Optical Engineering

SPIEDigitalLibrary.org/oe

Space telescope sensitivity and controls for exoplanet imaging

Richard G. Lyon
Mark Clampin

Space telescope sensitivity and controls for exoplanet imaging

Richard G. Lyon
Mark Clampin
NASA Goddard Space Flight Center
Greenbelt, Maryland 20771
E-mail: Richard.G.Lyon@nasa.gov

Abstract. We address design considerations and outline requirements for space telescopes with capabilities for high contrast imaging of exoplanets. The approach taken is to identify the span of potentially detectable Earth-sized terrestrial planets in the habitable zone of the nearest stars within 30 parsecs and estimate their inner working angles, flux ratios, SNR, sensitivities, wavefront error requirements, and sensing and control times parametrically versus aperture size. We consider 1, 2, 4, 8, and 16-m diameter telescope apertures. The achievable science, range of telescope architectures, and the coronagraphic approach are all active areas of research and are all subject to change in a rapidly evolving field. Thus presented is a snapshot of our current understanding with the goal of limiting the choices to those that appear currently technically feasible. We describe the top-level metrics of inner working angle, contrast and photometric throughput and explore how they are related to the range of target stars. A critical point is that for each telescope architecture and coronagraphic choice, the telescope stability requirements have differing impacts on the design for open- versus closed-loop sensing and control. [DOI: 10.1117/1.OE.51.1.011002]

Subject terms telescope; coronagraphy; exoplanets; wavefront sensing; wavefront control.

Paper 111192SS received Sep. 24, 2011; revised manuscript received Nov. 2, 2011; accepted for publication Nov. 3, 2011; published online Feb. 8, 2012; corrected Feb. 15, 2012.

1 Introduction

High-contrast exoplanet imaging refers to numerous variations of coronagraphy and occulters whereby the starlight is suppressed relative to the planet light, thereby increasing the planet-to-star contrast. Contrast, as defined herein, is expressed as a ratio of the of the star's luminosity to planet's luminosity. A coronagraph operating at a contrast of 10^{10} would imply that a star and planet differing by 10 orders of magnitude in luminosity would be detectable in the focal plane. An ideal coronagraph would suppress all the starlight leaving only the planet's light. The contrast at which a given coronagraph operates is a function of the angular separation of the planet to star, since a planet with a large angular separation from its parent star sees a natural reduction in the starlight due to its being concentrated into an Airy disk point spread function [(Fig. 1(a)]. The planet-to-star angular separation at which the contrast falls below a required value is known as the inner working angle (IWA). Thus planets at or outside the IWA are considered detectable, while those in the IWA are not. However, there is not a definitive break between inside and outside the IWA since the detectability is often a graceful function of the IWA [Fig. 1(b)]. The outer working angle (OWA) is the largest separation angle at which a planet could be detected and this is generally limited by the field-of-view of the optics and/or the number and density of actuators of the deformable mirrors within the coronagraph.

Light collected by the aperture is absorbed, diffracted, and/or scattered. It thus absorbs and spreads the focal plane planet light out over a larger region, and scatters starlight into the region of the planet; these effects lower the contrast. Throughput is generally lower for high-contrast imaging systems than for simple imaging systems. This is due to the larger number of optics and the introduction of focal planet masks (occulting masks) and pupil plane (Lyot) masks. Ideally the star- and planet-light throughput would be zero and unity, respectively. However, this is not generally true in practice due to a host of effects that will be described.

The target star, if we assume it has one or more planets, would likely have dust and debris at or near the equatorial plane of the star; this contributes stray light known as exozodiacal light (exoZodi). Additionally, dust in our solar system also contributes stray light known as local zodiacal light (Zodi).

The target star is generally relatively close to our solar system, from 1 to 30 parsecs, and thus is likely to have background stars and galaxies within the field-of-view that could initially be misinterpreted as planets. Relative motion and/or spectral differences can be used as discriminators. Additionally, since a planet orbits its star, conditions would not always be favorable to detection since it may be too close to the star, or at a phase angle or inclination angle where little or no light is reflected from the planet towards the telescope. A single observation can only determine if a planet exists, but not whether it does not exist. Thus a given star system may have to be observed multiple times to definitively determine whether a planet exists. This is known as "completeness."[1–3] Ensuring completeness to a given level links the number of observations for each star system, to overall mission lifetime.

The signal-to-noise ratio (SNR) is the ratio of the collected planetary photons to the integrated noise contribution, which includes stellar leakage from diffraction, imperfect

0091-3286/2012/$25.00 © 2012 SPIE

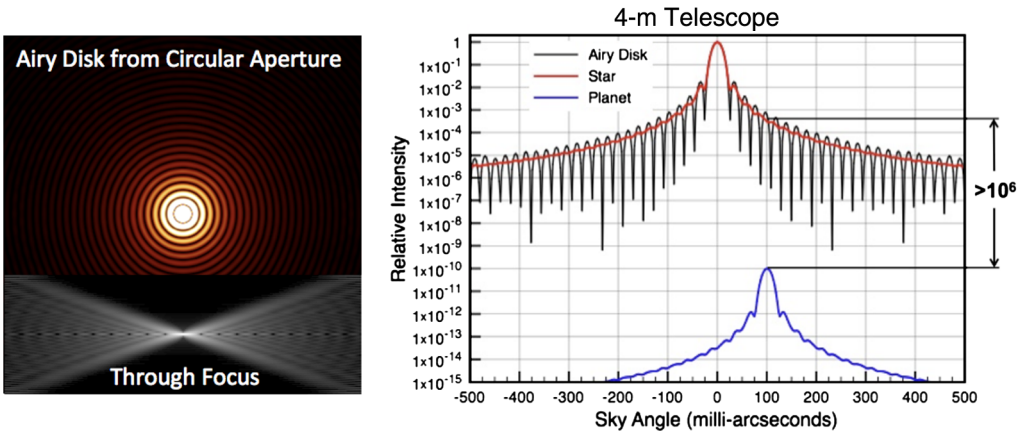

Fig. 1 Contrast and focal plane diffraction. (a) Example of a 4-m circular aperture telescope's ideal diffraction pattern at $\lambda = 500$ nm. (b) The star's Airy disk normalized to unity at the peak (red) and a planet with a luminosity of 10^{-10} of the star separated by 100 mas (blue) would appear only 10^6 to 10^7 times dimmer due to the angular separation.

optics, occulting masks, Lyot stops, stray light, pointing jitter, finite size of target star, thermal- and vibration-induced optical instabilities, photon noise, read noise, dark current noise, and out-of-field scatter. The SNR is a function of contrast, angular separation, and time. Solving for the time to a given SNR yields a viable sensitivity metric, e.g., solving for the time to a SNR = 5 (5-sigma detection) at the IWA.

Contrast, IWA, OWA, and optical throughput are purely instrumental (telescope + coronagraph) parameters whereas SNR and sensitivity depends on the target star, angular separation of planet to star, as well as the instrument's performance; thus, the discovery space of potential science targets plays a large role in defining the required telescope and coronagraph architecture.

The purpose of an exoplanet high-contrast system is to detect and characterize exosolar planets. This problem is manifestly difficult due to typical terrestrial planets being $\sim 10^{10}$ times dimmer than the parent star in reflected light and at angular separations as close as a few tens of milliarcseconds (mas). In designing a high-contrast imaging system, performance considerations have to include: aperture size; diffracting structure within the aperture; optical surface quality; optical stability; polychromatic effects; polarization; sensing and control approach; and control times relative to drift rates, pointing, and postprocessing such as point-spread function and background subtraction.

High-contrast imaging for exoplanets is not a new problem. The Hubble Space Telescope's (HSTs) original instrument complement contained a coronagraph, the Faint Object Camera (FOC) $f/288$ coronagraphic mode. However, few images were collected since HST's primary mirror conic constant rendered it impractical for coronagraphy. Subsequent HST instruments such as Near Infrared Camera and Multi-Object Spectrometer (NICMOS), Space Telescope Imaging Spectrograph (STIS) and Advanced Camera for Surveys/High Resolution Channel (ACS/HRC) had coronagraphs within them. However, contrasts corresponding to 10^{-9} to 10^{-10} suppression of the central star could only be achieved at significant OWA (e.g. Kalas et al.).[4] If a future servicing mission were available for the HST then an area where a new HST scientific instrument could have a high impact on exoplanet science is with a high-contrast instrument.[5] The James Webb Space Telescope (JWST) also will have significant high-contrast imaging capabilities.[6]

2 Mapping of Science Requirements to Telescope

2.1 *Candidate Stars*

A database (HIP30), from the Hipparcos mission[7,8] of 2350 stars out to 30 parsecs has become a standard for designing mission architectures for direct imaging exoplanet missions. The HIP30 catalog has been culled to a shorter list of 575 viable candidates, searchable for exoplanets by assuming that for each star a prospective planet exists at the orbital distance where liquid water could exist, i.e. in the habitable zone (HZ). The inner edge (water boils off and/or runaway greenhouse effect) of this zone is the innerHZ (*i*HZ) and outer edge (where water remains in ice phase) is the outerHZ (*o*HZ) for terrestrial planets. We define an average HZ (aveHZ) as aveHZ $= (1/2)(i\text{HZ} + o\text{HZ})$ and use this as its nominal value.

Figure 2 plots the 575 candidates versus distance from our solar system where the "All Stars" are the 2350 stars in the database and "Candidates" are the selected 575 candidate stars. The five other curves show the stars by spectral classes based on effective blackbody temperature, M (<3500 K), K, (3500 to 5000 K), G (5000 to 6000 K), F (6000 to 7500 K), and A (7500 to 10,000 K). The number to the right of the spectral class in the legend shows the number of stars for each spectra class. There are 24 stars labeled as U-stars (for unknown) that have no listed spectral class in the database. G-stars are similar in spectral class to our own Sun. Overall there are 163 M, 219 K 124 G, 27 F and 18 A candidates. If a terrestrial planet existed within the HZ, then based on the star's luminosity and spectral class a region around the star can be mapped to an angular separation of the planet from its parent star versus distance to the star as shown in Fig. 3. The angular scale of the HZ is an important consideration in the sizing of the telescope aperture and sets the IWA for high-contrast imaging. It is unlikely that all candidates will have a planet and the probability that a given planet has a terrestrial planet is known as η_{EARTH}. Various estimates are available for η_{EARTH} and NASA's ongoing Kepler mission and the European Space Agency's ongoing COROT mission will likely allow a more refined estimate in the near future.

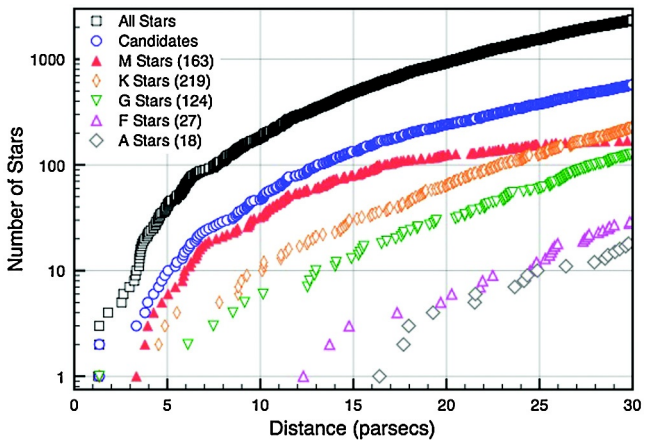

Fig. 2 Hipparcos stars <30 parsecs from Earth. All 2350 Hipparcos stars plotted versus distance from Earth (black) and then are filtered by selection criteria to a list of 575 candidates (blue) and separated by stellar spectral class.

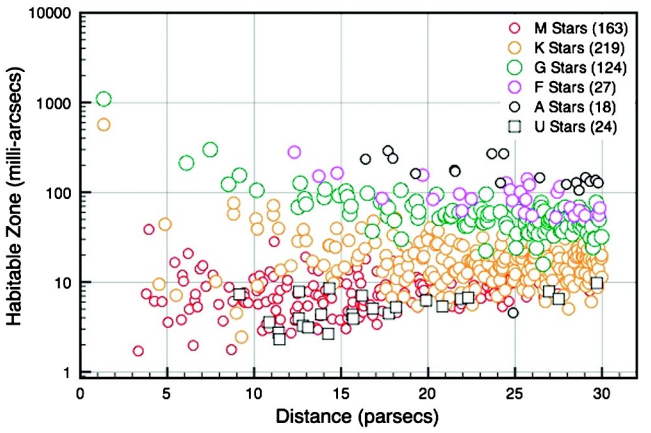

Fig. 3 Candidate stars mean habitable zone versus distance to star by spectral class. All candidate HZ's are <1 arcsec from the parent star and concentrated between 1 and 100 mas.

Scatter, diffraction, stray light, and stability of the sensing and control limit the IWA. An approach limited to, for example, IWA $= 2(\lambda/D)$ requires accurate sensing and control of wavefront and amplitude spatial frequencies at ~2 cycles per aperture, i.e., of periods on the order of ½ the diameter. This is due to the diffraction integral, in the small-angle approximation, mapping periodic wavefront errors at N cycles per aperture to localized speckle in the focal plane at $N(\lambda/D)$. Thus approaches operating at small inner working angle require more demanding tolerances at low spatial frequencies, i.e., for points on the primary mirror that are physically further apart and thereby more difficult to sense and control. While $4(\lambda/D)$ is considered as being conservative[9] and $1(\lambda/D)$ as being considered very aggressive, it is expected that in the nearer term, while still aggressive, $2(\lambda/D)$ is more reasonable, and exoplanets from the ground have been imaged at this IWA.[10] Thus, we assume a coronagraphic instrument with an IWA of $2(\lambda/D)$ where λ is the wavelength of visible band reflected planetary light ($\lambda = 550$ nm) and D is the diameter of the aperture.

The HZ is mapped to an aperture of a specific diameter, by setting the IWA to the mean HZ, IWA $= 2(\lambda D) =$ HZ or $D = 2(\lambda/\mathrm{HZ})$, and the integral number of candidates at or outside an IWA $= 2(\lambda/D)$ versus aperture diameter is plotted in Fig. 4. Scaling to a different IWA is accomplished by shifting the abscissa by the same factor, e.g., for an IWA $= 4(\lambda/D)$ would require the abscissa (on a log scale) to shift to the left such that on Fig. 4 a meter aperture is shifted to ½ m, etc., and thus the number of candidates would decrease. From Fig. 4 it is seen that a 120-m-diameter telescope operating at IWA $= 2(\lambda/D)$ would be required to assess all the candidate stars out to 30 parsecs. This excessively large aperture is driven by the M-stars that are colder and have peak luminosity in the NIR requiring that a planet in the HZ be at closer angular separation for liquid water to exist.

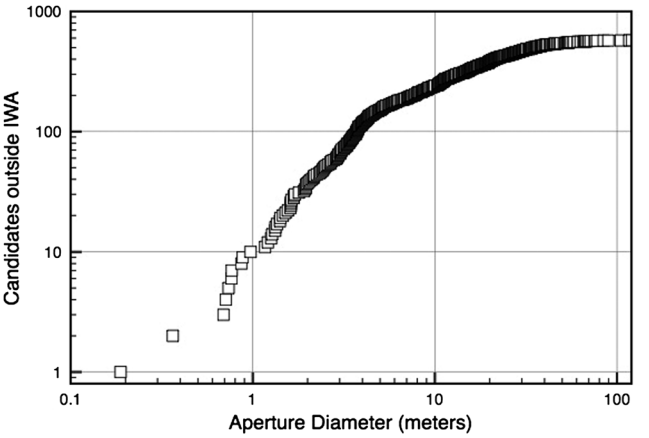

Fig. 4 HIP30 stars mapped to telescope diameter. Candidate HZ is mapped to high-contrast imaging system operating at IWA $= 2(\lambda/D)$. Assessment of *all* stars within 30 parsecs would require a 120-m telescope; however, a 4-m would allow for searching 94 stars, of which 50 are G-stars, within ~34 days per visit, or ~170 days total if each was visited five times for completeness.

A 120-m-aperture telescope operating at IWA $= 2 * \lambda/D$ is clearly difficult for a space mission, but filtering out those M-stars requiring close IWA allows for more reasonable sized apertures as shown in Table 1. If η_{EARTH} is assumed to be 20%, then the number of candidates would decrease by a factor of 5. Jovian (Jupiter analog) planets could also be searched for and would generally result in a larger IWA.

2.2 *Planet-to-Star Luminosity Ratio and Aperture Flux*

The aperture sizing is based purely on inner working angle, but the aperture collecting area is the primary consideration for SNR ratio and sensitivity. Detector integration time could become prohibitive for some candidates due to low planetary photon count rates in visible light. A photometric model can be used to assess the photon count rates for contrast, SNR, and sensitivity calculations.

The luminosity ratio[11,12] of the reflected planet light to emitted starlight is given by:

$$\text{Luminosity Ratio} = \frac{L_P}{L_S} = \alpha\left(\frac{R_P}{R_d}\right)^2 \phi(\beta). \tag{1}$$

L_P and L_S are the luminosities of the planet and star, respectively, α is the planet's geometric albedo, R_p is the radius of the planet, and R_d is the planet's radial distance

Table 1 Candidate stars versus aperture.

		Number of stars at or outside IWA							
Diameter (meters)	IWA (mas)	A (18)	F (27)	G (124)	K (219)	M (163)	U (24)	Total (575)	Δt to SNR = 5
1 m	226.9	5	1	2	1	0	0	9	159.19
2 m	113.4	16	8	6	1	0	0	31	120.74
4 m	56.7	17	22	50	5	0	0	94	33.76
8 m	28.4	17	27	119	30	1	0	194	6.08
16 m	14.2	17	27	124	132	9	0	309	0.79

from the star. $\phi(\beta)$ is the orbital phase function given by $\phi(\beta) = [\sin\beta + (\pi - \beta)\cos\beta]/\pi$ where $\beta \in [0, \pi]$ is the phase angle and where $\cos\beta = -\sin(i)\sin(2\pi\Phi)$ where $i \in [0, \frac{\pi}{2}]$ is the inclination angle. An inclination angle of 0 implies the star system is seen face-on, and an inclination of $\pi/2$ implies the star system is seen edge-on. The orbital phase is given by $\Phi \in [0, 1]$ and for an inclination angle of 0 then $\phi(\beta) \approx 0.32$ independent of phase angle. At an inclination angle of 90 deg the phase function reaches a maximum of ~0.32 and a value of 0.32 is used throughout, i.e., we assumed the best viewing conditions. The geometric albedo of Earth is 0.367 and this is also assumed throughout. The phase function is 0.32 for the point of longest elongation and this is not necessarily where the planet is brightest but where the planet is visible.[11] The radius of the Earth is 6378.1 km and the Sun–Earth distance is 149×10^6 km, yielding an Earth/Sun luminosity ratio of 2.2×10^{-10}. This value can be used to scale the luminosity ratio of an assumed Earth-sized terrestrial planet at the mean HZ around stars within the database via:

$$\frac{(L_P/L_S)}{(L_{\text{EARTH}}/L_{\text{SUN}})} = \frac{\alpha_P}{\alpha_{\text{EARTH}}}\left(\frac{R_P}{R_{P\text{-EARTH}}}\right)^2\left(\frac{R_{d\text{-EARTH}}}{R_d}\right)^2, \tag{2}$$

where $L_{\text{EARTH}}/L_{\text{SUN}} = 2.2 \times 10^{-10}$ is the luminosity ratio of Earth to the Sun and L_P/L_S is the luminosity ratio of the planet to star. The albedo and the radius of the planet are both not generally known before an observation and we assume they are both the same as for Earth, to arrive at:

$$\frac{L_P}{L_S} = 2.2 \times 10^{-10}\left(\frac{R_{d\text{-EARTH}}}{R_d}\right)^2. \tag{3}$$

The radial distance for Earth from our Sun is 1 AU, and as expected, the luminosity ratio scales inversely proportional to the square of the distance from the star. The inverse square is a consequence of assuming a fixed Earth diameter planet, independent of how it formed or the distance from its parent star, and, that its albedo is also constant; thus, its subtended solid angle as seen from its parent star scales inversely to the square of its distance from its parent star. The Earth-sized planet is a tenuous assumption since terrestrial planets are likely to span a range of sizes, masses, albedos inclination, and phase angles; it is however a reasonable assumption as a starting point for defining a range of potential telescope architectures.

Figure 5 plots the luminosity ratio, color-coded for spectral class, for each of the 575 candidates, assuming each has an Earth-sized terrestrial planet at the mean HZ. The ordinate is the number of stars with luminosity ratio greater than the value shown on the abscissa. The majority of star systems have luminosity ratios greater than 10^{-10} with most A stars having less favorable luminosity ratios. G, K, and M stars tend to have more favorable luminosity ratios. The luminosity ratio is independent of the telescope plus instrument and depends only on the star system.

Figure 6 plots the luminosity ratio for a planet in the HZ versus angular separation for each of the spectral classes. M and K stars tend to have more favorable luminosity ratios for detection; however, they tend to be colder stars with the HZ closer to the star requiring either a smaller IWA or larger aperture telescope to achieve the required contrast. G, F, and A stars have more favorable angular separations requiring larger IWA but smaller luminosity ratios, requiring higher contrast to detect them. Figure 6 relates the telescope aperture (via the IWA) and the desired contrast (via the luminosity ratio) over the span of candidate stars in the HIP30 database. Figure 6 shows that the luminosity ratio scales as approximately the inverse of the angular separation independent of spectral class. This implies that a coronagraph may not have as stressing contrast requirements for

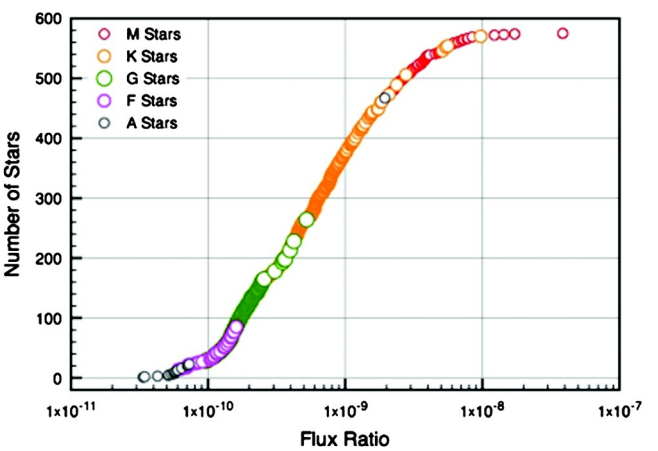

Fig. 5 Number of candidates versus luminosity ratio for each spectral class. Luminosity ratio is independent of the telescope and instrument and depends only on the star system.

HZ at smaller angular separations, and a compromise exists between IWA and contrast when the available candidate stars are incorporated.

The stellar photon count rate at the telescope aperture is estimated from the star's visual magnitude, aperture diameter, and spectral bandpass. The stellar and terrestrial aperture photon count rate for a planet in the HZ, with $F_0 = 10^8$ ph/sec-m^2-nm, are given by:

$$\begin{cases} F_{\mathrm{STAR}}(\mathrm{ph/sec}) = \pi F_0 \times 10^{-\frac{M_V}{2.5}} (D/2)^2 \Delta\lambda \\ F_{\mathrm{PLANET}}(\mathrm{ph/sec}) = (L_P/L_S)\pi F_0 \times 10^{-\frac{M_V}{2.5}} (D/2)^2 \Delta\lambda \end{cases}. \tag{4}$$

Figure 7 plots the total collected planetary photons per second (ph/sec) for an aperture area of 1-m^2 assuming each candidate has a planet in the HZ with a 20% spectral filter of full-width-half-max (FWHM) of $\Delta\lambda = 110$ nm centered on $\lambda = 550$. The estimates are from applying the second equation in Eq. (4). Scaling to other aperture sizes is by multiplying the aperture area in m^2. These aperture count rates are not the detector count rates, since diffraction and scatter spread the counts out, and absorption losses in the telescope optics, instrument, and detector reduce the focal plane count rates.

2.3 *Focal Plane Count Rates*

The photon counts incident upon the aperture are subsequently absorbed and diffracted in the optics and have imperfect conversion from photon to electron counts by the detector. The diffraction is a consequence of the finite aperture size and results in redistribution of the photons into an optical point spread function (PSF). The ideal PSF of a circular aperture is the Airy disk function, where $\theta = \sqrt{\theta_x^2 + \theta_y^2}$ is the sky angle in units of lam/D, given by:

$$\mathrm{PSF}(\theta) = \pi \left[\frac{J_1(\pi\theta)}{\pi\theta}\right]^2 \approx \pi^{\frac{3}{2}} \left(\frac{3\sqrt{3}}{16}\right) \left[\frac{1}{1 + (\pi^4/8)\theta^3}\right]. \tag{5}$$

The PSF is normalized such that its integral is unity. Its central core, to a radius of $\theta = 1.22\lambda/D$, contains ~84% of the energy and a detector pixel is sized to account for 84% of the energy.

The respective stellar and planetary energy are each distributed into separate PSFs centered on their sky locations. For an angular separation of θ_P, in units of λ/D, the distribution of focal plane photoelectron count rates is given by:

$$\begin{cases} F_{\mathrm{STAR}}(\theta) = 0.84 S F_0 10^{-\frac{M_V}{2.5}} (D/2)^2 (3\sqrt{3}/16)\pi^5 (1/1 + \frac{\pi^4}{8}\theta^3)\Delta\lambda T_O \mathrm{q.e.} \\ F_{\mathrm{PLANET}}(\theta) = 0.84 (F_P/F_S) F_0 10^{-\frac{M_V}{2.5}} (D/2)^2 (3\sqrt{3}/16)\pi^5 [1/1 + \frac{\pi^4}{8}(\theta - \theta_P)^3]\Delta\lambda T_O \mathrm{q.e.} \end{cases}. \tag{6}$$

In Eq. (6), T_O is the end-to-end transmission of the optics, *q.e.* is the quantum efficiency in units of photoelectrons/photon, and S is the starlight suppression factor of the coronagraph. The units of $F_{\mathrm{STAR}}(\theta)$ and $F_{\mathrm{PLANET}}(\theta)$ are photoelectrons per pixel per second. Based on the above model, the ratio of focal plane planet detector counts to star counts at the location of the planet $\theta = \theta_P$ is:

$$Q(\theta_P) = \frac{F_{\mathrm{PLANET}}(\theta_P)}{F_{\mathrm{STAR}}(\theta_P)} = \frac{1}{S}\left(\frac{F_P}{F_S}\right)\left(1 + \frac{\pi^4}{8}\theta_P^3\right), \tag{7}$$

where Q is a function of the planet-to-star angular separation, and since the model in Eq. (6) has a θ_P^3 dependency, so does Q. All coronagraphs do not necessarily have this same dependency with angle but this is assumed based on assumptions about diffraction in Lyot and occulter-type coronagraphs. It is likely that a phase-induced amplitude apodization (PIAA coronagraph), without an inverse PIAA, would have a higher order dependency; however, in order to put the field of view back, an inverse PIAA would be required and this would put the dependency back to $\sim\theta_P^3$. Nulling type coronagraphs perform beam nulling in the pupil plane but will ultimately be brought to focus with a limiting aperture introducing diffraction that is also $\sim\theta_P^3$. Thus based on limited information herein, we assume the θ_P^3 but with the caveat that other dependencies are likely but are not yet well understood; as further information becomes available this should be revisited since a clear trade exists between the background diffraction dependency and performance of a given coronagraph.

Defining the contrast at the IWA as $C = S[1 + (\pi^4/8)\mathrm{IWA}^3]^{-1}$ gives

$$Q(\theta_P) = \frac{1}{C}\left(\frac{F_P}{F_S}\right)\frac{1 + \frac{\pi^4}{8}\theta_P^3}{1 + \frac{\pi^4}{8}\mathrm{IWA}^3} \approx \frac{1}{C}\left(\frac{F_P}{F_S}\right)\left(\frac{\theta_P}{\mathrm{IWA}}\right)^3. \tag{8}$$

Thus a star system with a planet in the HZ for a coronagraph at the IWA and operating at $C = 10^{10}$ would give $Q = 1$, implying the count rate from the star and planet in a pixel centered on the planet would be the same. Q expresses the ratio of the planet counts to leaked stellar counts and it is linear in contrast and linear in luminosity ratio but grows approximately as the cube of the angular separation when expressed in units of λ/D, implying that doubling the aperture size moves the planet twice as far out in λ/D units and thus gives an 8-fold gain in Q. The optimal value of Q is discussed is a function of the sampling, detector noise characteristics, pointing jitter, finite size of a detectors' pixel and where the planet occurs within the pixel, spectral bandpass, and the postprocessing algorithms. Coronagraphs can operate with $Q < 1$ since matched filtering, Bayesian estimation[13] and nonlinear techniques can be employed to optimally estimate if a planet exists in a given dataset.

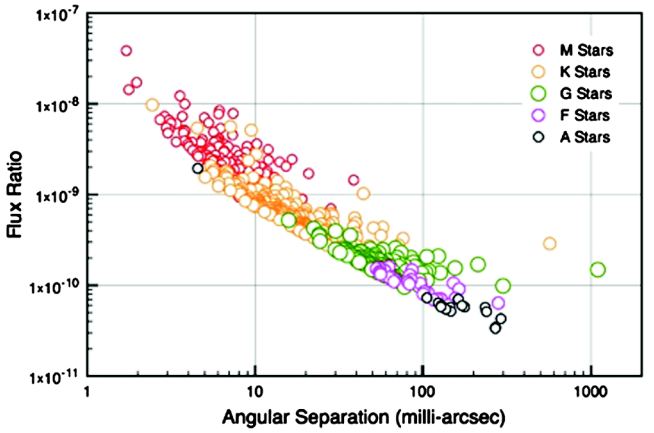

Fig. 6 Luminosity ratio for a terrestrial planet in the HZ for each spectral class versus angular separation. Relates IWA and contrast to science requirements.

2.4 *Local and Exo-Zodiacal Light*

The above does not contain the effects of local zodiacal and exo-zodiacal light. Model and observations of exoZodi generally show complex morphological behavior both radially and azimuthally and as function of the age of the stellar system.[14,15] Herein we are primarily concerned with the deleterious effects both Zodi and exoZodi would have on the detection of terrestrial planets and thus express both the Zodi and exoZodi as simplified models only at the assumed location of a terrestrial planet. Zodi and exoZodi are usually expressed in magnitude per arcsec square. Conversion to photoelectron count rate in a detector pixel of solid angle on the sky of $\Delta\Omega$ is given by:

$$\begin{cases} F_{\mathrm{Ap}:Z} = F_0 10^{-\frac{M_Z}{2.5}} \frac{\pi D^2}{4} \Delta\lambda\Delta\Omega T_0 \mathrm{q.e.} \\ F_{\mathrm{Ap}:EZ} = F_0 10^{-\frac{M_{EZ}}{2.5}} \frac{\pi D^2}{4} \Delta\lambda\Delta\Omega T_0 \mathrm{q.e.} \end{cases}. \tag{9}$$

The units of $\Delta\Omega$ are arcsec-squared. Assuming that a given detector pixel is square of width λ/D, gives $\Delta\Omega = (\lambda/D)^2 \cdot (3600 \times 180/\pi)^2$ arcsec^2, and gives $(\pi D^2/4)\Delta\lambda\Delta\Omega = (\pi/4)\lambda^2\Delta\lambda \cdot (3600 \times 180/\pi)^2$, which is just an expression of the radiance being conserved; i.e., the product of the aperture with solid angle is $\sim\lambda^2$.

The typical values of $M_Z = 22$ magnitudes/arcsec^2 and $M_{EZ} = 22$ magnitudes/arcsec^2 are used throughout to denote 1-Zodi. Using 22 magnitudes/arcsec^2 gives 0.15 photons/ sec /pixel per m^2 of aperture without including the optics transmission and quantum efficiency. Figure 7 also shows the Zodi plotted as a solid line. Even 1-Zodi has the net effect that some of the planets yield lower count rates per pixel than the Zodi. While the exoZodi is not generally known, models of it exist as a function of age of a given star system. Models show that Zodi is slowly varying on a spatial case with respect to the scale of planet; hence it can in principle be subtracted off, but it does contribute photon noise after subtraction. Young systems, <500 Myrs, can have dust density at thousands of times more dense that our solar system at 4.5 Gyrs age. The level of exoZodi represents an unknown for the design of any high-contrast exoplanet system at this time; however, ground-based astronomy and JWST will likely make significant inroads to the levels of dust in nearby candidate systems.

2.5 *Signal-to-Noise Ratio and Sensitivity*

The above set of calculations allows for the estimation of SNR parametrically as a function of aperture over the HIP30 database of candidate stars. With the above relations and SNR, in the form of sensitivity, we are in a better position to estimate the realizable science for a specific set of architectures and to error budget these architectures for various coronagraphic approaches. The SNR is given by:

$$\mathrm{SNR} = \frac{F_{\mathrm{PLANET}}\Delta t}{\sqrt{F_{\mathrm{PLANET}}\Delta t + F_{\mathrm{STAR}}\Delta t + F_Z\Delta t + F_{EZ}\Delta t + i_D\Delta t + \sigma_e^2}}, \tag{10}$$

where i_D is dark current in photoelectrons/ sec /pixel, σ_e is the detector read noise, and Δt is the detector integration time. The residual speckle noise is implicitly included in the term $F_{\mathrm{STAR}}\Delta t$ since this term is the noise variance of the residual leaked starlight due to incomplete starlight suppression and imperfect wavefront control. The term $F_{\mathrm{PLANET}}\Delta t$ is the photon noise due to the planet, and $F_Z\Delta t$ and $F_{EZ}\Delta t$ are the photon noise assuming perfect subtraction of the Zodi and exoZodi. In practice these will be unlikely to be perfectly subtracted and will show up in a manner analogous to flat fielding error. The photon limited SNR can be taken as an upper bound on performance and is given by assuming that $\sigma_e = 0$ and $i_D = 0$ to give:

$$\mathrm{SNR} = \frac{F_{\mathrm{PLANET}}\sqrt{\Delta t}}{\sqrt{F_{\mathrm{PLANET}} + F_{\mathrm{STAR}} + F_Z + F_{EZ}}}. \tag{11}$$

The terms in the denominator are the photon noise from each the planet, stellar leakage (speckle noise), Zodi, and exoZodi, respectively. Scatter from out-of-field sources and background sources are not included. Solving Eq. (11) for time yields the photon-limited time to a given SNR as:

$$\Delta t = \mathrm{SNR}^2 \left(\frac{F_{\mathrm{PLANET}} + F_{\mathrm{STAR}} + F_Z + F_{EZ}}{F_{\mathrm{PLANET}}^2} \right). \tag{12}$$

The terms within Eq. (12) were integrated over a V-band spectral filter for each of the 575 candidates for a 1-m diameter aperture and the "normalized" time versus angular separation plotted in Fig. 8. The actual time (seconds) to SNR are given by dividing the value on the ordinate of Fig. 8 by the factor $D^{\frac{10}{3}}$ where D is the aperture diameter in meters. Thus, for example, a normalized time of 10^6 s would scale

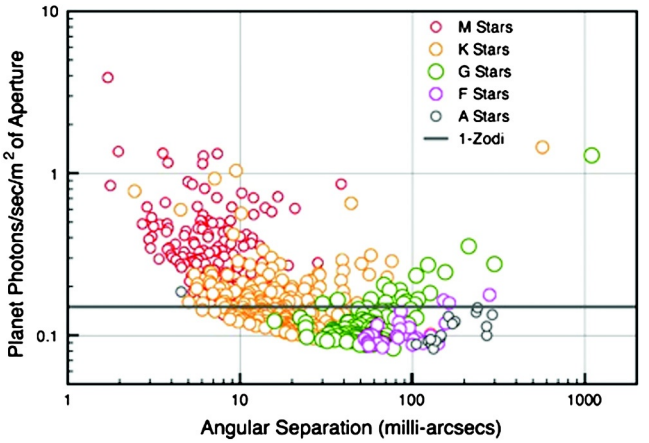

Fig. 7 Aperture flux of planet in HZ versus angular separation. Zodi shown as flux in pixel of solid angle $\sim(\lambda/D)^2$.

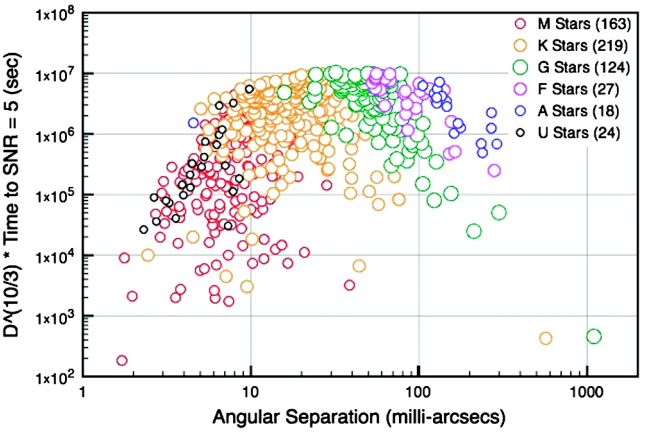

Fig. 8 Normalized V-band sensitivity. Divide the normalized sensitivity by $D^{10/3}$ to arrive at the actual time (in seconds) to achieve a SNR = 5 for each candidate star with a planet in the HZ for a V-band filter.

down by a factor of $2^{\frac{10}{3}} = 10.08$ or to $\sim 10^5$ s for a 2-m diameter aperture.

Times to different SNR's are given by multiplying by the ratio of squares of the SNRs per Eq. (12). Changing spectral filters is accomplished by multiplying time by the ratio of 110 nm to the new filters bandpass width—while this is an approximation, it is reasonable in visible light.

Figure 8 is significant in that it couples the integration time for each of the candidates from each spectral class to its angular separation from the parent star for a planet in the HZ. The choice of aperture size in Fig. 8 has been left unconstrained since we desire to estimate the science return parametrically versus aperture. In other words, the scaling shown in Fig. 8 now allows for parametrically assessing exoplanet detection performance as a function of aperture size and these results are shown in Table 1.

2.6 *Detection Summary*

Table 1 summarizes the number of stars, that if exoplanets exist within their HZ, could be detected for five aperture sizes and the amount of time to SNR = 5 in days to detect these planets using a spectral filter centered on 550 nm with

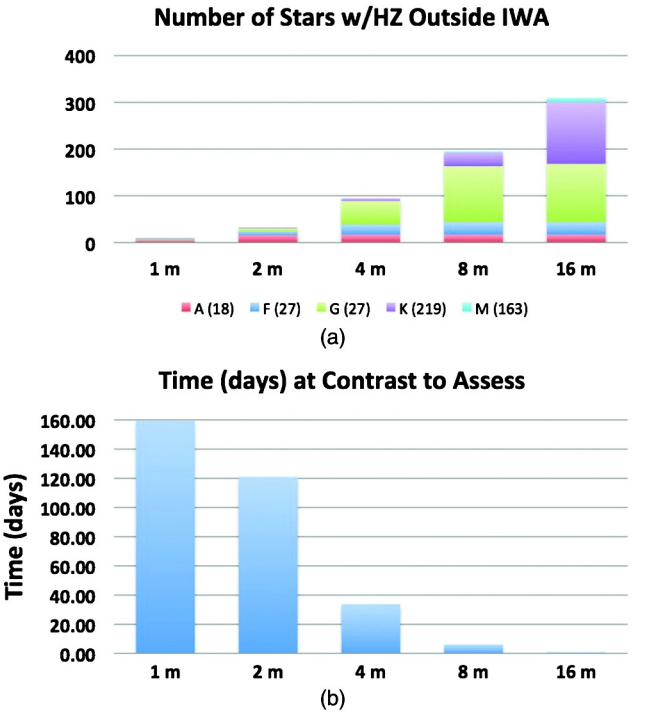

Fig. 9 (a) Number of stars with an HZ at or outside IWA versus aperture diameter (bar plot of second to last column in Table 1). (b) Time (days) to SNR = 5 for a single observation of each of candidates versus aperture in Table 1 (last column of Table 1).

FWHM of 110 nm. The number of stars increases with increasing aperture size [Fig. 9(a)], while the integration time decreases [Fig. 9(b)]. A 4-m aperture operating at $2(\lambda/D)$ IWA would be capable of detecting a total of 94 HZ exoplanets with 50 of those G-stars if exoplanets existed around them. If we assume that to completely search the system it would take five visits at 33.76 days per visit, it would yield a total of 168.8 days, and if η_{EARTH} is 20% then potentially $\sim$18 exoplanets in the HZ could be detected. At the time these 18 exoplanets were definitively detected they could be spectrally characterized to a SNR = 5 with a 10-nm-width spectral filter in (110 nm/10 nm) 33.8 days = 371.4 days, or approximately one year per planet. Lower spectral resolutions would give shorter times.

2.7 *Characterization and Spectroscopy*

Spectroscopic detection is one of the prime mission drivers since it is what would be used to determine if biomarkers existed within a planet's atmosphere. Spectroscopy is more time consuming since the light must be dispersed into spectral bands and integrated long enough to achieve the SNR per spectral band. These times can be long if the collection area is small, per the example in the previous section. Additionally the planet may not be visible throughout the time it takes to integrate enough light since the planet is moving in its orbit. Also, the atmospheric abundances and pressures may be such that the absorption lines are only small dips in the spectrum and if not sufficiently pressure-broadened could be very narrow, all of which when combined may make the spectral characterization difficult. An alternative approach is spectrophotometry, i.e., filter spectroscopy, whereby a set of filters centered on either specific spectral lines and/or

colors are used. If the spectral filters are used serially then the times would still be long; however, if dichroic beamsplitters are used then all the filters could be brought to bear at the same time, effectively multiplexing the time as a dispersing spectrometer would, but with each filter's width optimized for a specific spectral signature. While the range of options is large and ill-defined at this point, we do show the time to SNR, versus aperture size, versus spectral resolution to enable setting a bound on the time to spectrally characterize.

3 Architecture Considerations

3.1 *Generic Model of Internal Coronagraph*

In order to assess the tolerances on a given architecture a generic approach is needed that applies to viable coronagraphs. In order to impart insight and delineate the sensitivities to errors and control frequencies, an approximated analytic formalism is developed. An 8th-order Lyot-type coronagraph[16] is shown in Fig. 10 whereby propagation of light through the coronagraph can be broken down to: (i) propagation from the telescope's exit pupil to 1st focal plane, (ii) application of an occulting mask, (iii) propagation from this occulting plane to a re-imaged exit pupil, (iv) application of a Lyot stop, and (v) propagation to the final science focal plane. While not immediately obvious, such a model can encompass most coronagraphs, provided the correct terms are identified. We will first step through the sequence of steps as shown in Fig. 10 and subsequently approximate the terms that are important for wavefront and amplitude errors.

Let the complex electric field at the telescope's exit pupil be given by:

$$P(\vec{r}) = A(\vec{r})[1+\delta A(\vec{r})]e^{i\delta\phi(\vec{r})} \approx A_0[1+\delta A(\vec{r})+i\delta\phi(\vec{r})], \tag{13}$$

where the approximation is straightforward and based on retaining up to the 1st-order term in the small-angle approximation of the phasor term. This is a common approximation in coronagraphy since in order for a coronagraph to operate at high contrast it must operate in a regime where the small-angle approximation is valid. The effect of this approximation and its range of limitation has been explored in the context of coronagraphy by multiple authors.[17,18]

The normalized pupil plane coordinates are given by $\vec{r}=(x,y)$, such that $|\vec{r}| \le D/2$, where D is the diameter of the aperture. For elliptical or segmented or sparse apertures, it is the largest distance between any two points within the aperture, A_0 is the mean amplitude, and $\delta A(\vec{r})$ represents a fractional variation in the amplitude after traversing the telescope, $\delta\phi(\vec{r})$ is the phase errors incurred in passing through the telescope, and $\delta\phi(\vec{r}) = (2\pi/\lambda)W(\vec{r})$ relates the wavefront error $W(\vec{r})$ to the phase errors, and where λ is the wavelength. The errors in a coronagraph need to be small to detect a planet, and hence the small-angle approximation has been used in Eq. (13), whereby terms to only the first order have been retained.

Propagation through the sequence of five steps shown in Fig. 10 is accomplished by stepping through:

$$\begin{cases} \text{(i)} \rightarrow \text{Exit pupil to focal plane} \\ \text{(ii)} \rightarrow \text{Apply occulting mask} \\ \text{(iii)} \rightarrow \text{Re-image the exit pupil} \\ \text{(iv)} \rightarrow \text{Apply Lyot stop} \\ \text{(v)} \rightarrow \text{Propagate to science focal plane} \end{cases} = \begin{cases} \mathrm{ASF}(\vec{\theta}) = \int P(\vec{r}) e^{-i\frac{2\pi}{\lambda}\vec{\theta}\cdot\vec{r}}\mathrm{d}\vec{r} \\ t_{\mathrm{COR}}(\vec{\theta})\mathrm{ASF}(\vec{\theta}) = t_{\mathrm{COR}}(\vec{\theta}) \int P(\vec{r}) e^{-i\frac{2\pi}{\lambda}\vec{\theta}\cdot\vec{r}}\mathrm{d}\vec{r} \\ \int t_{\mathrm{COR}}(\vec{\theta})\mathrm{ASF}(\vec{\theta}) e^{i\frac{2\pi}{\lambda}\vec{\theta}\cdot\vec{r}}\mathrm{d}\vec{\theta} = \tilde{t}_{\mathrm{COR}}(\vec{r}) * * P(\vec{r}) \\ L(\vec{r})[\tilde{t}_{\mathrm{COR}}(\vec{r}) * * P(\vec{r})] \\ \int L(\vec{r})[\tilde{t}_{\mathrm{COR}}(\vec{r}) * * P(\vec{r})] e^{-i\frac{2\pi}{\lambda}\vec{\theta}\cdot\vec{r}}\mathrm{d}\vec{r} = \mathrm{ASF}_L(\vec{\theta}) * * [t_{\mathrm{COR}}(\vec{\theta})\mathrm{ASF}(\vec{\theta})] \end{cases}. \tag{14}$$

In Eq. (14) $\vec{\theta}=(\theta_x,\theta_y)$ is the focal plane coordinate mapped to sky angular coordinates, $\mathrm{ASF}(\theta)$ is the complex scalar field in the 1st focal plane known as the amplitude spread function, $t_{\mathrm{COR}}(\vec{\theta}) \in \mathbb{C}$ is the focal plane occulting mask and can either be real (amplitude mask) or complex (phase and/or amplitude and phase mask), $L(\vec{r})$ is the Lyot stop, and is usually a region of the reimaged exit pupil that is smaller, i.e. less relative area, than the original exit pupil, and $\mathrm{ASF}_L(\vec{\theta})$ is the amplitude spread function of the Lyot stop. "$**$" denotes 2-D convolution. Magnifications are implicitly taken into account by choosing to work in relative pupil coordinates and angular sky coordinates. Overall phase factors have been neglected, since they do not contribute the final result. The form of the propagators take the form of two-dimensional (2-D) Fourier transforms with conjugate variables given by $\vec{r}$ and $\vec{\theta}$.The science focal plane's complex electric field and intensity are respectively given by:

$$\begin{cases} E_F(\vec{\theta}) = \mathrm{ASF}_L(\vec{\theta}) * * (t_{\mathrm{COR}}(\vec{\theta})\mathrm{ASF}(\vec{\theta})) \\ I_F(\vec{\theta}) = |\mathrm{ASF}_L(\vec{\theta}) * * |[t_{\mathrm{COR}}(\vec{\theta})\mathrm{ASF}(\vec{\theta})]|^2 \end{cases}. \tag{15}$$

Equation (15) describes the basic propagation through an internal coronagraph and can be used to model most variations including Lyot coronagraphs with various occulting masks and Lyot stops,[16] quadrant phase masks,[19] vortex type coronagraphs,[20] and visible nulling coronagraphs.[21] External occulters[22–24] also approximately fit into this model since the starlight is suppressed prior to entering the telescope, and the residual starlight entering the aperture is subsequently diffracted into an Airy-like pattern. Phase induced amplitude apodization (PIAA)[25] performs pupil remapping and requires a different model. However, if a PIAA is used with an inverse PIAA, the field of view is reconstructed after suppression and subsequently brought

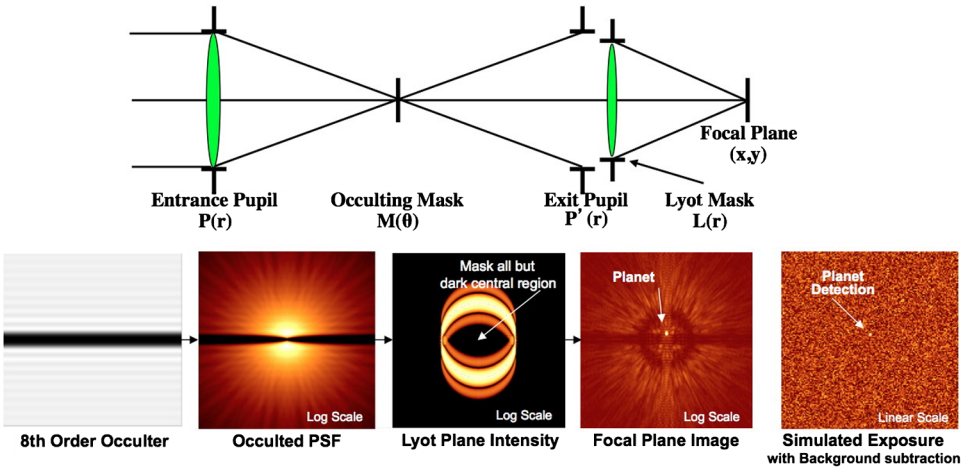

Fig. 10 Star and planet light propagation through 8th-order Lyot coronagraph.

to focus yielding a similar dependence, i.e., the residual will follow approximately an Airy disk-like dependence.

3.2 *Wavefront and Amplitude Errors*

The approach taken is to use the forms derived in Sec. 3.1 based on the linear expansion of the wavefront phasor to assess first-order requirements on sensing and control to achieve the contrast at the IWA for each of the candidate stars. Multiple authors have used various forms of these approaches in the context of coronagraphy to derive scaling relations,[18,26] as the linear approach imparts physical insight and is generally used for simplicity. Full-scale modeling ultimately is required to assess the deviations of the simpler models from the more rigorous models. The form we use is derived herein for completeness.

The approximated form of the telescope pupil can now be used in Eq. (15) to yield

$$I_{\mathrm{FOC}}(\vec{\theta}) \approx \langle |\mathrm{ASF}_L(\vec{\theta}) \\ **(t_{\mathrm{COR}}(\vec{\theta})\mathrm{ASF}(\vec{\theta}) \\ + t_{\mathrm{COR}}(\vec{\theta})\{\mathrm{ASF}_0(\vec{\theta}) \\ **[\delta\tilde{A}(\vec{\theta}) + i\delta\tilde{\phi}(\vec{\theta})]\})|^2\rangle, \tag{16}$$

where $\mathrm{ASF}_0(\vec{\theta})$ is the amplitude spread function without amplitude and phase errors, and $\delta\tilde{A}(\vec{\theta})$, $\delta\tilde{\phi}(\vec{\theta})$ are the 2-D Fourier transforms of the amplitude and phase errors, respectively. If $\mathrm{ASF}_0(\vec{\theta})$ is much more compact than $\delta\tilde{A}(\vec{\theta}) + i\delta\tilde{\phi}(\vec{\theta})$ then Eq. (16) can be approximated by:

$$I_{\mathrm{FOC}}(\vec{\theta}) \approx \langle |\mathrm{ASF}_L(\vec{\theta}) ** \{t_{\mathrm{COR}}(\vec{\theta})\mathrm{ASF}(\vec{\theta}) \\ + t_{\mathrm{COR}}(\vec{\theta})[\delta\tilde{A}(\vec{\theta}) + i\delta\tilde{\phi}(\vec{\theta})]\}|^2\rangle. \tag{17}$$

Let $\langle \delta\tilde{A}(\vec{\theta})\rangle = \langle \delta\tilde{\phi}(\vec{\theta})\rangle = 0$, since mean amplitude and phase errors do not contribute to a loss in contrast, to give:

$$I_{\mathrm{FOC}}(\vec{\theta}) \approx \mathrm{PSF}_L(\vec{\theta}) ** [T_{\mathrm{COR}}(\vec{\theta})\mathrm{PSF}(\vec{\theta})] \\ + \mathrm{PSF}_L(\vec{\theta}) ** \{T_{\mathrm{COR}}(\vec{\theta})[\mathrm{PSD}_{\delta A}(\vec{\theta}) \\ + \mathrm{PSD}_{\delta\phi}(\vec{\theta})]\}, \tag{18}$$

where $T_{\mathrm{COR}}(\vec{\theta}) = |t_{\mathrm{COR}}(\vec{\theta})|^2$ is the intensity transmittance of the coronagraph versus angle on the sky and where we have used $\mathrm{PSD}_{\delta A}(\vec{\theta}) = \langle |\delta\tilde{A}(\vec{\theta})|^2\rangle$ and $\mathrm{PSD}_{\delta\phi}(\vec{\theta}) = \langle |\delta\tilde{\phi}(\vec{\theta})|^2\rangle$ for the power spectral densities of the amplitude and phase errors, respectively.

The power spectral densities are based on ensemble averages of the Fourier transforms of the wavefront and amplitude errors in keeping with the formal definition of the power spectral density (PSD). Thus a given PSD does not represent a specific realization of wavefront or amplitude-induced speckle, but does represent an average over an ensemble of these speckles induced by different realizations of wavefront and amplitude errors. The use of PSDs to specify optical surfaces, wavefront, and amplitude errors is now a standardized approach and was used to specify the requirements on the Hubble Space Telescope, Chandra X-ray Observatory[27] and the James Webb Space Telescope, among others, and a PSD transfer function approach was developed for a visible nulling coronagraph.[28]

A perfect coronagraph is when the first term, $\mathrm{PSF}_L(\vec{\theta}) ** [T_{\mathrm{COR}}(\vec{\theta})\mathrm{PSF}_0(\vec{\theta})] = 0$, and the remainder gives the leakage as:

$$I_{\mathrm{LEAK}}(\vec{\theta}) \approx \mathrm{PSF}_L(\vec{\theta}) \\ ** \{T_{\mathrm{COR}}(\vec{\theta})[\mathrm{PSD}_{\delta A}(\vec{\theta}) \\ + \mathrm{PSD}_{\delta\phi}(\vec{\theta})]\}. \tag{19}$$

If we normalize the PSDs such that integral over the focal plane, i.e. all spatial frequencies, is unity, then the focal plane leakage is given by:

$$I_{\mathrm{LEAK}}(\vec{\theta}) \approx \mathrm{PSF}_L(\vec{\theta}) \\ ** \{T_{\mathrm{COR}}(\vec{\theta})[\sigma_{\delta A}^2\mathrm{PSD}_{\delta A}(\vec{\theta}) \\ + \sigma_{\delta\phi}^2\mathrm{PSD}_{\delta\phi}(\vec{\theta})]\}, \tag{20}$$

and $\sigma_{\delta A}^2$, $\sigma_{\delta\phi}^2$ are the amplitude and phase error variances, respectively, where $\sigma_{\delta\phi} = (2\pi/\lambda)\sigma_{\delta W}$, and $\sigma_{\delta W}$ is the rms wavefront error in the same units as the wavelength. Interchange between spatial frequencies, in cycles per aperture (cpa) and angle on the sky in units of λ/D, is trivial in these units; e.g., a spatial frequency of 4 cpa gives a

speckle centered on $4(\lambda/D)$ of width defined by the width of Lyot stop PSF, i.e., where D is the stopped-down D. The net effect is that, to 1st order, amplitude and phase errors both induce focal plane speckle; however, the focal plane location of the amplitude error-induced speckle does not change nor scale with wavelength, whereas both the focal plane location and brightness of a wavefront error speckle varies with wavelength.[29]

A fixed given integral, i.e., a white noise and a power law PSD, both with the same power integrated out to a given spatial frequency, yields a power-law PSD with more power in the lower spatial frequencies thereby requiring progressively tighter tolerances on the sensing and control at low spatial frequencies and correspondingly less control at mid- and high-spatial frequencies. Low spatial frequencies arise from points on the surface that are far apart, i.e. long correlation lengths. Thus the most stressing tolerances are for those that are the most difficult to control.

A white noise PSD is, however, unlikely; conventional polishing and coating practices tend to yield mirror surface PSDs[30,31] and residual coating error PSDs that follow power law functions of the form $\approx 1/[1+(f/f_n)^{\alpha}]$, where f_n is the knee spatial frequency, so called because on a log-log plot the PSD has a break-over point at the knee frequency. The exponent α is typically on the order of 2 to 3, and the point at which the PSD is larger than the PSF is the transition from diffraction to scatter limited.[32] If we assume a normalized PSD of the form:

$$\mathrm{PSD}(\vec{f}) = \frac{\alpha}{2\pi^2 f_n^2}\ \sin\left(\frac{2\pi}{\alpha}\right)\frac{\sigma_T^2}{1+(|\vec{f}|/f_n)^{\alpha}} \quad \text{for } \alpha > 2, \tag{21}$$

this functional form is normalized such that its integral over all 2-D spatial frequencies is equal to the variance σ_T^2 of the total amplitude or phase error and where $|\vec{f}| = \sqrt{f_x^2 + f_y^2}$. The PSD has units of power per spatial frequency squared, with the spatial frequency in cpa or equivalently in λ/D of sky angle.

3.3 *Temporal Drift and Control*

The amplitude and phase errors are also dynamic in that they vary with time due to thermal and structural variations in the optics and the structures that mount the optics. Even if the primary and secondary mirrors are dimensionally stable, the truss structure has eigenmodes which, when excited, cause the footprint of the primary mirror (PM) beam to shift or shear on the secondary mirror (SM). This implies that the spatial frequency on the PM will shear on the SM, causing temporal variations of both the amplitude and phase errors and this occurs throughout the optical system, i.e., modal excitations and thermal drift shift, and distort the optical surfaces. Thus the PSDs have a component in the temporal frequency domain and are actually three-dimensional (3-D) with respect to the frequencies (f_x, f_y, f_t) where the first two frequencies are spatial frequencies in cpa in x and y directions and the third is the temporal frequency in Hz. The 3-D PSD then gives the power at a given spatial-temporal frequency. This is a more natural way to specify both the critical spatial and temporal frequencies as well as the interplay between them, and lends itself well to the formalism of sensing and control via a transfer function approach for each contributor, and for wavefront and amplitude sensing and DM control.

Assume a temporal PSD of the form $\approx[1+(f_t/f_D)^2]^{-1}$ where f_D is the temporal knee or drift frequency. It is likely that the underlying form of the temporal PSD approximately follows this functional form but with sharp or high Q modes due to the natural frequencies of the structure. Excitation of these modes can occur due to reaction and momentum wheels; the assumption herein is that they are well damped, or isolated, and that this PSD form represents a reasonable approximation to the underlying model.

The overall error spatial-temporal PSDs become:

$$\mathrm{PSD}(\vec{f}, f_t) = \sigma_T^2 \frac{\alpha}{\pi^3 f_n^2 f_D}\ \sin\left(\frac{2\pi}{\alpha}\right)\frac{1}{1+(|\vec{f}|/f_n)^{\alpha}} \times \frac{1}{1+(f_t/f_D)^2} \quad \text{for } \alpha > 2, \tag{22}$$

where $f_t \in [0,\infty]$ and $|\vec{f}| = \sqrt{f_x^2 + f_y^2}$ and $(f_x, f_y) \in [0,\infty]$. The units of the total error PSD are now power per spatial frequency squared per Hz. The integral of total error PSD over all spatial and temporal frequencies is σ_T^2. There is an implicit assumption in this PSD form in that it assumes that the spatial errors are uncorrelated with the temporal errors. Strictly speaking this is not the most general case, since for example, a periodic temporal variation of the SM support structure will introduce time variations that are different at differing spatial frequencies. It is more proper to think of this separable functional form of the PSD as a bound on the errors.

Let the maximum frequency at which the deformable mirror (DM) is controlled be given by f_C, and let the spatial frequency exponent be $\alpha = 3$ and divide the spatial frequency into the 3 bands for low-, mid- and high-spatial frequency. Integrating over the three bands from 0 to the control frequency with a knee frequency of $f_n = 1$ cpa and $\alpha = 3$ gives the contribution from each band:

$$\begin{cases} \text{Low} \\ \text{Mid} \\ \text{High} \end{cases} \rightarrow \begin{cases} f \in [0,4] \text{ cpa} \\ f \in [4,32] \text{ cpa} \\ f \in [32,\infty] \text{ cpa} \end{cases} \rightarrow \begin{cases} \sigma_{\mathrm{LO}} \\ \sigma_{\mathrm{MID}} \\ \sigma_{\mathrm{HI}} \end{cases} = \sigma_T \sqrt{\tan^{-1}\left(\frac{f_C}{f_D}\right)} \begin{cases} 0.710992 \\ 0.338612 \\ 0.128267 \end{cases}. \tag{23}$$

If $f_C \gg f_D$, or equivalently the system is stable relative to the control, then $\tan^{-1}(f_C/f_D) \approx \pi/2$, and $\sigma_{\mathrm{LO}}^2 + \sigma_{\mathrm{MID}}^2 + \sigma_{\mathrm{HI}}^2 = \sigma_T^2$. Thus only as the ratio of the control to drift frequency exceeds unity does the control efficacy approach unity (Fig. 11).

Folding in the temporal PSD leads to a modified leakage term by multiplying the PSDs by the idealized control term to give a leakage contribution in the science focal plane of:

$$I_{\mathrm{LEAK}}(\vec{\theta}) \approx \frac{\pi}{2}\frac{1}{\tan^{-1}(f_C/f_D)}\mathrm{PSF}_L(\vec{\theta}) \\ **\{T_{\mathrm{COR}}(\vec{\theta})[\sigma_{\delta A}^2\mathrm{PSD}_{\delta A}(\vec{\theta}) \\ +\sigma_{\delta\phi}^2\mathrm{PSD}_{\delta\phi}(\vec{\theta})]\}. \quad (24)$$

Thus the leakage is a strong function of the ratio of the control to drift frequency, and for low control frequencies, $f_C \ll f_D$, then

$$\left(\frac{\pi}{2}\right)\left[\frac{1}{\tan^{-1}(f_C/f_D)}\right] \approx \frac{\pi}{2}\left(\frac{f_D}{f_C}\right)$$

implies leakage increases linearly with drift frequency, i.e., as a random walk, where for $f_C > f_D$ the upper limit is asymptotically approached.

The ratio of the leaked starlight, at the planet location $\vec{\theta}_P$, to the planet light is given by $(L_S/L_P)I_{\mathrm{LEAK}}(\vec{\theta}_P)$, where $\frac{L_S}{L_P}$ is the luminosity ratio of the star-to-planet. In general to 'see' the planet against this background requires $(L_P/[L_S I_{\mathrm{IEAK}}(\vec{\theta}_P)]) \geq 1$, i.e., the planet should appear brighter than the leaked speckle. This implies that the fractional leaked starlight is at, or lower, than the luminosity ratio. The ratio L_S/L_P is the inverse of the required operating contrast; therefore, to 'see' the planet requires:

$$I_{\mathrm{LEAK}}(\vec{\theta}_P) \leq \frac{L_P}{L_S} = \frac{1}{\langle C\rangle}. \quad (25)$$

The required operating contrast $\langle C\rangle$ is set by the science and $I_{\mathrm{LEAK}}(\theta_P)$ relates it to the instrument requirements. Thus for each star with a planet in the HZ at an angular separation of $\vec{\theta}_P$, with a given luminosity ratio, a requirement can be set on $I_{\mathrm{LEAK}}(\vec{\theta}_P)$ and thus a requirement on the rms amplitude and wavefront errors and implicitly on their PSDs.

Since the leakage is proportional to the inverse of the contrast, this implies that the contrast increases with increasing control frequency, or conversely, that the tolerances on the amplitude and wavefront errors are more stressing for control frequencies slower than the drift frequencies. Thus a telescope with faster control, relative to its drift, has more relaxed tolerances on amplitude and wavefront errors.

If we assume a cubic spatial frequency power-law PSD, $\alpha = 3$, and coronagraphic transmission at the IWA of $T(\theta) = 1$, and integrating the leakage over a region defined by a focal plane speckle to give $I_{\mathrm{LEAK}}(\vec{\theta}) \approx \sigma_{\delta\phi}^2\mathrm{PSD}_{\delta\phi}(\vec{\theta})$, and integrating the 3-D PSD out to the control frequency f_C yields

$$\mathrm{PSD}(\vec{f}) = \frac{3}{2\pi^2 f_n^2}\sqrt{\frac{3}{4}}\left[\frac{\sigma_T^2}{1+(|\vec{f}|/f_n)^3}\right].$$

Equating the two gives

$$\sigma_{\mathrm{WFE}} = \frac{\lambda}{\sqrt{\langle C\rangle}}\frac{f_n}{3^{\frac{3}{4}}}\sqrt{\frac{2}{\pi}\tan^{-1}(f_C/f_D)}[1+(|\vec{f}|/f_n)^3]^{\frac{1}{2}},$$

and

$$I_{\mathrm{LEAK}}(\vec{\theta}) \approx \frac{\sigma_T^2}{\sqrt{3\pi^2}}\tan^{-1}\left(\frac{f_C}{f_D}\right) = 10^{-10}. \quad (26)$$

Solving for $\sigma_T = 3^{\frac{1}{4}}\pi \times 10^{-5}/\sqrt{\tan^{-1}(f_C/f_D)}$ yields $\sigma_{\mathrm{WFE}} \approx 0.0088652\sqrt{(2/\pi)\tan^{-1}(f_C/f_D)}$ nm, where σ_{WFE} is the rms wavefront error at a single spatial frequency of 2 cpa but with the temporal control folded in. If the control frequency is much greater than the drift frequency then $(2/\pi)\tan^{-1}(f_C/f_D) \approx 1$ and an upper-bound of $\sigma_{\mathrm{WFE}} \approx 0.0089$ nm at 2 cpa. Note that this σ_{WFE} is the rms of a single spatial term of the PSD integrated over all temporal frequencies. Figure 11 plots the wavefront error at this single spatial frequency versus the ratio of control to drift frequency. Thus, as expected, the tolerance to achieve and hold contrast is more stressing as the control frequency falls below the drift frequency. Control at frequencies less than the drift frequency places increasingly demanding tolerances on the wavefront and amplitude errors; thus, the lower the control frequency, the lower the allowable errors, and a clear trade exists between allowable wavefront error and the ratio of control to drift frequencies. This trade has significant impact on the design of a flight coronagraphic system.

For each of the candidate stars in the HIP30 database, the expected HZ and flux ratio have been calculated (Fig. 6), and this therefore sets a requirement on the wavefront and amplitude errors at the spatial frequency defined by the HZ. A lower bound on the phase-sensing precision is inferred from the Heisenberg uncertainty relations. The product of the uncertainty in photon counts and phase is given by $\langle n\rangle\langle\phi\rangle \geq 1/2$ and is equal to ½ for all naturally occurring light,[33] and where $\langle n\rangle = \sigma_n$ and $\langle\phi\rangle = \sigma_\phi$. The lower bound on the standard deviation in sensing phase for a given photon rate is:

$$\sigma_\phi = \frac{1}{2\sqrt{n}} = \frac{1}{2\sqrt{R_n\Delta t}} \rightarrow \sigma_{\mathrm{WFE}} = \left(\frac{\lambda}{4\pi}\right)\left(\frac{1}{\sqrt{R_n\Delta t}}\right). \quad (27)$$

The number of detected photons is n, and $R_n\Delta t$ is the product of the photon count rate with integration time. This expresses a theoretical lower bound on the precision of phase sensing that in practice is usually not achievable due to sampling, detector quantization, and other noise sources. However, the theoretical precision can be used to estimate the bound on wavefront sensing precision. For example, sensing of 0.1 nm rms wavefront error requires ~200,000 photons per speckle (or per spatial frequency).

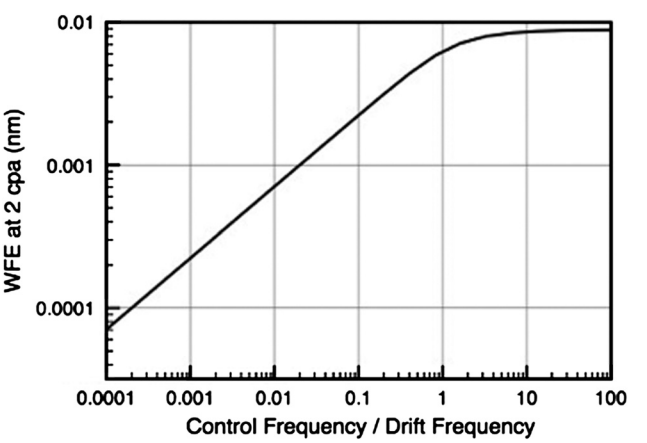

Fig. 11 WFE at 2 cpa versus control. Full-control effectiveness is not reached until the control frequency exceeds the drift frequency.

If only the photons that leak through the coronagraph are used for wavefront sensing, then the integration times are longer than if all the photons could be optimally used.

To derive a theoretical bound for coronagraphic wavefront error, we use Eqs. (21) and (28), and ignoring amplitude errors:

$$\mathrm{PSF}_L(\vec{\theta}) * *\{T_{\mathrm{COR}}(\vec{\theta})[\sigma^2_{\delta\phi}\mathrm{PSD}_{\delta\phi}(\vec{\theta})]\}|_{\vec{\theta}=\vec{\theta}_P} \leq \frac{L_P}{L_S} = \frac{1}{\langle C\rangle}. \tag{28}$$

If we assume the integral (convolution) is taken over the region defined by a single speckle, i.e. over $\mathrm{PSF}_L(\vec{\theta})$, and that the transmission of the coronagraph at the location of the planet is unity, i.e. $T_{\mathrm{COR}}(\vec{\theta}_p = 1)$, and use the phase PSD with a knee frequency of 1 cpa and exponent of 3, and solving for the wavefront precision at the IWA gives

$$\sigma_{\mathrm{WFE}} = \lambda \frac{3^{\frac{3}{4}}}{\pi\sqrt{2}} \frac{f_n}{\sqrt{\langle C\rangle}} \sqrt{\frac{2}{\pi}\tan^{-1}(f_C/f_D)}[1 + (\mathrm{IWA}/f_n)^3]^{\frac{1}{2}}, \tag{29}$$

The wavefront error requirement derived in Eq. (29) scales inversely as the square root of contrast and approximately as the cube of the angular separation due to the assumed form for the wavefront error PSD. For a knee frequency such that $f_n \ll \mathrm{IWA}$ the requirement scales as $\sim \mathrm{IWA}^{\frac{3}{2}}/\sqrt{f_n}$ and for a knee frequency such that $f_n \gg \mathrm{IWA}$ (white noise assumption) the requirement scales linearly with the knee frequency.

Equation (29) expresses the required rms wavefront error required to achieve a given contrast at a given IWA, folding in the knee frequency and ratio of control to drift frequency. The required rms wavefront error was estimated via Eq. (29) for each of the HIP30 candidates and plotted versus angular separation in Fig. 12. Figure 12 assumes a power law PSD with an exponent of 3 and a knee frequency of 1 cpa, a 4-m telescope at a wavelength of 550 nm, and assumes that the control frequency is fast with respect to the drift frequency.

The most stressing requirements are for the G stars and an angular separations of $1(\lambda/D)$. The results for coronagraphs with $\mathrm{IWA} = 2(\lambda/D)$ does not give significantly different results, yielding approximately wavefront error requirements of ~0.006 nm rms.

This requirement implicitly assumes the photons are available to sense to this level; however, there are clearly two cases: (1) all the photons from the target star, or (2) only the photons that leak through the occulter, i.e. dark photons. Folding in the time to sense to each wavefront error requirement for each of the HIP30 stars under these two assumptions results in the two extremes plotted in Fig. 13. Using only the dark photons implies only those stellar photons that are leaked by the occulting mask and Lyot stop, and using all the photons refers to using all the stellar photons collected by telescope. The time to sense the two extremes varies dramatically as shown in Fig. 13.

A third case arises, that of internal metrology. Internal metrology would generally require photon sources internal to telescope and optics to sense the relative placements, deformations, and drifts of the optics and components. In an internal metrology approach, the photons used for sensing and control are not from the stellar source but from a source internal to the spacecraft, such as one or more lasers, and hence not limited by the stellar photon rates and could result in a robust sensing and control but with the added complexity of additional optics to route the lasers to the optics and possibly separate detectors. Such an approach should be considered as part of the sensing and control trade space but is not further discussed herein.

Figure 13 plots the time to sense to each of the required wavefront errors versus angular separation for each of the HIP30 candidates, but assuming a V-band spectral filter centered on 550 nm with a 4-m aperture. A broader filter would shorten the time; however, it has not yet been shown that sensing and control through a coronagraph is feasible with broadband filters. It is seen that the times are prohibitively long using only the leaked photons at the final control step. Such an approach would likely require internal metrology, unless the photons which are reflected and/or absorbed at the occulting mask are also used (lower plot). Figure 13 assumes that effectively no drift occurs in the wavefront during the sensing time. If drift does occur, then tolerances are more demanding and it would scale as the functional form shown in Fig. 11.

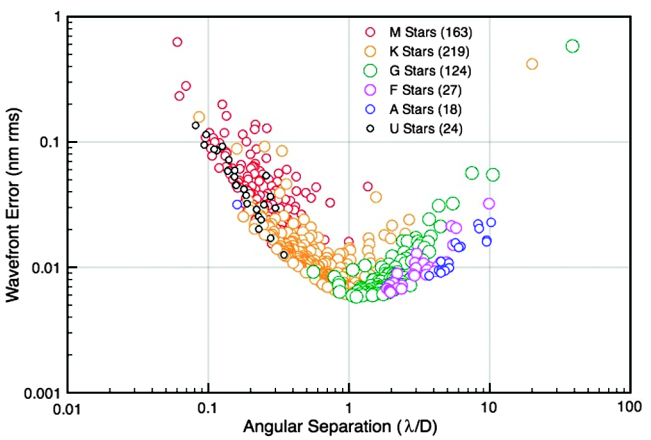

Fig. 12 Required WFE for each HIP30 candidate to yield stellar leakage, at the location of planet in the HZ, equal to the planet brightness, assuming the control frequency is fast with respect to drift.

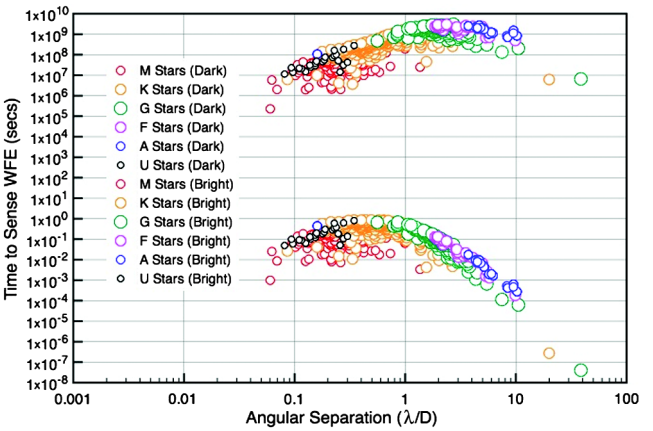

Fig. 13 Photon-limited time to sense WFE for each HIP30 candidate to yield leakage equal to planet brightness, at planet in HZ, assuming fast control with respect to the drift. (a) Time to sense assuming only photons that leak by the mask, and (b) all photons are used.

One of the underlying questions is whether enough photons are collected from a given target star to control to the contrast needed. The results shown in Fig. 13 attempt to answer that question. No reasonable mission could fly if only the leaked photons were used as the prime photon source for sensing and control. However, modulation schemes whereby the deformable mirror and/or other optics are deliberately moved to modulate the leaked photons would greatly shorten the time. A multitude of wavefront control approaches have been or are under development and many are based on variations of electric field conjugation. See, for example, Ref. 34 and the references therein. What is not clear in these approaches are the amount of photons actually collected to achieve the contrast results.

Control can be broken down into a hierarchy of regimes. In optical sensing and control, what is generally sensed are electrons from the detector, from which it is inferred that photons are counted via the quantum efficiency. Thus, photon counts are measured and a model and/or algorithmic approach subsequently used to convert photons to estimates of the amplitude and wavefront errors. However, wavefront error cannot ever be directly measured since it corresponds to phase of the optical field which is not observable[33] and only photons converted to electrons in the form of current is measured.

Figure 14 shows an example of stability and control for the proposed Extrasolar Planetary Imaging Coronagraph (EPIC) mission.[35] Figure 14(a) shows rms WFE in nm based on the rollup of a structural/thermal/optical model of the telescope and visible nulling coronagraph instrument versus time. The primary reason for growth in wavefront error is slow thermal drift due to spacecraft rolling with respect to sun angle during an observation from heliocentric orbit. The box shows the time window over which the requirement is met. Figure 14(b) shows the inverse of contrast (leakage) versus time. This is for a coronagraphic mission for Jovian planets and hence requires a less demanding 10^9 contrast than a terrestrial planet mission. Based on the EPIC model a 7000-s stability window is achieved that allows for an active control scheme whereby null control sequences, ~1200 s, are interlaced with observing sequences of approximately 7000 s and where at the outset of observing sequence the wavefront error is better than required but drifts like a random walk up to its requirement prior to the null control sequence. A terrestrial planet detection mission will have more demanding tolerances on achieving and maintaining the contrast, which will likely drive its short stability times and hence short observation windows if an active scheme is used. An adaptive scheme where continuous control occurs through the science observation may be more robust.

For many optical control systems the phase is fit to the control modes of the DM, i.e., either as modeled or measured influence functions of the DM, or as a linear superposition of these influence functions, implying a linear model. Each influence function is the change in wavefront due to a unit motion of a single control degree of freedom (DOF), and there are as many influence functions as control DOFs. Each influence function is controlled by a voltage or set of voltages on the actuators. Various approaches, such as

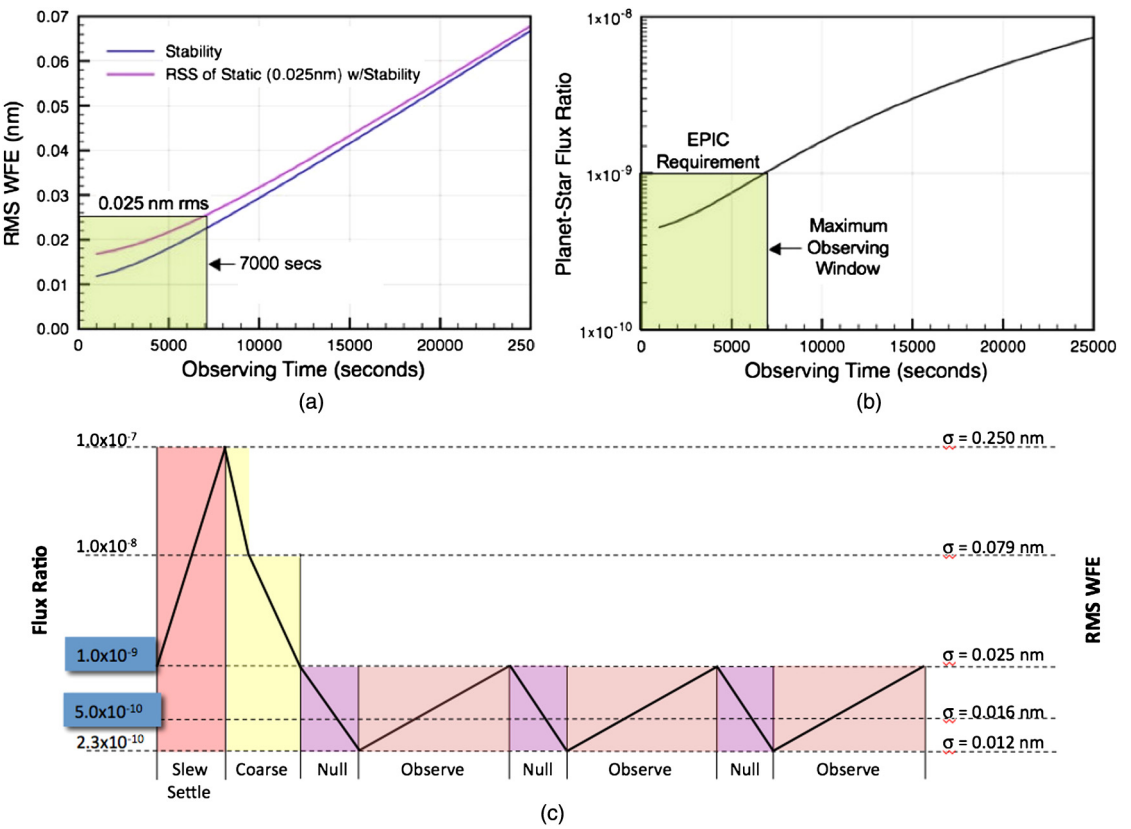

Fig. 14 Stability and control sequence for extrasolar planetary imaging coronagraph. (a) rms WFE in nm based on the rollup of a structural/thermal/optical model of the telescope and visible nulling coronagraph instrument versus time. (b) The inverse of contrast (leakage) versus time. (c) An active control scheme whereby null control sequences, ~1200 s, are interlaced with observing sequences of approximately 7000 s.

combining the influence functions and solving the eigenmodes to orthogonalize and reduce the DOFs, are available.[36] However, from a control point of view these approaches infer wavefront and/or amplitude and subsequently decompose it into control modes, and feedback to actuators in either an active or adaptive fashion. An alternative approach is to directly map what is measured, i.e. photons (or electrons), to the control voltages of the DM and remove the step of converting to wavefront and amplitude errors, an approach deemed "null control". Such an approach implies that no separation of wavefront and amplitude errors are required; the only thing that matters in a coronagraph are the leaked focal plane photons which reduce the contrast, and thus the control system works to reduce these directly by choosing combinations of actuator motions which minimize the "darkness" at a specific or range of locations in the focal plane.[21] This can also be performed in an active or adaptive fashion, where active refers to performing the control prior to each observation but not during it. If an out of specification condition occurred during the observation, then control could be restarted. Adaptive refers to control throughout the observation in a closed-loop approach. There are advantages and disadvantages of each.

Sensing and control defines a complex parameter space whereby square-law-type wavefront sensing (incoherent) versus interferometric (coherent), with using only the photons from the stellar source, or using internal metrology. If the drift rates are fast relative to the photon collection time to sense for control then there may be no choice but to add internal metrology to mitigate the starved photon problem.

Derivations and assessment of the entire sensing and control trade space is outside the scope of this work, but we attempted to address it at a conceptual level based upon expected stellar photon rates from the HIP30 catalog. The important parameter is the ratio of the drift to control rate; if the drift is slow with respect to the control, then a quasi-static condition occurs whereby control may be needed only at the outset of an observation. One possibility would be to specify the telescope requirements such that this condition is always true, but it may result in telescope that is either unbuildable or untestable, or prohibitively expensive. However, the telescope dynamic tolerances can only be relaxed to the point where the lowest photon rate science target still gives enough photons to achieve closed loop. Otherwise a degradation in performance will result. Alternatively, if only a few of the science sources drive the requirements, an alternative approach would be to remove those candidate sources deemed most stressing, alleviating the stressing requirements, or alternatively set the sensing and control separately for each star system such that it is optimal for the location of the HZ.

4 Summary and Conclusions

Herein we have mapped a candidate database of stars within 30 parsecs and deduced the flux ratio, contrast, inner working angle, and time to SNR for each star for each spectral class, assuming an Earth-sized terrestrial planet in its habitable zone, an IWA $= 2(\lambda/D)$), and diffraction dependency on the background that scales as the cube of the inner working angle. Based on these assumptions the primary conclusions are:

(1) Assessment of *all* stars within 30 parsecs would require a 120-m telescope operating at IWA $= 2(\lambda/D)$, however a 4-m telescope operating at IWA $= 2(\lambda/D)$ would allow for searching 94 stars, of which 50 are G-stars, within ~34 days per visit, or ~170 days total if each was visited five times for completeness.

(2) The set of 575 candidate stars have flux ratios which vary from 3×10^{-11} to 5×10^{-8} with most G-stars having flux ratios between 3×10^{-10} to 1×10^{-8}, implying contrast requirements at the IWA from 0.3×10^{10} to 10^{8}.

(3) The flux ratio has an approximately linear dependence versus planet-to-star angular separation with the majority of G-stars falling between 5 and 100 mas; however, the planets closer in HZ have more favorable contrast ratios. M-stars would have planets in the HZ and the smallest separation while F would be at the largest separations.

(4) The aperture flux of candidate planets in the HZ varies by only a factor of 3 over the span of angular separations and spectral classes.

(5) The V-band sensitivity scales as $\sim D^{10/3}$ and yields time to SNR $= 5$ for a planet in the HZ that varies by approximately four orders of magnitude, and for a 4-m telescope varies from ~10 to ~100,000 s with most G-stars in the range of 1000–100,000 s.

(6) The luminosity ratio scales as approximately the inverse of the angular separation independent of spectral class, implying that a clear trade/compromise exists between IWA and contrast when the available candidate stars are incorporated into the design approach.

(7) The most stressing stars, close IWA, requires wavefront tolerances of ~0.008 nm rms wavefront error for control times that are comparable to drift times; however, slow control, i.e., control times that are long relative to the drift times, requires significantly more demanding tolerances. Adaptive controls appear to be the most promising, since this appears feasible relative to the dynamics of realistic spacecraft and wavefront drifts.

(8) The wavefront error requirements are a strong function of the angular separation and are >0.008 nm rms wavefront error for the majority of G-stars.

(9) The times to sense and control is limited by the photon statistics which sets a bound on recovery of wavefront errors based on the photon count rates. Wavefront or null-control approaches based only the leaked starlight, without modulating to increase the counts, have prohibitively long sensing times. Modulation schemes, which deliberately change the wavefront errors in a known and deterministic way greatly increase the counts and hence shorten the sensing and control times. Approaches that use all the stellar photons have sensing times that are about eight orders of magnitude shorter than schemes that use only leakage photons. In practice, this can be accomplished with a visible nulling coronagraph that uses both the bright and dark output channels, or by

approaches which pick stellar photons off from the front of the occulting mask.

References

1. R. A. Brown, "Obscurational completeness," *ApJ* **607**(2), 1003–1013 (2004).
2. R. A. Brown, "Single-visit photometric and obscurational completeness," *Astrophys. J.* **624**(2), 1010–1024 (2005).
3. R. A. Brown and R. Soummer, "New completeness methods for estimating exoplanet discoveries by direct detection," *Astrophys. J.* **715**(1), 122–131 (2010).
4. Kalas et al., "Optical images of an exosolar planet 25 light-years from Earth", *Science*, **322**(5906), 1345–1348 (2008)
5. R. G. Lyon et al., "Feasibility of exoplanet coronagraphy with the Hubble Space Telescope," *Proc. SPIE* **7731**, 773111 (2010).
6. M. Clampin, "Coronagraphic detection of exosolar planets with the James Webb Space Telescope," http://www.stsci.edu/jwst/science/whitepapers/JWST_Coronagraph_WP.pdf.
7. J. Schneider, "The extrasolarplanets encyclopedia," (2011), http://exoplanet.eu/.
8. C. Turon et al., "Properties of the hipparcos input catalogue," *Astron. Astrophys.* **304**, 82–93 (1995).
9. J. T. Trauger and W. A. Traub, "A laboratory demonstration of the capability to image an Earth-like extrasolarplanet," *Nature* **446**, 771–773 (2007).
10. E. Serabyn, D. Mawet, and R. Burruss, "An image of an exoplanet separated by two diffraction beamwidths from a star," *Nature* **464**, 1018–1025 (2010).
11. D. Charbonneau et al., "An upper limit on the reflected light from the planet orbiting the star tau-Bootis," *Astrophys. J.* **522**(2), L145–L148 (1999).
12. J. N. Winn, "Transits and occultations,", in *Time*, S. Seager, Ed., 2010, 1001. Available at: http://arxiv.org/abs/1001.2010 (2011).
13. N. J. Kasdin and I. Braems, "Linear and Bayesian planet detection algorithms for the terrestrial planet finder," *Astrophys. J.* **646**(2), 1260–1274 (2006).
14. O. Absil et al., "Do we need to solve the exozodi question? If yes, how to best solve it?," *Proc. Pathways Towards Habitable Planets*, ASP Conference Series Vol. **430**, 293 (2010).
15. C. C. Stark and M. A. Kuchner, "A new algorithm for self-consistent three-dimensional modeling of collisions in dusty debris disks," *Astrophys. J.*, **707**(1), 543–553 (2008).
16. M. J. Kuchner and W. A. Traub, "A coronagraph with a band-limited mask for finding terrestrial planets," *Astrophys. J.* **570**(2), 900–908 (2002).
17. M. Perrin et al., "The structure of high Strehl ratio point-spread functions," *Astrophys. J.* **596**(1), 702–712 (2003).
18. A. Sivaramakrishnan et al., "Speckle decorrelation and dynamic range in speckle noise-limited imaging," *Astrophys. J.* **581**(1), L59–L62 (2002).
19. D. Rouan et al., "The four-quadrant phase-mask coronagraph. I. Principle," *PASP* **112**(777), 1479–1486 (2000).
20. D. Mawet et al., "The vector vortex coronagraph: laboratory results and first light at Palomar Observatory," *Astrophys. J.* **709**, 53–57 (2010).
21. R. G. Lyon et al., "Vacuum nuller testbed performance, characterization and null control: progress report," *Proc. SPIE* **8151**, 81510F (2011).
22. W. Cash, "Detection of Earth-like planets around nearby stars using a petal-shaped occulter," *Nature* **442**(7098), (2006).
23. R. J. Vanderbei, E. Cady, and N. J. Kasdin, "Optimal occulter design for finding extrasolar planets," *Astrophys. J.* **665**, 794–798 (2007).
24. R. G. Lyon et al., "Externally occulted terrestrial planet finder coronagraph: simulations and sensitivities," *Proc. SPIE* **6687**, 668719 (2007).
25. O. Guyon, "Phase-induced amplitude apodization of telescope pupils for extrasolar terrestrial planet imaging," *A&A* **404**, 379–387 (2003).
26. R. Soummer et al., "Speckle noise and dynamic range in coronagraphic images," *Astrophys. J.* **669**, 642–656 (2007).
27. J. E. Harvey, E. C. Moran, and W. P. Zmek, "Transfer function characterization of grazing incidence optical systems," *Appl. Opt.* **27**, (1988).
28. R. G. Lyon et al., "Visible nulling coronagraphy for exo-planetary detection and characterization," in *Direct Imaging of Exoplanets: Science and Technology, Proceedings of IAU Colloquium No. 200*, C. Aime and F. Vakili, Eds., 345–352 (2005).
29. S. B. Shaklan, "The Terrestrial Planet Finder Coronagraph optical surface requirements," *Proc. SPIE* **6265**, 62651I (2006).
30. J. E. Harvey et al., "Calculating BRDFs from surface PSDs for moderately rough surfaces," *Proc. SPIE* **7426**, 74260I (2009).
31. E. L. Church and P. Z. Takacs, "Light scattering from non-Gaussian surfaces," *Proc. SPIE* **2541**, 91–107 (1995).
32. P. Y. Bely, *The Design and Construction of Large Optical Telescopes*, Springer-Verlag, New York (2003).
33. R. Loudon, *The Quantum Theory of Light*, 2nd Ed., Clarendon Press, Oxford (1983).
34. A. Give'on, "The electric field conjugation: a unified formalism for wavefront correction algorithms," presented at *Adaptive Optics: Methods, Analysis and Application*, OSA, San Jose, CA (2009).
35. M. Clampin, "Extrasolar Planetary Imaging Coronagraph (EPIC)," *AAS Meeting #213, #234.06, BAAS* **41**, 506 (2009).
36. T. P. Murphy et al., "Sparse matrix approximation method for an active optical control system," *Appl. Opt.* **40**(35), 6505–6514 (2001).

Richard Lyon is an optical scientist in NASA/Goddard Space Flight Center's Laboratory for Exoplanets and Stellar Astrophysics and is involved in techniques and instrumentation for the direct detection and characterization of exosolar planets. Previously he was a scientist in NASA's Center of Excellence in Space Data and Information Sciences where he investigated maximum entropy deconvolution, symbiotic stellar jet systems, phase retrieval/diversity and imaging interferometry. He also was prior to that a member of the technical staff at Perkin-Elmer Corporation (now Goodrich) where he proposed and developed, prior to the Hubble Space Telescope launch, phase retrieval approaches as a backup method of wavefront sensing for aligning HST. He also originally proposed phase retrieval as the method of aligning the James Webb Space Telescope, then the Next Generation Space Telescope (NGST). He holds numerous awards for this work and NASA's Medal for Exceptional Service.

Mark Clampin is currently the James Webb Space Telescope (JWST) Observatory Project Scientist at GSFC. Prior to joining the JWST project, he was a member of the science staff at the Space Telescope Science Institute, where he served first as an Instrument Scientist for WFPC2, followed by STIS, and then as the Advanced Camera for Surveys Group manager from its inception to the completion of orbital verification of the instrument. He is also a co-investigator for the ACS science team and served as the team's Detector Scientist. Prior to joining STScI, he developed adaptive optics technologies for coronagraphs at the Johns Hopkins University.