

## Paths to robust exoplanet science yield margin for the Habitable Worlds Observatory

Christopher C. Stark,<sup>a,\*</sup> Bertrand Mennesson<sup>b,</sup> Steve Bryson,<sup>c</sup>  
Eric B. Ford<sup>d,e,f,g</sup> Tyler D. Robinson<sup>h,</sup> Ruslan Belikov<sup>i,c</sup> Matthew R. Bolcar,<sup>a</sup>  
Lee D. Feinberg,<sup>a</sup> Olivier Guyon<sup>i</sup> Natasha Latouf<sup>a,j</sup> Avi M. Mandell<sup>b,a</sup>  
Bernard J. Rauscher<sup>b,a</sup> Dan Sirbu<sup>c</sup> and Noah W. Tuchow<sup>b,a</sup>

<sup>a</sup>NASA Goddard Space Flight Center, Greenbelt, Maryland, United States

<sup>b</sup>California Institute of Technology, Jet Propulsion Laboratory, Pasadena, California, United States

<sup>c</sup>NASA Ames Research Center, Moffett Field, California, United States

<sup>d</sup>The Pennsylvania State University, Department of Astronomy and Astrophysics, University Park, Pennsylvania, United States

<sup>e</sup>The Pennsylvania State University, Center for Exoplanets and Habitable Worlds, University Park, Pennsylvania, United States

<sup>f</sup>The Pennsylvania State University, Institute for Computational and Data Sciences, University Park, Pennsylvania, United States

<sup>g</sup>The Pennsylvania State University, Center for Astrostatistics, University Park, Pennsylvania, United States

<sup>h</sup>University of Arizona, Lunar & Planetary Laboratory, Tucson, Arizona, United States

<sup>i</sup>University of Arizona, Steward Observatory, Tucson, Arizona, United States

<sup>j</sup>George Mason University, Fairfax, Virginia, United States

**ABSTRACT.** The Habitable Worlds Observatory (HWO) will seek to detect and characterize potentially Earth-like planets around other stars. To ensure that the mission achieves the Astro2020 Decadal's recommended goal of 25 exoEarth candidates (EECs), we must take into account the probabilistic nature of exoplanet detections and provide a "science margin" to budget for astrophysical uncertainties with a reasonable level of confidence. We explore the probabilistic distributions of yields to be expected from a blind exoEarth survey conducted by such a mission. We identify and estimate the impact of all major known sources of astrophysical uncertainty on the EEC yield. As expected,  $\eta_{\oplus}$  uncertainties dominate the uncertainty in EEC yield, but we show that sampling uncertainties inherent to a blind survey are another important source of uncertainty that should be budgeted for during mission design. We adopt the Large UV/Optical/IR Surveyor Design B (LUVOIR-B) as a baseline and modify the telescope diameter to estimate the science margin provided by a larger telescope. We then depart from the LUVOIR-B baseline design and identify six possible design changes that, when compiled, provide large gains in EEC yield and more than an order of magnitude reduction in exposure times for the highest priority targets. We conclude that a combination of telescope diameter increase and design improvements could provide robust exoplanet science margins for HWO.

© The Authors. Published by SPIE under a Creative Commons Attribution 4.0 International 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.10.3.034006](https://doi.org/10.1117/1.JATIS.10.3.034006)]

**Keywords:** telescopes; methods: numerical; planetary systems

Paper 24039G received Apr. 12, 2024; revised Aug. 8, 2024; accepted Aug. 23, 2024; published Sep. 14, 2024.

\*Address all correspondence to Christopher C. Stark, [christopher.c.stark@nasa.gov](mailto:christopher.c.stark@nasa.gov)

## 1 Introduction

A primary driving science case for the Habitable Worlds Observatory (HWO) is the high-contrast imaging of potentially Earth-like planets or exoEarth candidates (EECs). The Astro2020 Decadal Survey recommended a quantitative science goal of detecting and characterizing 25 EECs<sup>1</sup> with a ~6 m inscribed diameter (ID) telescope, roughly based on the expectation value of blind survey EEC yields from Ref. 2. However, assuming few EECs have been detected prior to launch,<sup>3</sup> HWO's blind survey detection rates will be probabilistic, with many factors affecting our chances of success. As such, the EEC yield for HWO cannot be known exactly in advance and is more accurately represented as a distribution.

There are multiple astrophysical uncertainties that will ultimately lead to uncertainties in HWO's direct imaging exoplanet yield for a blind survey. Some of these are unavoidable while others could conceivably be reduced with future observations. Previous studies have examined some of these sources of uncertainty, but have mostly treated them in isolation from one another. Reference 4 was the first to include exoplanet sampling uncertainties but did not account for some other uncertainties such as the exozodi distribution, whereas Ref. 5 looked only at the impact of median exoplanet albedo and exozodi independently. Other studies that have combined multiple sources of uncertainty have been incomplete and adopted rudimentary methods. References 6 and 7 attempted to simultaneously incorporate many sources of astrophysical uncertainty into the yield calculations, but ignored exoplanet albedo, simplified the treatment of exozodi uncertainty, and did not have a well-informed distribution of possible exoplanet occurrence rates. Here, we perform a more complete study of the impact of astrophysical uncertainties on EEC yields.

The Astro2020 Decadal Survey asserted that a sample size of 25 EECs “provides robustness against the uncertainties in the occurrence rate of Earth-sized worlds and against the vagaries associated with the particular systems near Earth.”<sup>1</sup> Here, we quantitatively assess this statement by estimating the EEC yield distribution for HWO and using this distribution to estimate our probability of achieving a given EEC yield. By adopting design choices that shift this distribution to higher yields, we show how building a “science margin” into the mission design can ensure HWO has a higher chance of achieving its goals. This same science margin, if designed properly, can also help budget against performance degradation.

We use the most recent version of the Altruistic Yield Optimizer (AYO), detailed in Ref. 8, to estimate EEC yield distributions for a blind exoEarth survey with a coronagraph-based mission in the family with HWO. In Sec. 2, we briefly review the AYO methods and present our baseline mission assumptions. We then discuss the sources of astrophysical uncertainty one by one in Sec. 5.1, show how each impacts the yield distribution, and present a final estimated yield distribution incorporating known sources of astrophysical uncertainties. In Sec. 4, we identify paths to improving EEC yields through tangible changes to our baseline mission, some of which are relatively straightforward and some of which require significant technological development. Finally, we discuss how the concept of “science margin” can also help budget for performance degradation in the mission or (relatedly) provide margin against cost growth by allowing relaxation of parameters that drive the mission cost.

The design of HWO will be informed by many metrics. Here, we focus on a single well-defined metric: the detection and characterization of EECs. The Astro2020 Decadal Report did not define “characterization,” so we make the same assumption as the LUVOIR and HabEx reports and budget for the spectral characterization time required to search each EEC for water vapor. As shown by previous studies,<sup>2,9,10</sup> a survey designed to detect and characterize EECs will, by its nature, detect many additional exoplanets. These additional exoplanets will be very diverse, spanning a broad range of phase space, and their yield will be an important metric for HWO design. However, the relative merits of diverse exoplanet yields are a topic beyond the scope of this paper.

## 2 Methods and Baseline Assumptions

We use AYO<sup>5,8,11</sup> to calculate exoplanet yields. Briefly, AYO distributes a large number ( $\sim 10^5$ ) of synthetic EECs around each star for thousands of nearby stars, sampling the range of possible orbits, phases, and planet radii consistent with the adopted definition of an EEC; calculates their

exposure times given a model of the observatory/instrument and background sources; and then numerically determines the completeness  $C$  as a function of exposure time  $t$  (and importantly, its derivative  $dC/dt$ ). AYO then uses an advanced version of the equal-slope method<sup>12</sup> to determine the optimal value of  $dC/dt$  for all observations, simultaneously optimizing the selected targets, the number of visits to each star, and the exposure and delay times for each visit. Briefly, the equal-slope method requires that  $dC/dt$  is equal for all observations, ensuring that they are equally productive per unit time, as expected for optimally distributed exposure time.<sup>5,11</sup>

We use the HWO Preliminary Input Catalog (HPIC) as our input target list.<sup>13</sup> This target list contains  $\sim 13\text{k}$  stars within 50 pc complete to a TESS magnitude of 12, formed from the union of the TESS Input Catalog<sup>14</sup> and the Gaia DR3 catalog.<sup>15</sup> The HPIC includes a wide range of stellar properties, including photometry, distance, effective temperature, spectral type, luminosity, radius, and mass. The HPIC also contains basic information on binarity from the Washington Double Star Catalog and the Gaia Catalog of Nearby Stars,<sup>16</sup> which we use to calculate a stray light background for each star in a manner identical to Ref. 2. We do not consider multi-star wavefront control techniques such that that being investigated for Roman CGI.<sup>17</sup> The HPIC has similar fidelity to the HWO Mission Stars List developed by the NASA Exoplanet Exploration Program office,<sup>18</sup> but provides the much larger sample of stars needed for accurate trade studies<sup>13</sup> (the latter provides just 160 stars, whereas we investigate scenarios that can survey as many as  $\sim 500$  stars).

Table 1 lists all high-level astrophysical assumptions we make. Unless otherwise stated, we adopt the same EEC definition as in the HabEx and LUVOIR final reports.<sup>6,7</sup> Specifically, the  $10^5$  EECs distributed around each star are placed on circular orbits within the conservative HZ spanning 0.95–1.67 AU for a solar twin, as described in Refs. 19 and 20. These planets have wavelength-independent geometric albedos of 0.2 (we address this assumption later in Sec. 3.2), have maximum radii of  $1.4 R_{\oplus}$ , and minimum radii given by  $0.8(a/EEID)^{-0.5} R_{\oplus}$ , where  $a$  is semi-major axis and EEID is the Earth-equivalent insolation distance. We adopt a baseline EEC occurrence rate  $\eta_{\oplus} = 0.24$ , consistent with the estimated occurrence rates for FGK stars integrated over our EEC boundaries,<sup>21,22</sup> and maintain constant  $\eta_{\oplus}$  independent of spectral type (we address uncertainty in  $\eta_{\oplus}$  in Sec. 3.5).

**Table 1** Baseline astrophysical parameters.

Parameter	Value	Description
$\eta_{\oplus}$	0.24 <sup>a</sup>	Fraction of Sun-like stars with an EEC
$R_p$	[0.6 <sup>b</sup> , 1.4] $R_{\oplus}$	EEC radius range
$a$	[0.95, 1.67] AU	EEC semi-major axis range <sup>c</sup>
$e$	0	Eccentricity (circular orbits)
$\cos i$	[-1, 1]	Cosine of inclination (uniform distribution)
$\Omega$	[0, $2\pi$ )	Argument of pericenter (uniform distribution)
$M$	[0, $2\pi$ )	Mean anomaly (uniform distribution)
$\Phi$	Lambertian	Phase function
$A_G$	0.2	Geometric albedo of EEC at 550 and 1000 nm
$z$	23 mag arcsec <sup>-2</sup>	Average V band surface brightness of zodiacal light <sup>d</sup>
$z'$	22 mag arcsec <sup>-2</sup>	V band surface brightness of 1 zodi of exozodiacal dust <sup>e</sup>
$n$	3.0	Median exozodi level

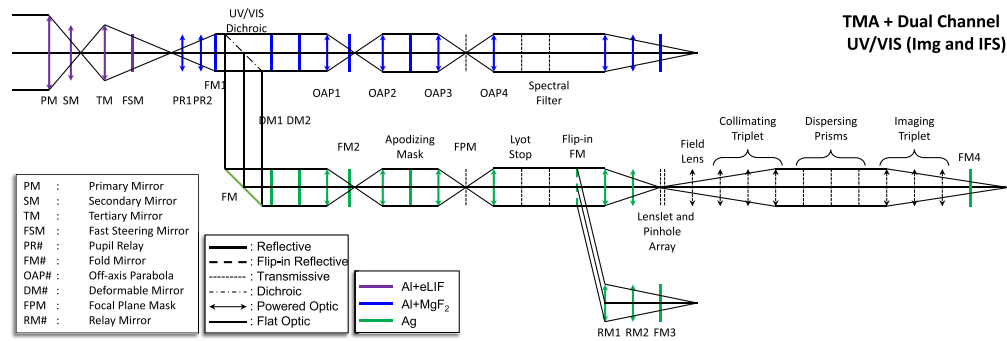
<sup>a</sup>Corresponds roughly to  $\Gamma_{\oplus} \sim 0.4$  for the adopted EEC definition.

<sup>b</sup>At the HZ outer edge. Minimum planet radius given by  $0.8(a/EEID)^{-0.5} R_{\oplus}$ .

<sup>c</sup>For a solar twin. The habitable zone is scaled by  $\sqrt{L_{\star}/L_{\odot}}$ .

<sup>d</sup>Varies with ecliptic latitude.

<sup>e</sup>For Solar twin. Varies with spectral type and planet-star separation—see Appendix C in Ref. 5.



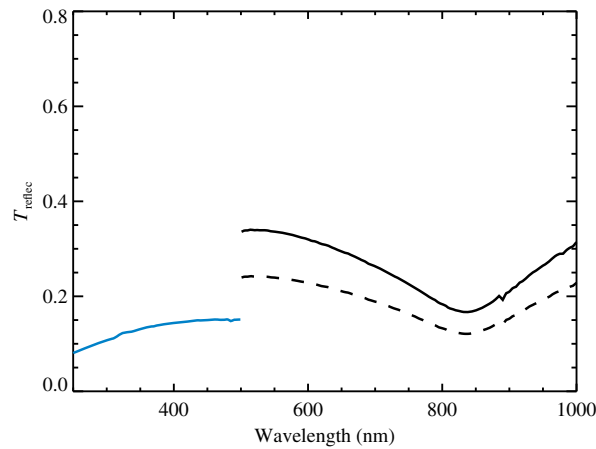
**Fig. 1** Optical layout for our LUVOIR-B baseline mission parameters. We do not explicitly show dual parallel polarization channels for each wavelength channel, which we assume for the baseline coronagraph design.

We adopt the same zodi and exozodi definitions as the LUVOIR and HabEx studies but reduce the median exozodi level from 4.5 zodi to 3.0 zodi, in line with the latest results from the Large Binocular Telescope Interferometer (LBTI) HOSTS survey.<sup>23</sup> The LUVOIR and HabEx studies adopted a distribution of exozodi values and assigned individual values to each star. Here, we assign the same 3 zodi to every star for our baseline calculations. We will show the impact of an exozodi distribution on yields in subsequent sections. The adopted exozodi model has the same color as the host star, no azimuthal dependence, no planet-induced structures, and has a flux that falls with the inverse square of the circumstellar distance, as described in Ref. 5.

For our baseline mission, we start with the same assumptions as LUVOIR-B, with the ID scaled down to 6 m (from 6.7 m). LUVOIR-B adopted an off-axis segmented primary mirror with three coronagraph channels operating in the UV, VIS, and NIR. Although all three coronagraph channels were parallelized, separated by dichroics, the LUVOIR study assumed two coronagraph channels could operate in parallel at a time. Given that the NIR channel would have a larger inner working angle (IWA), the LUVOIR study chose to operate the UV and VIS channels in parallel for detection. As such, the UV channel was designed to extend to a maximum wavelength of  $\sim 500$  nm, where there are more stellar photons. Figure 1 illustrates the end-to-end optical layout of our baseline mission. We assume dual polarization channels that can operate in parallel for both the UV and VIS wavelength channels, which we do not explicitly show in the illustration for the sake of clarity. The throughputs and reflectivities of all optics were calculated as functions of wavelength. The VIS channel was assumed to operate from 500 to 1000 nm to cover the water band short of 1000 nm. Because our yield analyses in this paper will not address wavelengths longer than 1000 nm, we largely ignore the NIR channel and leave discussion of it to future work.

Figure 2 shows the optical throughput for our baseline mission. The blue and black solid lines show the optical throughput for the UV and VIS imagers, respectively. The dashed line shows the throughput of the VIS integral field spectrograph (IFS) used for spectral characterizations. Our baseline quantum efficiency (QE) response curve is 0.9, independent of wavelength. When calculating detection exposure times, we adopt the throughput and QE evaluated at the central wavelength of the bandpass, an approximation that is reasonable for slowly varying responses such as that shown in Fig. 2. For spectral characterization exposure times, we adopt the throughput and QE at the long-wavelength edge of the bandpass, under the conservative expectation that characterizations will predominantly occur at wavelengths where the coronagraph's IWA plays an important role. One exception to this approach is in our treatment of the Skipper charge-coupled device (CCD) in Sec. 4.2.4, which exhibits a fast-varying QE response curve and warrants a bandpass-averaged approach. A more realistic handling of bandpass-varying response curves would require an understanding of how such variances affect spectral retrievals, something that has yet to be studied within the community.

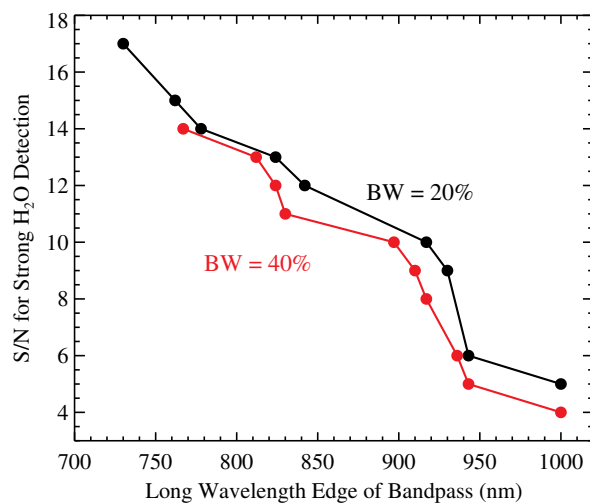
Building off of the work described in Ref. 8, we include bandpass optimization in our analyses. AYO optimizes bandpass selection on a star-by-star basis by calculating multiple possible exposure times for different bandpasses, then choosing the option that provides the maximum value of  $C/t$ . For our baseline mission assumptions, we allow for spectral characterization bandpass optimization. We budget for the detection of water vapor on all detected EECs at spectral



**Fig. 2** Wavelength-dependent optical throughput for our LUVOIR-B baseline mission. The optical throughput of the UV and VIS imagers are shown as solid blue and black lines, respectively, whereas the IFS is shown as a black dashed line.

resolution  $R = 140$  (notably larger than the  $R = 70$  assumed in the LUVOIR final report and motivated by the work of Ref. 24), allowing AYO to optimize the bandpass for each star using the  $S/N$  and wavelength options shown in Fig. 8 of Ref. 8, which we have reproduced as a black line in Fig. 3. Given these options, AYO typically chooses the  $S/N = 5$  at 1000 nm option.

We do not allow detection bandpass optimization for our baseline mission assumptions and require that all detections be performed at a wavelength of 500 nm. This limitation is motivated by the fact that the LUVOIR-B mission concept split the UV and VIS coronagraph channels at 500 nm to enable efficient parallel usage of both channels for exoplanet detections. Detection bandpasses must obviously stay within their respective channels, and the current bandpass optimization method does not currently allow for independent limits on each coronagraph channel. Ultimately, this restriction on detection bandpass will have negligible impact on the baseline mission yield, as the vast majority of targets prefer detections at 500 nm anyway.<sup>8</sup> We note that we relax this assumption later in Sec. 4.2, at which point we alter the coronagraph design from the baseline assumptions. For all broadband exoplanet detections, we require  $S/N > 7$ , a conservative estimate that provides a low mission-long probability of false positives<sup>26</sup> under



**Fig. 3** Continuum  $S/N$  required for a strong detection of water vapor on an Earth-like exoplanet as a function of the long-wavelength edge of the bandpass for 20% (black) and 40% (red) bandwidth.<sup>25</sup> A broader bandwidth covers more water lines, allowing for detection at a lower continuum  $S/N$ . We adopt 20% bandwidth for our baseline mission and examine doubling the number of visible wavelength coronagraph channels in Sec. 4.2.2.

the assumption of Gaussian noise (deviation from Gaussian noise would lead to even longer exposure times).

We adopt the Deformable Mirror-assisted Vortex Coronagraph (DMVC) used in the LUVOIR-B study. This coronagraph provides a 360 deg dark zone and is estimated to achieve an azimuthally averaged contrast  $\sim 10^{-10}$  in a single polarization, an IWA of  $\sim 3.5 \lambda/D$ , a core throughput of  $\sim 45\%$  at large separations, and a bandwidth of 20% (further details can be found in Ref. 2). As previously noted, we implicitly assume that the coronagraph design has parallel polarization channels. We note that the core throughput of the off-axis PSF is a smooth function of separation, such that exoplanets can be detected interior to the formal IWA.

We make the same detector assumptions as the LUVOIR study, namely a red-enhanced electron multiplying charge capture device (EMCCD) based on future improvements to the Roman Coronagraph's EMCCD. While the QE of the Roman Coronagraph EMCCD is just a few percent at 1000 nm,<sup>27</sup> the LUVOIR study adopted an optimistic wavelength-independent QE of 0.9. However, we note that Ref. 8 explicitly showed that if such a QE were not possible, we could search for water at shorter wavelengths at the expense of some EEC yield. We also adopted a clock-induced charge (CIC) of  $1.3 \times 10^{-3} \text{ e}^- \text{ pix}^{-1} \text{ frame}^{-1}$ , roughly an order of magnitude better than what has been demonstrated by the Roman Coronagraph EMCCD. CIC is a noise term that becomes apparent when operating an EMCCD in Geiger mode (also called photon-counting mode).

The EMCCD was paired with an IFS in the VIS channel to obtain exoplanet and debris disk spectra. An IFS requires additional optics, illustrated in Fig. 1, that reduce throughput and disperse the exoplanet's light over a large number of pixels, effectively amplifying the impact of the EMCCD detector noise. We carry forward the LUVOIR-B IFS assumptions, adopting a 30% reduction in throughput due to IFS optics and 96 pixels per PSF core at  $1 \mu\text{m}$  (16 lenslets at  $1 \mu\text{m}$  assuming Nyquist sampling at 500 nm,  $2 \times 3$  pixels per dispersed lenslet). Table 2 summarizes the high-level assumptions fed to AYO for our baseline mission parameters.

The HabEx and LUVOIR studies mandated six visits to every target to account for orbit determination of the EECs, assuming this results in  $\sim 3$  detections. Recent work has shown that two detections of a planet in reflected light may be adequate to constrain the orbit when including photometry.<sup>28</sup> As such, we drop the six-visit mandate for this study. For a coronagraph-based mission, the impact of such a mandate on yields is small,<sup>29</sup> so we expect our results to be approximately valid even when including a six-visit mandate.

### 3 Astrophysical Sources of Yield Uncertainty

Estimating yield uncertainties for a future HWO mission requires an understanding of how a given source of uncertainty will impact observations. Some uncertainties can be retired as the mission survey is conducted, which we dub as “actionable.” Others likely cannot be measured early enough in the survey to fully react, which we dub as “static.” For example, Ref. 11 showed that as long as we can measure the exozodi background of each star after the first visit, the achievable yield approaches that of having perfect prior exozodi knowledge—exozodi is therefore an actionable source of uncertainty. However, EEC albedo is likely to be more of a static source of uncertainty—with a target sample size of 25 EECs, we will not have much of an understanding of the albedo distribution until the majority of the survey has been conducted.

We address actionable and static sources of uncertainty differently in our yield calculations. To estimate actionable sources of uncertainty, we run a large number of independent yield calculations to estimate the yield distribution. For each calculation, we draw from a distribution of values describing the source of uncertainty and pass that information to AYO. AYO then optimizes the observations based on the information provided to it and returns an EEC yield. This process effectively assumes perfect prior knowledge of the parameter, an assumption that is approximately valid for actionable sources of uncertainty.<sup>11</sup> For static sources of uncertainty, we cannot pass any information about the astrophysical property to AYO. Instead, we use AYO to optimize observations under our baseline astrophysical assumptions and then vary the parameter after the observation plan has been “set in stone.”

**Table 2** Coronagraph-based mission parameters.

Parameter	Value	Description
—	—	<b>General parameters</b>
$\Sigma\tau$	2 years	Total exoplanet science time of the mission
$\tau_{\text{slew}}$	1 h	Static overhead for slew and settling time
$\tau_{\text{WFC}}$	2.7 h <sup>a</sup>	Static overhead to dig a dark hole
$\tau'_{\text{WFC}}$	1.1	Multiplicative overhead to touch up a dark hole
$X$	0.7	Photometric aperture radius in $\lambda/D_{\text{LS}}$ <sup>b</sup>
$\Omega$	$\pi(X\lambda/D_{\text{LS}})^2$ radians	Solid angle subtended by photometric aperture <sup>b</sup>
$\zeta_{\text{floor}}$	$10^{-10}$	Raw contrast floor
$\Delta\text{mag}_{\text{floor}}$	26.5	Noise floor (faintest detectable point source at $S/N_d$ )
$T_{\text{contam}}$	0.95	Effective throughput due to contamination
—	—	<b>Detection parameters</b>
$\lambda_{d,1}$	450 nm <sup>c</sup>	Central wavelength for detection in SW coronagraph
$\lambda_{d,2}$	550 nm <sup>c</sup>	Central wavelength for detection in LW coronagraph
$S/N_d$	7	$S/N$ required for detection (summed over both coronagraphs)
$T_{\text{optical},1}$	0.15 <sup>c</sup>	End-to-end reflectivity/transmissivity at $\lambda_{d,1}$
$T_{\text{optical},2}$	0.34 <sup>c</sup>	End-to-end reflectivity/transmissivity at $\lambda_{d,2}$
$\tau_{d,\text{limit}}$	2 mos	Detection time limit including overheads
—	—	<b>Characterization parameters</b>
$\lambda_c$	1000 nm <sup>c</sup>	Wavelength for characterization in LW coronagraph IFS
$S/N_c$	5 <sup>c</sup>	Signal to noise per spectral bin evaluated in continuum
$R$	140	Spectral resolving power
$T_{\text{optical,IFS}}$	0.23 <sup>c</sup>	End-to-end reflectivity/transmissivity at $\lambda_c$
$\tau_{c,\text{limit}}$	2 mos	Characterization time limit including overheads
—	—	<b>Detector parameters</b>
$n_{\text{pix},d}$	4 <sup>c</sup>	# of pixels in photometric aperture of each imager at $\lambda_{d,\#}$
$n_{\text{pix},c}$	96 <sup>c</sup>	# of pixels per spectral bin in LW coronagraph IFS at $\lambda_c$
$\xi$	$3 \times 10^{-5} \text{ e}^- \text{ pix}^{-1} \text{ s}^{-1}$	Dark current
RN	$0 \text{ e}^- \text{ pix}^{-1} \text{ read}^{-1}$	Read noise
$\tau_{\text{read}}$	N/A	Time between reads
CIC	$1.3 \times 10^{-3} \text{ e}^- \text{ pix}^{-1} \text{ frame}^{-1}$	CIC
$T_{\text{QE}}$	0.9	Raw QE of the detector at all wavelengths
$T_{\text{dQE}}$	0.75	Effective throughput due to bad pixel/cosmic ray mitigation

<sup>a</sup>See Eq. (17) from Ref. 2.<sup>b</sup> $D_{\text{LS}}$  is the diameter of Lyot stop projected onto the primary mirror.<sup>c</sup>Example provided at most likely bandpass; AYO optimizes bandpass and adjusts values accordingly.

Below, we inspect each prominent source of astrophysical uncertainty one at a time, model it as an actionable or static source of uncertainty, and discuss its impact on the yield. We compile these uncertainties as we go, building up to a final combined yield uncertainty.

### 3.1 Exoplanet Sampling Uncertainty

The most fundamental source of uncertainty for a blind survey is due to exoplanet sampling or the uncertainty that results from “luck of the draw.” Occurrence rates describe the mean rate of planets per star. Even if we knew the occurrence rates perfectly, we are not guaranteed to find the exact expectation value of planets in our population of target stars.<sup>4</sup> In addition, the chances of a planet occurring around one star versus another can affect yields.

Exoplanet sampling uncertainty is a fundamental limit of a blind survey—the only way to fully retire it is to have perfect prior knowledge of every EEC. Precursor extreme-precision radial velocity (EPRV) surveys could identify, which stars host EECs and help reduce exoplanet sampling uncertainty while improving the efficiency of HWO.<sup>3</sup> Simulations of EPRV surveys suggest exoEarths could be detectable around nearly  $\sim 100$  high-priority stars for a range of future dedicated EPRV telescope architectures.<sup>30</sup> However, these simulations represent best-case scenarios and would take at least a decade after EPRV instrument commissioning. While such precursor information will be useful when conducting HWO’s EEC survey, allowing it to achieve faster EEC yields,<sup>3</sup> it will come too late to affect the early stages of HWO design when the scale of the mission and key telescope/instrument trades will ultimately dictate the range of accessible targets (addressed in Sec. 4).

We note that a Poisson draw treatment for exoplanet sampling uncertainty is not strictly correct, as it would assume the presence of one planet in a given system does not affect the presence of another planet, which Newton would roundly reject. On the one hand, the presence of a planet should rule out nearby planets that would be gravitationally unstable, such that a Poisson draw would tend to concentrate planets around fewer stars and thus be a conservative choice. On the other hand, exoplanets may “flock together,” such that the presence of a planet implies a higher likelihood of another planet, making a Poisson draw an optimistic choice. Ideally, we would distribute planets consistent with known multiplicity rates and check for orbital stability, but empirical multiplicity rates in the HZs of FGK stars are unknown. We therefore proceed with Poisson draws and note the possibility that we will underestimate the exoplanet sampling uncertainty. We note that compared to the simple numerical alternative of a Monte Carlo success-based draw, in which a maximum of one planet per star is assigned, our method is conservative.

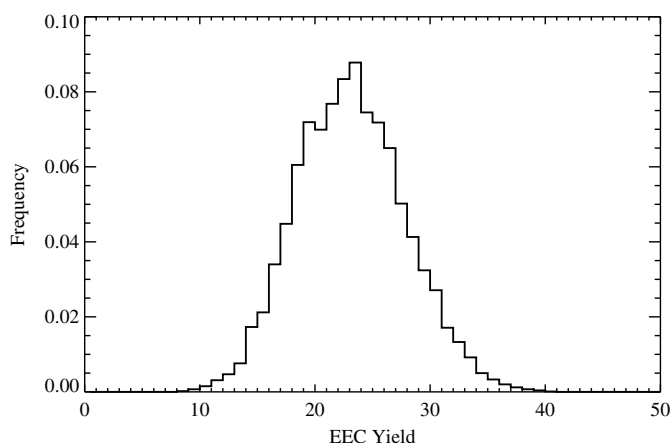
To estimate the exoplanet sampling uncertainty for an HWO blind survey, we first ran a single AYO calculation using our baseline mission and astrophysical parameters. This resulted in an expected yield of 22.5 EECs, in agreement with Ref. 2. As part of the calculation, we saved many properties of the  $10^5$  EECs injected around each star: their orbital elements, fluxes, positions, exposure times, and critically, the visit during which they were detected. With this information in hand, we then performed a Poisson draw on each individual star using an occurrence rate equal to  $\eta_{\oplus}$ . We then randomly selected the appropriate number of planets from the star’s population of  $10^5$  EECs. Randomly selected planets with a valid visit record were counted as detected, whereas those without a valid visit record were treated as undetected. We repeated this Poisson draw 10 k times, building up a distribution of yields.

Figure 4 shows the impact of sampling noise for our baseline mission. Given our assumption of Poisson draws that ignore multiplicity, our estimate for exoplanet sampling uncertainty should be considered a lower limit. With a mean-normalized standard deviation of 0.21, the spread in yields is substantial. HWO will need to contend with relatively large uncertainties inherent to a blind exoplanet survey.

### 3.2 Exoplanet Albedo Uncertainty

The yield distribution above assumes all EECs have a uniform geometric albedo  $A_G = 0.2$ , equivalent to that of an Earth-twin.<sup>31</sup> In reality, EEC albedos will vary. Unfortunately, we have no way of knowing the distribution of EEC albedos in advance. Given an expected sample size of  $\sim 25$  EECs, it is likely that we would not understand the distribution of albedos until well into the survey. We therefore treat albedo uncertainty as a static uncertainty.





**Fig. 4** EEC yield distribution of our baseline mission considering only “luck of the draw” exoplanet sampling uncertainty. Without prior knowledge of which stars host EECs, there is no way to reduce this fundamental uncertainty. This distribution assumes Poisson draws ignoring exoplanet multiplicity and should be regarded as a lower limit on the amount of uncertainty.

To constrain the effect of a static albedo uncertainty, we implement a more detailed exoplanet sampling treatment than previously described, allowing for randomly assigned albedos among the drawn EECs. To do so, we first use AYO to optimize observations under the assumption that all EECs have  $A_G = 0.2$ . We then perform a Poisson draw on each star and randomly select the appropriate number of random EEC orbits and phases, such as in Sec. 3.1. Next, we assign each randomly drawn planet an albedo that differs from the AYO-assumed  $A_G = 0.2$ . With this difference in albedo and the saved planet fluxes from AYO, we can determine the adjusted flux of every randomly selected EEC under the assumption of Lambertian phase functions. For each visit to the star, we advance the randomly drawn planet along its orbit based on the orbital properties saved by AYO, determine its visit-updated separation and albedo-adjusted flux, and determine if it would have been detected during the visit.

To determine if the randomly drawn planet would have been detected during a visit, we take advantage of the fact that AYO resolves every orbit into 100 evenly spaced mean anomalies. Each AYO observation effectively detects planets along a segment of an orbit, such that any planets detected along the orbit segment occupy a finite range of fluxes. The faintest detected planet flux along the orbit segment (usually corresponding to the crescent phase) is limited by the exposure time, whereas the brightest planet flux along the segment (usually corresponding to the gibbous phase) is constrained by the IWA of the coronagraph. Therefore, we can approximately determine whether a randomly drawn EEC with differing albedo is detectable during a given visit by requiring (1) its albedo-adjusted flux to be greater than the minimum exoplanet flux detected during that visit by AYO along the same orbit and (2) its stellar separation to be greater than the minimum separation of any EEC on the same orbit detected by AYO during that visit. To verify that this new approximate sampling treatment produced adequate results, we first randomly drew planets, all with  $A_G = 0.2$ , and compared results with the simple procedure described in Sec. 3.1. After 10 k random draws, the mean and standard deviation of the distributions were statistically identical, validating our method.

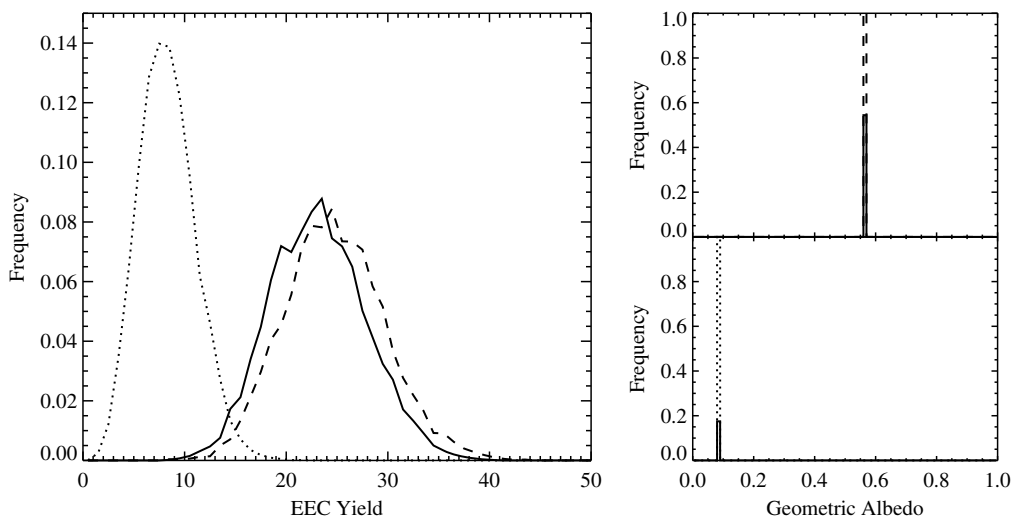
We note one significant limitation of our method for estimating the impact of albedo uncertainties. AYO “designs” the survey under the assumption that  $A_G = 0.2$ . Detection times are used to determine whether a planet of differing albedo would have been detected, but characterization times are not considered. Characterization times are budgeted for by AYO under the assumption that the planets have a single  $A_G = 0.2$ . Obviously, darker planets would require longer characterization times while brighter planets would require less. However, because the observation plan is already “set in stone” by AYO when doing the albedo draw, we cannot adjust the characterization times after the fact. To first order, we do not expect this issue to be a significant driver of yields as long as the albedo distribution is roughly symmetric about  $A_G = 0.2$  (which will be true for our final preferred albedo distribution). We do expect this limitation to overestimate the yield of faint planets in systems that are already near the 2-month characterization limit—planets

with  $A_G < 0.2$  in these systems could require characterization times in excess of the limit and should not count toward yield. However, planets with  $A_G < 0.2$  and characterization times close to the 2-month limit will represent a minority of the planets contributing to the yield (see right panel of Fig. 6). We leave refining this method to future work and note that we may be underestimating the yield degradation due to the exoplanet albedo distribution.

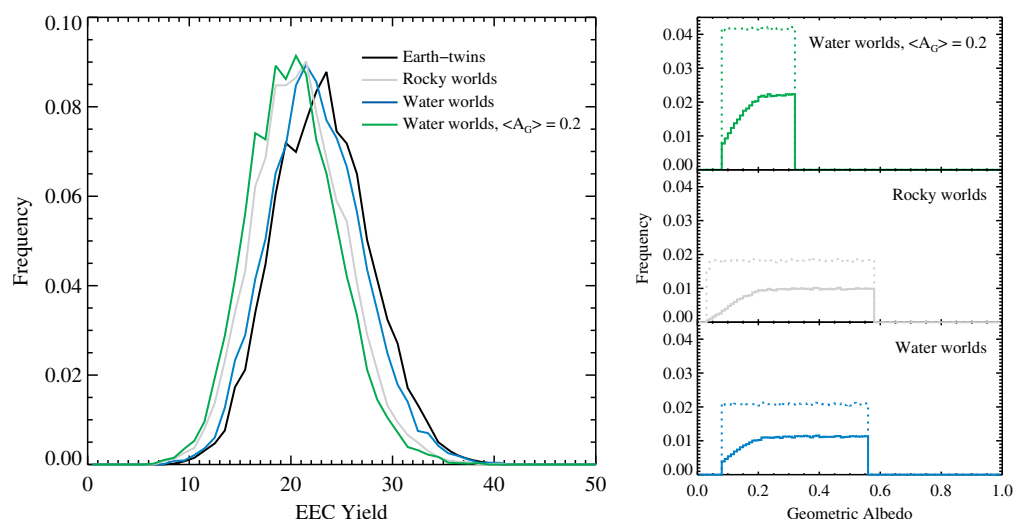
We consider two extreme scenarios to constrain the impact of exoplanet albedo: relatively dark, completely cloud-free water worlds, and relatively bright, completely cloud-covered water worlds. Both models were 100% ocean-covered. The dark extreme imagines a cloudless ocean world whose full-phase brightness would be relatively small, owing to the low reflectivity of deep ocean water seen in backscatter. At the opposite (but still habitable) extreme, a completely cloud-covered ocean is relatively bright at full phase and has a phase function that is distinct from that generated by ocean glint. These extremes bound the expected reflectance behaviors for ocean worlds that, more realistically, would present cloud-covered and cloud-free scenes across the disk. Phase-dependent reflectance models were computed using an existing 3D tool for producing disk-integrated synthetic observations of a pixelated planetary disk.<sup>31–33</sup> We assume Earth-like liquid water clouds and ocean wind speeds (which are a necessary input to the ocean specular reflectance model<sup>34</sup>).

The phase functions of these models are relatively close to Lambertian up to phase angles of  $\sim 100$  deg. Given that the majority of detections occur near phase angles of 90 deg (i.e., quadrature), we choose to treat these models as Lambertian spheres and calculate albedos that reproduce the proper reflectance at quadrature. This results in  $A_G = 0.08$  for the cloud-free model and  $A_G = 0.56$  for the cloud-covered model.

The left panel of Fig. 5 shows the yield distributions that result when observations are tuned to  $A_G = 0.2$ . The solid line shows a benchmark: the results when 100% of EECs are drawn with  $A_G = 0.2$ . The dotted and dashed lines show the results when 100% of EECs are drawn with  $A_G = 0.08$  and  $A_G = 0.56$ , respectively. While the width of the distribution does decrease in the case of  $A_G = 0.08$ , to first order, the dominant effect of changing the exoplanet albedo is that the peak of the distribution shifts. Given the limitations of our method discussed above, we note that we are likely overestimating the yield in the case  $A_G = 0.08$ . The right panel shows the albedo distribution of injected planets (dotted and dashed lines), as well as the albedo distribution of detected planets (solid lines). These distributions are effectively Dirac functions, broadened by



**Fig. 5** Left: EEC yield distribution of our baseline mission with exoplanet sampling and albedo uncertainties for the extreme scenarios in which all EECs turn out to be water-covered cloudless planets (dotted line) or cloud-covered water worlds (dashed line), compared to Earth-twins (solid line). Right: Geometric albedo distributions of injected planets (dotted and dashed) compared to detected planets (solid). The detection rate for dark water worlds is  $<20\%$ , explaining the shift in the yield distribution to lower values.



**Fig. 6** Left: EEC yield distribution of our baseline mission with exoplanet sampling and albedo uncertainties assuming a distribution of albedo values. Yields are shown for a uniform distribution of water worlds (blue), a uniform distribution of rocky worlds (gray), a uniform distribution of water worlds with  $\langle A_G \rangle = 0.2$  (green), and Earth-twins for comparison (black). Right: Geometric albedo distributions of injected planets (dotted) compared to detected planets (solid) for each of the three albedo distributions assumed. Observations are “tuned” to  $A_G = 0.2$ , resulting in lower detection rates for  $A_G < 0.2$ . We adopt the green curves as our fiducial albedo distribution.

our choice of histogram binning. The distributions show that while the detection rate for bright cloud-covered worlds is  $\sim 55\%$ , the detection rate for dark water worlds is  $< 20\%$ . This highlights the impact of an observational bias: planets brighter than expected do not substantially help yields, but planets fainter than expected can go undetected. This sort of observational bias will creep up again in subsequent sections as we consider additional sources of astrophysical uncertainty.

The dotted and dashed lines shown in Fig. 5 are extremes that set rough constraints on the impact of albedo uncertainty on exoplanet yield. In reality, we expect EEC albedos to occupy a distribution of values, but we do not yet know the nature of that distribution. Despite our ignorance, we look into the effects of adopting three different uniform distributions. First, we adopt our full range of water worlds with  $0.08 < A_G < 0.56$ . Second, we adopt an even broader range of  $0.03 < A_G < 0.58$ , roughly representing the range of reflectances one might expect near quadrature for rocky worlds as dark as Mercury and as bright as Venus. Both of these uniform distributions have  $\langle A_G \rangle > 0.2$ , so we adopt a third distribution with  $0.08 < A_G < 0.32$ , which has a mean geometric albedo equal to that of an Earth-twin. The left panel in Fig. 6 shows the results of these uniform distributions, with the blue, gray, and green curves corresponding to our full range of water worlds, full range of rocky worlds, and narrow range of water worlds, respectively. The  $A_G = 0.2$  benchmark is shown in black. The blue curve coincidentally mirrors that of the benchmark, whereas the gray and green curves are shifted to lower yield values.

The dotted lines in the panels on the right show the distribution of injected planets as a function of albedo for each of our three assumed distributions. Solid lines show the distribution of detected planets. The detection rate of planets with  $A_G > 0.2$  is relatively flat, but decreases linearly with albedo for  $A_G < 0.2$ . This explains the shift in the yield distribution: many planets at the faint end of the albedo distribution will go undetected. In all scenarios, a minority of the yield is comprised of planets with  $A_G < 0.2$ , and only a fraction of those would exceed the 2-month exposure time limit, suggesting that the limitations of our albedo draw method would have a relatively small effect on the estimated yields.

We experimented with “tuning” the observations to different exoplanet albedo. First, we ran AYO with  $A_G = 0.15$  for all EECs (tuning observations to fainter-than-Earth-twin planets), then drew the same three distributions. EEC yields were lower in all three cases, as AYO devoted more time to searching for fainter planets and ended up observing fewer stars during the 2-year time

budget. We note that the limitations of our albedo draw method should have an even smaller impact in this case, as our observations are tuned closer to the faint edge of the albedo distribution. Next, we ran AYO with  $A_G = 0.25$  for all EECs (tuning observations to brighter-than-Earth-twin planets). EEC yields were slightly larger for the black, blue, and gray curves, as AYO opted to “pick off” the brighter portions of the distribution, whereas the green curve with  $\langle A_G \rangle = 0.2$  remained approximately the same. However, we have less confidence in these results as the limitations of our albedo draw method should become more pronounced as observations are tuned to brighter planets, which should tend to overestimate yields to a greater degree. It is possible that if we are willing to assume that there are EECs with albedo greater than that of Earth, we could gain some yield at the expense of finding fewer planets fainter than the Earth. However, this assumption seems both poorly founded and risky. We conclude that observation optimization cannot significantly improve the yields of planets fainter than the Earth—we would need to improve the mission performance parameters to accomplish this.

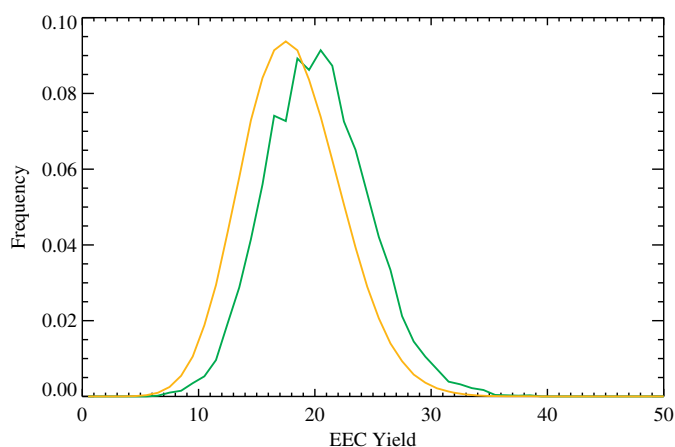
None of the yield distributions shown in Fig. 6 are correct. Completely cloud-free water worlds are probably unlikely, as are completely cloud-covered water worlds. This suggests that our uniform albedo distributions are pessimistic. However, some EECs may in fact be as dark as Mercury or as bright as Venus. The actual albedo distribution may even be multi-modal. Given a need to budget for albedo uncertainty at some level, we choose to move forward with the most pessimistic uniform albedo distribution ( $0.08 < A_G < 0.32$ ), which produces the green yield distribution shown in Fig. 6. The green yield distribution in Fig. 6 has a mean value of 19.8 EECs; budgeting for albedo uncertainty decreases the expected yields by  $\sim 12\%$ . Notably, the standard deviation of this green distribution is 22% of the mean; the albedo uncertainty did not significantly increase the fractional width of the yield distribution, i.e., exoplanet sampling dominates the uncertainty in yield. We note that the general resilience of EEC detections against albedo uncertainties does not imply that the characterization time for the EECs is also resilient against uncertainties in target spectra. Characterization time—in terms of, e.g., exposure time required to detect key atmospheric species—will depend strongly on the details of atmospheric composition, cloud distributions, and surface reflectivity.

### 3.3 Exozodi Sampling Uncertainty

So far we have combined exoplanet sampling uncertainty with albedo uncertainty, but we have assumed all stars are assigned the same amount of exozodiacal dust. In reality, each star will have a different brightness of exozodiacal dust. While we may know some individual exozodi levels in advance of the HWO survey, we will not know all of them. However, we can learn the rest of the individual exozodi levels “on the fly” and adapt to them as the survey progresses. Reference 11 showed that the EEC yield when adapting to exozodi levels after the first observation is nearly equal to the yield if they were all known in advance. Real-time adaptation to exozodi levels should lead to even higher yields. Exozodi sampling uncertainty is therefore an actionable uncertainty.

To estimate the impact of exozodi sampling uncertainty on yield distributions, we adopt the best-fit exozodi distribution from the LBTI HOSTS survey, which has a median exozodi level of three zodi and is multi-modal, with several peaks at higher zodi levels.<sup>23</sup> From this distribution, we randomly draw exozodi levels, assign them to individual stars, provide that information to AYO, and then calculate an optimized yield. We repeat this process 500 times. We include the exoplanet sampling and albedo uncertainties previously discussed by performing 1000 draws of random EECs with  $0.08 < A_G < 0.32$  for each of the 500 exozodi draws. The LBTI HOSTS survey detected dust around four potential HWO targets:  $297 \pm 56$  zodi around Eps Eri,  $148 \pm 28$  zodi around Tet Boo,  $588 \pm 121$  zodi around 72 Her, and  $235 \pm 45$  zodi around 110 Her; for yield calculations, we assigned these stars their LBTI-measured nominal exozodi levels.<sup>23</sup>

Figure 7 shows the new EEC yield distribution including exozodi sampling uncertainty in orange, along with the previous green distribution from Fig. 6. The yield distribution again maintains roughly the same width but shifts to the left, with a mean of 17.6 EECs. This is due to another observational bias analogous to the albedo distribution discussed in Sec. 3.2. If a high-priority target is assigned a zodi value less than three zodi, there is little yield to be gained, as detection exposure times were already short. However, randomly assigning a larger exozodi



**Fig. 7** EEC yield distribution of our baseline mission with exoplanet sampling, albedo, and exozodi sampling uncertainties (orange), compared with the green yield distribution calculated in Sec. 3.2. Drawing exozodi values from a distribution (as opposed to assigning all stars the same median value) shifts the yield distribution to lower values, as some high priority targets are assigned higher exozodi values.

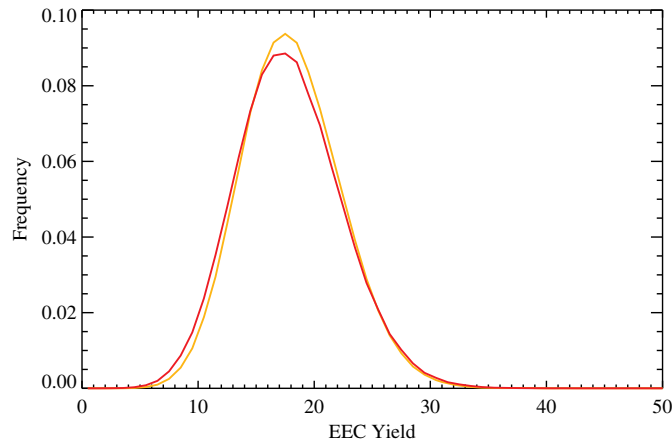
value to a high-priority star can significantly extend exposure times. Even if the yield code replaces these high-zodi stars with other low-zodi stars, the limited pool of targets means the code must replace a previously productive target with a lower-productivity target, driving the yield distribution systematically to lower values. Roughly one third of this shift can be explained by the four specific stars discussed above being assigned high zodi values—these otherwise high-priority stars are effectively scrubbed from the target list, reducing the expected EEC yield by one.

We note that the mean-normalized standard deviation of the orange curve in Fig. 7 is 0.24, not too dissimilar from the 0.22 mean-normalized standard deviation of the green curve. This shows that exozodi sampling uncertainty is a minor contribution to the total uncertainty budget, which at this point remains dominated by exoplanet sampling uncertainty.

### 3.4 Exozodi Distribution Uncertainty

Not only is the individual exozodi level of each star unknown, but our understanding of the exozodi level *distribution* is uncertain. While the LBTI HOSTS survey fit the data to derive a single maximum likelihood distribution,<sup>23</sup> many other distributions are also consistent with the data, albeit at lower likelihoods. Here, we add exozodi distribution uncertainty to the planet sampling, albedo, and exozodi sampling uncertainties already estimated.

To estimate the impact of exozodi distribution uncertainty, we must first form a set of possible exozodi distributions and calculate their likelihoods. We start by considering the maximum likelihood methods used in the LBTI HOSTS analysis, which do not formally rely on Bayesian priors (discussed later). To do this, we generate a series of 300 k exozodi values from the maximum likelihood LBTI HOSTS exozodi distribution<sup>23</sup> following the iterative approach described in Sec. 4.6.3 of Ref. 35. We then generate 30 k “perturbed” distributions that differ from the best-fit distributions. We note that in practice, the method for perturbing the maximum likelihood distribution is not well defined, but we have examined multiple methods, all producing similar results. We then calculate the likelihood  $L$  of having observed the data from each of those distributions using Eqs. (15) and (16) from Ref. 35 and compare with the maximum likelihood  $L_{\max}$ . For a given likelihood ratio, there are many possible perturbed distributions, and the perturbed distributions we generated do not follow a normal distribution. We therefore enforce  $X = (-2 \ln(L/L_{\max}))^{1/2}$  to follow a unit normal distribution by defining 31 bins ranging from  $-3 \leq X \leq 3$  and randomly drawing the correct number of distributions with the proper likelihood value. The end result is a “set” of  $\sim 20$  k distributions that follow the LBTI HOSTS maximum likelihood approach, with the statistics for the set following a normal distribution based on likelihood.



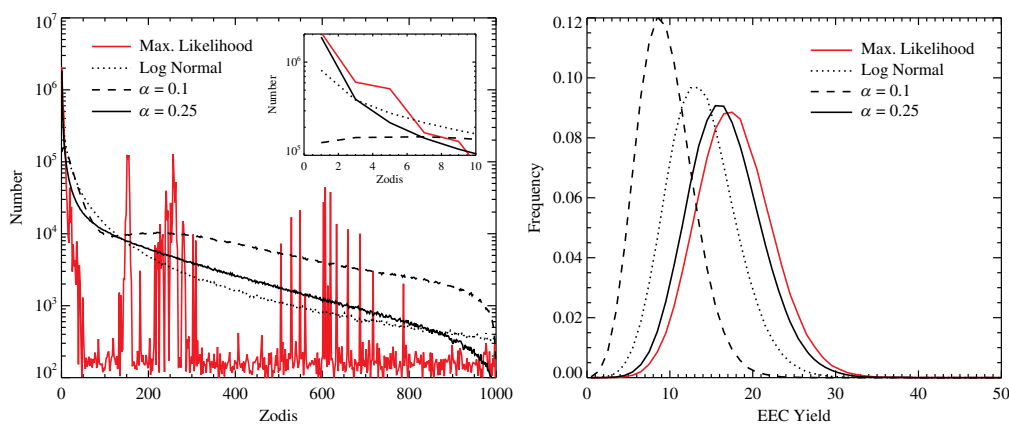
**Fig. 8** EEC yield distribution of our baseline mission with exoplanet sampling, albedo, exozodi sampling, and exozodi distribution uncertainties (red), compared with the orange yield distribution calculated in Sec. 3.3. Drawing exozodi values from all distributions consistent with the LBTI HOSTS data negligibly impacts yield uncertainty.

With these distributions in hand, we perform 498 distinct yield calculations. For each yield calculation, we draw a random exozodi distribution and then assign each star a random exozodi level drawn from that distribution. We continue to include exoplanet albedo and exoplanet sampling uncertainties following the methods previously described. Figure 8 shows the results of including the uncertainty in the exozodi distribution as a red line, compared with our previous yield distribution without it in orange. While the red curve with exozodi distribution uncertainty is slightly shifted to lower values and slightly broader (mean-normalized standard deviation of 0.26 compared with 0.24), there is remarkably little difference in the two yield distributions. This suggests that following the LBTI HOSTS formalism and assuming the LBTI HOSTS results are not systematically biased, any remaining uncertainty in the exozodiacal dust distribution has a smaller impact than the inherent exoplanet sampling uncertainty.

While it has been previously shown that EEC yield is a weak function of median exozodi level,<sup>5</sup> this is the first estimate suggesting that the uncertainty in the distribution has a negligible impact on an HWO blind survey. As such, we briefly consider alternative approaches to deriving a set of exozodi distributions consistent with the LBTI data set. To do so, we adopt a Bayesian approach to fitting the LBTI data set. We consider two possible functional forms: a log-normal distribution and a non-parametric 40th degree Bernstein polynomial basis distribution. For the non-parametric approach, we adopt two different priors described by a Dirichlet distribution with parameter  $\alpha = [0.1, 0.25]$ , where  $\alpha = 0.1$  weights the result toward a smooth fit and  $\alpha = 0.25$  allows for a higher degree of modality in the distribution (i.e., multi-peaked).

The left panel of Fig. 9 shows the result of 10k exozodi draws from 500 randomly drawn distributions for all four of our approaches. The red curve shows the LBTI HOSTS maximum likelihood approach, whereas the black curves show the log-normal and non-parametric approaches. The right panel of Fig. 9 shows the yield distribution for each of these approaches, calculated in a similar fashion as previously described. The log-normal approach, which does not accommodate multi-modality, and the  $\alpha = 0.1$  prior approach, which allows only for modest multi-modality, predict higher exozodi levels on average and significantly shift the yield curve to lower numbers. The  $\alpha = 0.25$  prior approach, which does accommodate some degree of multi-modality, produces results that are very similar to the LBTI HOSTS maximum likelihood approach. We note that this is despite the median exozodi level for the  $\alpha = 0.25$  prior approach being twice that of the maximum likelihood approach—the reason for this is that while the approaches have different medians, both have similar fractions of the distribution  $\lesssim 3$  zodis, as shown by the inset panel in Fig. 9.

We caution that the choice of priors appears to significantly affect the implied exozodi distribution and ultimately the yield distribution, and the yield distribution is most sensitive to the fraction of the exozodi distribution at very low exozodi levels. Determining which approach is best is difficult without more/better data and is therefore beyond the scope of this paper.



**Fig. 9** Left: 10 k randomly drawn exozodi levels from 500 randomly drawn distributions when fitting the LBTI HOSTS data with maximum likelihood (solid red), log-normal (dotted black), non-parametric with  $\alpha = 0.1$  (dashed black), and non-parametric with  $\alpha = 0.25$  (solid black) approaches. Right: Yield distributions including exozodi distribution uncertainty for each of these approaches; the red curve is identical to the red curve in Fig. 8. Adopting a multi-modal non-parametric fit ( $\alpha = 0.25$ ) produces similar yields to the LBTI HOSTS maximum likelihood approach<sup>23</sup> because the fraction of distributions  $\lesssim 3$  zodis is similar.

However, given that the LBTI data set appears to clearly be multi-modal<sup>23</sup> and that the  $\alpha = 0.25$  and maximum likelihood approaches largely agree, we proceed with the LBTI HOSTS maximum likelihood approach to estimate exozodi distribution uncertainty, i.e., the solid red curve shown in Fig. 8.

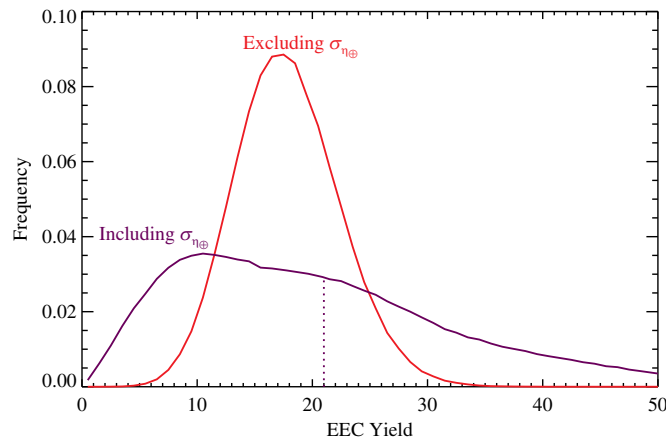
We note that all of the yield distribution curves shown in Figs. 8 and 9 come with several major caveats. For example, our calculations explicitly assume that we can subtract exozodi to the Poisson noise limit without impacting the planet’s signal. While this has been shown to be true for inclined, smooth exozodi less dense than a few tens of zodis,<sup>36</sup> it may be difficult for smooth edge-on disks<sup>36</sup> as well as edge-on disks with structure.<sup>37,38</sup> In addition, the presence of hot dust near the star<sup>39,40</sup> or cold pseudo-zodi at small projected distances in edge-on disks<sup>41</sup> may cause contrast degradation and make PSF subtraction difficult. We leave these issues for future investigations.

Of all the sources of astrophysical uncertainty we have considered thus far, the exoplanet sampling uncertainty inherent to a blind survey appears to be the dominant term. Our estimate of the impact of this source of uncertainty, which should be regarded as a lower limit, resulted in a mean-normalized standard deviation of 0.21, whereas all other terms combined only increased the mean-normalized standard deviation to 0.26. Assuming uncertainties add in quadrature, this suggests all other terms combined result in a mean-normalized standard deviation of 0.15. The dominant effect of these other sources of uncertainty has been to reduce the expectation value of the yield by nearly 25%, from 22.5 EECs to 17.3 EECs due to observational biases. Assuming we do not find the majority of EECs or measure the majority of target star’s exozodi levels in advance of the HWO mission, most of these uncertainties cannot be mitigated prior to the mission and thus must be budgeted for with yield margin. Next, we consider the one remaining source of astrophysical uncertainty, which will dominate over all other terms, but also has the potential to be somewhat mitigated prior to launch.

### 3.5 Uncertainty in $\eta_{\oplus}$

The final source of astrophysical uncertainty we consider is the occurrence rate of EECs,  $\eta_{\oplus}$ . While  $\eta_{\oplus}$  uncertainty may seem like a static source of uncertainty, for a fixed mission lifetime we must budget spectral characterization time for each detected EEC. For example, in the event we detect more planets than expected, we must devote time to characterize them, reducing the time we can spend searching for EECs. We therefore treat  $\eta_{\oplus}$  as an actionable source of uncertainty.

To incorporate the uncertainty in  $\eta_{\oplus}$ , which we refer to as  $\sigma_{\eta_{\oplus}}$ , we use the  $\eta_{\oplus}$  values from Ref. 22. Reference 22 calculated occurrence rates using the Kepler DR25 exoplanet data



**Fig. 10** EEC yield distribution of our baseline mission with all known sources of astrophysical uncertainty: exoplanet sampling, albedo, exozodi sampling, exozodi distribution, and  $\eta_{\oplus}$  uncertainties (purple). The red yield distribution is the same as was calculated in Sec. 3.4, which excludes uncertainty in  $\eta_{\oplus}$ . The dotted purple line marks the mean of the purple distribution. Uncertainties in  $\eta_{\oplus}$  substantially broaden the EEC yield distribution.

catalog<sup>42</sup> supplemented by Gaia-based stellar and exoplanet properties<sup>43,44</sup> and corrected for catalog completeness and reliability. Reference 22 used a power-law population model with a rate that depended on exoplanet radius, exoplanet insolation flux, and host star effective temperature, with the power law parameters inferred using a Poisson likelihood. To compute  $\eta_{\oplus}$ , we integrate the power law model using the posterior power law parameter values from their analysis of planets in the conservative habitable zone of quiet, isolated FGK main sequence dwarfs. Our domain of integration, defining our rocky habitable zone population, is

- The exoplanet radius range  $0.8R_{\oplus}\sqrt{EEID/a} < R < 1.4R_{\oplus}$ , where  $R$  is radius,  $a$  is the semi-major axis, and EEID is the Earth-Equivalent Insolation Distance (the distance at which the planet would have the same insolation as Earth), chosen to be consistent with the radius range adopted for yield calculations herein
- The exoplanet insolation flux range defined for each stellar effective temperature by the conservative habitable zone<sup>19</sup>
- The stellar effective temperature range  $3900 \text{ K} < T_{\text{eff}} < 7300 \text{ K}$ .

Reference 22 accounted for the lack of information about DR25 catalog completeness beyond 500-day orbital periods by computing power law posteriors for two bounding cases: case 1 assumed completeness was zero beyond 500 days, and case 2 assumed the completeness beyond 500 days was equal to the completeness at 500 days. We computed the  $\eta_{\oplus}$  distribution for each of these cases and created our final  $\eta_{\oplus}$  distribution by uniformly randomly drawing from both cases.

Performing this computation for each element of our posterior, we get an  $\eta_{\oplus}$  with mean and 86% confidence interval  $\eta_{\oplus} = 0.26^{+0.29}_{-0.14}$ . The large uncertainties are primarily due to the very small number of detections of habitable zone planets orbiting FGK stars whose planet radius is within our desired range.

To include  $\sigma_{\eta_{\oplus}}$  in our yield calculations, we repeat the 498 calculations performed in Sec. 3.4, but draw a unique value of  $\eta_{\oplus}$  for each one using the methods described above. Figure 10 shows the results in purple compared to our previous results from Sec. 10 shown in red. While the median of the purple distribution is similar to that of the red distribution, the mean (shown as a vertical purple dotted line) is substantially higher due to a tail of large  $\eta_{\oplus}$  values, including uncertainty in  $\eta_{\oplus}$  leads to higher mean yields. Our uncertainty in  $\eta_{\oplus}$  has a major impact on the breadth of the expected yield distribution. In short, even for missions with an expectation value close to two dozen, the uncertainties in  $\eta_{\oplus}$  are large enough to produce non-negligible chances of single-digit EEC yields.



We note that the mean of the red distribution shown in Fig. 10 is  $\sim 20\%$  lower than the yield predicted for a 6 m ID telescope by Ref. 2. Although we made several changes to assumed inputs, notably  $R = 140$  instead of  $R = 70$  to detect  $\text{H}_2\text{O}$  and an updated LBTI best-fit exozodi distribution, these changes mostly offset. The majority of the decrease in expected yields is due to the geometric albedo distribution (which decreases yield by  $\sim 15\%$ ) and exozodi distribution uncertainty (which decreases yield by  $\sim 4\%$ ), which were not included in Ref. 2.

In summary, assuming plausible distributions in exoplanet albedo, uncertainty in the HWO EEC yield appears to be dominated by two sources of astrophysical uncertainty. The first is simply exoplanet sampling, which is inherent to a blind survey and may only be partially overcome by precursor detection of the EECs. To be maximally useful, this must happen prior to the design of the mission. The second and most significant source of astrophysical uncertainty is  $\eta_{\oplus}$ . We note that a goal of 25 EECs ignoring uncertainties in  $\eta_{\oplus}$  is therefore not equivalent to a goal of 100 cumulative HZs. A goal of 100 cumulative HZs implicitly ignores *both* dominant sources of uncertainty,  $\sigma_{\eta_{\oplus}}$  and exoplanet sampling uncertainties.

The Astro2020 Decadal Survey asserted that a sample size of 25 EECs “provides robustness against the uncertainties in the occurrence rate of Earth-sized worlds and against the vagaries associated with the particular systems near Earth.”<sup>1</sup> Here, robustness can be defined as the probability of achieving a given yield goal. With some minimum yield goal defined, we can use the distributions shown in Fig. 10 to calculate this probability. As the Astro2020 Decadal Report did not define the minimum acceptable yield goal, the meaning of spectral characterization, nor the vagaries of particular systems, we must define them here. We therefore adopt two minimum yield goals throughout the rest of this study: the detection of 25 EECs and subsequent search for water vapor including all sources of astrophysical uncertainty, and the detection of 25 EECs and subsequent search for water vapor including all sources of astrophysical uncertainty except  $\eta_{\oplus}$  uncertainties. These two goals correspond to the two yield distribution curves shown in Fig. 10.

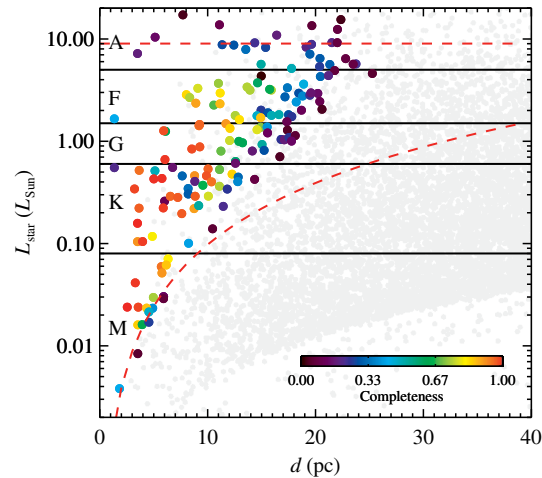
We define the probability of detecting and searching 25 EECs for water vapor,  $P_{25}$ , as the fraction of the yield distribution that exceeds 25 EECs. We calculate this quantity for each of the distributions shown in Fig. 10. For our baseline mission parameters with a 6 m inscribed diameter, we find  $P_{25}$  is just 6% when ignoring  $\eta_{\oplus}$  uncertainties and 32% when including  $\eta_{\oplus}$  uncertainties. In the following section, we investigate modifications to the LUVOIR-B design that could improve scientific performance.

## 4 Paths to Budget for Yield Uncertainty

Here, we explore paths to shift the final two distributions shown in Fig. 10 to larger values and increase the confidence in achieving the goal of 25 EECs. We will explore multiple possible improvements to our baseline mission assumptions. For each, we will describe the impact of the change on the fundamental mission parameters, EEC yield, and data quality.

To understand how to improve yields, we must first understand the astrophysical performance of our baseline mission. Figure 11 shows the targets selected for observation by AYO for a single representative simulation of our baseline LUVOIR-B scenario. Targets are color-coded by HZ completeness, and the full input target list is shown in gray. Black horizontal lines mark the luminosity boundaries for different stellar types. The red dashed lines roughly mark the boundaries of accessible targets. EEC contrast becomes more challenging for early-type stars, reducing HZ completeness for targets at higher luminosity. The upper horizontal line marks where a  $1.4 R_{\oplus}$  planet at the EEID has a contrast equal to the systematic noise floor, a rough visual guide marking a limit imposed on target accessibility by the noise floor.

The curved red line indicates the luminosity at which the outer edge of the HZ is located at  $1.5 \lambda/D$  when  $\lambda = 1000$  nm. Because the angular scale of the HZ decreases with distance and for late-type stars observing these targets requires operation at smaller working angles where coronagraph throughput is lower and contrast is degraded. The lower curved dashed line therefore represents another visual guide marking limits imposed by operating at small working angles. Of course, operating near  $1.5 \lambda/D$  means the exoplanet is only marginally resolved and would likely blend together with other planets in the scene.<sup>45</sup> We expect that spatial resolution limitations will ultimately place a strict limit on the working angle that we do not enforce here. Yield estimates would benefit from future studies firmly establishing this working angle limit.



**Fig. 11** HZ completeness of selected targets for one representative run of the LUVOIR-B baseline scenario. Star-to-star variation in completeness is due to the random assignment of exozodi levels. Red dashed lines roughly mark the boundaries of the observable targets, whereas gray dots indicate the range of the input target list.

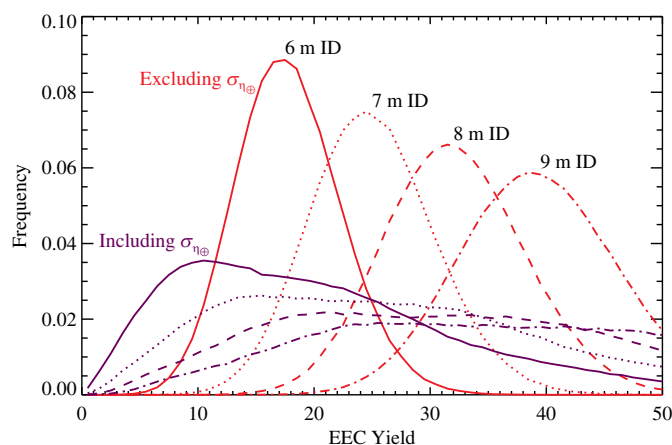
Based on the red dashed boundaries and the selected targets shown in Fig. 11, it is clear that the baseline mission has many accessible targets that go unobserved or are under-observed. If additional mission time were available or exposure times were shortened, the additional observations would increase completeness toward the upper right corner of the plot. Throughout this section, we will therefore highlight the importance of achieving shorter exposure times.

In addition to improved yields, there are other important motivators for decreasing exposure times. The AYO calculations here adopt the same exposure time limit as the HabEx and LUVOIR studies: 2 months. While exposure times are budgeted for properly in the yield code, exposure times this long are problematic. First, exposure times lasting several weeks can make additional spectral characterization, beyond just the H<sub>2</sub>O detection that we include, unlikely, and a single 20% bandpass to search for water vapor may not constitute adequate “characterization” of EECs for HWO. Further, exposure times on the order of a month will be complicated by the motion of the planet, which may disappear behind the IWA or move into the crescent phase. Finally, long exposure times present real-world scheduling constraints that may be difficult to overcome and optimize. The shorter we can make exposure times, the easier the observations become and the more spectral information we can obtain on each EEC.

We note that extending the mission lifetime is fundamentally different from shortening exposure times. The former provides more time to observe any target that currently meets the 2-month limit—this would increase the yield, but would not extend the range of accessible targets. The latter changes the exposure time of every star, extending the range of accessible targets such that more targets are compliant with the 2-month limit. The targets selected for observation (colored dots) in the upper right corner of Fig. 11 have exposure times approaching the 2-month limit. An extended mission lifetime would increase the completeness of targets already selected for observation (colored dots), whereas a reduction in exposure times would allow more of the unobserved targets (gray dots to the upper-right) to be observed.

#### 4.1 Build a Bigger Telescope

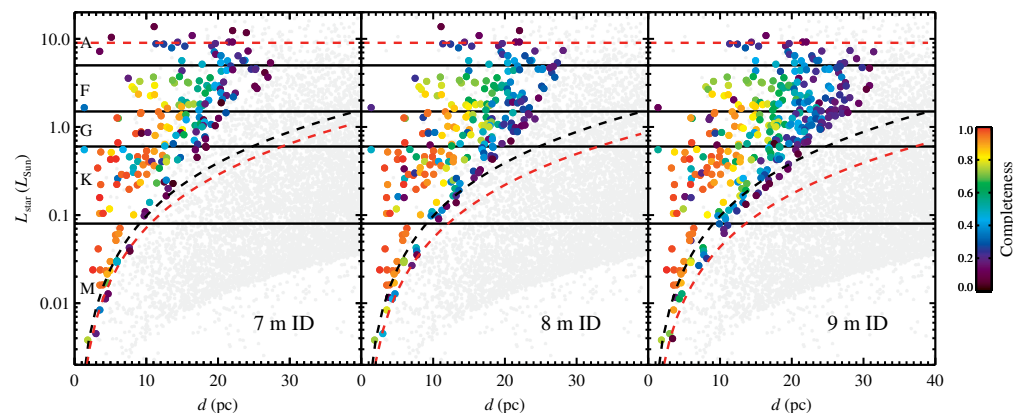
As shown by Ref. 5, EEC yield is most sensitive to telescope diameter. Therefore, we study how the distributions shown in Fig. 10 vary with telescope diameter. To do so, we repeat the calculations from Secs. 3.4 and 3.5 for inscribed diameters, IDs, ranging from 6 to 9 m. Figure 12 shows the resulting EEC yield distributions, with solid, dashed, dotted, and dash-dotted lines corresponding to 6, 7, 8, and 9 m IDs, respectively. Purple distributions correspond to those with all known astrophysical uncertainties included and red lines correspond to those without  $\sigma_{\eta_{\oplus}}$  included. Excluding  $\sigma_{\eta_{\oplus}}$ , a 9 m ID telescope increases yield by a factor of 2.2, in agreement with power law relationships established by previous works.<sup>5,11</sup>



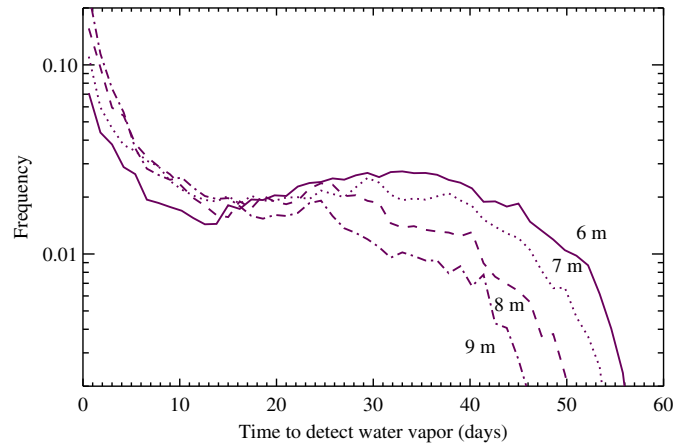
**Fig. 12** EEC yield distributions for four different telescope diameters when adopting our baseline mission parameters including  $\eta_{\oplus}$  uncertainty (purple) and excluding it (red). Solid, dotted, dashed, and dot-dashed lines correspond to IDs of 6, 7, 8, and 9 m, respectively.

Figure 13 shows the targets selected for observation as a function of telescope diameter. These plots assume the same single representative exozodi draw as in Fig. 11. The black dashed line roughly marks the working angle limit of a 6 m ID telescope. As the telescope diameter increases, the working angle limit shifts downward, as shown by the curved red line. Larger telescope diameters allow targets with smaller angular HZs to be selected, whether they are at larger distances or lower luminosities. Notably, none of the scenarios in Fig. 13 “use up” all of the potentially accessible targets. Regardless of aperture size, the LUVOIR-B baseline parameters therefore lead to missions that are limited by exposure times. Only by reducing exposure times can a mission take advantage of the full extent of the target list, a concept we explore in Sec. 4.2.

Figure 14 shows the distribution of possible spectral characterization times to detect water vapor absorption for all 498 yield calculations performed. We only show the scenario in which  $\eta_{\oplus}$  uncertainty is excluded. Of course, as we shorten exposure times by going to larger telescopes, more observations can be included that will necessarily be around more challenging targets and extend the distributions, making it difficult to see the exposure time impacts for the highest priority stars. Therefore, to make the distributions shown in Fig. 14, we consider only the first 18 EECs of any simulation.



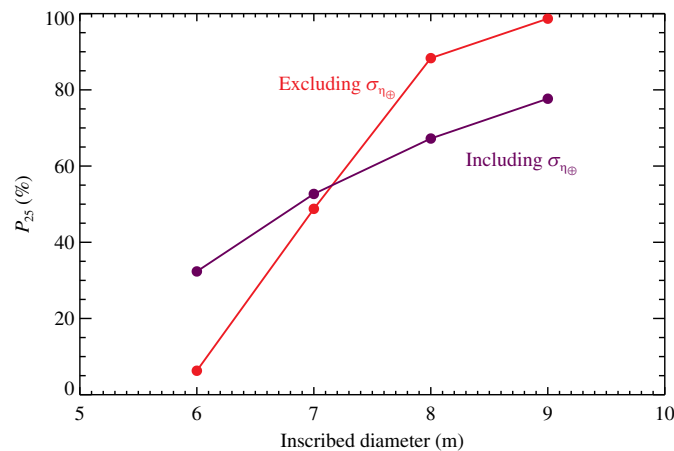
**Fig. 13** Targets selected for observation for one simulation of the 7, 8, and 9 m ID telescope scenarios, color-coded by total completeness, assuming the same single representative exozodi draw as in Fig. 11. The red dashed lines indicate the boundaries of the accessible targets—the horizontal line roughly marks the assumed astrophysical noise floor while the curved line roughly indicates the working angle limit. The black dashed line indicates the working angle limit of a 6 m ID telescope for reference. Larger telescopes expand the range of accessible targets.



**Fig. 14** Possible spectral characterization time distributions of the first 18 EECs for our baseline mission parameters with a 6, 7, 8, and 9 m ID telescope shown in solid, dotted, dashed, and dash-dotted lines, respectively. A 9 m ID telescope reduces spectral characterization times by a factor of 6.

Not surprisingly, the spectral characterization time distribution of the telescope with a 9 m ID is more sharply peaked toward shorter exposure times. In the non-background-limited regime, exposure times should scale as  $D^{-2}$ , whereas in the background-limited regime they should scale as  $D^{-4}$ . If all else were equal, we would therefore expect exposure times for our highest priority targets to be reduced by a factor of 2.3 to 5.1 when increasing the telescope diameter from 6 to 9 m. However, all else being equal, larger telescopes also provide higher coronagraphic throughput at a fixed angular separation for the DMVC, as well as an expanded target list to choose from, further reducing exposure times. We find the mean exposure time decreases by a factor of 6.2 going from 6 to 9 m ID; the mean spectral characterization time of the first 18 EECs is 22 days for a 6 m ID telescope but can be shortened to just 3.5 days for a 9 m ID telescope.

We calculate  $P_{25}$  for each of the distributions shown in Fig. 12 and plot it as a function of telescope diameter in Fig. 15. For our baseline mission parameters, an  $\sim 8.2$  m ID telescope can achieve  $P_{25} > 90\%$  when ignoring uncertainties in  $\eta_{\oplus}$ , but even a 9 m ID telescope cannot achieve  $P_{25} \gtrsim 80\%$  when budgeting for  $\eta_{\oplus}$  uncertainties. Unless we consider diameters significantly larger than 6 m, telescope diameter alone cannot build in robust science margin for uncertainty in  $\eta_{\oplus}$  under our baseline mission parameters. In the following section, we investigate modifications to the LUVOIR-B design that could improve scientific performance without increasing the primary mirror diameter.



**Fig. 15** Fraction of yield distribution  $>25$  EECs,  $P_{25}$ , when budgeting for detection and water vapor characterization, as a function of telescope diameter for our baseline mission parameters. The purple line includes  $\eta_{\oplus}$  uncertainty and the red line excludes it.

## 4.2 Improve the Mission Design

There are a number of changes to the LUVOIR-B baseline design that could improve the quantity and quality of data. Some of these are relatively straightforward changes, whereas others require the development of new technologies. Here, we highlight six possible improvements, noting that many others likely exist. We examine each design change one at a time, starting from the simplest and adding them up as we go. All of these changes will focus on reductions in exposure time. While previous works have shown that improvements to parameters controlling exposure time have only a modest impact on the EEC yield,<sup>5,11</sup> these impacts will ultimately compile, resulting in significant changes to the expected EEC yield. The end result demonstrates that some design changes are truly synergistic.

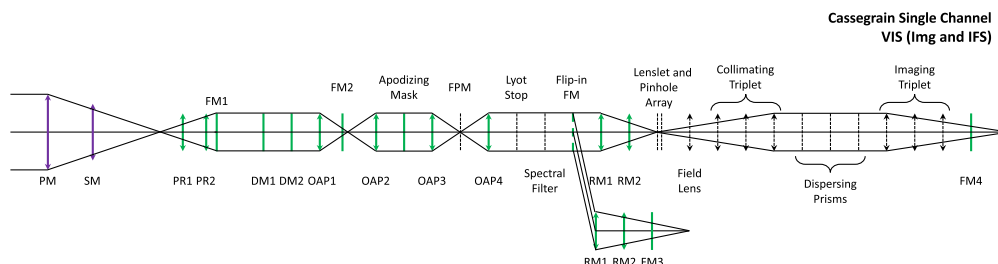
### 4.2.1 Scenario A: minimize aluminum reflections

The LUVOIR-B design adopted a three-mirror anastigmat optical telescope assembly (OTA) with a fourth fast-steering mirror. Three additional pre-coronagraph optics were necessary prior to the UV-VIS channel dichroic. All seven of these mirrors were aluminum coated. Typical reflectivities for protected aluminum-coated mirrors are  $\sim 90\%$ ,  $87\%$ , and  $92\%$  at 500, 760, and 1000 nm, respectively. Silver-coated mirrors have reflectivities of  $\sim 98\%$ ,  $97\%$ , and  $96\%$  at the same three wavelengths. Just five aluminum-coated mirrors will reduce throughput at 760 nm, a key wavelength for detection of molecular oxygen (a biosignature gas), by almost a factor of two compared to silver. As every exoplanet photon is precious, we should strive to reduce the number of aluminum-coated mirrors. We note that Ref. 2 estimated the impact of some of these changes already; here, we break these choices down in detail.

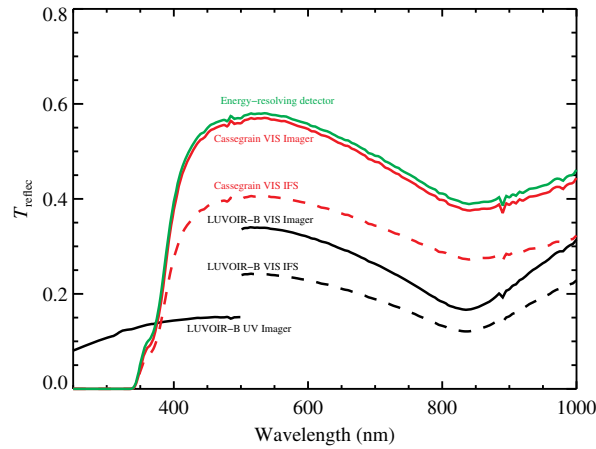
Our first potential design change is to adopt a Cassegrain telescope with only two aluminum-coated telescope mirrors. These two mirrors preserve UV science for any other instrument in the observatory, including a UV coronagraph. While it may still be possible to operate a UV coronagraph in parallel with a VIS coronagraph under this assumption, here we make the conservative assumption that the UV coronagraph cannot be parallelized, simply to illustrate a point: in terms of EEC yield, the increased throughput will more than make up for the lack of a parallelized UV channel. Figure 16 shows the optical layout adopted for this design change and includes both IFS and broadband imaging modes for the single VIS channel. A possible UV channel is not shown. We assume dual polarization operation, but do not show the potential split needed for separate polarization channels; a detailed study of the need for separate polarization channels is beyond the scope of this paper.

Without a static dichroic present to split the parallel UV and VIS wavelengths, we are now able to implement the detection bandpass optimization feature within AYO. As shown by Ref. 8, this usually results in V band detections for the majority of stars but can select longer wavelengths for nearby and late-type stars. We carry this feature forward through the rest of the analyses in this study.

Figure 17 shows the end-to-end optical throughput (not including the coronagraph's core throughput) of this Cassegrain VIS channel (red) compared with the LUVOIR-B baseline (black). At 500 nm, where most detections will be performed, the optical throughput is now



**Fig. 16** Optical layout for a Cassegrain design with a single VIS channel. We do not explicitly show dual parallel polarization channels, which we assume for the baseline coronagraph design. See Fig. 1 for legend.

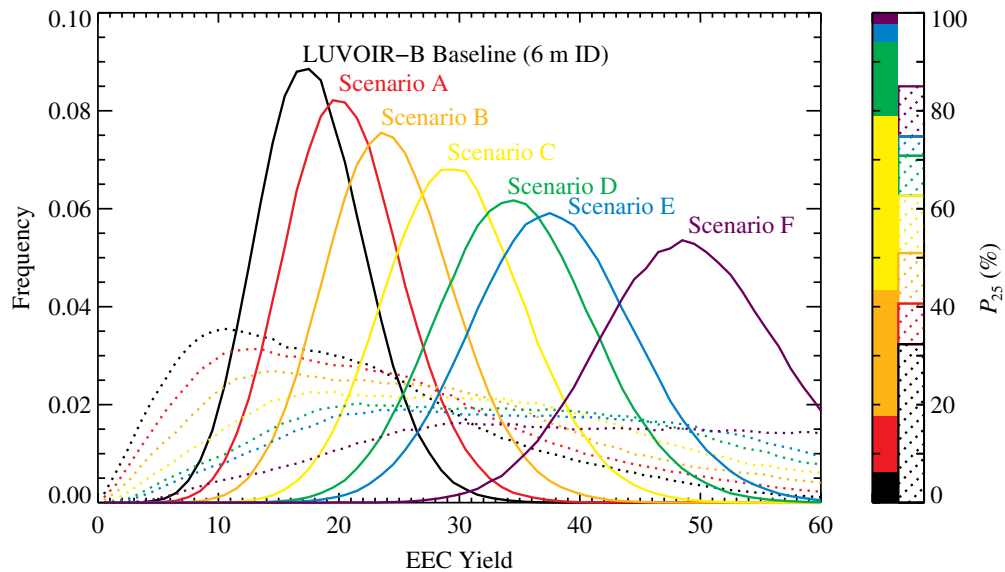


**Fig. 17** Wavelength-dependent optical throughput for a Cassegrain design with a single VIS channel imager and IFS (red), an ERD (green), and the LUVOIR-B baseline (black). Minimizing aluminum reflections can increase the throughput by 50% at 1000 nm and nearly double it at  $\sim 700$  nm. Opting for an ERD can increase the throughput by an additional 40% compared with an IFS.

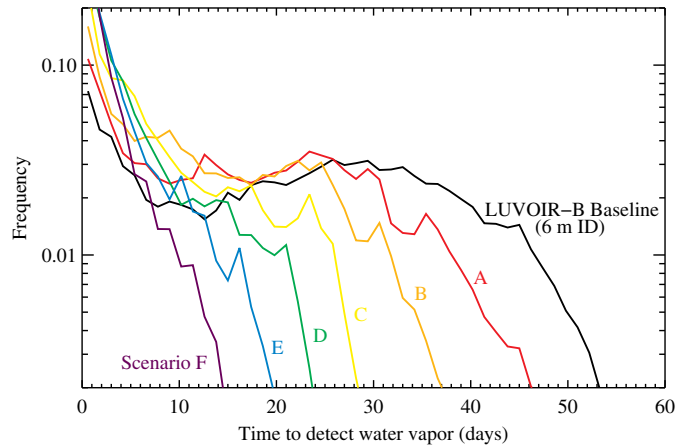
$\sim 1.7\times$  that of our baseline design’s VIS coronagraph and is  $1.2\times$  the effective throughput of the baseline design’s combined parallel UV and VIS coronagraphs.

Figure 18 shows the impact of this change on EEC yield. The red curve shows the estimated yield distribution for scenario A compared to the LUVOIR-B baseline yield distribution in black. Dashed and solid lines correspond to including and excluding  $\eta_{\oplus}$  uncertainties, respectively. Despite assuming only a single coronagraph channel, by minimizing the number of aluminum reflections, the EEC yield increases by 15%. The bar plots on the right in Fig. 18 show  $P_{25}$  increases as well to 0.18 and 0.41 when excluding and including  $\sigma_{\eta_{\oplus}}$ , respectively.

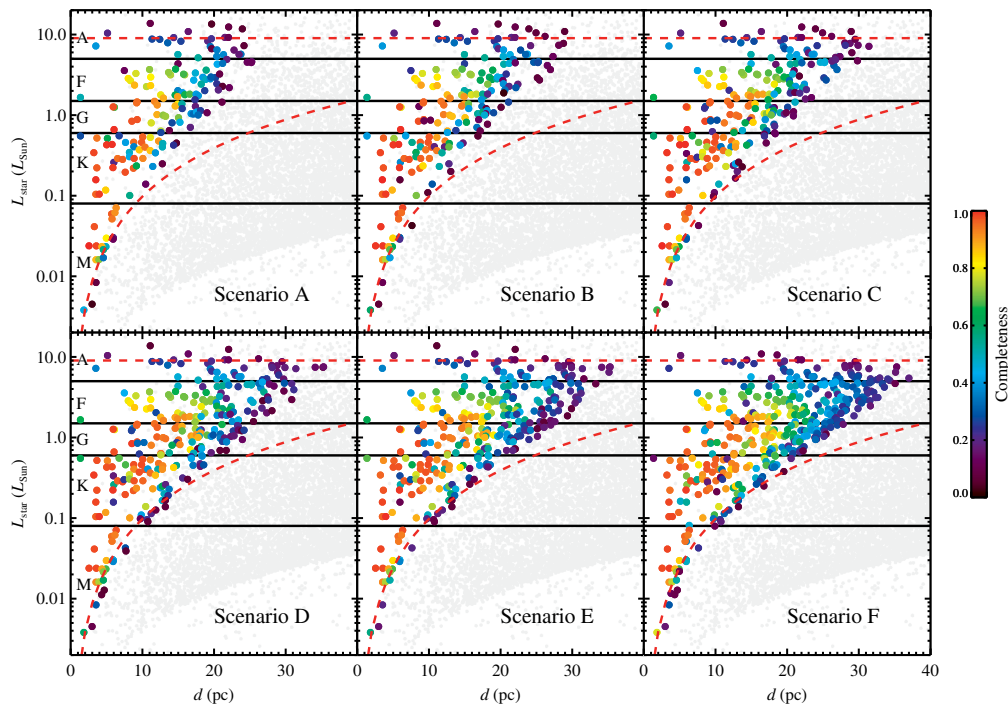
This change also improves exposure times in the visible channel. Figure 19 shows the spectral characterization time distribution for scenario A (red) compared to the LUVOIR-B baseline (black). The throughput at  $\sim 700$  nm is nearly twice that of our baseline design and is 50% greater



**Fig. 18** Left: Yield distributions for all design change scenarios considered, with dotted and solid lines including and excluding  $\eta_{\oplus}$  uncertainty, respectively. Table 4 summarizes the design changes included in each scenario. Right: Fraction of yield distribution  $>25$  EECs,  $P_{25}$ , when budgeting for detection and water vapor characterization. Dotted and solid bar plots correspond to including and excluding  $\eta_{\oplus}$  uncertainty, respectively.



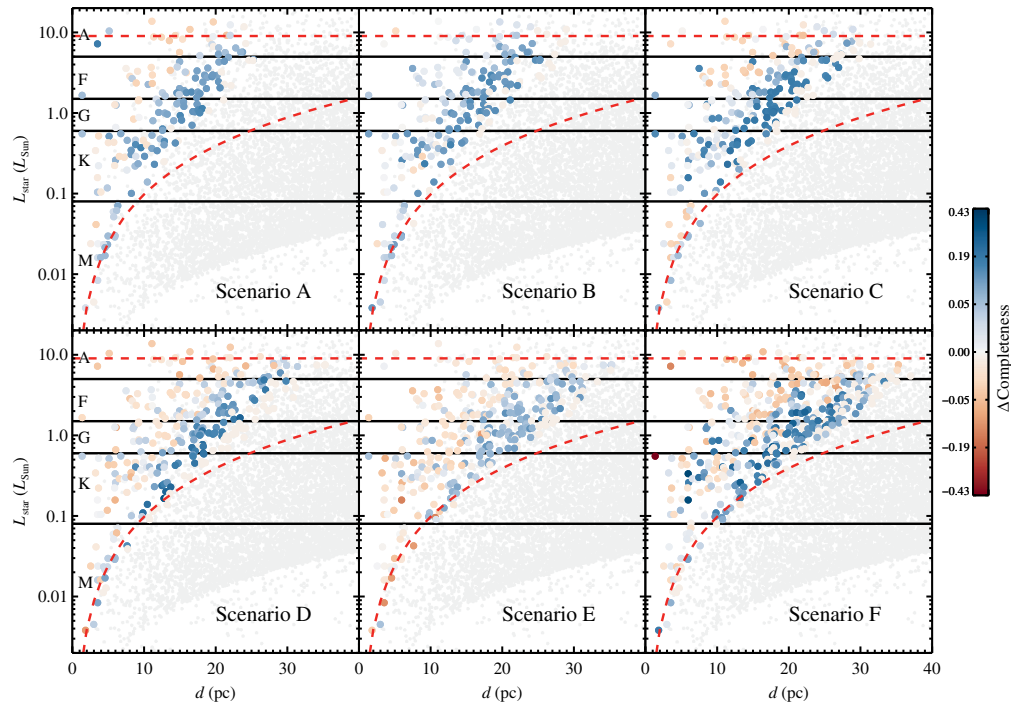
**Fig. 19** Spectral characterization time distributions for all design change scenarios considered when excluding  $\eta_{\oplus}$  uncertainty. Table 4 summarizes the design changes included in each scenario.



**Fig. 20** The targets selected for each scenario color-coded by total completeness. The scenarios are cumulative, in that we compile multiple design improvements. The dashed lines indicate the boundaries of the accessible targets—the horizontal line roughly marks the assumed astrophysical noise floor while the curved line roughly indicates the working angle limit.

at 1000 nm (cf. Fig. 17). Ultimately, this translates into mean characterization exposure times  $1.4\times$  shorter.

The top-left panel of Fig. 20 plots each target selected for observation in scenario A, color-coded by the total completeness achieved on each target. The horizontal dashed line indicates the stellar luminosity at which a  $1.4 R_{\oplus}$  planet at quadrature has a  $\Delta\text{mag}$  equal to the assumed astrophysical noise floor,  $\Delta\text{mag}_{\text{floor}}$ . The curved dashed line indicates the luminosity at which the outer edge of the HZ (1.67 AU) is located at  $1.5 \lambda/D$  for  $\lambda = 1000$  nm. These two dashed lines roughly indicate the boundaries of the accessible target list. The stars at the top-right corner of this “wedge” of targets require longer exposure times. Therefore, expanding the yield without reducing the IWA requires shorter detection times.



**Fig. 21** The incremental change in completeness between successive scenarios. Each design improvement allows the mission to access more distant stars.

The top-left panel of Fig. 21 shows the incremental change in the completeness of each star for scenario A compared with the baseline LUVOIR-B scenario. Blue indicates an increase in completeness, whereas red indicates a decline in completeness. The scale has been stretched to emphasize the sign of the change, as shown by the color bar on the right. As shown in Fig. 21, as the mission becomes more capable, the optimal redistribution of exposure time toward what were previously more challenging stars slightly reduces the completeness of nearby stars. While it may seem counter-intuitive to reduce exposure times of nearby stars in favor of more distant stars, this effect is the result of the less capable missions having “over-invested” time in nearby stars due to exposure time limitations.

#### 4.2.2 Scenario B: operate two visible coronagraphs in parallel

The LUVOIR-B study adopted a single visible wavelength coronagraph channel covering 500 to 1000 nm. In scenario B, we consider adding a second, parallel VIS coronagraph channel to scenario A, as shown in Fig. 1. This design change will significantly improve the yield of EECs that we can search for water vapor and will also substantially improve spectral data quality.

There are a number of reasons why two parallel VIS channels would be an improvement. First, detection efficiency would increase. If the two coronagraph channels could both observe near 500 nm, where detections are efficient, then the bandpass would effectively double, decreasing detection exposure times by a factor of two (ignoring overheads) and providing color information in every detection.

Second, the broader bandpass could improve the instantaneous spectral coverage for characterization. There are a number of key absorption features in the  $\sim 750$  to 1000 nm range, including molecular oxygen at 760 nm and water at 950 nm. These features may not be simultaneously observable with a single coronagraph channel without using advanced WFSC methods that sacrifice the field of view for bandwidth. This means that under the assumption of a single VIS channel, the spectrum from  $\sim 750$  to 1000 nm may have to be pieced together using observations at different epochs. This will be complicated by the fact that the planet changes phases during its orbit; stitching together the spectrum of an orbiting planet may be a challenging task. By covering the majority of the desired spectrum in a single observation, this challenge is eased.



Accessing more water absorption lines simultaneously via a broader bandwidth can substantially reduce the time needed to detect water. Reference 25 showed that by doubling the coronagraphic bandpass to 40%, the  $S/N$  needed to detect water near  $1\ \mu\text{m}$  is reduced from 5 to 4, which should equate to a  $\sim 40\%$  reduction in exposure time.

Finally, two visible coronagraphs would provide redundancy of HWO's primary science instrument, a critical requirement of a Class A mission. This redundancy could also provide the capability to observe two polarizations simultaneously in the event that a coronagraph channel requires a single polarization to achieve the necessary contrast.

To operate two VIS coronagraph channels efficiently for EEC detections, we would desire both coronagraphs to operate near 500 nm. On the other hand, efficient spectral characterizations mean we would want to operate the coronagraphs in the  $\sim 750$  to 1000 nm range. One way to enable both is to split the channels with a *selectable* dichroic, allowing us to split channels near  $\sim 500$  nm during detections or  $\sim 850$  nm during characterizations.

To estimate the impact of a dual VIS design, we perform the same calculations as described in Sec. 4.2.1, but double the bandwidth of the detection coronagraph and the number of detector pixels under the assumption of a selectable dichroic. As such, we assume that channels must be observed in adjacent bandpasses. To account for the reduced  $S/N$  required to detect water with 40% bandwidth, we adopt the red curve shown in Fig. 3 for spectral characterizations. The orange curve in Fig. 18 shows the distribution of yields for scenario B. This change has a substantial impact on the EEC yield, increasing it by 19%.

Doubling the number of visible coronagraph channels has a significant impact on yield because it effectively doubles the photon collection rate for photometric detections and reduces spectral characterization time by 40% by lowering the  $S/N$  required to detect  $\text{H}_2\text{O}$ , as shown by the orange curve in Fig. 19. We note that  $\text{H}_2\text{O}$  may be one of few atmospheric species that can take advantage of the broader bandwidth (another notably being  $\text{CH}_4$ ). In comparison,  $\text{O}_2$  has a single sharp feature near 760 nm that would not benefit from a broader bandwidth.

Dual visible coronagraph channels have several other scientific benefits that are not reflected in the yield or exposure time numbers. Specifically, dual photometric detections provide rudimentary color information that can help distinguish between planets from epoch to epoch,<sup>46</sup> which may be important as planets shift in position relative to the host star. Second, dual visible coronagraphs can provide an enhanced ability to simultaneously detect multiple atmospheric species in a single observation. Reference 25 showed that two 20% bandpasses can simultaneously detect  $\text{H}_2\text{O}$ ,  $\text{O}_2$ , and  $\text{O}_3$  at  $R = 140$  and  $S/N = 11$  assuming present atmospheric levels for an Earth twin, something a single 20% bandpass simply cannot do in a single observation.

#### 4.2.3 Scenario C: adopt model-based PSF subtraction

The LUVOIR study baselined coronagraphs with raw contrast of  $10^{-10}$ . However, HWO will need to detect planets with contrasts more challenging than  $10^{-10}$  at  $S/N > 10$ .<sup>6</sup> This means that the speckle noise floor must, at least, be better than  $10^{-11}$ . Designing a coronagraph with raw contrast better than  $10^{-11}$  would be very challenging, as restricting the raw contrast can limit the rest of the design phase space, potentially resulting in low throughput, a narrow bandpass, and/or greater sensitivities to wavefront aberrations.<sup>47</sup> Therefore, we will need to perform PSF subtraction to reduce speckles to better than  $10^{-11}$ . There are many potential PSF subtraction methods, each with benefits and challenges. Here, we address several key methods and discuss how adopting a model-based PSF subtraction method could substantially improve the science return.

The Roman Coronagraph will baseline reference differential imaging (RDI). Under the RDI approach, the instrument observes a science target and a reference target back-to-back (or potentially interleaved). The reference target is ideally an exact match to the science target, but bright and isolated, with no astrophysical scene around it. However, at the  $10^{-10}$  contrast level, we expect most stars to have some level of astrophysical contamination around them. Further, even though the bandpass for HWO may be relatively narrow, it will operate at wavelengths where the color of stars can be significantly different. Some coronagraphs are also sensitive to stellar diameter, meaning our reference star and science star would also have to match in terms of angular diameter. For these reasons, we suggest that RDI may be challenging for HWO.

One alternative is angular differential imaging (ADI), in which the science target is observed twice at two different roll angles. The speckles rotate with the telescope while the astrophysical scene remains fixed in the sky. As a result, we can co-align each exposure in the instrument frame to subtract the speckles, producing positive and negative copies of the astrophysical scene. This would provide a much better match in terms of the reference star and may help empirically subtract exozodiacal dust,<sup>36</sup> but comes at a cost. First, the differential roll angle to displace a PSF near the IWA would be  $\sim 40$  deg for HWO, placing strict constraints on wavefront stability as a function of roll angle. Second, the empirical ADI subtraction multiplies the count rate of all noise components by a factor of two. The LUVOIR and HabEx studies adopted ADI as the baseline PSF subtraction method and most yield calculations to date have included this factor of two on background count rates (c.f., Eq. (11) in Ref. 48 and Eq. (5) in Ref. 2).

Model-based PSF subtraction operates differently. By combining and correlating high-cadence wavefront telemetry with the bright unobscured starlight, we may be able to reconstruct the coronagraphic PSF at any point in time during the science exposure.<sup>49</sup> This could partially relax some telescope stability requirements, as we do not need to maintain PSF stability/repeatability at the  $10^{-11}$  level—only to a level that provides the desired raw contrast. Model-based PSF subtraction could also reduce the systematic speckle noise floor,<sup>49</sup> potentially to the Poisson noise limit, or to a level governed by the incoherence of the stellar leakage. Here, we ignore this potential benefit of model-based PSF subtraction and maintain the same noise floor described by  $\Delta\text{mag}_{\text{floor}} = 26.5$ , as prior studies have shown little gain in EEC yield when improving the noise floor.<sup>5,11</sup> Most pertinent to this study, because no empirical background subtraction is required, *model-based PSF subtraction removes the factor of two on all background count rates, effectively cutting all exposure times in half.*

We define scenario C as scenario B with model-based PSF subtraction added. For this scenario, we repeat the calculations performed in Sec. 4.2.2, but remove the factor of two in front of all background count rates in our exposure time calculator. The yellow curve in Fig. 18 shows the results for scenario C. Model-based PSF subtraction is estimated to improve yields by 23%, from 23.8 EECs for scenario B to 29.3 EECs for scenario C.  $P_{25}$  increases accordingly to 0.79 and 0.63 when excluding and including  $\sigma_{\eta_{\text{e}}}$ , respectively. As shown in Fig. 19, the characterization exposure times are reduced by a factor of 1.6. These significant improvements are the result of reducing *all* noise count rates by a factor of two, including detector noise.

HWO's PSF subtraction method may ultimately end up being a combination of multiple approaches. This could range from using a library of empirical PSFs<sup>50</sup> to spectral differential imaging.<sup>51</sup> Regardless of the technique, reducing the background noise associated with some empirical methods would be a fruitful endeavor.

#### 4.2.4 Scenario D: improve detector performance

The LUVOIR and HabEx studies baselined an EMCCD as the visible wavelength coronagraph detector. The adopted parameters for this EMCCD were optimistic, based on assumed future improvements to the Roman Coronagraph EMCCD. We have carried these assumptions through to this study, as shown in Table 2. Some of those assumptions have proven true. Roman's EMCCD dark current values are on par with the LUVOIR-B baseline assumption of  $3 \times 10^{-5}$  counts  $\text{pix}^{-1} \text{s}^{-1}$  and the dQE terms budgeting for photon counting efficiency, cosmic ray efficiency, hot pixel efficiency factors, etc., as defined by Ref. 52, are estimated to be  $\sim 0.77$  at end of life,<sup>52</sup> consistent with LUVOIR assumptions. Other assumptions remain optimistic. These include CIC, which remains a factor of 10 higher than the LUVOIR assumptions, and most notably, the raw QE near 1000 nm, which is only a few percent for Roman's EMCCD<sup>27</sup> but was assumed to be 90% in the LUVOIR study. Despite this, there remain paths forward using an EMCCD. The desire for high QE near 1000 nm was motivated by the desire for efficient detection of water vapor, but Ref. 8 showed that low QE near 1000 nm can be partially mitigated by searching for water at shorter wavelengths. Alternatively, a dedicated ultra-low-noise NIR detector could be more appropriate for the detection of water vapor.

Several alternative detector technologies may improve performance beyond the LUVOIR-B assumptions. Here, we examine the potential benefits of such detectors. A thorough examination of the impact of different detector technologies, critical to the success of HWO, would require

**Table 3** Detector parameters.

Parameter	Units	EMCCD	Skipper	ERD
QE	—	0.9	See Fig. 23	0.9
dQE	—	0.75	0.99	1.0
DC	counts pix <sup>-1</sup> s <sup>-1</sup>	$3 \times 10^{-5}$	$6.8 \times 10^{-9}$	0
CIC	counts pix <sup>-1</sup> frame <sup>-1</sup>	$1.3 \times 10^{-3}$	$1.5 \times 10^{-4}$	0
RN	counts pix <sup>-1</sup> read <sup>-1</sup>	0	0	0
Scenarios	—	0, A, B, C	D	E, F

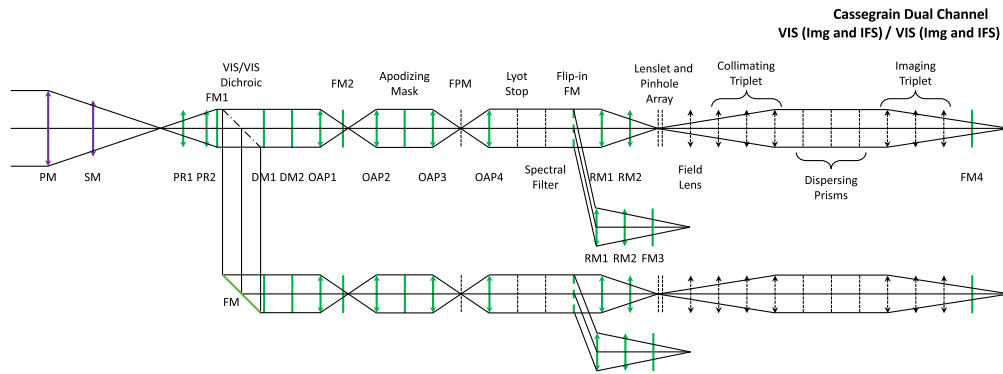
an exhaustive study comparing all detector options, which is beyond the scope of this paper. We therefore choose to adopt parameters consistent with two possible detector options. We start with performance parameters that may be possible with a photon-counting Skipper CCD. Table 3 summarizes the performance parameters we adopted compared to the LUVOIR-B baseline. We adopt a dark current one order of magnitude lower than the LUVOIR assumptions, which has been demonstrated for the Skipper CCD.<sup>53,54</sup> We note that this reduction in dark current stems from a combination of a high degree of shielding and cosmic ray identification and removal, which may be possible for a traditional EMCCD as well. However, the Skipper CCD also has a clock-induced charge 10× better than LUVOIR assumptions and has negligible dQE. As a result, the parameters we adopt provide ~30% higher effective throughput, as well as noise properties that are effectively negligible.

Skipper CCDs use multiple non-destructive reads to average read noise down to deeply sub-electron levels and thereby count photons.<sup>55</sup> Recent advances in semiconductor fabrication technology have made possible thick, fully depleted, photon-counting p-channel Skipper CCDs such as those used by Ref. 55. These p-channel Skippers, developed at the Lawrence Berkeley National Laboratory (LBNL), have important advantages for space astrophysics including excellent radiation tolerance and QE greater than 80% at 940 nm. However, the primary challenge for using LBNL’s Skipper CCDs in space is radiation nonetheless. Although p-channel Skipper CCDs do not degrade in space like n-channel CCDs, they require short exposure times to minimize cosmic ray disturbance. This is on account of the  $\approx 200 \mu\text{m}$  thick silicon that is used to achieve good near-IR QE. As a practical matter, p-channel Skipper exposure time will need to be on the order of 1 ms to limit cosmic ray disturbance to about 10% of pixels, requiring additional amplifier outputs—the recently developed Multi-Amplifier Sensing CCD<sup>56</sup> is a step toward this.

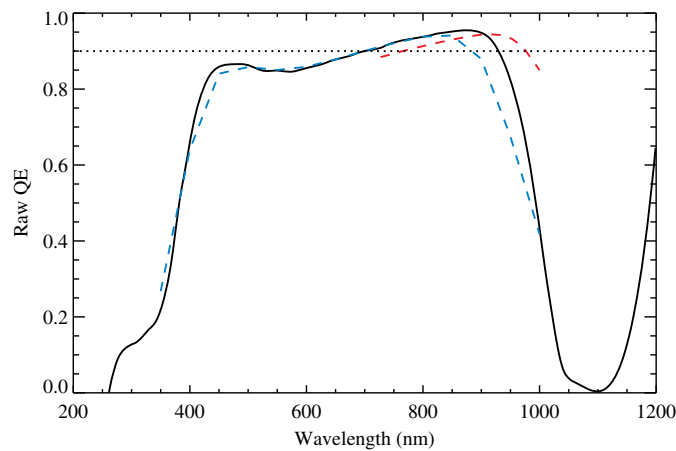
We define scenario D as scenario C with the EMCCD’s performance parameters replaced with the Skipper CCD parameters listed in Table 3. Here, we define QE as the traditional QE, i.e., the number of photoelectron groups created per photon received. We define dQE as the “detective” QE, a factor specific to EMCCDs discussed in Ref. 52.

Figure 22 shows the optical layout of scenario D and Fig. 23 shows our adopted raw QE curve for the Skipper CCD as a black solid line. For yield calculations, we are interested in a bandpass-integrated QE. For exoplanet detections, we assume the detection wavelength is centered in the bandpass as usual and integrated over a bandwidth of 20%, resulting in the blue dashed line that we adopt for detection raw QE. For spectral characterizations, we work with the longest wavelength of the bandpass to ensure the exoplanet is exterior to the coronagraph IWA over the whole bandpass. Using the single QE value at this wavelength would be doubly conservative, as the QE curve drops rapidly near 1000 nm. Thus, for spectral characterizations, we integrate the QE over a 20% bandpass with the longest wavelength given by the  $x$ -axis of Fig. 23, resulting in the red dashed line. We note that this detail is critical: the bandpass-integrated QE over the water vapor absorption feature near 1000 nm is twice that of the QE at 1000 nm. Future studies should investigate the impact of realistic QE curves on water vapor retrievals near 1000 nm.

Figure 18 shows the resulting yield distribution for scenario D in green. EEC yield increases by 17%, and  $P_{25}$  increases to 0.94 and 0.71 when excluding and including  $\sigma_{\eta_{\oplus}}$ , respectively. Figure 19 shows that the characterization times decrease as well by an additional factor of 1.5.



**Fig. 22** Optical layout for a Cassegrain design with two parallel VIS channels. We do not explicitly show dual parallel polarization channels for each wavelength channel, which we assume for the baseline coronagraph design. See Fig. 1 for legend.



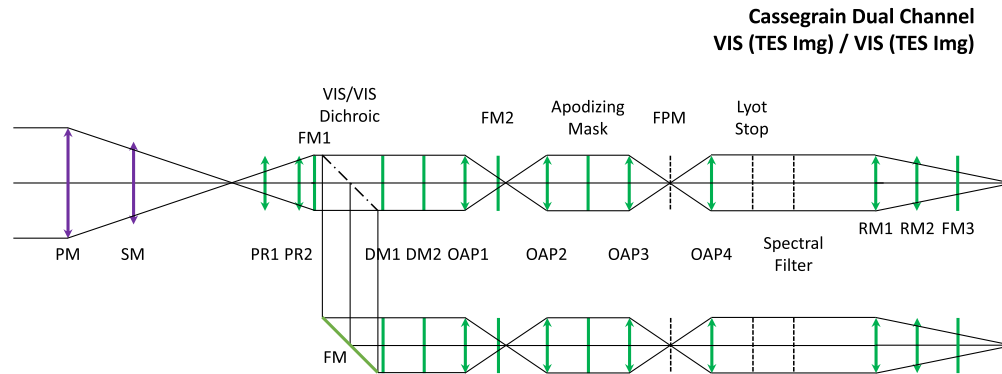
**Fig. 23** Skipper CCD raw QE (solid line) and the raw QE values adopted for the yield code when integrating over a 20% bandpass for detection (blue dashed) and spectral characterization (red dashed). The raw QE for the LUVOIR-B baseline EMCCD and ERD (scenario E) is shown for comparison (dotted line).

#### 4.2.5 Scenario E: adopt an energy-resolving detector

Next, we show the impact of swapping the Skipper CCD and IFS with a noiseless ERD. An IFS uses a lenslet at each “pixel” in the image plane to focus light onto a dispersing element, ultimately producing spectra of each image plane pixel in the final detector plane. While this instrument provides spatially resolved spectra over the entire field of view, there are some disadvantages. First, the additional IFS optics have a throughput estimated to be a factor of 0.7 (c.f., Table 2). Second, the PSF’s core is spread over potentially hundreds of pixels at the long wavelength end of the channel, effectively amplifying the detector’s per-pixel noise properties.

With an ERD, read noise manifests as part of the energy resolution budget<sup>57</sup> and there is no need for an IFS, eliminating throughput-reducing optics. Reference 10 showed that this, combined with negligible dQE, could result in a 30% increase in EEC yield compared with the LUVOIR-B baseline. Here, we examine the incremental improvement of an ERD compared with a Skipper CCD-based IFS, which also has negligible dQE and noise properties, so we will not see the same 30% increase in EEC yield.

Figure 24 illustrates the instrument layout with an ERD. We note that Fig. 24 does not show any optics to thermally isolate the detector from the rest of the instrument, which will be required for an ERD. These optics will reduce the throughput of the system, but a detailed assessment of thermal isolation and the transmissivities of the necessary optics is beyond the scope of this paper. Our estimate of the impact of an ERD should therefore be considered an upper limit.



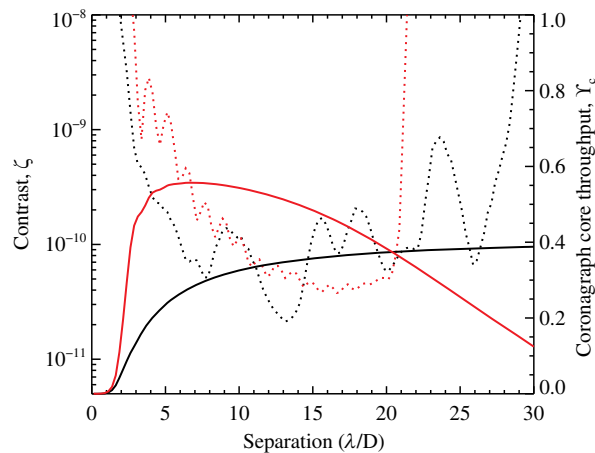
**Fig. 24** Optical layout for a Cassegrain design with two parallel VIS channels with ERD. See Fig. 1 for legend.

Figure 17 shows the assumed optical throughput using an ERD (green). Because there is no separate imaging mode for an ERD, the green line illustrates the throughput for detections and for spectroscopy. An ERD can potentially increase spectroscopic throughput by 40% compared with an IFS. Table 3 summarizes the adopted performance parameters for our ERD, based on the expected performance of a transition edge sensor.<sup>57</sup> To date, TES arrays have demonstrated  $R = 90$  at 485 nm<sup>58</sup> and MKIDS have demonstrated  $R = 52$  at 402 nm.<sup>59</sup> We assume that a future ERD can achieve the  $R = 140$  requirement that we adopt.

Figure 18 shows the resulting yield distribution for scenario E in blue. The ERD increases EEC yield by 9% compared with the Skipper CCD scenario and  $P_{25}$  increases to 0.98 and 0.75 when excluding and including  $\sigma_{\eta_{\oplus}}$ , respectively. The ERD also reduces spectral characterization times compared with the Skipper CCD scenario by another factor of 1.4 $\times$ , as shown in Fig. 19. Most of these improvements are the result of the increased throughput due to lack of IFS optics, not reduced detector noise, as the adopted noise parameters for the Skipper were already very low. We note that when compared with the LUVUOIR-B baseline detector assumptions, the ERD gains are larger, with a 2.2 $\times$  reduction in characterization times and a 1.3 $\times$  gain in EEC yield.

#### 4.2.6 Scenario F: adopt a high-throughput coronagraph

The LUVUOIR-B study baselined a DM-assisted charge six vortex coronagraph (DMVC6). Figure 25 shows the azimuthally averaged contrast for a  $0.1 \lambda/D$  star as a dotted black line over a 20% bandwidth for a single polarization. We note that while the DMVC6 only works for a single



**Fig. 25** Azimuthally averaged contrast for a star with diameter  $0.1 \lambda/D$  (dotted line) and core throughput (solid line) for the two coronagraphs included in this study. The x-axis is in units of  $\lambda/D$ , where  $D$  is the circumscribed diameter of the telescope. The DMVC6 and PIAA-FPM2.5 are shown in black and red, respectively.

polarization, we have implicitly assumed so far that the design allows for parallel polarization channels. The solid black line shows the core throughput of this coronagraph. The IWA for the DMVC6 is  $\sim 3.5 \lambda/D$ , where the core throughput reaches half its maximum value. We remind the reader that  $D$  in this study and that adopted for the  $x$ -axis of Fig. 25, is the circumscribed diameter of the telescope. The DMVC6 is limited to using the inscribed diameter of the telescope, making its IWA in Fig. 25 larger than that of a circular aperture. Notably, there is a useful throughput interior to the IWA, which the yield code takes advantage of for nearby, later-type stars (see Fig. 11). While the core throughput reaches a relatively high maximum value of  $\sim 0.45$ , it rises fairly slowly with working angle, such that it is  $\sim 5\%$  at  $2 \lambda/D$ .

Here, we examine one possible alternative coronagraph design that has higher throughput at small working angles: the phase-induced amplitude apodizer (PIAA) coronagraph. We adopt a PIAA design created for the LUVOIR-B aperture, which incorporated a focal plane mask with radius  $2.5 \lambda/D$  and a DM-assisted solution robust to stellar diameters as large as  $0.1 \lambda/D$  (we note that this DM solution may help other coronagraph designs as well). We refer to this design as PIAA-FPM2.5. Figure 25 shows the azimuthally averaged performance of the PIAA-FPM2.5 in red. In comparison with the DMVC6, the contrast near  $\sim 4 \lambda/D$  is a factor of  $\sim 4$  worse and the effective OWA is notably limited by contrast degradations to  $\sim 20 \lambda/D$ . Notably, we maintain the same uniform noise floor used throughout this study, which is not proportional to the raw contrast. We choose this single, non-ideal coronagraph design as an example to illustrate a point: despite the degraded contrast, if the noise floor remains the same, the increase in core throughput at small working angles will more than compensate on a survey scale, resulting in significantly improved yields. This is because exposure times are reduced for targets in which the leaked starlight does not dominate the background count rate (e.g., distant stars).

Unlike the DMVC6, the PIAA-FPM2.5 works for both polarizations simultaneously. This has the potential to reduce the complexity of the instrument by eliminating parallel polarization channels. In principle, the complexity could be maintained and the parallel polarization channels could be replaced by another dichroic split, doubling the total instrument bandwidth. However, here, we make the conservative assumption that dual polarization channels are still required to minimize polarization cross-talk and do not adopt any increase in total instrument bandpass.

The purple curve in Fig. 18 shows the yield distribution for Scenario F. In spite of the degraded contrast, the higher throughput of the PIAA-FPM2.5 coronagraph design increases EEC yield by 30%. Combined with all previously discussed changes,  $P_{25}$  is now estimated to be 0.99 and 0.85 when excluding and including  $\sigma_{\eta_{\oplus}}$ , respectively. The PIAA-FPM2.5 reduces spectral characterization times compared with scenario E by another factor of 1.6 $\times$ , as shown in Fig. 19.

We note that many other coronagraph designs with smaller IWA and higher core throughput exist in addition to the PIAA-FPM2.5 examined here. Such coronagraphs may also lead to higher yields to varying degrees. A simple example that we do not consider in this paper is combining our baseline charge 6 DMVC with a charge 4 design. A future trade study building off of the Coronagraph Design Survey<sup>60</sup> and examining a broad range of coronagraphs designed specifically for HWO would be highly valuable.

#### 4.2.7 Summary of design changes

We investigated six possible design changes from the LUVOIR-B baseline design. As shown in Fig. 19, these improvements significantly reduced exposure times. This in turn allowed the simulated mission to observe a larger portion of the accessible targets at larger distances, as illustrated in Figs. 20 and 21, and achieve much higher yields.

Notably, all of these design changes produced modest incremental improvements to yield and all are roughly consistent with the scaling relationships between exposure time and yield reported in previous studies.<sup>2,11</sup> Here, we showed that while yield is only moderately sensitive to changes in exposure time factors (throughput, bandwidth, etc.), there are many such factors. If multiple factors are improved simultaneously, the impact on exposure time and yield can be large.

**Table 4** Summary of design change impacts.

Scenario	Description	Refer to Section	Reduction in char. time to detect <sup>a,b*</sup>	Increase in instantaneous VIS bandwidth <sup>a</sup>	Increase in EEC yield <sup>a</sup>	$P_{25}$ when ignoring $\sigma_{\eta_{\oplus}}$	$P_{25}$ when including $\sigma_{\eta_{\oplus}}$
0	LUVOIR-B baseline (6 m ID)	2	1×/1×	1×/1×	1×/1×	0.06	0.32
A	Minimize Al coatings	4.2.1	1.4×/1.4×	1×/1×	1.15×/1.15×	0.18	0.41
B	A + Dual VIS	4.2.2	1.4×/1.9×	2×/2×	1.19×/1.37×	0.43	0.51
C	B + Model-based PSF sub.	4.2.3	1.6×/3.1×	1×/2×	1.23×/1.69×	0.79	0.63
D	C + Skipper CCD	4.2.4	1.5×/4.8×	1×/2×	1.17×/1.98×	0.94	0.71
E	D + Energy-resolving det.	4.2.5	1.4×/6.8×	1×/2×	1.09×/2.16×	0.98	0.75
F	E + High-throughput coron.	4.2.6	1.6×/11.1×	1×/2×	1.30×/2.81×	0.99	0.85

<sup>a</sup>Columns with more than one value separated by a slash indicate the incremental/cumulative change in the quantity for each scenario compared to the previous/baseline scenario.

<sup>b</sup>For the first <18 EECs, excluding static overheads.

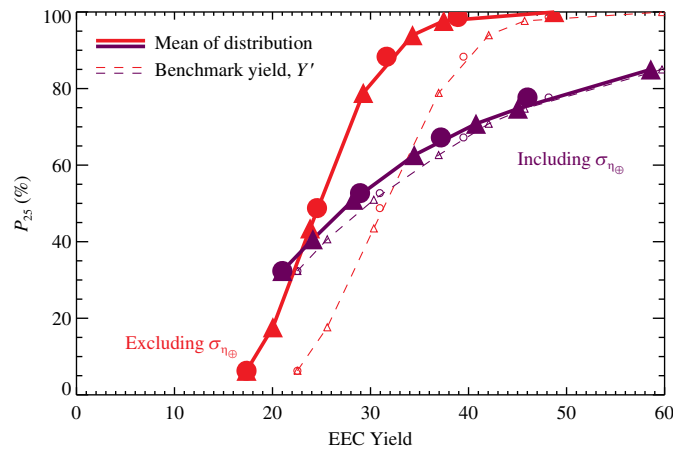
Table 4 lists the incremental and cumulative reductions in exposure times and increases in EEC yield, as well as the value of  $P_{25}$  associated with each design change. Incremental changes to yield are broadly consistent with scaling relationships from previous works.<sup>2,5,11</sup> Overall, spectral characterization times can be reduced by more than an order of magnitude while doubling the characterization bandwidth over the visible spectrum. This results in nearly a tripling of the EEC yield. The probability of detecting and characterizing >25 EECs increases dramatically, with  $P_{25}$  increasing from 6% to 99% in the case in which  $\eta_{\oplus}$  uncertainty is ignored. Given that these six examples are not an exhaustive list, and none of these changes includes increasing the telescope diameter, we conclude that it is possible to establish EEC science margins substantial enough to offset most astrophysical uncertainties for HWO.

## 5 Budgeting for Uncertainties with Science Margin

### 5.1 Budgeting for Astrophysical Uncertainty

Precisely how much science margin is required for HWO to budget for astrophysical uncertainty? This depends in large part on HWO's risk posture, formalized minimum yield goal, whether HWO's formal science goals should account for uncertainty in  $\eta_{\oplus}$ , the magnitude of that uncertainty, and whether the uncertainty can be reduced with precursor science observations or analyses. These decisions will likely come out of future formalized HWO modeling efforts and are beyond the scope of this paper. However, we can use the results of Secs. 4.1 and 4.2 to provide basic guidance for these future decisions by relating the expectation value of EEC yield to  $P_{25}$ . The solid lines in Fig. 26 show  $P_{25}$  as a function of mean EEC yield excluding and including uncertainty in  $\eta_{\oplus}$  in red and purple, respectively. The yields obtained by increasing telescope diameter (Sec. 4.1) are shown as filled circles, whereas the yields from improving mission design (Sec. 4.2) are shown as filled triangles with a connecting line. The agreement between the lines and circles suggests that  $P_{25}$  is independent of the specific means used to obtain higher yields. Figure 26 can therefore be used to estimate  $P_{25}$  for a broad range of HWO trade studies by calculating the expectation value of the EEC yield distribution.

We note that calculating the yield distribution including all astrophysical noise sources is critical, as it includes shifts in the expectation value of the yield due to observational biases induced by exoplanet albedo and exozodi uncertainties. However, in practice, calculating a yield distribution is numerically taxing, as it requires hundreds of independent yield calculations to sample the range of possible exozodi and  $\eta_{\oplus}$  values, and calculating yield distributions including



**Fig. 26** Fraction of yield distributions  $>25$  EECs,  $P_{25}$ , as a function of expected EEC yield excluding (solid red) and including (solid purple) uncertainty in  $\eta_{\oplus}$ . Filled circles indicate yields obtained by changes to telescope diameter while filled triangles with a connecting line indicate those due to design changes. The agreement between lines and circles suggests  $P_{25}$  is independent of the means used to obtain higher yields. The unfilled symbols and dashed lines show the expectation value of the benchmark yield,  $Y'$ , which can be used to estimate  $P_{25}$  without calculating a yield distribution.

all sources of astrophysical noise could slow trade studies. To aid with this, we “translate” each of the filled points in Fig. 26 to a much simpler quantity that only requires a single yield calculation, a benchmark yield,  $Y'$ . We define  $Y'$  as the yield assuming 3 zodis of dust around all stars,  $\eta_{\oplus} = 0.24$ , and no albedo or exozodi observational biases included. The empty circles and triangles in Fig. 26, along with the thin dashed line, show  $Y'$  for each of the scenarios shown as filled symbols. Using the dashed curves in Fig. 26, one can design a mission to achieve a given  $P_{25}$  via a single yield calculation, knowing that when astrophysical uncertainties are included the mean yields will shift to the solid curves.

## 5.2 Budgeting for Performance Uncertainty

Science margin can also help reduce risk by budgeting for performance uncertainties. There are unlimited potential causes of on-sky performance degradations, manifesting as, e.g., poor line of sight jitter, coating or detector degradations, and unexpected stray light. Here, we do not focus on root causes. Instead, we discuss performance degradation in terms of bulk parameters used in our yield analyses, such as raw contrast and throughput.

As originally shown by Ref. 5 and later verified by Ref. 2 with updated fidelity, EEC yield decreases relatively gracefully with degradations in most parameters. For the LUVOIR-B baseline parameters, yield is only moderately sensitive to throughput-related terms (scaling roughly to the  $\sim 0.37$  power<sup>2</sup>); a relatively large factor of two reductions in effective throughput would only reduce yield by  $\sim 25\%$ , though it would increase exposure times by roughly a factor of two. Assuming a noise floor independent of raw contrast, LUVOIR-B was also very insensitive to raw contrast (scaling as raw contrast to the  $\sim -0.07$  power<sup>2</sup>); a factor of two degradation in raw contrast would only reduce yield by  $\sim 5\%$ . To first order, assuming the noise floor is not coupled to raw contrast, degradations in these quantities slow the progress of observing the full target list, but do not limit the boundaries of accessible targets (indicated by the red dashed lines shown in Fig. 20). Because this can be partially mitigated by additional survey time, we do not consider these parameters as substantial drivers of performance risk.

We posit that the astrophysical noise floor,  $\Delta\text{mag}_{\text{floor}}$ , is a primary driver of performance risk. The noise floor is ultimately determined by one of the largest technological “tall poles:” the ability to precisely estimate the coronagraphic speckle pattern, which may require picometer stability,<sup>6</sup> very precise wavefront sensing if model-based PSF subtraction is used, or a combination of the two. As shown by Fig. 8 in Ref. 11, for  $\Delta\text{mag}_{\text{floor}} = 26.5$  as assumed in this study, degradations of a factor of two in the flux associated with the noise floor could lead to reductions in EEC yield of  $\sim 25\%$ , on par with throughput factors. However, the impacts of the noise floor



differ from throughput factors in two important ways. The first is that the yield’s sensitivity to the noise floor appears to be a “cliff” (c.f. Fig. 8 in Ref. 11), suggesting that additional degradations beyond the initial factor of two become significantly more costly—i.e., it is a slippery slope.

The second is more fundamental: degradations in the noise floor directly limit the range of accessible targets. The horizontal dashed lines in Fig. 20 mark the luminosity at which a  $1.4 R_{\oplus}$  planet at quadrature at the EEID has a flux equal to the astrophysical noise floor. This is a rough visual guide. In reality, because planets could occupy a range of phase angles, semi-major axes, and radii, the effects of the noise floor can be seen in Fig. 20 as a broad horizontal band that reduces the completeness of targets with luminosities  $\gtrsim 2 L_{\odot}$ . Therefore, reductions in the noise floor will move this band to lower luminosities, directly removing targets that cannot be accessed in any other way. Further, reductions in the noise floor would begin to remove the most Sun-like stars, G-type stars, from the target list, substantially affecting the mission’s ability to survey for Earth-like planets around Sun-like stars.

One approach to mitigating the risk of degradation in the noise floor is to adopt technologies that relax the system level requirements needed to achieve the desired noise floor. Any technology that relaxes optical stability requirements would aid in this. Of particular note is the model-based PSF subtraction method we considered in scenario C. Model-based PSF subtraction monitors the wavefront error at a high cadence so that we can reconstruct the instantaneous science PSF in high fidelity. This would allow us to discard photons collected at times with poor WFE or at least accurately model how the science PSF varies with WFE to a level better than the Poisson noise. It also would remove the need to maintain an ultrastable WF in the presence of a spacecraft slew or roll, which would be required for the RDI and ADI PSF subtraction methods, respectively. Much work is needed to determine if model-based PSF subtraction is viable for HWO, but the benefit could be significant.

Another approach to mitigating the risk of noise floor degradation is to budget for it with a science margin. If degrading the noise floor moves the horizontal dashed lines in Fig. 20 downward, we could maintain our access to a large pool of stars by shifting the curved dashed line downward as well. In other words, the target list would shift toward later type stars. The curved dashed line in Fig. 20 marks where a working angle of  $1.5 \lambda/D$  is at the outer edge of the HZ. This working angle is admittedly fairly extreme already, and it is unlikely that alternative coronagraph designs could reduce it further in units of  $\lambda/D$ . In addition, as discussed in Sec. 3.4, there is likely a minimum “useful” working angle set by spatial resolution requirements interior to which planets blend together too often<sup>45</sup>—this may be larger than the  $1.5 \lambda/D$  shown in Fig. 20. There are therefore only two ways of shifting the curved dashed line downward: by operating at shorter  $\lambda$  or larger  $D$ . Reference 8 showed that detecting water at shorter  $\lambda$  is possible, but not preferable, as it requires significantly longer exposure times. Thus, we conclude that an increase in telescope diameter should be considered as an option to budget for noise floor degradation.

## 6 Conclusions

We identified and estimated the impact of all major sources of astrophysical uncertainty on the EEC yield from a blind exoEarth survey by HWO, where “yield” was defined as the detection and search for water vapor on all EECs. We find that while  $\eta_{\oplus}$  uncertainties dominate the uncertainty in EEC yield, the sampling uncertainties inherent to a blind exoplanet survey are another important source of uncertainty and should be accounted for in mission design. We caution against adopting a science goal of 100 cumulative HZs, which is not equivalent to ignoring only the uncertainty in  $\eta_{\oplus}$ ; such a science goal would effectively ignore *both* dominant sources of astrophysical uncertainty.

We find that exoplanet albedo uncertainty and exozodi sampling uncertainties shift the expectation value of the yield to lower values. The effect of the uncertainty in the exozodi distribution is less clear. We performed a re-analysis of fits to the LBTI exozodi observations. We find that assuming the exozodi distribution is multi-modal, the uncertainty in the exozodi distribution appears to have a relatively minor impact on yield uncertainty. However, this is only true if the fraction of the exozodi distribution  $\lesssim 3$  zodis (which is a better predictor of exoplanet science yield than the median exozodi level), is well-constrained. Details of our ability to

precisely subtract exozodi, which we do not address here, may still have a large impact on mission yield as it sets a systematic noise floor for the mission.

By including all astrophysical uncertainties, we estimated the yield distribution for a given mission design scenario and calculated the fraction of the distribution  $>25$  EECs, defined as  $P_{25}$ . Adopting the LUVOIR-B baseline design, we find that an  $\sim 8$  to 9 m ID telescope is needed to produce large science margins, resulting in  $P_{25} \sim 95\%$  when ignoring uncertainties in  $\eta_{\oplus}$  and  $P_{25} \sim 75\%$  when including  $\eta_{\oplus}$  uncertainties. We identified six possible design changes from the LUVOIR-B baseline, each of which focuses on reducing exposure times and provides a modest gain in yield on its own. However, when combined, *these improvements compile to provide significant performance gains*, nearly tripling EEC yield and reducing spectral characterization times by more than an order of magnitude for the highest priority targets. We find that with these changes it is possible for a  $\sim 6$  m ID telescope to produce substantial science margins, providing  $P_{25} > 99\%$  when ignoring uncertainties in  $\eta_{\oplus}$  and  $P_{25} \sim 85\%$  when including  $\eta_{\oplus}$  uncertainties. We conclude that a combination of telescope diameter increase and instrument design changes could provide robust exoplanet science margins for HWO.

We discussed how science margin can help mitigate on-sky performance degradations. Whereas degradation of contrast and throughput lengthen exposure times, we showed that degradation of the noise floor can lead to a large fraction of the target list being permanently unobservable. We identified increasing telescope diameter as a promising path to reducing the risk associated with noise floor degradation.

---

## Data and Code Availability

NASA regulations govern the release of source code, including what can be released and how it is made available. Readers should contact the corresponding author if they would like copies of the visualization software or data produced for this study.

## Acknowledgments

This project was supported by the NASA HQ-directed ExoSpec work package under the Internal Scientist Funding Model (ISFM). N.L. gratefully acknowledges financial support from an NSF GRFP. N.W.T. is supported by an appointment with the NASA Postdoctoral Program at the NASA Goddard Space Flight Center, administered by Oak Ridge Associated Universities under contract with NASA. N.W.T. acknowledges support from the GSFC Sellers Exoplanet Environments Collaboration (SEEC), which is supported by NASA's Planetary, Astrophysics, and Heliophysics Science Divisions' Research Program. The Center for Exoplanets and Habitable Worlds and the Penn State Extraterrestrial Intelligence Center are supported by the Pennsylvania State University and the Eberly College of Science. E.B.F.'s contributions focused on Bayesian analysis of the distribution of exozodi levels and interpretation of  $\eta_{\oplus}$  studies and did not extend to §4. C.S. acknowledges the contributions of two anonymous reviewers whose feedback substantially improved this manuscript.

## References

1. E. National Academies of Sciences and Medicine, *Pathways to Discovery in Astronomy and Astrophysics for the 2020s*, The National Academies Press (2021).
2. C. C. Stark et al., "ExoEarth yield landscape for future direct imaging space telescopes," *J. Astron. Telesc. Instrum. Syst.* **5**, 024009 (2019).
3. R. Morgan et al., "Faster Exo-Earth yield for HabEx and LUVOIR via extreme precision radial velocity prior knowledge," *J. Astron. Telesc. Instrum. Syst.* **7**, 021220 (2021).
4. D. Savransky and D. Garrett, "WFIRST-AFTA coronagraph science yield modeling with EXOSIMS," *J. Astron. Telesc. Instrum. Syst.* **2**, 011006 (2016).
5. C. C. Stark et al., "Maximizing the ExoEarth candidate yield from a future direct imaging mission," *Astrophys. J.* **795**, 122 (2014).
6. The LUVOIR Team, "The LUVOIR Mission concept study final report," arXiv:1912.06219 (2019).
7. B. S. Gaudi et al., "The Habitable Exoplanet Observatory (HabEx) mission concept study final report," arXiv:2001.06683 (2020).
8. C. C. Stark et al., "Optimized bandpasses for the Habitable Worlds Observatory's ExoEarth Survey," arXiv:2404.05654 (2024).

9. R. K. Kopparapu et al., “Exoplanet classification and yield estimates for direct imaging missions,” *Astrophys. J.* **856**, 122 (2018).
10. A. R. Howe, C. C. Stark, and J. E. Sadleir, “Scientific impact of a noiseless energy-resolving detector for a future exoplanet-imaging mission,” *J. Astron. Telesc. Instrum. Syst.* **10**(2), 025008 (2024).
11. C. C. Stark et al., “Lower limits on aperture size for an ExoEarth detecting coronagraphic mission,” *Astrophys. J.* **808**, 149 (2015).
12. S. L. Hunyadi, S. B. Shaklan, and R. A. Brown, “The lighter side of TPF-C: evaluating the scientific gain from a smaller mission concept,” *Proc. SPIE* **6693**, 66930Q (2007).
13. N. W. Tuchow, C. C. Stark, and E. Mamajek, “HPIC: the Habitable Worlds Observatory preliminary input catalog,” *Astron. J.* **167**, 139 (2024).
14. K. G. Stassun et al., “The revised TESS input catalog and candidate target list,” *Astron. J.* **158**, 138 (2019).
15. A. Vallenari et al., “Gaia data release 3. Summary of the content and survey properties,” *Astron. Astrophys.* **674**, A1 (2023).
16. R. L. Smart et al., “Gaia early data release 3. The Gaia catalogue of nearby stars,” *Astron. Astrophys.* **649**, A6 (2021).
17. D. Sirbu et al., “Multi-star wavefront control for the wide-field infrared survey telescope,” *Proc. SPIE* **10698**, 106982F (2018).
18. E. Mamajek and K. Stapelfeldt, “NASA Exoplanet Exploration Program (ExEP) mission star list for the Habitable Worlds Observatory (2023),” arXiv:2402.12414 (2024).
19. R. K. Kopparapu et al., “Habitable zones around main-sequence stars: new estimates,” *Astrophys. J.* **765**, 131 (2013).
20. R. K. Kopparapu et al., “Habitable zones around main-sequence stars: dependence on planetary mass,” *Astrophys. J.* **787**, L29 (2014).
21. R. Belikov et al., “ExoPAG ExoPAG SAG13: exoplanet occurrence rates and distributions,” 2017, [https://exoplanets.nasa.gov/system/presentations/files/67\\_Belikov\\_SAG13\\_ExoPAG16\\_draft\\_v4.pdf](https://exoplanets.nasa.gov/system/presentations/files/67_Belikov_SAG13_ExoPAG16_draft_v4.pdf) (accessed 12 April 2024).
22. S. Bryson et al., “The occurrence of rocky habitable-zone planets around solar-like stars from Kepler data,” *Astron. J.* **161**, 36 (2021).
23. S. Ertel et al., “The HOSTS survey for exozodiacal dust: observational results from the complete survey,” *Astron. J.* **159**, 177 (2020).
24. N. Latouf et al., “Bayesian analysis for remote biosignature identification on exoEarths (BARBIE). I. Using grid-based nested sampling in coronagraphy observation simulations for H<sub>2</sub>O,” *Astron. J.* **166**, 129 (2023).
25. N. Latouf et al., “Bayesian analysis for remote biosignature identification on exoEarths (BARBIE). II. Using grid-based nested sampling in coronagraphy observation simulations for O<sub>2</sub> and O<sub>3</sub>,” *Astron. J.* **167**, 27 (2024).
26. B. Nemati et al., “Method for deriving optical telescope performance specifications for Earth-detecting coronagraphs,” *J. Astron. Telesc. Instrum. Syst.* **6**, 039002 (2020).
27. “Nancy grace roman space telescope spacecraft and instrument parameters,” 2023, [https://roman.ipac.caltech.edu/sims/Param\\_db.html](https://roman.ipac.caltech.edu/sims/Param_db.html) (accessed 12 April 2024).
28. M. Bruna et al., “Combining photometry and astrometry to improve orbit retrieval of directly imaged exoplanets,” *Mon. Not. R. Astron. Soc.* **519**, 460–470 (2023).
29. C. C. Stark et al., “A direct comparison of exoEarth yields for starshades and coronagraphs,” *Proc. SPIE* **9904**, 99041U (2016).
30. J. Crass et al., “Extreme precision radial velocity working group final report,” arXiv:2107.14291 (2021).
31. T. D. Robinson et al., “Earth as an Extrasolar planet: earth model validation using EPOXI earth observations,” *Astrobiology* **11**, 393–408 (2011).
32. G. Tinetti et al., “Disk-averaged synthetic spectra of mars,” *Astrobiology* **5**, 461–482 (2005).
33. G. Tinetti et al., “Detectability of planetary characteristics in disk-averaged spectra. I: the Earth model,” *Astrobiology* **6**, 34–47 (2006).
34. C. Cox and W. Munk, “Measurement of the roughness of the sea surface from photographs of the sun’s glitter,” *J. Opt. Soc. Amer.* **44**, 838 (1954).
35. B. Mennesson et al., “Constraining the exozodiacal luminosity function of main-sequence stars: complete results from the Keck Nuller mid-infrared surveys,” *Astrophys. J.* **797**, 119 (2014).
36. J. Kammerer et al., “Simulating reflected light coronagraphy of earth-like exoplanets with a large IR/O/UV Space Telescope: impact and calibration of smooth exozodiacal dust,” *Astron. J.* **164**, 235 (2022).
37. D. Defrère et al., “Direct imaging of exoEarths embedded in clumpy debris disks,” *Proc. SPIE* **8442**, 84420M (2012).
38. M. H. Currie et al., “Mitigating worst-case exozodiacal dust structure in high-contrast images of Earth-like exoplanets,” *Astron. J.* **166**, 197 (2023).
39. O. Absil et al., “A near-infrared interferometric survey of debris-disc stars. III. First statistics based on 42 stars observed with CHARA/FLUOR,” *Astron. Astrophys.* **555**, A104 (2013).

40. S. Ertel et al., “A near-infrared interferometric survey of debris-disk stars. IV. An unbiased sample of 92 southern stars observed in H band with VLTI/PIONIER,” *Astron. Astrophys.* **570**, A128 (2014).
41. C. C. Stark, M. J. Kuchner, and A. Lincowski, “The Pseudo-Zodi problem for edge-on planetary systems,” *Astrophys. J.* **801**, 128 (2015).
42. S. E. Thompson et al., “Planetary candidates observed by Kepler. VIII. A fully automated catalog with measured completeness and reliability based on data release 25,” *Astrophys. J.* **235**, 38 (2018).
43. T. A. Berger et al., “The Gaia-Kepler Stellar properties catalog. II. Planet radius demographics as a function of stellar mass and age,” *Astron. J.* **160**, 108 (2020).
44. T. A. Berger et al., “The Gaia-Kepler stellar properties catalog. I. Homogeneous fundamental properties for 186,301 Kepler stars,” *Astron. J.* **159**, 280 (2020).
45. P. Saxena, “Photobombing Earth 2.0: diffraction-limit-related contamination and uncertainty in Habitable Planet Spectra,” *Astrophys. J.* **934**, L32 (2022).
46. J. Krissansen-Totton et al., “Is the pale blue dot unique? Optimized photometric bands for identifying Earth-like exoplanets,” *Astrophys. J.* **817**, 31 (2016).
47. K. St. Laurent et al., “Apodized pupil Lyot coronagraphs designs for future segmented space telescopes,” *Proc. SPIE* **10698**, 106982W (2018).
48. R. A. Brown, “Single-visit photometric and obscurational completeness,” *Astrophys. J.* **624**, 1010–1024 (2005).
49. O. Guyon et al., “High contrast imaging at the photon noise limit with self-calibrating WFS/C systems,” *Proc. SPIE* **11823**, 1182318 (2021).
50. R. Soummer et al., “Orbital motion of HR 8799 b, c, d using Hubble Space Telescope Data from 1998: constraints on inclination, eccentricity, and stability,” *Astrophys. J.* **741**, 55 (2011).
51. R. Fergus et al., “S4: a spatial-spectral model for speckle suppression,” *Astrophys. J.* **794**, 161 (2014).
52. P. Morrissey et al., “Flight photon counting electron multiplying charge coupled device development for the Roman Space Telescope coronagraph instrument,” *J. Astron. Telesc. Instrum. Syst.* **9**(1), 016003 (2023).
53. C. Bebek et al., “Ccd development for the dark energy spectroscopic instrument,” *J. Instrum.* **10**, C05026 (2015).
54. L. Barak et al., “SENSEI: characterization of single-electron events using a Skipper charge-coupled device,” *Phys. Rev. Appl.* **17**, 014022 (2022).
55. J. Tiffenberg et al., “Single-electron and single-photon sensitivity with a silicon Skipper CCD,” *Phys. Rev. Lett.* **119**, 131802–131806 (2017).
56. A. M. Botti et al., “Fast single-quantum measurement with a multi-amplifier sensing charge-coupled device,” arXiv:2308.09822 (2023).
57. B. J. Rauscher et al., “Detectors and cooling technology for direct spectroscopic biosignature characterization,” *J. Astron. Telesc. Instrum. Syst.* **2**, 041212 (2016).
58. J. E. Sadleir, “Ultra-high efficiency noiseless quantum sensors for HWO and QIS,” NASA Goddard Space Flight Center, <https://techport.nasa.gov/view/146757> (accessed 17 July 2024).
59. P. J. de Visser et al., “Phonon-trapping-enhanced energy resolution in superconducting single-photon detectors,” *Phys. Rev. Appl.* **16**, 034051 (2021).
60. R. Belikov et al., “Coronagraph design survey for future exoplanet direct imaging space missions: interim update,” *Proc. SPIE* **12680**, 126802G (2023).

Biographies of the authors are not available.