stahl, and S. B. Shaklan, "Contrast leakage as a function of telescope motion," in *Mirror Technol./SBIR/STTR Workshop*, hosted by the International Society for Optical Engineering, NASA Marshall Space Flight Center, Bellingham, WA (2016).
14. B. Nemati et al., "Effects of space telescope primary mirror segment errors on coronagraph instrument performance," *Proc. SPIE* **10398**, 103980G (2017).
15. B. Nemati et al., "HabEx telescope WFE stability specification derived from coronagraph starlight leakage," *Proc. SPIE* **10743**, 107430G (2018).
16. L. Lebouilleux et al., "Pair-based analytical model for segmented telescopes imaging from space for sensitivity analysis," *J. Astron. Telesc. Instrum. Syst.* **4**(3), 035002 (2018).
17. G. Ruane et al., "Vortex coronagraphs for the Habitable Exoplanet Imaging Mission concept: theoretical performance and telescope requirements," *J. Astron. Telesc. Instrum. Syst.* **4**(1), 015004 (2018).
18. L. E. Coyle et al., "Large ultra-stable telescope system study," *Proc. SPIE* **11115**, 111150R (2019).
19. B. P. Crill and N. Siegler, "Space technology for directly imaging and characterizing exo-Earths," *Proc. SPIE* **10398**, 103980H (2017).
20. W. A. Traub and B. R. Oppenheimer, "Exoplanets direct imaging of exoplanets," in *Exoplanets*, S. Seager, Ed., pp. 111–156, University of Arizona Press, Tucson, AZ (2010).
21. A. Mallama, B. Krobusek, and H. Pavlov, "Comprehensive wide-band magnitudes and albedos for the planets, with applications to exo-planets and Planet Nine," *Icarus* **282**, 19–33 (2017).
22. B. Nemati, J. E. Krist, and B. Mennesson, "Sensitivity of the WFIRST coronagraph performance to key instrument parameters," *Proc. SPIE* **10400**, 1040007 (2017).
23. J. Krist, B. Nemati, and B. Mennesson, "Numerical modeling of the proposed WFIRST-AFTA coronagraphs and their predicted performances," *J. Astron. Telesc. Instrum. Syst.* **2**(1), 011003 (2015).
24. D. Mawet et al., "The vector vortex coronagraph: sensitivity to central obscuration, low-order aberrations, chromaticism, and polarization," *Proc. SPIE* **7739**, 773914 (2010).
25. J. Trauger et al., "Hybrid Lyot coronagraph for WFIRST-AFTA: coronagraph design and performance metrics," *J. Astron. Telesc. Instrum. Syst.* **2**(1), 011013 (2016).
26. M. N'Diaye and R. Soummer, "ATLAST APLC coronagraph design," courtesy of Mamadou N'Diaye and Remi Soummer, STScI, private communication.

27. J. E. Harvey and C. Ftaclas, "Diffraction effects of telescope secondary mirror spiders on various image-quality criteria," *Appl. Opt.* **34**(28), 6337–6349 (1995).
28. J. V. Baliga and B. D. Cohn, "Simplified method for calculating encircled energy," *Simulation Model. Opt. Syst.* **0892**, 152 (1988).
29. L. Pogorelyuk and N. J. Kasdin, "Dark hole maintenance and a posteriori intensity estimation in the presence of speckle drift in a high-contrast space coronagraph," *Astron. J.* **873**(1), 95 (2019).
30. G. Ruane et al., "Reference star differential imaging of close-in companions and circumstellar disks with the NIRC2 vortex coronagraph at the W. M. Keck Observatory," *Astron. J.* **157**(3), 118 (2019).
31. M. Ygouf et al., "Data processing and algorithm development for the WFIRST-AFTA coronagraph: reduction of noise free simulated images, analysis and spectrum extraction with reference star differential imaging," *Proc. SPIE* **9605**, 96050S (2015).
32. D. Mawet et al., "Fundamental limitations of high contrast imaging set by small sample statistics," *Astron. J.* **792**(2), 97 (2014).
33. A. Give'on, B. D. Kern, and S. Shaklan, "Pair-wise, deformable mirror, image plane-based diversity electric field estimation for high contrast coronagraphy," *Proc. SPIE* **8151**, 815110 (2011).
34. B. J. Seo et al., "Hybrid Lyot coronagraph for WFIRST: high-contrast broadband testbed demonstration," *Proc. SPIE* **10400**, 104000F (2017).
35. S. R. Martin and A. J. Booth, "Strong starlight suppression sufficient to enable direct detection of exoplanets in the habitable zone," *Astron. Astrophys.* **511**(2), L1 (2010).
36. L. K. Harding et al., "Technology advancement of the CCD201-20 EMCCD for the WFIRST coronagraph instrument: sensor characterization and radiation damage," *J. Astron. Telesc. Instrum. Syst.* **2**(1), 011007 (2015).
37. S. Gaudi et al., "The Habitable Exoplanet Observatory (HabEx): exploring new worlds, understanding our Universe," NASA, p. 21, 2019, <https://www.jpl.nasa.gov/habex/pdf/HabEx-Final-Report-Public-Release-LINKED-0924.pdf>.
38. W. H. Press et al., *Numerical Recipes in C: The Art of Scientific Computing*, 1st ed., Cambridge University Press, Cambridge, England (1989).
39. J. E. Krist, "PROPER: an optical propagation library," SourceForge, <http://proper-library.sourceforge.net/>.

Bijan Nemati received his PhD in high-energy physics from the University of Washington in 1990. Since 2001, he has been working on modeling and technology development of advanced space-based astronomical instruments, such as interferometers and coronagraphs, first at the Jet Propulsion Laboratory and recently at the University of Alabama in Huntsville. On the Roman coronagraph, he has served as a detector lead, modeling lead, and system engineer in charge of the coronagraph error budget.

H. Philip Stahl received his BA degree in physics and mathematics from Wittenberg University in 1979, his MS degree in optical science, and his PhD from the University of Arizona Optical Sciences Center, in 1983 and 1985, respectively. He is a senior optical physicist at NASA Marshall Space Flight Center currently leading the telescope design study for the potential Habitable Exoplanet Observer Mission and is a principle investigator developing advanced technology to manufacture, test, and thermally control large mirror systems. Previous NASA assignments include: principle investigator for the Advanced Mirror Technology Development project; mirror technology lead for the James Webb Space Telescope, including development of JWST mirror fabrication and testing technology and providing insight/oversight of JWST optical components; and various technology roadmap studies. Additionally, he was responsible for fabrication and test of the Spitzer secondary mirror, alignment, and operation of the Air Force 4-meter 7-segment LAMP mirror, absolute calibration of a LIGO test flat, testing of the Keck primary mirror segments, and designing human-operated optical systems for Shuttle micro-gravity experiments. He is a leading authority in optical metrology, optical engineering, and phase-measuring interferometry. Many of the world's largest telescopes have been fabricated with the aid of high-speed and infrared phase-measuring interferometers developed by him,

including the Keck, VLT, and Gemini telescopes. He is a fellow of SPIE, fellow of OSA, and member of AAS and IAU. He was the 2014 SPIE president and an ICO vice-president (2005 to 2011).

Mark T. Stahl received his BS degree in physics and mathematics from Wittenberg University in 2010 and his master's degree in electrical engineering from the University of Alabama in Huntsville in 2015. He has been an optical physicist at NASA Marshall Space Flight Center, since 2012, where he supports the x-ray telescope manufacturing process, including polishing, metrology, and telescope assembly. Previous assignments include: lead systems engineer on an x-ray telescope proposal, test conductor of flight hardware verification, and principle investigator on a surface roughness study. He has eight SPIE publications and has served on the SPIE Information Technology Committee from 2018 to 2019.

Garreth J. Ruane is an optical engineer at NASA's Jet Propulsion Laboratory with a background in astronomical instrumentation, especially for high contrast imaging and characterization of exoplanets. He received his PhD from Rochester Institute of Technology in 2016 and he was an NSF Astronomy and Astrophysics Postdoctoral Fellow from 2016 to 2019.

Leah J. Sheldon received her bachelor of science degree in mathematics from the University of Alabama in Huntsville (UAH) in 2009 and is currently a graduate student in mathematics at UAH. She is also a research assistant at the UAH Center for Applied Optics, where she supports the Nancy Grace Roman Space Telescope coronagraph project on a contract with the Jet Propulsion Laboratory.