ANNUAL REVIEWS

Annual Review of Astronomy and Astrophysics

High-Contrast Coronagraphy

Matthew A. Kenworthy[1] and Sebastiaan Y. Haffert[1,2]

[1] Leiden Observatory, Leiden University, Leiden, The Netherlands;
email: kenworthy@strw.leidenuniv.nl

[2] Steward Observatory, University of Arizona, Tucson, Arizona, USA

Annu. Rev. Astron. Astrophys. 2025. 63:179–216

First published as a Review in Advance on
May 19, 2025

The *Annual Review of Astronomy and Astrophysics* is
online at astro.annualreviews.org

https://doi.org/10.1146/annurev-astro-021225-022840

Keywords

optics, coronagraphs, exoplanets, high contrast, computational methods

Abstract

Imaging terrestrial exoplanets around nearby stars is a formidable technical challenge, requiring the development of coronagraphs to suppress the stellar halo of diffracted light at the location of the planet. In this review, we discuss the science requirements for high-contrast imaging, present an overview of diffraction theory and the Lyot coronagraph, and define the parameters used in our optimization. We discuss the working principles of coronagraphs both in the laboratory and on-sky with current high-contrast instruments, and we describe the required algorithms and processes necessary for terrestrial planet imaging with extremely large telescopes and proposed space telescope missions:

- Imaging terrestrial planets around nearby stars is possible with a combination of coronagraphs and active wavefront control using feedback from wavefront sensors.
- Ground-based 8–40 m class telescopes can target the habitable zone around nearby M-dwarf stars with contrasts of 10^{-7}, and space telescopes can search around solar-type stars with contrasts of 10^{-10}.
- Focal plane wavefront sensing, hybrid coronagraph designs, and multiple closed loops providing active correction are required to reach the highest sensitivities.
- Polarization effects need to be mitigated in order to reach 10^{-10} contrasts while keeping exoplanet yields as high as possible.
- Recent technological developments, including photonics and microwave kinetic inductance detectors, will be folded into high-contrast instruments.

Contents

1. INTRODUCTION

Initially developed to image the Sun's corona without the need for a solar eclipse (Lyot 1933), one of the most significant science applications of the latest coronagraphs involves the detection and characterization of circumstellar material and planets around nearby stars. Young self-luminous gas giant exoplanets have been directly imaged at IR wavelengths around young stars (for a review of these detections, see Zurlo 2024), both in the nearby Galactic field and further away in young stellar OB associations (out to 400 pc). Typically these exoplanets have luminosities of 10^{-4}–10^{-6} of their parent star at angular separations up to a few arcseconds.

The search for life beyond Earth has focused on the idea that water is an essential part of life cycles elsewhere in the Universe, as it is a polar solvent formed from two elements that are abundant throughout the Galaxy. Places where water can exist in its liquid state, notably Earth-like planets and ice moons with a liquid-water ocean under an ice layer, are prime locations for these searches. The region around a star where liquid water can exist on the surface of a terrestrial planet (with appropriate atmospheric pressure) is referred to as the habitable zone (HZ). For the Sun, the HZ ranges from 0.9 to 1.2 au, but this range can shift as the luminosity of stars evolves over time.

At visible light wavelengths, the flux from exoplanets that can be directly imaged is overwhelmingly dominated by the reflected light from their parent star. For an Earth analog with a similar radius, albedo, and effective temperature orbiting around a solar-type star 10 pc away, the typical amount of reflected light in the optical range is 10^{-10} of the central star at a separation of 0.1 arcsec at maximum elongation. The technical challenge is to distinguish the light of the parent star from the light of the planet. Planetary systems that are closer to the Sun have two benefits. First, one that is twice as close as the other will have double the angular separation between the star and planet, and second, according to the inverse square law, four times more flux will be received from the planet. For direct imaging, therefore, the stars closest to the Sun are the ones that are studied for the presence of directly imaged exoplanets. On average, there are more M-dwarf stars close to the Sun than other solar-type stars: In a volume-limited (20 pc) sample around the Sun, there are ~140 solar G-type stars and on the order of 1,900 M-dwarf stars (Kirkpatrick et al. 2024). For solar-type stars, the contrast required in the optical wavelengths is on the order of 10^{-10} for terrestrial planets in the HZ, but for lower-mass stars with lower luminosities, the contrast is 10^{-7} for M dwarfs. The contrast for stars around the Sun is shown in **Figure 1**.

The markers for biosignatures also set the parameters (e.g., wavelength and bandwidth) that form part of the design decision. Spectra of earthshine (sunlight that is reflected off the Earth and illuminates the dark portion of the Moon) from 0.5 to 2.4 μm show O_2, O_3, CO_2, CH_4, and H_2O (Turnbull et al. 2006), and the literature offers many discussions about the possible biomarker molecules that should be searched for and characterized as evidence for biosignatures (for reviews, see Seager et al. 2016, Kaltenegger 2017, Schwieterman et al. 2018). It is clear that the unambiguous detection of life will require not a single detection of a molecule but rather the combination of several different lines of evidence. We focus on the technical challenges for imaging a rocky terrestrial planet around a nearby star in our Galaxy.

2. FROM MAXWELL'S EQUATIONS TO WAVE FRONTS

The vast majority of energy from astrophysical objects arrives at our telescopes in the form of electromagnetic radiation. This time-dependent interaction is described by Maxwell's equations (with relevant physical quantities defined in **Table 1** and following Lavrinenko et al. 2014):

$$\begin{aligned} \frac{\partial \mathcal{D}}{\partial t} &= \nabla \times \mathcal{H} && \text{(Faraday's law)}, \\ \frac{\partial \mathcal{B}}{\partial t} &= -\nabla \times \mathcal{E} && \text{(Ampère's law)}, \\ \nabla \cdot \mathcal{B} &= 0 && \text{(Gauss's law)}, \\ \nabla \cdot \mathcal{D} &= 0 && \text{(Coulomb's law)}. \end{aligned} \qquad 1.$$

Here, Maxwell's equations are in the material form without any charge and current sources. The material form is used to describe the propagation of electromagnetic fields inside matter. This set of equations is completed by describing the particular matter of the medium with the constitutive

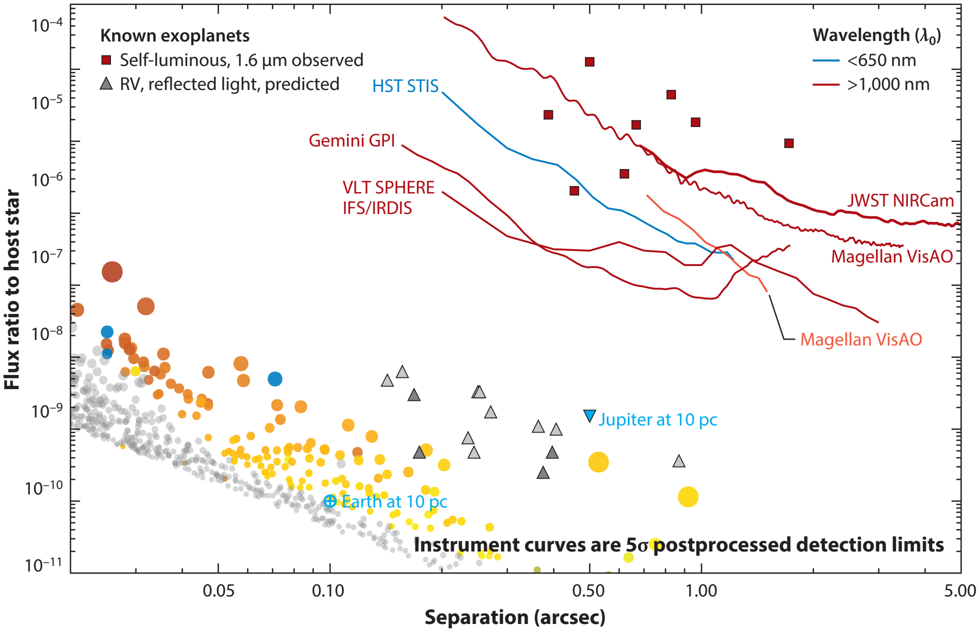


Figure 1

Angular separation versus contrast for directly imaged exoplanets. Known self-luminous gas giant planets are shown on the top right (Lacy & Burrows 2020). The lines represent 5σ point source contrast limits from recent instruments and surveys. Gray triangles represent the expected contrasts for reflected light from known radial velocity (RV)-detected exoplanets (Batalha et al. 2018). Circles represent reflected light estimates for planets with 1 Earth radius, assuming one planet per star within 20 pc of the Sun in the habitable zone at maximum elongation. Gray points represent planets that are fainter than the thirtieth magnitude in V. The color represents the effective temperature of the star. The larger the point is, the closer the stellar system is to the Sun. Reflected light data are from Guyon (2023). Figure adapted from Bailey & Hildebrandt Rafels (2024) (Apache 2.0); see associated references for the contrast curves.

Table 1 Definitions of physical quantities

Unit	Definition	Unit	Definition
i	Imaginary unit	c	Speed of light in vacuum
$\mathcal{E}$	Electric field	$\mathcal{H}$	Magnetic field
$\mathcal{D}$	Electric displacement	$\mathcal{B}$	Magnetic induction
ϵ	Relative electric permittivity	ϵ_0	Electric permittivity of vacuum
μ	Relative magnetic permeability	μ_0	Magnetic permeability of vacuum
ω	Angular frequency	$\lvert\vec{k}\rvert$	Wave number
λ	Wavelength	λ_0	Central wavelength in the bandpass
Π	Telescope pupil function	$\Delta\lambda$	Width of the bandpass
D	Diameter of the pupil	n	(Complex) refractive index of macroscopic media
$\mathcal{F}_{x,y}[\cdot]$	Fourier transform operator	$\mathcal{F}^{-1}_{x,y}[\cdot]$	Inverse Fourier transform operator
$C_\lambda[\cdot]$	General coronagraph propagation operator	$\Psi_\lambda[\vec{k}]$	Coronagraphic image
ϕ	Phase of the electric field	α	Amplitude of the electric field

relations $\mathcal{D} = \epsilon\mathcal{E}$ and $\mathcal{B} = \mu\mathcal{H}$, where ϵ and μ are the permittivity and magnetic permeability, respectively. The wave equation for electromagnetic waves can be derived by taking the curl of Ampère's law. Doing so yields the classic wave equation if we assume that the electromagnetic wave propagates in isotropic and homogeneous materials:

$$\mu\epsilon\frac{\partial^2\mathcal{E}}{\partial t^2} - \nabla^2\mathcal{E} = 0. \qquad 2.$$

We consider a purely monochromatic electromagnetic wave with angular frequency ω. The wave function is then $\mathcal{E} = \psi(r)e^{iwt}$, where $\psi(r)$ describes the spatial distribution. Substituting this relation into Equation 2, we find

$$\nabla^2\mathcal{E} + \mu\epsilon\omega^2\mathcal{E} = 0. \qquad 3.$$

The usual definitions of permittivity and permeability are $\epsilon = \epsilon_\mathrm{r}\epsilon_0$ and $\mu = \mu_\mathrm{r}\mu_0$, where ϵ_r and μ_r are the relative permittivity and permeability compared with those of vacuum, ϵ_0 and μ_0. In optics, most glasses and materials are defined by their refractive index, n, which is related to permittivity as $\epsilon_\mathrm{r} = n^2$. Many of them are also nonmagnetic, which means that $\mu_\mathrm{r} = 1$. Substituting this relation into Equation 3 yields the classic Helmholtz equation:

$$\nabla^2\mathcal{E} + n^2k^2\mathcal{E} = 0. \qquad 4.$$

Here, we have made use of the facts that the speed of light is $c = 1/\sqrt{\epsilon_0\mu_0}$ and that $k = c\omega$ is the wave number. Propagation through an optical system has a preferential direction that is usually defined along the z axis. The z-axis evolution can be derived by separating the spatial components into the z component and the perpendicular components (x and y):

$$\frac{\partial^2\mathcal{E}}{\partial z^2} = -\nabla_\perp^2\mathcal{E} - n^2k^2\mathcal{E}. \qquad 5.$$

This differential equation can be solved by assuming a plane-wave expansion, $\mathcal{E}(x, y, z) = e^{i(k_x x+k_y y)}f(z)$, which results in

$$\frac{\partial^2\mathcal{E}}{\partial z^2} = -\left(n^2k^2 - k_\perp^2\right)\mathcal{E}. \qquad 6.$$

The solution to this equation is the so-called angular spectrum propagator, which relates the electric field at any one plane to the electric field at any other:

$$\mathcal{E}(x, y, z') = \mathcal{F}^{-1}_{x,y}\left\{e^{-ik_z(z'-z)}\mathcal{F}_{x,y}[\mathcal{E}(x, y, z)]\right\}. \qquad 7.$$

Here, the z component of the wave vector is defined as $k_z = \sqrt{n^2k^2-k_\perp^2}$, and $\mathcal{F}^{(-1)}_{x,y}$ is the (inverse) Fourier transform over the x and y coordinates. While Equation 7 describes the full propagation from one plane to another, it is quite unwieldy to use and does not provide much physical insight. For many optical systems, it is sufficient to analyze the paraxial performance. The paraxial approximation assumes that the plane waves make small angles with respect to the z axis, which means that the x and y wave-vector components are k_x, $k_y \ll 1$. In this regime, the propagation factor $k_z = \sqrt{n^2k^2-k_\perp^2} \approx nk-(k_\perp^2/2nk)$, which simplifies to

$$\mathcal{E}(x, y, z') = e^{-ink(z'-z)}\mathcal{F}^{-1}_{x,y}\left\{e^{-i\frac{k_\perp^2}{2nk}}\mathcal{F}_{x,y}[\mathcal{E}(x, y, z)]\right\}. \qquad 8.$$

The real-space solution that is derived from this propagation equation is the classical Fresnel diffraction equation:

$$\mathcal{E}(x, y, z') = \frac{e^{-ink(z'-z)}}{\lambda z}\iint_{-\infty}^{\infty}\mathcal{E}(u, v, z)e^{\frac{ik\left[(x-u)^2+(y-v)^2\right]}{2(z'-z)}}\,\mathrm{d}u\mathrm{d}v. \qquad 9.$$

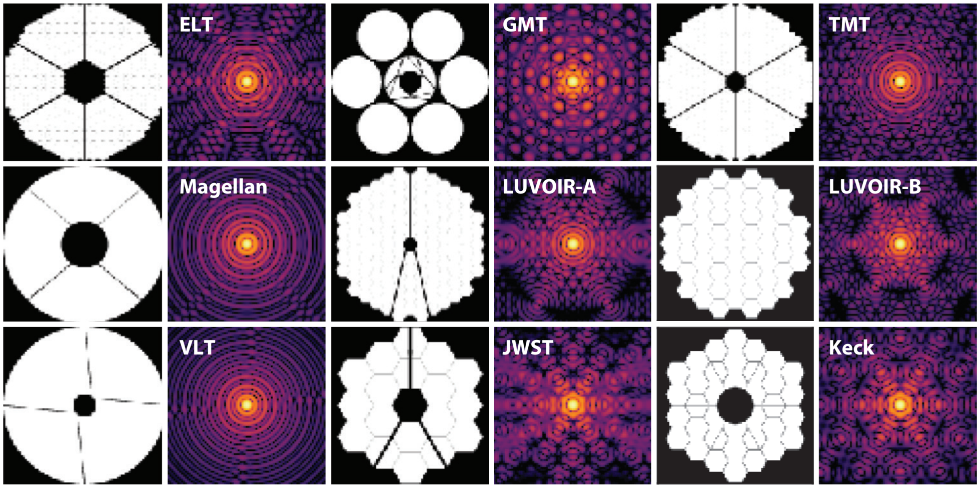


Figure 2

Telescope pupils and their point spread functions (PSFs) for several ground- and space-based telescopes. The PSFs have been normalized to the peak intensity.

The far-field approximation ($z \gg x, y$) and the substitutions $\theta_x = x/\lambda z$ and $\theta_y = y/\lambda z$ result in the Fraunhofer diffraction integral:

$$\mathcal{E}(\theta_x, \theta_y) \propto \iint_{-\infty}^{\infty} \mathcal{E}(u, v) e^{-i2\pi(\theta_x u + \theta_y v)} \mathrm{d}u \mathrm{d}v. \qquad 10.$$

This equation shows that the far-field distribution of an electromagnetic wave is its Fourier transform. A perfectly aligned lens, or the mirrors of a telescope, brings rays that come from infinity to the focal point. This is identical to a far-field transform; therefore, the operation of a lens can be described by the Fraunhofer diffraction integral.

The point spread function (PSF) is defined as the response function of an optical system to a point source at infinity. **Figure 2** shows the PSFs (defined as the square of the Fourier transform of the telescope pupil, Π) equal to $|\mathcal{F}_{x,y}[\Pi]|^2$ for several telescope pupils.

3. THE LYOT CORONAGRAPH

Stellar coronagraphs have now been in use for several decades. However, the first coronagraph was developed in the 1930s by Bernard Lyot (1939) to observe the corona of the Sun. It took more than half a century for astronomers to apply coronagraphs to image the faint circumstellar environment by blocking starlight.

The first coronagraph to successfully image a debris disk was a Lyot coronagraph, built by Vilas & Smith (1987), that imaged the edge-on circumstellar disk around Beta Pictoris in 1984 (Smith & Terrile 1984). **Figure 3** depicts the optical layout of the Lyot coronagraph; panels *a*–*f* represent the images present at that location in the coronagraph light path. The telescope pupil (**Figure 3*a***) is reimaged into a focal plane of the sky (**Figure 3*b***), where a focal plane mask (FPM) that has high absorptivity and low reflectivity blocks the light from any on-axis source (**Figure 3*c***). Optics then form an image of the resultant pupil (**Figure 3*d***) to an intermediate pupil plane, where a pupil plane mask (PPM)—the Lyot stop—is located. The diffraction of starlight around the FPM results in a ring of light around the diameter of the reimaged telescope pupil (**Figure 3*e***), and the Lyot stop blocks this ring of light. A second optical system then reimages this light onto the science camera focal plane (SCFP) (**Figure 3*f***). Any circumstellar objects outside the radius of

FPM: focal plane mask

PPM: pupil plane mask

SCFP: science camera focal plane

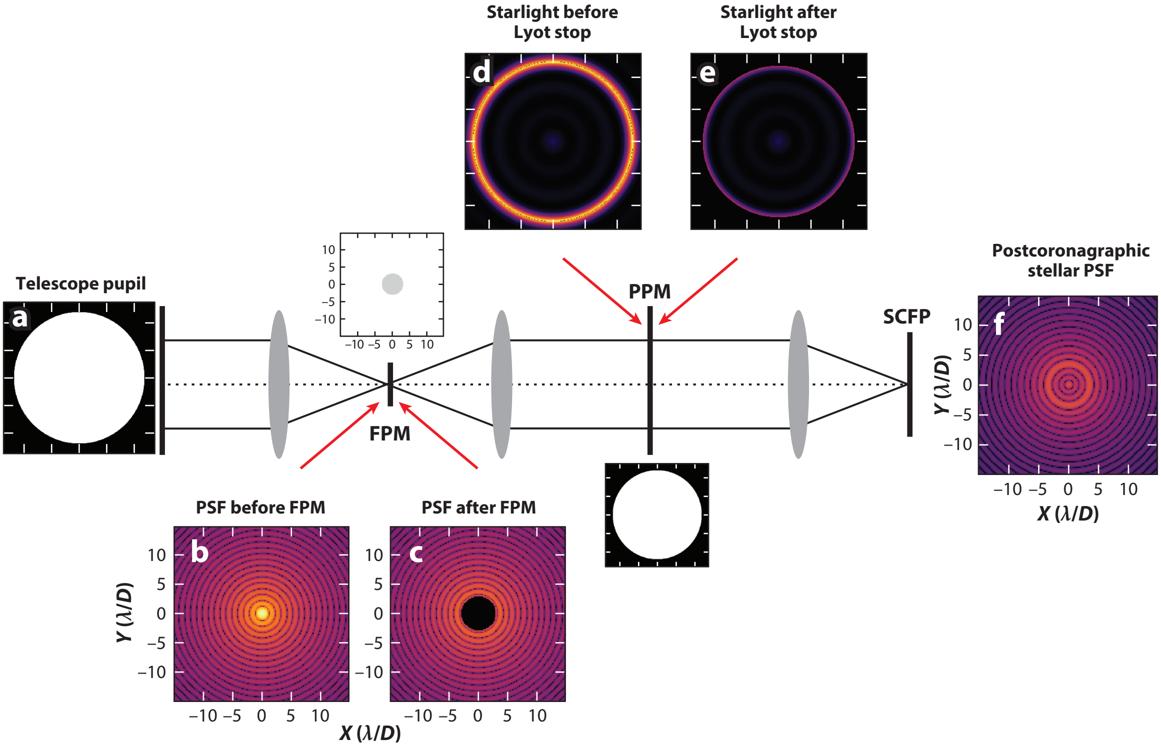


Figure 3

The Lyot coronagraph. A telescope with a circular unobstructed pupil (*a*) points at a star along the optical axis of the telescope and instrument. The telescope pupil is reimaged through the instrument optics to a focal plane (*b*), showing an Airy core surrounded by diffraction rings. A focal plane mask (FPM) then blocks the central region of the stellar image (*c*). Another set of optics reimages *c* to the pupil plane in panel *d*. The removal of the Airy core at the FPM redistributes light into a broad ring around the edge of the reimaged telescope pupil. The Lyot stop [more generally referred to as a pupil plane mask (PPM)] blocks this starlight (*e*), at a cost of reduced off-axis throughput and decreased angular resolution. A final set of relay optics then forms the science camera focal plane (SCFP) (*f*). Abbreviation: PSF, point spread function.

the FPM then pass unimpeded through the coronagraph and are subsequently reimaged in the SCFP. The light rays pass through the coronagraph optics to form an image (**Figure 3*f***) with only a minor modification: The reimaged pupil in **Figure 3*d*** is superficially very similar to the telescope pupil in **Figure 3*a***.

For an on-axis source, the removal of the Airy core plus attendant diffraction rings significantly modifies the wave front passing through the coronagraph. The result is the flux redistribution in **Figure 3*d***, where the flux is concentrated in a ring whose peak brightness lies along the perimeter of the reimaged telescope pupil, extending both beyond the radius of the pupil and into the center of the pupil. The purpose of the Lyot stop is to remove as much of this ring of light as possible while maximizing the throughput of the pupil image in **Figure 3*d*** for off-axis sources. Decreasing the diameter of the FPM changes the full width half maximum (FWHM) of the ring of light in **Figure 3*d***, which requires a smaller Lyot stop to block—but the throughput of the pupil for off-axis sources then decreases. Decreasing the Lyot stop aperture has another effect: The reduced pupil diameter increases the FWHM of the images in the final focal plane (**Figure 3*f***), spreading the flux from the off-axis sources over a larger area in the detector and degrading the angular resolution of the telescope and instrument. The optimal diameters of the FPM and Lyot stop

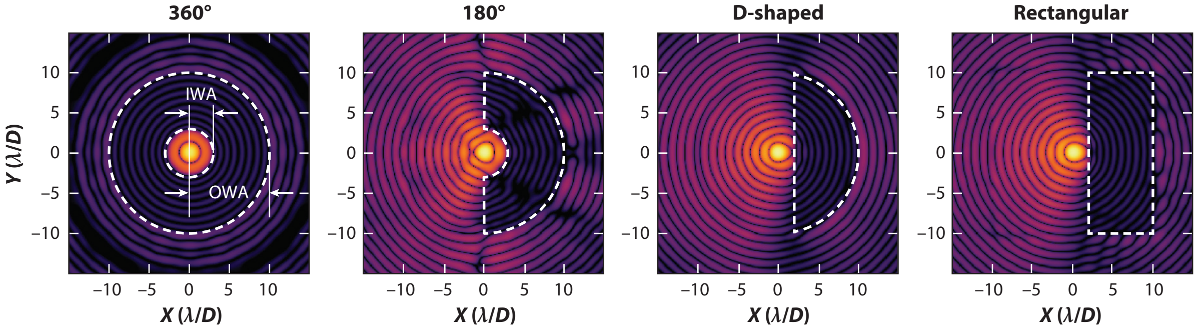


Figure 4

Geometries of dark holes commonly used in high-contrast imaging.

aperture are then determined by the science requirements—how close to the central star (measured in diffraction widths at the lower spatial resolution) the coronagraph should be in order to transmit light from off-axis objects in the field of view.

4. PARAMETERS TO OPTIMIZE

Clear apertures (i.e., with no secondary obscurations that block light from a simple circular aperture) can have solutions that perfectly remove any on-axis light. However, real systems are not ideal and generally do not have a clear aperture. Even more importantly, stars are not ideal point sources. Solutions that work only for point sources, therefore, are out of the question. Current and future generations of coronagraphs are designed to take on nonideal environments, meaning that they are optimized for a specific list of parameters during the design process.

The first set of parameters consists of the inner working angle (IWA) and the outer working angle (OWA). The IWA is the angular separation where the throughput is 50% of the peak off-axis throughput, and the OWA is set by the design and optical properties of the coronagraph. The IWA and OWA set the smallest and largest angular separation where the coronagraph will suppress the stellar halo, producing a dark hole region in which to image planets and/or circumstellar material. Some coronagraphs, such as the optical vortex coronagraph (OVC) or four-quadrant phase mask (FQPM), have only an IWA. **Figure 4** shows the different types of coronagraphic dark hole geometries.

IWA: inner working angle

OWA: outer working angle

OVC: optical vortex coronagraph

FQPM: four-quadrant phase mask

The contrast and throughput constitute the next set of parameters that we need to optimize for. The contrast sets the amount of starlight that is left after the coronagraph. It is important to define the term contrast, as it can mean many different things. Ruane et al. (2018) provide a thorough overview of the different metrics and their definitions. They define the contrast as

$$C = \frac{\eta_*(\vec{r})}{\eta_\mathrm{p}(\vec{r})}, \qquad 11.$$

where $\eta_*(\vec{r})$ is the fractional throughput of the star at focal plane position $\vec{r}$ integrated over a photometric aperture. This value is then divided by $\eta_\mathrm{p}(\vec{r})$, the fractional throughput of the planet in the same photometric aperture (**Figure 5**). This procedure normalizes the contrast with regard to the throughput of the planet, which is important because the planet throughput usually varies as function of angular separation. Both the contrast C and $\eta_\mathrm{p}(\vec{r})$ need to be included in the optimization process, the first to make sure that the starlight is nulled and the second to make sure that the planet light is maintained. This optimization must be done over a certain spectral bandwidth, $\Delta\lambda$.

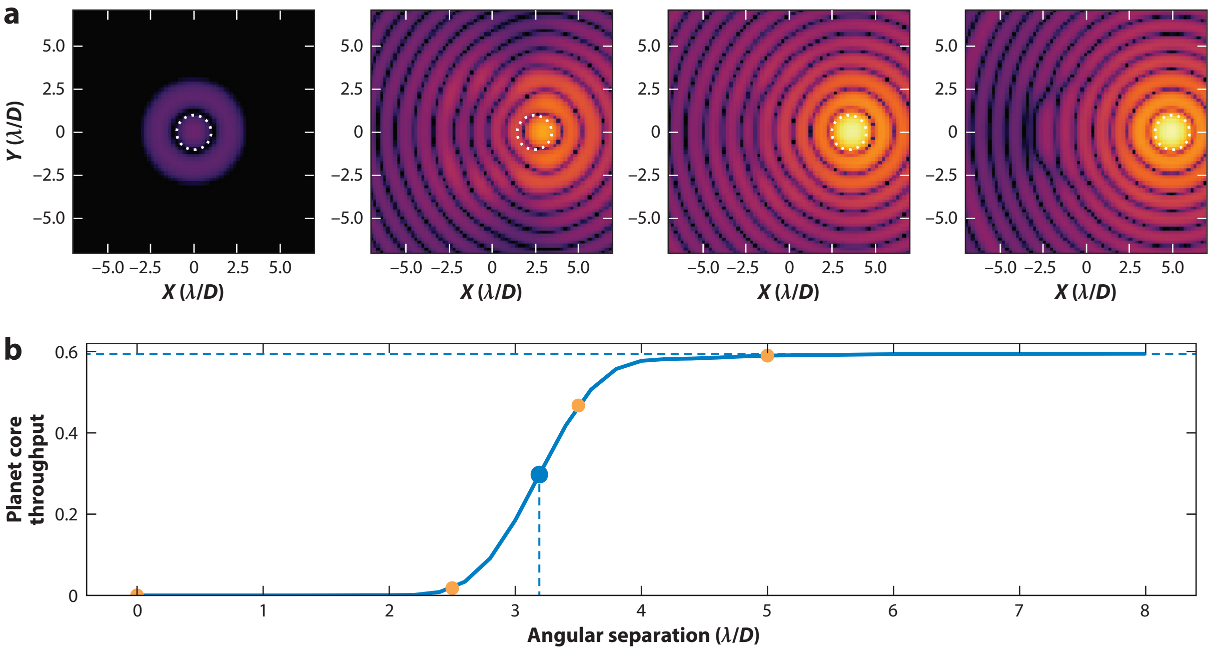


Figure 5

Definition of planet throughput, showing the integrated flux over a 2 λ/*D* aperture with increasing distance from the star. (*a*) The four postcoronagraphic images correspond to the four orange points on the throughput curve. (*b*) The planet's postcoronagraphic image is strongly warped at/close to the edge of the focal plane mask. The dotted line represents the maximum throughput at an infinite off-axis angle. The vertical dashed line and blue dot show the inner working angle at 50% throughput of the maximum throughput. The simulations presented here correspond to the apodized Lyot coronagraph described in Section 6.

Coronagraphs that are designed with only the previous set of optimization targets are not optimal in real environments. Real instrument environments contain wavefront aberrations and small instrumental drifts. These cause light to leak around the coronagraph and generate residual stellar speckles, which are λ/*D*-sized spots that form granular patterns arising from diffraction and interference effects when coherent beams undergo random scattering. In this case, the speckles originate from residual turbulence or noncommon path aberrations (NCPAs). The coronagraphs must be made robust against low-order wavefront errors and other instrumental drifts. Another, more practical thing to consider is the precision with which we can align an instrument. For example, how well can the Lyot stop be aligned? The performance of a coronagraph might be extremely sensitive to the Lyot stop alignment, which means that theoretically the coronagraph delivers the contrast but practically it will never reach it. Therefore, alignment tolerancing must be included in the coronagraph design to make sure the target performance is achieved.

Many other nuisance parameters can be included. However, numerical optimization will take significantly longer with a greater number of parameters. A good coronagraph designer will therefore make a trade-off between the parameters to include: either good to have or not significant.

5. BEYOND LYOT WITH COMPLEX PUPIL AND FOCAL PLANE MASKS

The Lyot coronagraph was designed at a time when it was still difficult to precisely manipulate the phase of light with optics. The advent of more advanced manufacturing capabilities enabled very precise phase control, a significant boost to the design toolbox for coronagraphs. The major downside of the classic Lyot coronagraph is that only the light that falls on the opaque FPM gets blocked. Any light that is off-axis will pass through the system. That is true not only for light that comes from off-axis sources but also for aberrations that cause light to end up outside the FPM. A bigger mask can block a larger fraction of the light and therefore achieve a deeper contrast. However, if the mask is made larger, then the IWA also becomes bigger, so fewer planets will be accessible. A smaller IWA is crucial for both extremely large ground-based telescopes and space-based telescopes. We note that the coronagraphic designs described below are not optimized but are shown in **Figures 6** and **7** for illustrative purposes.

AGPM: annular groove phase mask

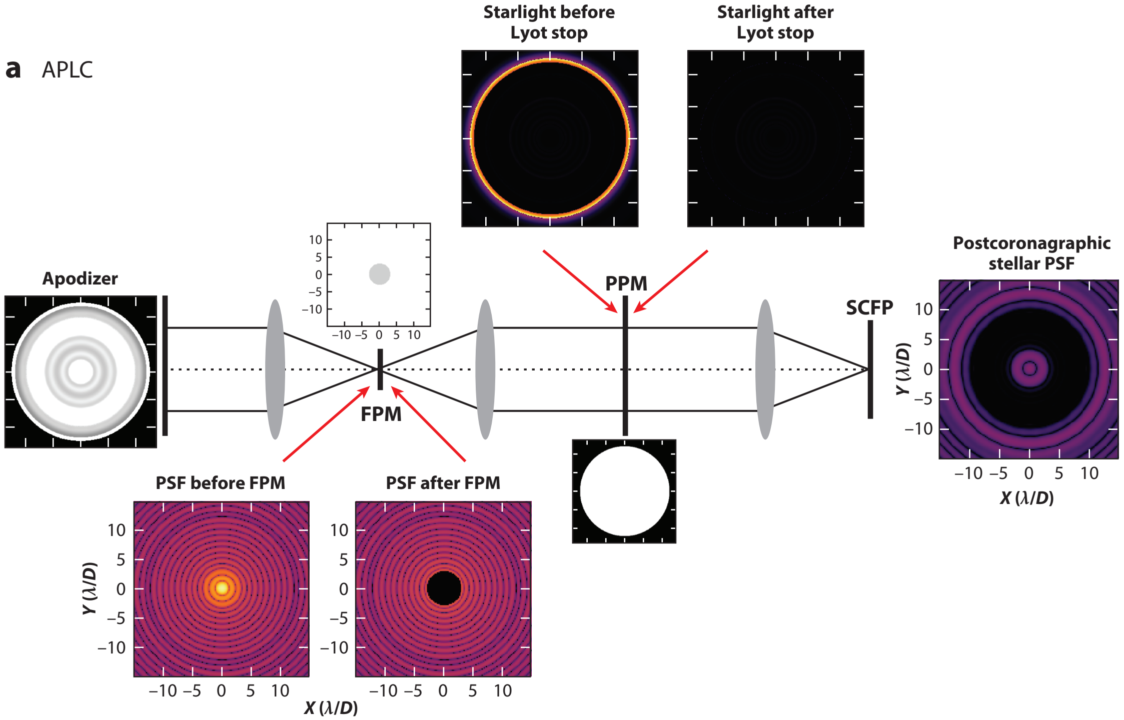
5.1. Focal Plane Phase Mask Coronagraphs

Focal plane phase masks offer a solution by phase-shifting a part of the PSF (usually the core), which leads to destructive interference at the Lyot plane. The Lyot stop blocks the areas where light does not destructively interfere. The 50% encircled energy radius is on the order of $\sim\lambda/D$ for almost all aperture shapes. As a result, perfect destructive interference can be achieved with a mask that has a size on the order of λ/D. This is the central idea that was used to design the Roddier & Roddier (1997) phase mask. This mask covers the core of the Airy pattern that contains 50% of the encircled energy and phase-shifts the core by π to cause destructive interference. It works well for monochromatic light, but over broad spectral bandwidths it degrades in contrast. Diffraction causes wavelength scaling of the PSF, making the FPM either too big or too small compared with the PSF. Masks made out of multiple concentric rings, such as the dual-zone phase mask (Soummer et al. 2003b), have been proposed to increase the spectral bandwidth.

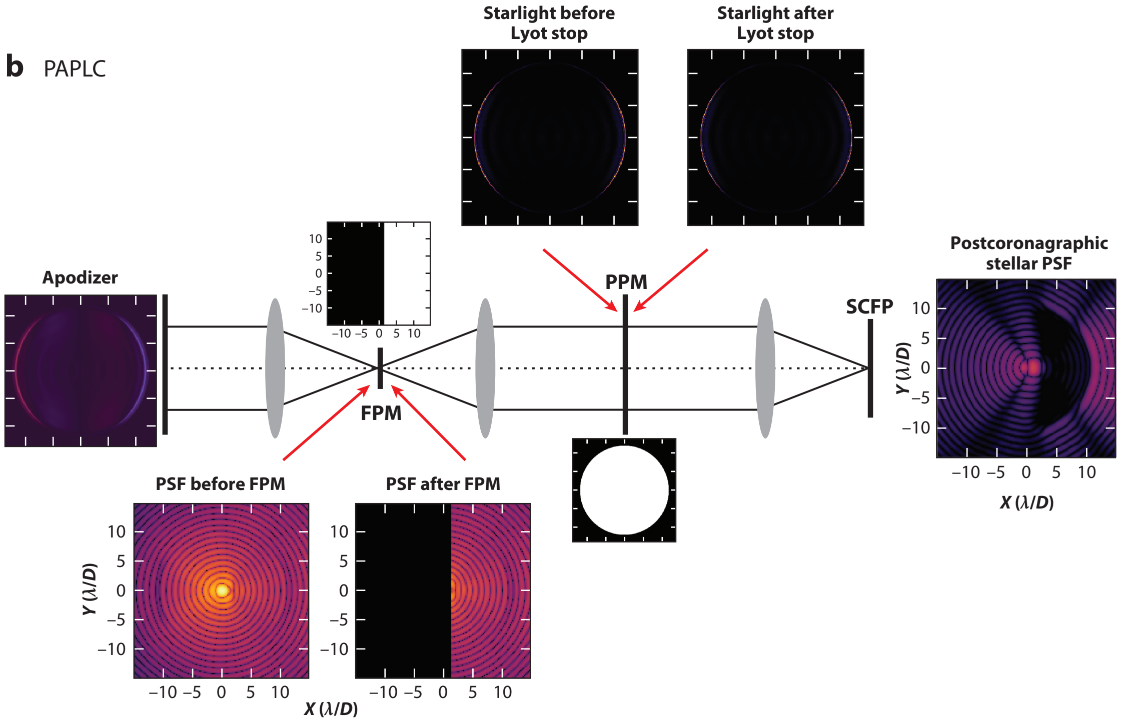
Further achromatization for larger spectral bandwidths is possible through the design of inherently achromatic phase masks. Inherent achromatic masks are scale invariant, which means that they always look the same regardless of the size of the PSF. The FQPM splits the focal plane into four quadrants and applies a checkerboard $[0, \pi]$ phase pattern (Rouan et al. 2000). The FQPM was implemented on the Spectro-Polarimetric High-contrast Exoplanet REsearch (SPHERE) instrument (Boccaletti et al. 2004). It is currently the coronagraph on the *James Webb Space Telescope* (JWST) with the smallest IWA and has characterized a planet at 1.8 λ/D (Franson et al. 2024).

If the planet falls on one of the four transition lines between the phases, its transmission is significantly reduced. Increasing the number of phase steps to the continuous limit results in a phase ramp about the optical axis, namely the OVC. The OVC uses a phase mask with a vortex pattern where the phase changes with the azimuthal angle: $\phi = q \cdot \theta$, where q is the charge of the vortex (the number of times the phase wraps around) and θ is the azimuth angle. One implementation of an OVC is the annular groove phase mask (AGPM) (Mawet et al. 2005). The AGPM uses subwavelength gratings to impart a charge-2 phase ramp in an FPM. The grating manufacture limits the AGPM to a charge-2 vortex, but higher charges are preferred in order to make a better match to the small but finite diameter of stellar disks for nearby stars. A more general challenge is that the OVC diffracts all the light out of the pupil plane only if the telescope pupil is unobstructed. Secondary support structures and a centrally obscured pupil scatter significant amounts of light back into the pupil plane. A suitable Lyot stop can block this stellar leakage, but at the cost of planet throughput. The impact of the central obscuration can be mitigated with the use of a telescope pupil gray apodizer to make a ring-apodized vortex coronagraph (RAVC) (Mawet et al. 2013).

Both phase masks completely null out all on-axis light from a clear aperture. This property is why both the FQPM and the OVC have been extensively studied over the past several decades. The IWA of both coronagraphs approaches 1 λ/D, which makes coronagraphy possible at the diffraction limit.



a APLC



b PAPLC

(*Caption for* ***Figure 6*** *appears on following page*)

Figure 6 (*Figure appears on preceding page*)

Layouts of the apodized phase Lyot coronagraph (APLC) and phase-apodized pupil Lyot coronagraph (PAPLC), showing the pupil, focal plane images, and masks. Abbreviations: FPM, focal plane mask; PPM, pupil plane mask; PSF, point spread function; SCFP, science camera focal plane.

5.2. Pupil Plane Mask Coronagraphs

AO: adaptive optics

All stars have a small but finite angular diameter on the sky, typically λ/100 or less but up to λ/10 for the closest stars. In the visible range, median stars for the Habitable Worlds Observatory (HWO) will be λ/10, with some a little larger than that; their numbers will increase significantly with the next generation of 30 m class telescopes. Proxima Centauri has a diameter of 1 mas, which is 1/3 to 1/6 λ/D for the extremely large telescopes (ELTs) and the Giant Magellan Telescope (GMT). The disk of the star can be treated as a set of incoherent point sources, so the small but finite angular size of the star means that the sensitivity of contrast of focal plane coronagraphs varies as a function of angle. Given that the star is tens of thousands to millions of times brighter than the target planet, even a small amount of stellar leakage can overwhelm the flux from the planet.

Apodizing the telescope pupil provides an opportunity to redistribute the light in the SCFP to form "dark zones" around the target star where an exoplanet can be imaged. Strictly speaking, such an instrument is not so much a coronagraph as a modification to the PSF of the instrument—all objects in the focal plane, including stars and exoplanets, have the same PSF. As long as the angular diameter of the star is smaller than the Airy core of the PSF, pupil plane coronagraphs are not affected by the diameter of the star or by residual tip/tilt vibrations that are not removed by the adaptive optics (AO) control loop, making them a robust alternative to the more efficient, smaller IWA focal plane coronagraphs. Suppressing diffraction in the PSF requires destructive interference using coherent light from the Airy core, but it decreases the encircled energy of the

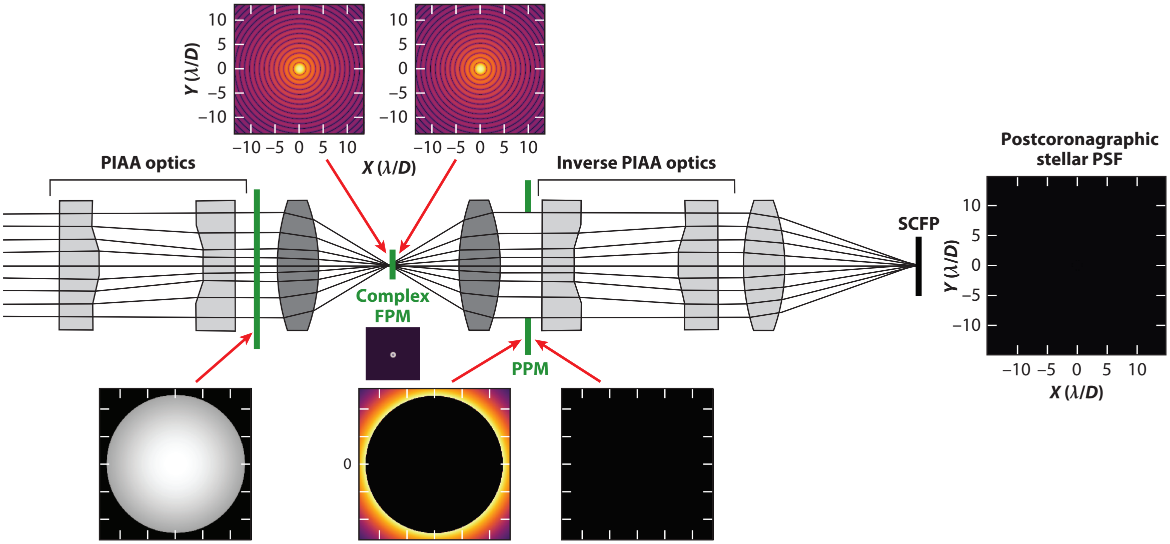


Figure 7

Principle of the phase-induced amplitude apodization (PIAA) complex mask coronagraph (PIACMC). Abbreviations: FPM, focal plane mask; PPM, pupil plane mask; PSF, point spread function; SCFP, science camera focal plane.

core of the PSF. Since the exoplanet PSF is identical to the stellar PSF, the planet throughput decreases, too.

SPP: shaped pupil plate

APP: apodizing phase plate

The earliest pupil apodizations, referred to as shaped pupil masks or shaped pupil plates (SPPs), were with binary amplitude masks (Jacquinot & Roizen-Dossier 1964, Kasdin et al. 2005), but they had low throughputs, a significant increase in the FWHM of the resultant PSF (affecting the encircled energy of the planet and the IWA), and a narrow opening angle (<45°) of the dark zone. Improvements in searches for the high-dimensional space of possible solutions resulted in throughput improvements of 50%, working angles of 2.5 to 15 λ/D, and contrasts of 10^{-6} (Carlotti et al. 2011). Optimizations have also been generalized so that they can be designed for arbitrary telescope pupils, allowing the secondary obscuration, support structures, and even gaps between segmented mirrors to be accounted for and avoided. Along with their achromatic performance, these optimizations make pupil apodization suitable for ELT pupils and allow dynamic coronagraphs using micromachine mirrors (Leboulleux et al. 2022a, Carlotti et al. 2023) to account for a dynamically changing pupil (e.g., missing and swapped mirror segments).

Another approach is to apodize in phase only, with the PPM set by the geometry of the telescope pupil (Codona & Angel 2004). This approach was demonstrated on-sky with the apodizing phase plate (APP) (Kenworthy et al. 2007), which used variations in the thickness of a piece of diamond-turned zinc selenide to impart a phase shift across the pupil with a central wavelength of 4 µm. An APP built and installed in the NAOS/CONICA Very Large Telescope (VLT) camera (Kenworthy et al. 2010) led to the first coronagraphic image of Beta Pictoris b (Quanz et al. 2010) and the direct imaging discovery of the exoplanet HD 100546 b (Quanz et al. 2013). The original algorithms found solutions with 180° dark D-shaped regions next to the star, but a more general theory of APP optimization (Por 2017) finds both 180° and 360° solutions as well as solutions consisting of regions of integer multiples of $\pi/2$ rad.

The first APP optics were chromatic. The optical path difference (OPD) was a function of the refractive index of the transmissive material, with the suppression decreasing with increasing bandwidth. Achromatic phase shifts can be implemented using the principle of geometric phase. The vector APP (vAPP) (Snik et al. 2012) replaces the classical phase pattern, $\phi_c(\vec{r}) = n(\lambda)\Delta d(\vec{r})$, with the geometric phase (known as the Pancharatnam–Berry phase; Pancharatnam 1956, Berry 1984). The vAPP phase pattern is imposed by a half-wave retarder with a patterned fast axis orientation $\theta(\vec{r})$. The geometric phase is imprinted on incident beams decomposed according to circular polarization state: $\phi_g(\vec{r}) = \pm 2 \cdot \theta(\vec{r})$, where the sign depends on the circular polarization handedness. As this fast axis orientation pattern does not vary as a function of wavelength (with the possible exception of an inconsequential offset/piston term), the geometric phase is strictly achromatic.

vAPP devices are produced through the application of two liquid-crystal techniques. The desired phase pattern is applied onto a substrate glass through a direct-write procedure (Miskiewicz & Escuti 2014), which applies the orientation pattern $\theta(\vec{r})$ by locally polymerizing the alignment layer material in the direction set by the controllable polarization of a scanning UV laser. Consecutive layers of birefringent liquid crystal are deposited on top of this alignment layer, which subsequently self-aligns (multitwist retarders; Komanduri et al. 2013) with predetermined parameters (birefringence dispersion, thickness, nematic twist) to yield a linear retardance that is close to a half-wave over the specified wavelength range. Additional layers broaden the wavelength range to more than an octave, at a cost of an absorption feature due to the carbon–carbon bonds within the liquid-crystal and glue layers. vAPP devices require additional optics (typically a half-wave plate and a Wollaston prism) to isolate the two circular polarizations and produce two separate PSFs with dark holes on opposing sides of the central star (Snik et al. 2012).

When a phase diffraction grating is added to the APP phase pattern to make a grating vector APP (gvAPP) (Snik et al. 2012, Otten et al. 2014), the other optics are no longer required. The two coronagraphic PSFs are diffracted into the $m = \pm 1$ order, with an $m = 0$ "leakage term" noncoronagraphic PSF that has a flux of a few percent of the original star left in the undeviated beam, acting as both an astrometric and a photometric reference (Otten et al. 2017, Sutlieff et al. 2024). The grating effect means that the PSF centroids vary as a function of wavelength; therefore, gvAPPs are ideal for imaging onto integral field units and image slicers (Sutlieff et al. 2021, 2023). This liquid-crystal technology has enabled coronagraphic designs that were previously impossible to manufacture, including the coronagraphic modal wavefront sensors (WFSs) (Wilby et al. 2017), sparse aperture masking with multiple holograms (Doelman et al. 2021b), complex amplitude vector vortex coronagraphs (VVCs) (Snik et al. 2014), and triple-grating coronagraphs (Doelman et al. 2020) that redisperse the PSFs back into white-light coronagraphic PSFs for VVCs (Doelman et al. 2023, Laginja et al. 2024). A comprehensive review of the coronagraphs enabled by the liquid-crystal technology is presented by Doelman et al. (2021a).

gvAPP: grating vector apodizing phase plate

WFS: wavefront sensor

VVC: vector vortex coronagraph

6. APODIZED LYOT CORONAGRAPHS

The fundamental goal of a coronagraph is to block light from the star while allowing light from the planet to pass through. FPM coronagraphs, like the classic Lyot coronagraph, find that it is very difficult to completely null out the star as a result of a mismatch between the modes of the incoming electric field from the aperture and the FPM filtered electric field. Light through a classic Lyot coronagraph can be described using two separate consecutive propagations. The first is propagation through the area that is covered by the mask itself, m_1, and the second is the negative of the mask m_2. Together, these two masks cover the full focal plane. Propagation of the electric fields then follows as

$$\mathcal{E}_{\text{out}} = t\mathcal{F}^{-1}\{m_1\mathcal{F}\{\mathcal{E}_{\text{in}}\}\} + \mathcal{F}^{-1}\{m_2\mathcal{F}\{\mathcal{E}_{\text{in}}\}\}. \tag{12}$$

We can simplify this equation by substituting $\hat{m} = \mathcal{F}\{m_1\}$ and $m_2 = 1 - m_1$ and by using the Fourier convolution theorem, yielding

$$\mathcal{E}_{\text{out}} = \mathcal{E}_{\text{in}} + (t-1)\hat{m} * \mathcal{E}_{\text{in}}. \tag{13}$$

The output is the sum of two electric fields, the original input electric field and a filtered electric field. The filtered light is allowed to diffract light outside the geometric pupil because the light is blocked by the Lyot stop. Therefore, the condition for perfect nulling is obtained by setting the output over the geometric pupil to zero:

$$\mathcal{E}_{\text{in}} + (t-1)\hat{m} * \mathcal{E}_{\text{in}} = 0. \tag{14}$$

This condition can be fulfilled only if the incoming electric field is an eigenfunction of the filtering operator:

$$\hat{m} * \mathcal{E}_{\text{in}} = \gamma\mathcal{E}_{\text{in}}. \tag{15}$$

Then,

$$\mathcal{E}_{in} + (t-1)\gamma\mathcal{E}_{\text{in}} = 0, \tag{16}$$

and a perfect null is achieved when $(t-1)\gamma = -1$. The normal pupil illumination is uniform and is not an eigenfunction of the filtering operator. Therefore, it is not possible to perfectly null the starlight with any type of Lyot-style FPM. These eigenfunctions are prolate spheroids, and they generate theoretically perfect nulls if combined with Roddier & Roddier coronagraphs (Soummer et al. 2003a). Therefore, the incoming pupil amplitude must be apodized. This insight led to the development of various coronagraphs with different combinations of phase/amplitude pupil apodization and opaque/phase-shifting FPMs.

6.1. The Apodized Phase Lyot Coronagraph

The first modification of the Lyot coronagraph to be experimented with was the apodized phase Lyot coronagraph (APLC). The APLC uses achromatic gray-scale pupil apodizers to better match the entrance amplitude distribution to the FPM (**Figure 6**). These coronagraphs are quite robust and are used in many different high-contrast imaging (HCI) instruments, such as SPHERE (Beuzit et al. 2019). Research on optimization strategies for the APLC is ongoing, including for future segmented space telescopes (Zimmerman et al. 2016) and for upgrades of ground-based instruments (Nickson et al. 2022). The major downside of the APLC is the lower transmission due to absorption in the apodizer. A variation on the APLC will be used by the Coronagraph Instrument on the *Nancy Grace Roman Space Telescope* (Krist et al. 2023). This instrument uses a hybrid Lyot coronagraph (HLC) whose design process accounts for the phase effects of the FPM. The metallic coatings on the FPM substrates are not completely opaque. The mask will leak starlight at a certain level, causing residual speckles. For broad-spectral-bandwidth dark holes, the exact transmission and phase shift need to be taken into account for the pupil apodization mask (Kuchner & Traub 2002). The HLC will be the first coronagraph in space to have active wavefront control for dark hole digging (Krist et al. 2023).

6.2. The Phase-Apodized Pupil Lyot Coronagraph

Phase apodization is twice as effective as amplitude apodization at improving the performance of the Lyot coronagraph (Por 2020) because of the range of allowed apodization values. It is possible to use phase offsets to increase the range from [0, 1] to [−1, 1] and allow for complex value apodization. The solutions for circular dark holes are discretized; the phase is either zero or π (Por 2020). This is reminiscent of the solutions found for optimal APPs (Por 2017) and binary shaped pupils (Carlotti et al. 2011). Full circular dark holes are not very efficient because they require strong phase patterns to create the dark holes. The result is significant Strehl loss (Por 2017): One-sided dark holes are much more efficient in terms of Strehl.

The phase-apodized pupil Lyot coronagraph (PAPLC) is the baseline for the GMagAO-X (Males et al. 2024). It has been implemented in the Space Coronagraph Optical Bench in Arizona (Ashcraft et al. 2022, Van Gorkom et al. 2022).

6.3. The Phase-Induced Amplitude Apodization Complex Mask Coronagraph

Phase-induced amplitude apodization (PIAA) remaps the telescope pupil such that a star on the optical axis forms a PSF with no diffraction rings—typically a 2D Gaussian profile (Guyon 2003; Guyon et al. 2005, 2014). The pupil remapping optics can be either transmissive or reflective; reflecting optics are more amenable to achromatization but also more challenging to manufacture. The optics induce aberrations for off-axis sources that are strong functions of increasing distance from the optical axis, significantly decreasing the Strehl ratio of these sources and lowering their effective sensitivity. A reimaging system that reverses the optical aberrations of the first set of PIAA optics then reforms a final focal plane image where all off-axis sources form diffraction-limited images. An on-axis FPM then blocks the starlight while allowing off-axis sources to propagate through to the SCFP. The original PIAA coronagraph uses a hard-edged apodizer, but when the design includes other coronagraphs (e.g., amplitude-apodized Lyot or complex mask coronagraphs), it can approach the ideal coronagraph in terms of suppression (**Figure 7**).

PIAA was originally designed for unobscured circular apertures. Telescope pupils with secondary obscurations require reformatting of the pupil to make a continuous, diffraction-free PSF in the coronagraphic focal plane. The introduction of complex masks that can be manufactured to the required tolerances enabled the development of PIAA for complex and segmented telescope pupils, which are suitable for space-based telescope concepts such as HWO and LUVOIR. Results from a laboratory demonstration of a phase-induced amplitude apodization complex mask coronagraph (PIAACMC) with a segmented aperture (Marx et al. 2021) show contrasts of 3×10^{-8} from 4 to 9 λ/D with a 2% bandwidth. The recent laboratory demonstration of high contrast with a PIAACMC on an obstructed and segmented aperture (Belikov et al. 2022) shows 1.9×10^{-8} contrast achieved in a 10% bandwidth between 3.5 and 8 λ/D. Ultimately, the rejected light can form the basis for a WFS to keep the PIAA pointed and aligned with the science target, and a WFS and coronagraph integrated with a PIAACMC have been demonstrated (Haffert et al. 2023a).

7. SPATIAL MODE DEMULTIPLEXING

The principle of spatial mode demultiplexing (SMD) is to use wave guides as modal filters on the complex electric field to separate them into different optical pathways. An early example is the filtering of the input of a nulling interferometer. Two subapertures from a wave front are brought to focus with a π rad phase shift between them, forming a set of Young's fringes on the sky with an on-axis stellar source nulled out quadratically. The electric field of the fringes is antisymmetric, and the complex electric field changes sign across the central null. With the use of a single-mode fiber at the focus, which permits transmission of only the lowest electromagnetic mode of the circular aperture of the fiber (HE_{11}), the sign change of the (point-antisymmetric) electric field results in a deep null, and wavefront aberrations are minimized (Haguenauer & Serabyn 2006, Serabyn & Mennesson 2006). Any planet that is sitting in an adjacent transmissive region of the fringe pattern (and is on the face of the fiber) has a point-symmetric electric field and will couple into the fiber, albeit with reduced transmission.

This principle has been used to minimize the contribution of speckles in the focal plane at the location of the planet, where the Airy core of the planet is injected into a single-mode fiber (Mawet et al. 2017). High coupling efficiency is possible (with a theoretical maximum of 81% for an Airy core into the Gaussian HE_{11} mode; Shaklan & Roddier 1988), and reflected light around the fiber can be used to sense and minimize speckles while the incoherent light of the planet remains constant and injected into the fiber.

The OVC has a planet throughput of 0.7 at 1 λ/D of the on-axis null (Mawet et al. 2005). Placement of an optimally matched single-mode fiber at the location of the null provides a suppression of the star much greater than the transmission loss of the coronagraph for a planet at 0.5 to 1 λ/D from the star, resulting in a peak transmission of 0.2 at 0.9 λ/D. This principle is called vortex fiber nulling (VFN) (Ruane et al. 2018). Different telescope pupils with both pupil plane and focal plane OVC have been explored by Ruane et al. (2019), and on-sky results are presented by Echeverri et al. (2024).

VFN provides almost no information about the azimuthal position of the planet, and because it is off-axis it does not couple with the highest efficiency. One solution is to use a mode-selective photonic lantern (MSPL) (Leon-Saval et al. 2013). An MSPL replaces a single-mode fiber with a cluster of fibers in a close-packed configuration that optimally matches the focal plane with an OVC. The properties of the coupling increase the planet throughput and partial localization of the planet (Xin et al. 2022). VFN and MSPLs feed high-dispersion spectrographs to perform high-spectral-dispersion HCI (Snellen et al. 2015), where the speckle field changes slowly with wavelength and can be approximated as a constant background over limited wavelength ranges.

SMD can also suppress the diffraction halo of a star by using a single-mode fiber that has a diffraction null crossing the face of the fiber. The sign change in the electric field across the null means that a fiber centered on the null has a significantly reduced electric field propagated through the fiber, but the Airy core of the planet will couple with high efficiency. This technique works in

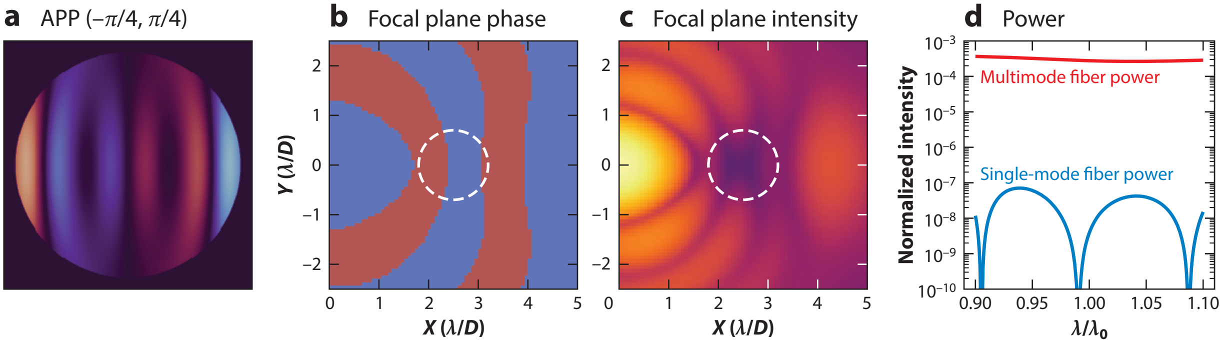


Figure 8

Demonstration of single-mode complex amplitude refinement. An apodizing phase plate (APP) modifies the distribution of the diffraction rings in the focal plane (*a*), moving two null crossing points to be within the same aperture (*white dashed circles*) (*b*,*c*). A multimode fiber propagates all the energy within the circle, but for a single-mode fiber the electric fields cancel out, resulting in a suppression of $\sim 10^{-4}$ over a broad bandwidth around a central wavelength of λ_0 (*d*).

the narrow band, but because the diffraction halo scales radially with λ/D, the null line no longer passes through the center of a fiber that is offset from the optical axis defined by the star. The effect can be made to work across a significantly wider bandwidth by having two successive nulls cross the fiber area: As the two nulls move out radially with wavelength, the two sign changes across each null compensate each other to first order (see **Figure 8**). An APP with modest apodization can be made to squeeze two null crossings closer to each other, and the use of a hexagonal lenslet array matched to the circular single-mode fibers enables high-efficiency transmission of planet flux (Haffert 2021) through to the spectrograph, but with a large bandwidth called single-mode complex amplitude refinement (SCAR) (Haffert et al. 2020, Por & Haffert 2020). This principle has been demonstrated in laboratory experiments with 360° and 180° dark regions from 0.8 to 2.4 λ/D around the star (Haffert et al. 2020). In these experiments, the 360° SCAR was designed for an unobscured telescope pupil and created a measured stellar null of $2\text{–}3 \times 10^{-4}$, and the 180° SCAR was designed for a telescope aperture with central obscuration and spiders and reached a null of 1×10^{-4}.

The field of quantum metrology uses quantum mechanics to enhance sensitivity and resolution in sensing devices. The tools from quantum metrology can be used to derive fundamental measurement limits through the determination of the quantum Cramér–Rao (QCR) bounds (Braunstein & Caves 1994). Limits on the variance of a measurement can be derived from the QCR bounds. These principles have been used to show that an optimal detection scheme can actually resolve incoherent point sources well within the classical Rayleigh diffraction limit (Tsang et al. 2016). SMD that sorts the incoming wave fronts into separate modal channels with photon-counting detectors saturates the QCR bound. Therefore, SMD is an optimal detection scheme for equal-brightness point sources.

Most quantum optimal imaging research focuses on equal-brightness point sources or low-contrast extended objects. Exoplanet imaging is a situation with an extreme contrast ratio. The optimal imaging limit could be different. Deshler et al. (2024a) derive the quantum limit for exoplanet imaging and find that SMD with the telescope's eigenmode basis is quantum optimal. However, they also find that both the PIACMC and OVC approach this fundamental limit for subdiffraction-limited separations within a factor of two. Beyond separations of 1 λ/D, the PIAACMC approaches the limit within 10%, making it nearly quantum optimal for resolved objects.

Photonic lanterns are devices that perform SMD, albeit not in the eigenmode basis of the telescope. Another avenue is to use multiplane light converters (MPLCs). An MPLC uses multiple phase plates that are separated by some distance to implement arbitrary unitary operations. That also means that MPLCs could implement optimal SMD (Deshler et al. 2024b).

DM: deformable mirror

8. WAVEFRONT SENSING AND CORRECTION

Coronagraph designs assume that incoming wave fronts from all astrophysical sources in the field of view are flat and that the optics in the coronagraph are ideal, propagating and modifying these wave fronts without distortion to the final SCFP. In reality, however, several factors cause wave fronts to deviate from this ideal: (*a*) optical manufacturing limitations, (*b*) environmental conditions (both static and dynamic) within the instrument and the telescope, and (*c*) in the case of ground-based telescopes, wavefront residuals from Earth's turbulent atmosphere that are partially corrected with a high-order AO system.

8.1. Adaptive Optics

AO senses the turbulence introduced by the Earth's atmosphere (ϕ_{ATM}) by using WFSs that measure a wave front (ϕ_{WFS}), reconstruct an estimate of this turbulence, and apply it to an electronically actuated deformable optical element—typically a deformable mirror (DM)[1]—within the instrument to modify the incoming turbulent wave front and flatten it. With the DM upstream of the WFS, and with an AO computer providing the calculation of wave front measured by the WFS and applying this correction (ϕ_{DM}) to the DM, a closed loop is formed wherein the response of the DMs correction is observed by the WFS in the instrument (**Figure 9**).

Incoming light is split using a dichroic or gray beam splitter, which sends some of the light to the science camera and the rest to the WFS camera. Many AO systems exploit the fact that the OPD introduced by Earth's atmosphere above ground-based telescopes is achromatic, despite OPD amplitudes of several tens of micrometers across large telescope apertures. A consequence is that a wavefront measurement at a shorter wavelength (typically at optical or near-IR wavelengths) will provide correction for all longer (science) wavelengths. For 8 m class telescopes looking at bright stars, when wavefront sensing is carried out at the same wavelength as the science camera ($\lambda_{\mathrm{WFS}} = \lambda_{\mathrm{SCI}}$), the theoretical contrast is 10^{-6}–10^{-7}, but when λ_{WFS} is in the optical range and λ_{SCI} is in the IR range, the limit is set by the scintillation chromaticity induced by Fresnel propagation through the atmosphere to 10^{-4}–10^{-5} within 1 arcsec (for a discussion of the limits to AO for HCI, see Guyon 2005).

The Earth's atmosphere is highly dynamic and changes on a timescale of milliseconds, but wavefront reconstruction and correction on the DM are not instantaneous, leading to a small but significant time lag between the measurement and the application of the correction. One measure of the turbulence is the Fried length, r_0, which is the radius of a circle where the mean variance at a wavelength λ is equal to 1 rad^2. The Fried length is a function of wavelength, $r_0 \propto \lambda^{6/5}$, and is typically quoted at $\lambda = 0.5$ µm. AO is a complex and mature field in its own right, covering atmospheric turbulence, optomechanics, engineering control theory, wavefront sensing, and information theory (each of these topics could be a review in its own right, so we refer the reader to Guyon 2018 for overviews).

Using `HCIpy` (Por 2020), we simulate an ELT AO system that feeds a high-contrast instrument containing an ideal coronagraph: The DM has 128 actuators across its diameter, resulting in a DM

---

[1] Several other optomechanical devices exploit different optically active principles to modify a wave front.

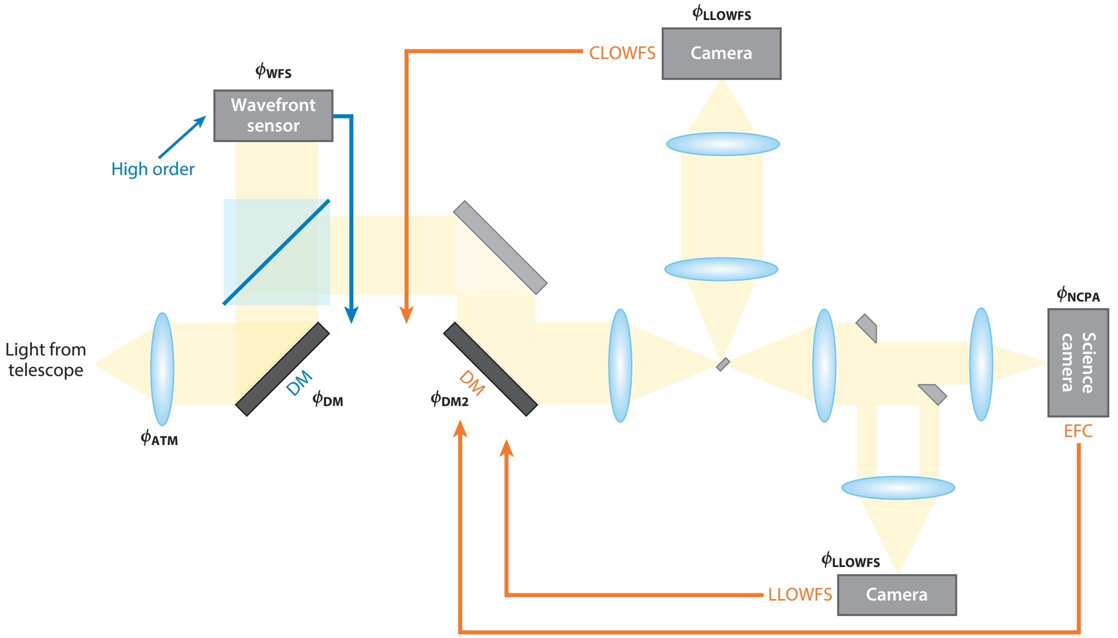


Figure 9

A schematic overview of a high-contrast imaging instrument (GMagAO-X), showing the optical components, closed loops, and wavefront sensors (WFSs). Light from the telescope is corrected by a high-order adaptive optics (AO) system with a high-order deformable mirror (DM), forming a closed loop that corrects a large fraction of the atmospheric turbulence. A beam splitter or dichroic sends science light into a second AO system. This second DM provides lower-order noncommon path aberration (NCPA) corrections via feedback from three possible sources: a coronagraphic low-order wavefront sensor (CLOWFS), a Lyot-based low-order wavefront sensor (LLOWFS), and electric field conjugation (EFC) from the science camera focal plane.

control halo that is 128 diffraction widths in diameter (**Figure 10**). The system speed is 2 kHz, with WFSs observing at 0.5 μm and a science camera wavelength of 800 nm through an atmosphere with $r_0 = 0.16$ m and a wind speed of 37 m s^{-1} (exaggerated to emphasize the wind-driven halo). A second 60 × 60 actuator DM provides NCPA correction.

Figure 10 shows a plume of speckles around the central image, a result of the unsensed atmosphere along the leading edge of the telescope crossing into the pupil before it is corrected by the AO system. Instantaneous speckles average out over time to an azimuthally symmetric halo, with an extended wind-driven halo that can be time varying in orientation and even asymmetric in the presence of strong turbulence (Cantalloube et al. 2018).

8.2. Noncommon Path Aberrations

Aberrations can be sensed and corrected to the point of the last WFS in the optical path in the high-contrast instrument. Ideally the sensed wave front, ϕ_{WFS}, is identical to the wave front delivered to the science camera, ϕ_{SCI}, but because the wave front is divided at the beam splitter, aberrations introduced in the two separate optical paths will result in differences between the two. The differential aberrations between this WFS and the final wave front delivered to the science camera are referred to as NCPAs. These are characterized by using the Pearson coefficient, $\rho(t)$, which quantifies the changes in intensity in the speckle field in the SCFP between two images taken at times t seconds apart. The Pearson coefficient is zero for completely uncorrelated

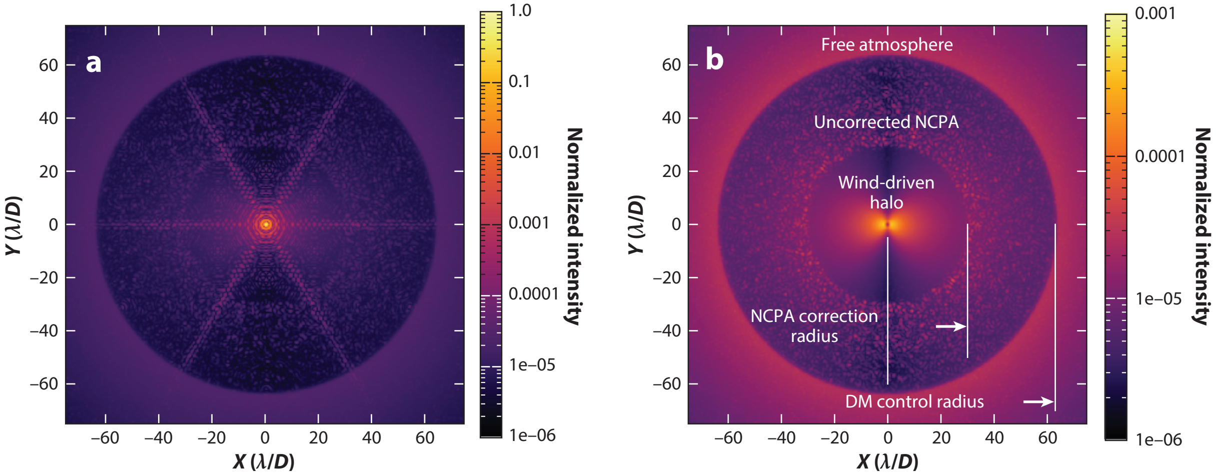


Figure 10

Closed loop of an extremely large telescope adaptive optics system with (*a*) the point spread function (PSF) and (*b*) the ideal coronagraphic PSF. In the coronagraphic PSF, the outermost radius is due to the deformable mirror (DM) control bandwidth, with the free atmosphere outside this radius. Residual speckles are observed in this annulus, down to the noncommon path aberration (NCPA) correction radius. Close to the star, the PSF is dominated by the wind-driven halo.

intensities between the two images, and a coefficient of unity means that the speckle field is static between the two images.

Studies using the internal calibration source of the SPHERE instrument (Martinez et al. 2012, Milli et al. 2018, Vigan et al. 2022) and SCExAO internal sources and on-sky data (Goebel et al. 2018) show three characteristic timescales of decorrelation. The first and shortest timescale is on the order of the speed of the closed AO loop and is attributable to the evolving atmosphere above the telescope aperture (∼2 ms for the 1.68 kHz SCExAO system). The second and third timescales observed in the on-sky SCExAO system are expressed as $\rho(t) = 0.018e^{-t/0.40} - 5.6 \times 10^{-5}t + C$, where an exponential decay of a few seconds is added to a linear decrease over several minutes. Both effects were observed with on-sky and internal calibration source measurements (with different timescales for the two cases). The SPHERE analysis of NCPAs (Vigan et al. 2022) observed similar effects and timescales; the shorter timescale has an amplitude of a few nanometers and is attributed to fast internal turbulence within the enclosure generated by heat from actuation motors. The other is a slow, quasi-linear decorrelation on the order of a few 10^{-3} nm rms s^{-1} that acts on timescales from minutes to hours and is due to changing optical figures on the internal optics caused by temperature gradients and other sources of mechanical flexure.

First-generation high-contrast instruments have provided few to no active NCPA measurement and in situ mitigation strategies, but the presence of NCPAs has had a significant impact on the sensitivity of these instruments at smaller IWAs, reducing the predicted contrast in their designs. More recent instruments have taken multiple approaches to NCPA at every stage of the instruments' life cycle.

During instrument design, minimizing the numbers of optics and making the optomechanics mechanically and thermally stable decrease the total NCPA (Absil et al. 2024). During science operation, methods for estimating the wave front at the final SCFP and then providing active feedback further reduce the speckle field in the dark hole. After the data are obtained, algorithms that provide estimates of the science camera PSFs at all times during the observations can minimize and/or remove the speckle field in port processing.

9. CHALLENGES FOR SEGMENTED TELESCOPES

Monolithic mirror telescopes are limited by their transport from the manufacturing point to the observatory location. Therefore, segmented telescope designs are used for diameters greater than 8 m on the ground. ELT projects include the European ELT, the Thirty Metre Telescope (TMT), and the GMT. All three telescopes have altitude/azimuth mounts, with segmented primary mirrors that have support structures holding a secondary mirror in front of the primary, blocking the central part of the telescope pupil. **Figure 2** depicts the telescope pupils. The large apertures mean that the IWAs are on the order of 10 mas for H-band imaging, enabling direct imaging searches and characterization upgrade paths and allowing new instruments to be built on the basis of the experience of the first-generation instruments. For the ELT, all three first light instruments—METIS (Brandl et al. 2022, Absil et al. 2024), MICADO (Huby et al. 2024, Sturm et al. 2024), and HARMONI (Houllé et al. 2021, Thatte et al. 2022)—have HCI modes that include the coronagraphs mentioned above. The challenges of atmospheric correction arising from the wind-driven halo, atmospheric dispersion, and the water vapor content of Earth's atmosphere mean that ground-based telescopes are limited mostly to HZ planets around nearby M dwarfs.

9.1. Missing/Tilted Segments

Segmented mirror telescopes present a challenge in that they require periodic cleaning, which results in varying transmission across the telescope pupil and occasionally requires segments to be removed entirely for realuminization. For the ELT, a baseline of three to eight segments per night are removed from the telescope pupil according to the realuminization schedule.

9.2. Atmospheric Dispersion (Ground-Based Only)

The wavelength-dependent differential refraction introduced by Earth's atmosphere, called atmospheric dispersion, increases $\propto \sec(z)$, where z is the zenith distance to an astronomical target. In units of diffraction widths, the atmospheric dispersion increases for larger-diameter telescopes, making it a challenge for ELTs to observe science targets far from the zenith (Kendrew et al. 2008, Skemer et al. 2009, van den Born & Jellema 2020). High-order atmospheric dispersion correctors are needed to produce diffraction-limited imaging over wide bandwidths (Kopon et al. 2013). The mid-IR ELT instrument METIS presents an additional complication due to the nonlinear and variable nature of the atmospheric dispersion around the water bands, which make atmospheric dispersion correction far more challenging (Absil et al. 2022).

9.3. Low-Wind Effect

When the wind speed within large telescope domes drops below 3 m s^{-1}, low-order large-amplitude wavefront distortions are observed in the science camera PSFs that are not sensed or removed by the WFSs. These wavefront distortions were initially discovered and characterized on SPHERE (Sauvage et al. 2016). The likeliest explanation is that air temperature gradients form next to the secondary support structure, the temperature of which is anomalously deviant from the nighttime air temperature. These temperature gradients then form piston-like aberrations within each sector of the telescope pupil, which the Shack–Hartmann WFS is insensitive to detecting. Milli et al. (2018) discuss mitigating this problem by adjusting the instrument design, and Pourré et al. (2022) present control solutions for the Shack–Hartmann WFS. Fast low-order algorithms that can sense these modes, such as Fast and Furious (F&F) (Wilby et al. 2018), then provide feedback to the AO system to remove this effect. Several other mitigation strategies have been tested and verified on-sky with the SCExAO/VAMPIRES system (Vievard et al. 2019).

9.4. Petal Modes

The large secondary mirror units of ELTs require large secondary support structures whose projected thickness, as seen on the telescope pupil, can be several Fried lengths. This blockage across a chord of the telescope pupil creates a discontinuity in the wavefront reconstruction, fragmenting the pupil into distinct separate regions that make reconstruction challenging because the thickness of the secondary support is several times greater than r_0.

Even monolithic mirrors have this problem with thick enough secondary support structures. Together with the low-wind effect, this problem is known as the island effect. Approaches to correcting it include apodizing each giant mirror segment individually (redundant apodized pupils; Leboulleux et al. 2022b,c) and measuring the effect with a spatially filtered unmodulated pyramid WFS (Levraud et al. 2024) or using the F&F algorithm (demonstrated on Subaru/SCExAO in Bos et al. 2020).

The GMT design has seven large mirrors, each 8 m in diameter. Circular mirrors are arranged in a hexagonal pattern, and the large gaps between the edges of the mirrors give rise to differential piston errors between the mirrors. The current phasing sensor for the GMT is a pyramid WFS (Quirós-Pacheco et al. 2022). However, the pyramid WFS can sense a differential piston within only half a wave, so larger signals are phase-wrapped. In addition, atmospheric turbulence reduces the pyramid WFS's sensitivity (Bertrou-Cantou et al. 2022). To solve these issues, GMT uses a two-stage system consisting of the pyramid WFS for fine phasing and a holographic dispersed fringe sensor for coarse phasing of the differential piston between the seven segments (Haffert et al. 2022). A phase plate is placed in a pupil plane, which produces radially dispersed fringes that encode the relative phase between pairs of GMT mirrors. The sensor has a dynamic range of 30 μm and can measure relative piston differences of 10 nm rms. The two-stage approach has been validated in simulations and lab tests on the high-contrast AO testbed (Hedglen et al. 2022, Quirós-Pacheco et al. 2024).

9.5. Segment Phasing

The ELT and TMT have another phasing challenge: Their primary mirrors comprise several hundred segments. Therefore, each telescope has developed its own dedicated phasing sensor. The TMT's phasing subsystem is similar to Keck's. The Keck 10 m telescopes relied on active control of their primary mirror segments by using both metrology from sensors between adjacent segments and a modified Shack–Hartmann WFS that looked at the PSF formed from the apertures straddling two adjacent mirror segments. The original phasing algorithm for the Keck mirror segments could go from a 30 μm piston down to 30 nm (Chanan et al. 1998, 2000). However, the original approach could not routinely reach 30 nm rms, which affected the performance of the HCI instruments on Keck. A promising phasing approach using a Zernike wavefront sensor (ZWFS) has demonstrated phasing to 46 nm precision on-sky (van Kooten et al. 2022), and a vector ZWFS has demonstrated an increase of 10% in Strehl ratio (Salama et al. 2024).

The ELT includes a ZEUS sensor (Dohlen et al. 2006), which is a ZWFS-style sensor that has been adjusted to work in seeing-limited conditions. This approach has been demonstrated both on testbeds (Pfrommer et al. 2018) and on-sky (Gonté et al. 2009) to reach the required levels.

9.6. Specific Challenges for Space Telescopes

The advantages of putting a telescope into space for HCI are immediately obvious: The turbulence, dispersion, and transmission of Earth's atmosphere no longer limit the achievable contrast, but the mirror sizes are limited by the rocket farings and their capacity. The instruments can use monolithic mirrors with a diameter of up to 2 to 3 m, whose diffraction limit drives coronagraphs with the highest throughput and smallest possible IWA so that a larger number of star systems can be imaged in comparison to an equivalent-diameter segmented mirror. OVC-based designs that include additional hybrid elements (see Section 6.1) can produce exoplanet yields of approximately 10 to 20, assuming an unobstructed telescope aperture (Stark et al. 2024). Larger-aperture telescopes can be realized with segmented mirror designs that can be folded into rocket farings and subsequently deployed, but they have a more stringent requirement for mirror segment alignment due to the higher contrasts required—picometer precision is necessary to achieve 10^{-10} contrasts. However, alignment of the segments in these designs is susceptible to temporal drifts, giving rise to aberrations in the focal plane that are in the scientific region of interest. Wavefront sensing and subsequent correction of these aberrations are therefore an important part of HCI.

Algorithms such as Coronagraphic Focal Plane Wavefront Estimation for Exoplanets (COFFEE) have been demonstrated for the JWST segmented primary pupil geometry (Leboulleux et al. 2020). The deployment of the JWST mirror has shown drifts of 9.0 nm rms per 48 h (Lajoie et al. 2023), demonstrating the need for active correction to obtain picometer precision for HWO (Laginja et al. 2022). Mechanical flexure in the segmented mirror support structure presents separate challenges for alignment and stability of the mirrors within the frame. Target stars are too faint for wavefront measurement and alignment of the mirror structure, so the telescope must tune up on a much brighter star before slewing to the science target. Ultrastable structures are therefore needed to keep the structure rigid during the slewing maneuver. The key component-level technologies have matured to a level where this goal is now feasible (Coyle et al. 2021).

When observations of target stars begin, a question arises as to what should be done if drifts in the optics start to fill in the dark hole. Is it better to actively maintain the dark hole at a cost of signal-to-noise ratio while sensing is carried out, or is it better to let the drifts occur over the duration of the observation (Pogorelyuk & Kasdin 2019, Redmond et al. 2020)?

10. POLARIZATION EFFECTS

The field of HCI is always hammering away at one noise floor after another. It started with the common phase aberrations that dominate at low to moderate contrast. After that, amplitude aberrations started to limit the contrast. This problem was solved with the use of multiple DMs. The contrasts that are achieved on-sky and on testbeds have now revealed another limit: polarization (Schmid et al. 2018, Millar-Blanchaer et al. 2022, van Holstein et al. 2023, Baudoz et al. 2024). Polarization is an often underappreciated property of light. The derivation presented in Section 2 actually ignores the effects of polarization for mathematical clarity. However, light consists of two orthogonal polarization states that do not interfere with each other. As a result, at any time, two beams of light are always propagating through the coronagraph that might interact in a different way with the instrument. A more detailed treatment of polarization and physical optics propagation can be found in the literature (McLeod & Wagner 2014).

The Fresnel equations describe how light is either reflected off or transmitted through an interface. **Figure 11** presents the definitions of all variables for the incident, reflected, and transmitted waves. The Fresnel equations for plane-wave interfaces are

$$r_{\mathrm{s}} = \frac{n_1 \cos\theta_{\mathrm{i}} - n_2 \cos\theta_{\mathrm{t}}}{n_1 \cos\theta_{\mathrm{i}} + n_2 \cos\theta_{\mathrm{t}}}, \qquad 17.$$

$$t_{\mathrm{s}} = \frac{2 n_1 \cos\theta_{\mathrm{i}}}{n_1 \cos\theta_{\mathrm{i}} + n_2 \cos\theta_{\mathrm{t}}}, \qquad 18.$$

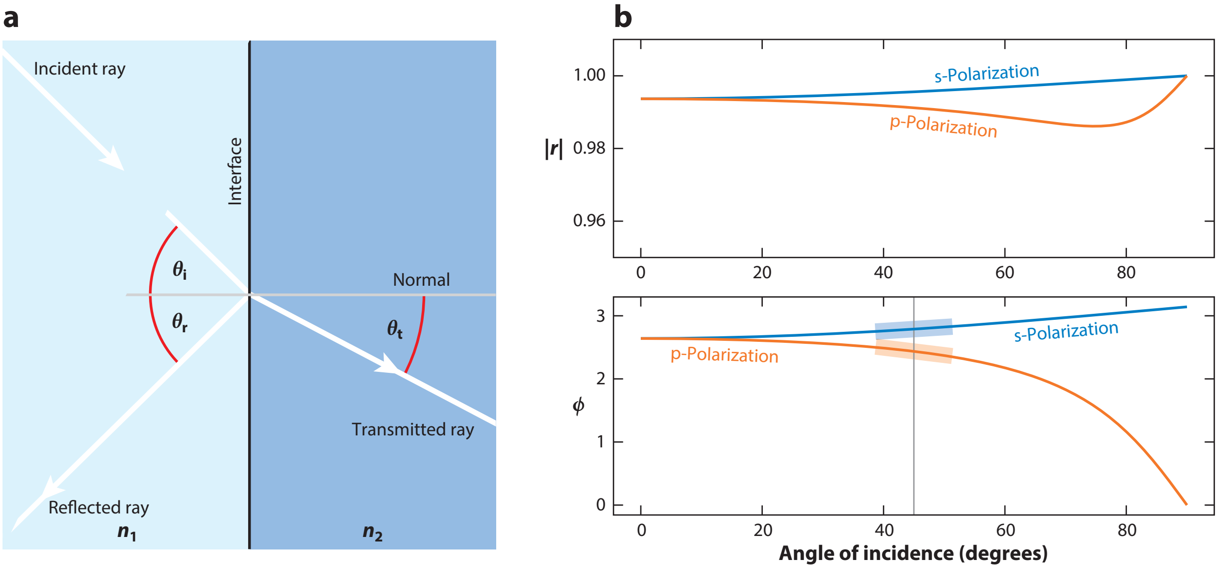


Figure 11

(*a*) Reflectance and transmittance at an optical interface. (*b*) The differences between s- and p-polarization reflections at the interface.

$$r_p = \frac{n_2 \cos\theta_i - n_1 \cos\theta_t}{n_2 \cos\theta_i + n_1 \cos\theta_t}, \tag{19}$$

$$t_p = \frac{2n_1 \cos\theta_i}{n_2 \cos\theta_i + n_1 \cos\theta_t}. \tag{20}$$

Here, r_x and t_x are the reflected and transmitted amplitudes for polarization state x. The Fresnel equations depend on the angle of incidence, θ_i. While the transmitted angle, θ_t, is part of the Fresnel equations, it depends on the incoming angle through Snell's law: $n_2 \sin\theta_t = n_1 \sin\theta_i$. Therefore, the Fresnel equations are a function only of the material and the incoming angle. They show that different polarization states have different coefficients. That would not be a problem if the Fresnel equations did not also depend on the angle of incidence. Any finite-sized beam, which means any physically plausible beam, has an angular spread because it consists of a linear combination of plane waves. Each wave reflects off the interface at a slightly different angle and with a different amplitude. This effect creates so-called polarization aberrations (Chipman 1989, Breckinridge et al. 2015).

Figure 11 shows the phase change when light reflects off a silver mirror. At a 45° incidence angle, there is a substantial difference between s- and p-polarization; it becomes even greater at larger angles of incidence. More importantly, there is a difference in slope, which means that a finite-sized beam with a certain angular width will have a different phase tilt aberration depending on the polarization state that enters the system. This effect is referred to in the literature as the Goos–Hänchen shift. In HCI systems, such effects create beam shifts that limit coronagraphic performance (Schmid et al. 2018, Millar-Blanchaer et al. 2022).

Polarization aberrations can be estimated using polarization ray tracing, where the Fresnel equations are applied for every surface that is encountered during the ray trace (Ashcraft et al. 2023). A convenient way to represent the aberrations is in the Jones pupil format. The Jones matrix is determined for each pixel in the pupil, which means that we end up with four pupil images

(xx, xy, yx, and yy). The Jones pupil can then be used in HCI physical optics simulations to estimate the interaction between the polarization aberrations and coronagraphs (Anche et al. 2023) or AO residuals (Millar-Blanchaer et al. 2022). Recent simulations suggest that polarization aberrations could affect ELTs at 1 to 3 λ/D (Anche et al. 2023). However, the strongest effects are observed in space-based coronagraphic systems that require a deep raw contrast of 10^{-10} to 10^{-8}. Current systems typically encounter polarization aberrations at the 10^{-8} level (Mawet et al. 2011, Seo et al. 2019, Baudoz et al. 2024). A common strategy is to put the whole instrument between polarizers to ensure that only one polarization state is propagated. Doing so ensures that the aberrations are controllable. However, this approach leads to loss of 50% of the light. Research is now focusing on finding solutions to mitigate the effects, by optimizing coatings of the optics (Balasubramanian et al. 2005, Miller et al. 2022), by active wavefront control (Mendillo et al. 2021), or by improving the polarization leakage of the coronagraph (Doelman et al. 2020, 2023).

11. FOCAL PLANE WAVEFRONT SENSING

Imperfections in the manufacture of the optics for a high-contrast instrument and changing environmental conditions result in speckles in the final SCFP. Several techniques employed for optical sensing of these residual aberrations, using telemetry or metrology within the instrument, have partly succeeded in sensing and removing these aberrations with closed loops by using actively deformable optics to provide correction for the sensed modes. Ultimately, these methods cannot sense the time-varying aberrations within the last optical elements before the SCFP. As a result, several ways to measure and characterize optical aberrations using the images from the SCFP have been developed. The fundamental challenge is that the vast majority of the science focal plane detectors are photodetectors, so they do not record the complex amplitude of the incoming electric field in the wave front, only the intensity.

The result is that an intensity image of the PSF cannot be uniquely inverted to yield the phase and amplitude of the wave front in the pupil of the system. An arbitrary wave front can be represented as the weighted sum of a series of even [$f(r) = -f(r)$] and odd [$f(r) = -f(-r)$] point-symmetric functions. Odd functions produce PSFs with point symmetry (e.g., tip/tilt, coma), but even functions produce the same PSFs with the same amplitude regardless of sign. Consider a wave front with focus, which has $\psi(r) = a\sqrt{3}(2r^2 - 1)$; both a and $-a$ will result in the same intensity distribution in the focal plane. Phase retrieval directly from the PSF is therefore an underconstrained inverse problem.

One of the earliest methods for phase retrieval is an iterative method known as the GS algorithm (named for Gerchberg & Saxton 1972). The GS algorithm first picks a random phase for each pixel in the pupil, then iterates between the focal plane and pupil plane, where it replaces the amplitude with either the square root of the measured PSF or the known pupil function. By constantly iterating, it solves the nonlinear phase retrieval problem. However, the GS algorithm cannot fundamentally solve the sign degeneracy problem of even modes. What happens during the iterations is that the algorithm uses the nonlinear cross talk from various modes to estimate the sign of the even modes. Therefore, the GS algorithm would be unable to measure the sign of defocus, for example, if there is only a defocus wavefront error.

In order to measure the complex amplitude of the PSF, a diversity must be introduced into the focal plane image, either temporally or spatially (for a review, see Fienup 2013, Gonsalves 2014). A natural diversity is the introduction of a focus offset, which can easily be implemented either by mounting a camera on a controllable stage or by using a DM (Van Gorkom et al. 2021). The classic phase diversity algorithm works only with normal PSF imaging. The COFFEE algorithm (Sauvage et al. 2012, Paul et al. 2013, Herscovici-Schiller et al. 2018) removes this requirement by estimating the aberrations in a coronagraphic instrument. Another diversity involves directly modifying the pupil by adding obstructions (Martinache 2013, Brooks et al. 2016, Bos et al. 2019, Gerard et al. 2023). Algorithms like phase diversity and COFFEE try to solve the full nonlinear inverse problem. However, this is a time-consuming problem, and during actual observations the wavefront errors might evolve faster than the required computing time.

Many variations of phase diversity that can work at real-time frame rates are in development. One such method is the F&F algorithm (Keller et al. 2012), which has been verified in the lab (Wilby et al. 2018) and demonstrated on-sky (Bos et al. 2020) on the SCExAO system. The F&F algorithm is a modified GS algorithm that uses the measurement and control command from the previous time step, allowing the F&F algorithm to solve the degeneracies while it is in a closed-loop configuration.

Other methods try to linearize phase diversity in either imaging mode or coronagraphy mode because linear methods allow for short computational times. These methods include the linearized focal plane technique (Meimon et al. 2010) and the linearized analytical phase diversity (Vievard et al. 2020). Certain coronagraphs also use algorithms that are specific to their behavior, such as QACITS (Huby et al. 2017) for the OVC. QACITS is a tip/tilt centering algorithm with the goal of keeping the star on top of the center of the OVC mask, resulting in a pointing stability of 2.4 mas over several hours. Most phase diversity methods aim to control only the low-order Zernike modes. However, most coronagraphic dark holes are dug at medium spatial frequencies. Linearized dark field control (Miller et al. 2017) is a method that extends beyond a few low-order modes. It has been demonstrated on-sky with the vAPP coronagraph, and whether it can work with arbitrary coronagraphs is under investigation (Miller et al. 2019, Ahn et al. 2023).

The push for linear methods is due to the computational complexity of nonlinear reconstruction algorithms. Recent advances in machine learning have made it possible to run deep neural networks with short inference times at real-time rates. These often data-driven methods can be used to solve nonlinear phase retrieval methods. Variations on the neural network architecture enable normal imaging (Orban de Xivry et al. 2021, Zhang et al. 2021), postcoronagraphic imaging (Quesnel et al. 2022), or use within an iterative scheme like F&F (Bottom et al. 2023). All of these phase retrieval algorithms are used to estimate the wavefront error and apply a correction; therefore, the loop is repeated until it converges to a flat wave front.

11.1. Dark Hole Digging

A flat wave front would ideally result in a deep and dark coronagraphic image. However, coronagraphic dark holes are not completely free of starlight after wavefront flattening because of the existence of amplitude aberrations and higher-order frequency-folding effects. Instead of aiming to flatten the wave front, HCI should aim to remove as much light as possible from the dark hole region. This approach is informally referred to as dark hole digging (Malbet et al. 1995, Bordé & Traub 2006).

Several control algorithms dig dark holes. They include energy minimization (Bordé & Traub 2006), speckle nulling (Martinache et al. 2012, 2014), electric field conjugation (EFC) (Give'On 2009), and stroke minimization (Pueyo et al. 2009). The major difference between them is the exact cost function that is optimized. The HCI algorithms are presented by Give'On (2009) with a unified formalism and by Give'On et al. (2011) under EFC. The fundamental concept behind EFC is that it is possible to create speckles with the DM that exactly cancel out the stellar speckles by destructive interference (see **Figure 12**). The DM commands are found by minimizing the electric field within the dark hole region S:

$$\hat{v} = \arg\min_{v} \int_{S} |\mathcal{E}_{\mathrm{s}}(x,y) + \mathcal{E}_{\mathrm{DM}}(x,y,v)|^2 \mathrm{d}A. \qquad 21.$$

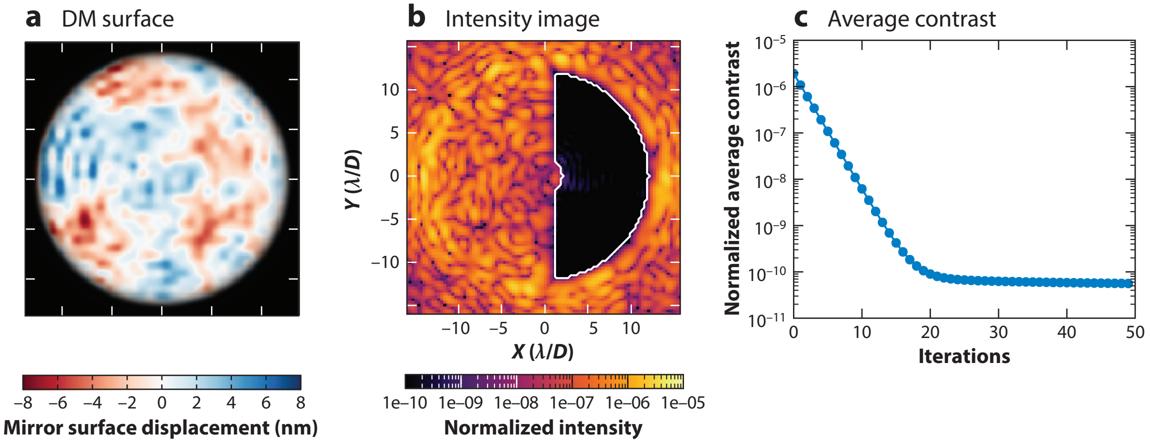


Figure 12

Electric field conjugation (EFC) generation of a dark hole. The dark hole is iteratively generated using the EFC algorithm. (*a*) The final deformable mirror (DM) surface. (*b*) The final science camera focal plane image with the target dark hole region is outlined in white. (*c*) The average contrast in the dark hole as a function of iteration number.

Here, v is the DM command, $\mathcal{E}_s$ are the stellar speckles in the coronagraphic focal plane, and $\mathcal{E}_{\mathrm{DM}}$ are the speckles created by the DM. Effectively, EFC finds the command that injects speckles with the exact same amplitude but opposite phase in the focal plane. EFC is currently the most widely used and best-optimized algorithm on testbeds and instruments (Mennesson et al. 2024).

The electric field in the focal plane must be measured before it can be canceled. Therefore, all dark hole digging algorithms are combined with an electric field sensing approach. There are two main approaches: pairwise probing (PWP), where the DM is used to introduce diversity, and the self-coherent camera (SCC), which uses starlight that has been rejected by the coronagraph to create interference fringes. The former method uses temporal diversity, while the latter uses spatial diversity. Originally, EFC was combined with PWP (Give'On 2009).

The deformable element within the instrument can introduce phase shifts into the pupil image that then change the resultant complex amplitude for each location in the SCFP. For a single point source, all the light in the focal plane is coherent with respect to the core of the PSF. Each pixel in the focal plane then becomes an intensity interferometer, and if four phase shifts that reasonably sample between 0 and 2π rad are introduced into the instrument, then the four recorded PSF intensity images can be fitted to provide an estimate of the complex amplitude at each location in the SCFP. Two pairs of complementary DM actuations can provide enough phase diversity to measure the complex amplitude across the focal plane out to the spatial sampling of the mirror actuators.

The SCC (Baudoz et al. 2006) takes light from the telescope, splits it into two beams, filters one beam through a pinhole, and recombines the two beams in a Fizeau configuration in the SCFP. All the speckles in the SCFP are subsequently modulated by a set of cosine fringes, where the relative position of the fringes encodes the complex amplitude of the electric field of the PSF. Any exoplanets or other point sources emit photons that are incoherent with the speckles; therefore, fringes do not appear at the location of the planet. Simulations (Galicher et al. 2010) demonstrate bandwidths of up to 5% and possible contrasts down to 10^{-10}.

Two flavors of control appear in the literature. The first, called implicit EFC (iEFC) (Haffert et al. 2023), is a model-free approach where the interaction of the DM with the speckles is empirically calibrated. In the second, more traditional approach, the interaction is determined from a numerical model of the optical system. Every combination of control (EFC or iEFC) and sensor (PWP or SCC) leads to a similar performance in dark hole contrast (Desai et al. 2024).

One DM enables correction of phase-only aberrations in the science camera PSF. If the aberrations include amplitude as well as phase, then correction with one DM will result in one side of the PSF becoming dark while the other does not. This problem can be solved with two DMs (with one DM between a focal plane and a pupil plane to allow for both amplitude and phase control), which can act in tandem to correct both amplitude and phase aberrations and correct a 360° region in the field of view (Pueyo et al. 2009).

EFC has been tested on the internal sources of SPHERE (Potier et al. 2020), SCExAO (Ahn et al. 2023), and MagAO-X (Haffert et al. 2023). Each of these instruments achieved a contrast of $\sim 10^{-7}$ at 100 to 200 mas from the optical axis within a few minutes. The on-sky implementation took longer because of the increased complexity introduced by atmospheric turbulence. The first on-sky demonstration of EFC, with the APLC coronagraph on SPHERE (Potier et al. 2020, 2022), used an iterative scheme to clear a dark hole on one side of the science camera PSF down to an intensity of 10^{-6}, an improvement of a factor of a few from 250 to 500 mas separations. A second technical run took place with the FQPM coronagraph on SPHERE (Galicher et al. 2024). Additionally, iEFC (Haffert et al. 2023) has now also been demonstrated on-sky (Haffert et al. 2024, Kueny et al. 2024). A similar performance gain of a factor of a few has been achieved with iEFC on MagAO-X. These algorithms can now clear out dark holes in coronagraphic images but are limited by the wind-driven halo from the AO system, which can be filtered out with a low-pass filter or postprocessing algorithms.

11.2. Modifications of the Self-Coherent Camera for Ground-Based Telescopes

The downside of the SCC is that it uses a pinhole far away from the pupil edge. The amount of light that a coronagraph then diffracts into the pinhole is very low. The NCPA and the wind-driven halo in ground-based instruments are typically on the order of 10^{-6} to 10^{-3}, significantly greater than the amount of light passing through the pinhole. As a result, there is a low fringe contrast for the classic SCC, which also implies a low signal-to-noise ratio.

Several modifications to send more light to the pinhole used to generate fringes in the SCFP have been proposed in an effort to make the SCC more suitable for ground-based telescopes. The first is a modified FPM. Instead of blocking light with a classic absorbing Lyot coronagraph FPM, the light is steered away with the use of a phase ramp. The ramp then centers the light on top of the pinhole, and the throughput can be further improved by adding a small amount of power in the FPM within 3 λ/D. This approach, termed fast atmospheric SCC (FAST) (Gerard et al. 2018), increases the throughput by a hundredfold to a thousandfold, enabling shorter exposures to capture the speckles in the SCFP and provide correction. Laboratory measurements have demonstrated 3×10^{-4} contrast (Gerard et al. 2022). A similar approach can be used to create a FAST-style PIAACMC (Haffert et al. 2023a).

Almost all coronagraphs diffract most of the starlight close to the pupil edges that then falls off rapidly. Moving the pinhole of the SCC closer in to the pupil edge will increase the throughput by a similar factor of a hundred to a thousand. The downside is that the pinhole needs to be modulated to disentangle the fringes from the stellar speckles. Several approaches to modulation have been proposed: temporal modulation (Martinez 2019), polarization modulation (Bos 2021), and spectral modulation (Haffert 2022).

11.3. Probing the Electric Field with the Atmosphere

A closed-loop AO system leaves time-varying wavefront aberrations propagating through to the SCFP. Short-exposure images freeze these speckles and show that, for a given location in the

SCFP, the flux changes as a function of time. For thermal IR science cameras on ground-based telescopes, the science images saturate rapidly due to the thermal background of Earth's atmosphere, so the rapidly changing speckles naturally provide a "free" source of phase diversity. Under the assumption that the NCPA changes on a much slower timescale than the speckles generated by the wind-driven halo, the WFS can provide an estimate of the complex amplitude of the SCFP, plus a fixed NCPA term. Phase-sorting interferometry (Codona & Kenworthy 2013) enables the construction of a virtual interferometer for each location in the SCFP and allows the NCPA to be calculated. The calculation of the NCPA was successfully demonstrated as a postprocessing technique at thermal IR wavelengths at the MMT Observatory with the Clio camera (Codona & Kenworthy 2013).

It is possible to go a step further and estimate the electric field in each frame by using WFS telemetry. Doing so enables both speckle estimation and the detection of incoherent (with respect to the star) sources (Frazin & Rodack 2021, Rodack et al. 2021). This approach, while computationally intensive, would enable a duty cycle of 100% while probing for incoherent planet light at the same time.

12. REJECTED LIGHT WAVEFRONT SENSING

Focal plane wavefront sensing requires starlight to sense wavefront errors. However, after digging a dark hole, the starlight is gone, so there is no longer any starlight to sense wavefront errors. Ongoing studies are investigating how to modify coronagraphs and algorithms to better utilize the light that has been rejected by the coronagraph to filter starlight. Furthermore, measuring the centroid of the star (or other low-order aberrations) is not possible in the coronagraph SCFP due to the insensitivity to low-order aberrations. Low-order aberrations can, however, be measured from the light rejected from the FPM.

The Lyot-style coronagraphs that use opaque FPMs can be naturally extended to include low-order wavefront sensing capability by making the FPMs reflective (**Figure 9**). The reflected light is reimaged onto a separate detector. The encircled energy at a radius of 1 λ/D already contains two-thirds of all the incoming starlight. The rejected light from a larger FPM (2–3 λ/D) will therefore contain almost all starlight. This approach is called coronagraphic low-order WFS (CLOWFS) (Guyon et al. 2009). The large number of photons enables high-speed control, and stabilization of the low-order modes shows that it can keep tip/tilt to around $10^{-3}\lambda/D$ for a baseline telescope and observation of a sixth-magnitude star (Guyon et al. 2009).

For coronagraphs that use a phase mask in the focal plane (e.g., a vortex mask), another method is required. Putting in a reflective Lyot stop allows the rejected light from the pupil plane to be reimaged to a separate camera. The rejected light forms the error signal for low-order aberration measurements and is called a Lyot-based low-order WFS (LLOWFS) (Singh et al. 2014, 2015). The LLOWFS can measure tip/tilt down to $\sim 10^{-2}\lambda/D$ (equivalent to 2 to 12 nm at 1.6 μm) per mode on the FQPM; on-sky results are somewhat larger (Singh et al. 2015). Both methods deliberately introduce defocus into the rejected light image so that the focus ambiguity is removed and the tip/tilt and focus modes can be simultaneously measured.

A space telescope mission such as HWO, which is based on Habitable Exoplanet Observatory and LUVOIR studies, requires picometer stability to achieve the 10^{-10} contrast requirement. This goal can be achieved only through active control, but that means that the wavefront errors must be measured with picometer-level precision. The segmented geometry of the HWO concept challenges the low-order WFSs because the segments create high-spatial-frequency errors. The rejected light from the coronagraph masks or Lyot stops does not always contain enough diversity to recover wavefront error signals such as segment misalignment.

The sensing can be significantly improved with the addition of a ZWFS. A ZWFS is a common path interferometer that interferes a $\pi/2$ phase-shifted core of the Airy pattern with the rest of the light. The phase dimple typically has a diameter of 1–2 λ/D. The resultant pupil image intensity directly encodes the phase of the wave front and, thus, is ideal for recovering wavefront errors at all spatial frequencies that are passed by the mask. The ZWFS measurements are fed into the control loop for the DM in the coronagraph to stabilize the wave front, as demonstrated by Ruane et al. (2020) for low and mid-spatial frequencies with a ZWFS up to the DM control radius.

A modification of the HLC uses a dual-purpose mask for the FPM to yield a dual-purpose Lyot coronagraph. This dual-purpose mask is a tiered metallic focal plane occultor that suppresses starlight in the transmitted coronagraph channel as well as a dichroic-coated substrate to reflect out-of-band light to a wavefront sensing camera. It acts as a ZWFS in reflection, sending out-of-band light to a CLOWFS to maintain high contrast in the science focal plane (Ruane et al. 2023). A similar concept was tested on the HiCAT testbed, where light rejected from the FPM was reimaged onto a ZWFS (Pourcelot et al. 2022, 2023). The ZWFS was able to control and reject the low-order wavefront errors that slowly crept into the HiCAT dark zone and kept it clean of starlight (Soummer et al. 2022). By putting a ZWFS at the location of the FPM in a PIAA coronagraph, this combination of algorithms and optics can approach fundamental sensitivity limits within the instrument (Haffert et al. 2023b).

13. HIGH-CONTRAST INSTRUMENTS FOR EXTREMELY LARGE TELESCOPES

For the ELT, the first light instruments are METIS (Brandl et al. 2022), MICADO (Sturm et al. 2024), and HARMONI (Thatte et al. 2022). All of them have associated high-contrast modes. MICADO will have three classical Lyot coronagraphs, a gvAPP, and two sparse aperture masking modes (Huby et al. 2024). METIS was designed with HCI in mind (Kenworthy et al. 2016, Absil et al. 2024) and will have a classical Lyot coronagraph, OVC designs with beam-switching capabilities, a RAVC design to accommodate the large secondary mirror in the ELT pupil, and gvAPP coronagraphs for both direct imaging (2.9–5.3 μm) and a high-spectral-resolution ($R \sim 100{,}000$) integral field spectrograph (IFS). METIS is unique in that it is the only imager and spectrograph working beyond 3 μm out to 19 μm on an ELT. HARMONI has an HCI mode (Houllé et al. 2021) with an SPP based on binary masks (Carlotti et al. 2023) and a dedicated ZELDA WFS at 1.175 μm to measure and correct NCPAs in the system. The bands are H and K, and they use an IFS with a fixed-elevation atmospheric dispersion corrector. The second generation of ELT instruments includes the Planetary Camera and Spectrograph (Kasper et al. 2021), which will use an extreme AO system to reach its goal of 10^{-8} at 15 mas angular separation from the star and 10^{-9} at 100 mas and beyond.

The GMT has the HCI GMagAO-X (Males et al. 2024) with a science requirement of photometry at a signal-to-noise ratio of five on a point source, with a flux ratio of 10^{-7} or better with respect to its host star. The instrument should work at any wavelength from 0.6 to 1.9 μm over a 10% bandwidth filter at a separation of 4 λ/D from the star. The baseline coronagraph design is a PAPLC, and a stretch goal uses a PIAACMC and transmissive complex FPMs. Focal plane low-order WFSs and LLOWFSs will make use of light rejected by the coronagraphs.

The Planetary Systems Imager (PSI) (Fitzgerald et al. 2022) is a proposed instrument suite for the TMT. PSI would be optimized for high-contrast exoplanet science from 0.5 to 13 μm. It would have a near-IR AO system that feeds other systems, notably PSI-Red (2–5 μm), PSI-Blue (0.5–1.8 μm), and PSI-10 (8–13 μm) subsystems. The PSI-Red system would have coronagraphs including a gvAPP, a VVC with a Lyot stop, and an SPP (Jensen-Clem et al. 2021).

14. PHOTONIC VERSUS BULK OPTICS

Optics change the complex amplitudes of wave fronts as they propagate through coronagraphs. Classical optics (referred to as bulk optics) are typically many thousands of times larger than the wavelength of light they shape and require precise and stable optomechanical components to accurately modify these wave fronts. Integrated (or photonic) optics enable direct manipulation of complex electric fields at the scale of the wavelengths used. Miniaturization of previously discrete macro-optics and their manufacture within a single homogeneous substrate remove the requirements for both separate optomechanical alignment and temperature-related misalignment associated with the mechanical mounts.

Beam combiners that are required for optical and near-IR interferometers require temperature- and vibration-controlled optical tables with subwavelength stability tolerances and alignment for the beam splitters and associated optics. Manufacture of wave guides within optical materials that perform beam division and combination considerably simplifies the optomechanical requirements, but the challenge then involves coupling the light from the macro-optics to the substrates while keeping the transmitted efficiency high; diffraction-limited optics are necessary to form PSFs that couple efficiently to the near-single-mode-sized micro-optics. Early examples include beam combiners for optical astronomical interferometers (e.g., the IOTA/IONIC beam combiner; Berger et al. 2001) and photonic lanterns (see Leon-Saval et al. 2010 and references therein). Typical coupling efficiencies are on the order of 10% but are as high as 90% for more recent designs, such as efficient injection from large telescopes into single-mode fibers (Jovanovic et al. 2017). Full electromagnetic propagation is needed to design and evaluate these photonic systems, but their complexity also enables new optical designs that could be combined to form compact, robust instrumentation (for reviews, see Minardi et al. 2021, Jovanovic et al. 2023).

Coupling multimode light into monomode photonics is done using photonic lanterns, where a multimode input is converted into the areal equivalent of a number of single-mode optical channels (Norris et al. 2022). A device equivalent to a Fabry–Pérot etalon can be constructed by etching an elongated loop with one-half of the loop parallel to the wave guide; frustrated transmission between the wave guide and the closed loop is modulated as a function of the number of integer wavelengths around the closed loop, imparting a precise frequency comb into the light. A small heating element on top of the closed loop can change the length of the loop. Due to the small physical size of the heating element, this modulation can be in the kilohertz range. All of these photonic concepts are being considered for the design of coronagraphs for next-generation space telescopes in order to image and characterize exoplanets, with the goal of exploring concepts of different combinations of photonic and bulk optics (Desai et al. 2023).

15. ALGORITHMS FOR ESTIMATING THE INSTANTANEOUS POINT SPREAD FUNCTION

Deviations from the ideal optical prescription of telescope and instrument optics result in wavefront errors that manifest as intensity deviations from the theoretical PSF. Furthermore, these deviations change in intensity and position with time in the SCFP, and they can be equal to or larger than the flux from astrophysical objects next to the star. The question, then, is how to estimate the science camera PSF for every single science camera exposure and subtract this estimate from the science camera image, leaving only the flux from astrophysical objects adjacent to the target star. Estimation of the PSF becomes more complicated when the position and brightness of the exoplanet are not known. Several diversities—properties of the exoplanet that are not the same as the stellar halo—can differentiate between them. The most important are the following:

- Angular diversity. For an altitude/azimuth telescope, the planet on the sky has a predictable angular position and velocity with respect to the orientation of the instrument optics.
- Spectral diversity. The planet has a different spectral energy distribution, meaning that the relative flux between star and planet changes with wavelength.
- Polarimetric diversity. The light from the star is almost completely unpolarized, but reflected light from clouds or dust around the exoplanet becomes polarized under single scattering (Gledhill et al. 1991).
- Wavelength diversity. The stellar halo scales with λ/D, but the planet remains at the same location on the sky.
- Coherence diversity. The exoplanet flux is not coherent with the stellar halo and so does not interfere with it.
- Stochastic speckle discrimination. The intensity fluctuations of the Airy core have a different statistical distribution from the fluctuations in the diffraction halo (Gladysz & Christou 2009).

Many algorithms have been developed to use one or more of these diversities to provide estimates of the science camera PSF, using different linear combinations of the science camera images to estimate the instantaneous science camera PSF. *Hubble Space Telescope* images of circumstellar material show residual speckles that obscure the faint circumstellar environment, even after the subtraction of an image of a nearby star used as a reference PSF. The concept of "roll subtraction" (Schneider et al. 1998) has been used to estimate and remove these residual speckles. Two or more images of the science target were taken with the telescope set at different angles about the target axis, so that the astronomical field would be rotated with respect to the (almost static) speckle field. Roll subtraction was demonstrated by Schneider et al. (1999) with the image of the disk around HR 4796A. Even with the *Hubble Space Telescope*, the roll observations were taken within 25 min of one another to minimize changes in the telescope's optical path resulting from the "breathing" of the telescope optical assembly as it passed from day to night in its low Earth orbit (Bély et al. 1993).

With a ground-based telescope, the speckle field changes on shorter timescales and with increased complexity because of (*a*) a continuously changing gravity vector on the telescope and instrument, (*b*) temperature and mechanical variations in the optomechanics within the instrument, and (*c*) changes in the performance of the AO system due to changing atmospheric conditions. Angular differential imaging (Marois et al. 2006) has become a fundamental algorithm for many ground-based telescope observations where significant sky rotation occurs during the observations of the planet.

The use of narrowband absorption features in the gas giant exoplanet spectrum has enabled methane spectral imaging. TRIDENT (Marois et al. 2005) was one of the first cameras built to employ this technique, along with the spectral differential imaging cameras at the MMT and VLT (Biller et al. 2007). With AO systems reaching optical wavelengths, differencing the images at two adjacent wavelengths has had a renaissance in Hα imaging with the MUSE integral field spectrograph and the discovery of the accreting protoplanet PDS 70c (Haffert et al. 2019).

Estimation of the stellar halo with images at nearby wavelengths has been generalized with the use of integral field spectrographs, where many science camera PSFs are sampled at different wavelengths simultaneously to form (x, y, λ) data cubes. Resampling the image slices into the same λ/D spatial scale radially smears out any exoplanet signal, so subtracting a median of these images removes the stellar halo but keeps most of the planet flux intact. As a result, the exoplanet is visible when the median subtracted cube is resampled into the sky coordinates and recombined into a single image to produce the PSF subtracted image, generalized as spectral differential imaging (Sparks & Ford 2002) and demonstrated on-sky with SINFONI (Thatte et al. 2007). For higher spectral resolutions, the spectrum of the exoplanet can begin to resolve the individual rotational-vibrational transitions, enabling high-spectral-contrast imaging (including the detection of HD 209458b; Snellen et al. 2010) and then generalizing to the principle of molecule mapping in directly imaged exoplanets such as Beta Pictoris b (Hoeijmakers et al. 2018).

Stochastic speckle discrimination (Gladysz & Christou 2009) can be achieved using short exposure images with the Airy core unsaturated. Photon-counting devices allow this detection method to work, as demonstrated with the use of a microwave kinetic inductance detector, and the use of this technique has led to the discovery of a substellar companion (Steiger et al. 2021). Each diversity improves sensitivity from a factor of a few to ten or more. Combining different diversities results in an even greater cumulative effect.

16. CONCLUSIONS

Since the first detection of planets outside our Solar System with the pulsar planets (Wolszczan & Frail 1992) using pulsar timing, and the discovery of the first exoplanet around a solar-type star (51 Peg b; Mayor & Queloz 1995) using radial velocity measurements on the star, we have indirectly detected thousands of planets with radial velocity and transit methods. Direct imaging of exoplanets have revealed dozens of young, self-luminous gas giant planets (Currie et al. 2023, Chauvin 2024). With the minimized IR background accessible with JWST, we are entering an era of direct detection of sub-Jupiter-mass planets.

ELTs with extreme AO systems and the next generation of space telescopes enable the reflected light detection of planets around the nearest stars. The direct imaging of exoplanets is a dynamic and rapidly changing field, with each decade of suppression bringing new challenges and researchers searching for (and finding) solutions to them. We have developed the mathematical theory that describes coronagraphs, verified them in the laboratory, and demonstrated them on-sky; we have also developed algorithms to tease out these faint signals against the almost overwhelming glare of their parent stars. It is perhaps inevitable that in the next decade we will be imaging and characterizing pale blue dots around our nearest neighbors, and we will take one further step toward learning whether the Earth is truly unique.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This review benefited from the 2023 Exoplanet Summer Program in the Other Worlds Laboratory at the University of California, Santa Cruz, a program funded by the Heising-Simons Foundation and NASA. Our research made use of NASA's Astrophysics Data System Bibliographic Services. M.A.K. acknowledges useful conversations with several people (most notably Phil Hinz, Eric Mamajek, and Andrew Skemer) at the Humble Sea Brewing Company.

LITERATURE CITED

Absil O, Delacroix C, Orban de Xivry G, et al. 2022. *Proc. SPIE* 12185:1218511
Absil O, Kenworthy M, Delacroix C, et al. 2024. *Proc. SPIE* 13096:1309652
Ahn K, Guyon O, Lozi J, et al. 2023. *Astron. Astrophys.* 673:A29
Anche RM, Ashcraft JN, Haffert SY, et al. 2023. *Astron. Astrophys.* 672:A121
Ashcraft JN, Choi H, Douglas ES, et al. 2022. *Proc. SPIE* 12180:121805L

Ashcraft JN, Douglas ES, Kim D, et al. 2023. *Proc. SPIE* 12664:1266404

Bailey V, Hildebrandt Rafels S. 2024. *CGI-flux-ratio-plot*. Software Package, GitHub. **https://github.com/nasavbailey/DI-flux-ratio-plot**

Balasubramanian K, Hoppe DJ, Mouroulis PZ, Marchen LF, Shaklan SB. 2005. *Proc. SPIE* 5905:59050H

Batalha NE, Smith AJRW, Lewis NK, et al. 2018. *Astron. J.* 156(4):158

Baudoz P, Boccaletti A, Baudrand J, Rouan D. 2006. *Proc. IAU* 200:C553–58

Baudoz P, Desgrange C, Galicher R, Laginja I. 2024. *Proc. SPIE* 13092:130926L

Belikov R, Sirbu D, Marx D, et al. 2022. *Proc. SPIE.* 12180:1218025

Bély P, Hasan H, Miebach M. 1993. *Orbital focus variations in the Hubble Space Telescope*. Instrum. Sci. Rep. SESD 93-16, Space Telesc. Sci. Inst., Baltimore, MD

Berger JP, Haguenauer P, Kern P, et al. 2001. *Astron. Astrophys.* 376:L31–34

Berry MV. 1984. *Proc. R. Soc. A* 392(1802):45–57

Bertrou-Cantou A, Gendron E, Rousset G, et al. 2022. *Astron. Astrophys.* 658:A49

Beuzit JL, Vigan A, Mouillet D, et al. 2019. *Astron. Astrophys.* 631:A155

Biller BA, Close LM, Masciadri E, et al. 2007. *Astrophys. J. Suppl.* 173:143–65

Boccaletti A, Riaud P, Baudoz P, et al. 2004. *Publ. Astron. Soc. Pac.* 116:1061–71

Bordé PJ, Traub WA. 2006. *Astrophys. J.* 638(1):488

Bos SP. 2021. *Astron. Astrophys.* 646:A177

Bos SP, Doelman DS, Lozi J, et al. 2019. *Astron. Astrophys.* 632:A48

Bos SP, Vievard S, Wilby MJ, et al. 2020. *Astron. Astrophys.* 639:A52

Bottom M, Walker SA, Cunnyngham I, Guthery C, Delorme JR. 2023. In *Adaptive Optics for Extremely Large Telescopes, 7th Edition* (*AOELT7*), art. 125. Avignon, Fr.: Palais Pap.

Brandl BR, Bettonvil F, van Boekel R, et al. 2022. *Proc. SPIE* 12184:1218421

Braunstein SL, Caves CM. 1994. *Phys. Rev. Lett.* 72(22):3439–43

Breckinridge JB, Lam WST, Chipman RA. 2015. *Publ. Astron. Soc. Pac.* 127:445–68

Brooks KJ, Catala L, Kenworthy MA, Crawford SM, Codona JL. 2016. *Proc. SPIE* 9912:991203

Cantalloube F, Por EH, Dohlen K, et al. 2018. *Astron. Astrophys.* 620:L10

Carlotti A, Baccar S, Leboulleux L, et al. 2023. In *Adaptive Optics for Extremely Large Telescopes, 7th Edition* (*AOELT7*), art. 45. Avignon, Fr.: Palais Pap.

Carlotti A, Vanderbei R, Kasdin NJ. 2011. *Opt. Express* 19(27):26796–809

Chanan G, Ohara C, Troy M. 2000. *Appl. Opt.* 39(25):4706–14

Chanan G, Troy M, Dekens F, et al. 1998. *Appl. Opt.* 37(1):140–55

Chauvin G. 2024. *C. R. Phys.* 24(Suppl. 2):139

Chipman RA. 1989. *Opt. Eng.* 28:90–99

Codona JL, Angel R. 2004. *Astrophys. J. Lett.* 604:L117–20

Codona JL, Kenworthy M. 2013. *Astrophys. J.* 767:100

Coyle LE, Knight JS, Pueyo L, et al. 2021. *Proc. SPIE* 11820:118200C

Currie T, Biller B, Lagrange A, et al. 2023. *Astron. Soc. Pac. Conf. Ser.* 534:799–838

Desai N, König L, Por E, et al. 2023. *Proc. SPIE* 12680:126801S

Desai N, Potier A, Redmond SF, et al. 2024. *J. Astron. Telesc. Instrum. Syst.* 10:035001

Deshler N, Haffert S, Ashok A. 2024a. arXiv:2403.17988 [quant-ph]

Deshler N, Ozer I, Ashok A, Guha S. 2024b. arXiv:2407.12776 [astro-ph.IM]

Doelman DS, Ouellet M, Potier A, et al. 2023. *Proc. SPIE* 12680:126802C

Doelman DS, Por EH, Ruane G, Escuti MJ, Snik F. 2020. *Publ. Astron. Soc. Pac.* 132(1010):045002

Doelman DS, Snik F, Por EH, et al. 2021a. *Appl. Opt.* 60(19):D52

Doelman DS, Wardenier JP, Tuthill P, et al. 2021b. *Astron. Astrophys.* 649:A168

Dohlen K, Langlois M, Lanzoni P, et al. 2006. *Proc. SPIE* 6267:626734

Echeverri D, Xuan JW, Monnier JD, et al. 2024. *Astrophys. J. Lett.* 965(2):L15

Fienup JR. 2013. *Appl. Opt.* 52(1):45

Fitzgerald MP, Sallum S, Millar-Blanchaer MA, et al. 2022. *Proc. SPIE* 12184:1218426

Franson K, Balmer WO, Bowler BP, et al. 2024. *Ap. J. Lett.* 974(1):L11

Frazin RA, Rodack AT. 2021. *J. Opt. Soc. Am. A* 38(10):1557–69

Galicher R, Baudoz P, Rousset G, Totems J, Mas M. 2010. *Astron. Astrophys.* 509:A31
Galicher R, Potier A, Mazoyer J, et al. 2024. *Astron. Astrophys.* 686:A54
Gerard BL, Dillon D, Cetre S, Jensen-Clem R. 2022. *J. Astron. Telesc. Instrum. Syst.* 8:039001
Gerard BL, Dillon D, Cetre S, Jensen-Clem R. 2023. *Publ. Astron. Soc. Pac.* 135(1044):024502
Gerard BL, Marois C, Galicher R. 2018. *Astron. J.* 156(3):106
Gerchberg RW, Saxton WO. 1972. *Optik* 35(2):237–46
Give'On A. 2009. *Proc. SPIE* 7440:74400D
Give'On A, Kern BD, Shaklan S. 2011. *Proc. SPIE* 8151:815110
Gladysz S, Christou JC. 2009. *Astrophys. J.* 698:28–42
Gledhill TM, Scarrott SM, Wolstencroft RD. 1991. *MNRAS* 252:50–54
Goebel SB, Guyon O, Hall DNB, et al. 2018. *Publ. Astron. Soc. Pac.* 130(992):104502
Gonsalves RA. 2014. *Proc. SPIE* 9148:91482P
Gonté F, Araujo C, Bourtembourg R, et al. 2009. *Messenger* 136:25–31
Guyon O. 2003. *Astron. Astrophys.* 404:379–87
Guyon O. 2005. *Astrophys. J.* 629:592–614
Guyon O. 2018. *Annu. Rev. Astron. Astrophys.* 56:315–55
Guyon O. 2023. *Input star catalog*. Res. Rep., Univ. Ariz., Tucson. **https://subarutelescope.org/staff/guyon/04research.web/14hzplanetsELTs.web/catalog.web/content.html**
Guyon O, Hinz PM, Cady E, Belikov R, Martinache F. 2014. *Astrophys. J.* 780(2):171
Guyon O, Matsuo T, Angel R. 2009. *Astrophys. J.* 693(1):75–84
Guyon O, Pluzhnik EA, Galicher R, et al. 2005. *Astrophys. J.* 622:744–58
Haffert S, Liberman J, Males J, et al. 2024. *Proc. SPIE* 13097:130971K
Haffert S, Males J, Ahn K, et al. 2023. *Astron. Astrophys.* 673:A28
Haffert SY. 2021. *J. Opt. Soc. Am. B* 38(7):A27–35
Haffert SY. 2022. *Astron. Astrophys.* 659:A51
Haffert SY, Bohn AJ, de Boer J, et al. 2019. *Nat. Astron.* 3:749–54
Haffert SY, Close LM, Hedglen AD, et al. 2022. *J. Astron. Telesc. Instrum. Syst.* 8:021513
Haffert SY, Males JR, Guyon O. 2023a. arXiv:2310.10892 [astro-ph.IM]
Haffert SY, Males JR, Guyon O. 2023b. In *Adaptive Optics for Extremely Large Telescopes, 7th Edition* (*AOELT7*), art. 79. Avignon, Fr.: Palais Pap.
Haffert SY, Por EH, Keller CU, et al. 2020. *Astron. Astrophys.* 635:A56
Haguenauer P, Serabyn E. 2006. *Appl. Opt.* 45(12):2749–54
Hedglen AD, Close LM, Haffert SY, et al. 2022. *J. Astron. Telesc. Instrum. Syst.* 8:021515
Herscovici-Schiller O, Mugnier LM, Baudoz P, et al. 2018. *Astron. Astrophys.* 614:A142
Hoeijmakers HJ, Schwarz H, Snellen IAG, et al. 2018. *Astron. Astrophys.* 617:A144
Houllé M, Vigan A, Carlotti A, et al. 2021. *Astron. Astrophys.* 652:A67
Huby E, Baudoz P, Vidal F, et al. 2024. *Proc. SPIE* 13097:1309776
Huby E, Bottom M, Femenia B, et al. 2017. *Astron. Astrophys.* 600:A46
Jacquinot P, Roizen-Dossier B. 1964. II. Apodisation. In *Progress in Optics*, Vol. 3, ed. E Wolf, pp. 29–186. Amsterdam: North-Holland
Jensen-Clem R, Hinz P, von Kooten M, et al. 2021. *Proc. SPIE* 11823:1182309
Jovanovic N, Gatkine P, Anugu N, et al. 2023. *J. Phys. Photonics* 5:042501
Jovanovic N, Schwab C, Guyon O, et al. 2017. *Astron. Astrophys.* 604:A122
Kaltenegger L. 2017. *Annu. Rev. Astron. Astrophys.* 55:433–85
Kasdin NJ, Vanderbei RJ, Littman MG, Spergel DN. 2005. *Appl. Opt.* 44(7):1117–28
Kasper M, Cerpa Urra N, Pathak P, et al. 2021. *Messenger* 182:38–43
Keller CU, Korkiakoski V, Doelman N, et al. 2012. *Proc. SPIE* 8447:844721
Kendrew S, Jolissaint L, Mathar RJ, et al. 2008. *Proc. SPIE* 7015:70155T
Kenworthy M, Quanz S, Meyer M, et al. 2010. *Messenger* 141:2–4
Kenworthy MA, Absil O, Agócs T, et al. 2016. *Proc. SPIE* 9908:9908A6
Kenworthy MA, Codona JL, Hinz PM, et al. 2007. *Astrophys. J.* 660:762–69
Kirkpatrick JD, Marocco F, Gelino CR, et al. 2024. *Astrophys. J. Suppl.* 271(2):55

Komanduri RK, Lawler KF, Escuti MJ. 2013. *Opt. Express* 21(1):404–20
Kopon D, Close LM, Males JR, Gasho V. 2013. *Publ. Astron. Soc. Pac.* 125(930):966
Krist JE, Steeves JB, Dube BD, et al. 2023. *J. Astron. Telesc. Instrum. Syst.* 9:045002
Kuchner MJ, Traub WA. 2002. *Astrophys. J.* 570(2):900–8
Kueny JK, Van Gorkom K, Kautz M, et al. 2024. *Proc. SPIE* 13097:1309761
Lacy B, Burrows A. 2020. *Astrophys. J.* 892(2):151
Laginja I, Doelman D, Snik F, et al. 2024. *Proc. SPIE* 13092:1309280
Laginja I, Sauvage JF, Mugnier LM, et al. 2022. *Astron. Astrophys.* 658:A84
Lajoie CP, Lallo M, Meléndez M, et al. 2023. *OTE science performance memo 2—a year of wavefront sensing with* JWST *in flight. Cycle 1 telescope monitoring & maintenance summary*. Tech. Rep. JWST-STScI-008497, Space Telesc. Sci. Inst., Baltimore, MD
Lavrinenko A, Lægsgaard J, Gregersen N, Schmidt F, Søndergaard T. 2014. *Numerical Methods in Photonics*. Boca Raton, FL: CRC
Leboulleux L, Carlotti A, Curaba S, et al. 2022a. *Proc. SPIE* 12188:121884J
Leboulleux L, Carlotti A, N'Diaye M. 2022b. *Astron. Astrophys.* 659:A143
Leboulleux L, Carlotti A, N'Diaye M, et al. 2022c. *Astron. Astrophys.* 666:A91
Leboulleux L, Sauvage JF, Soummer R, et al. 2020. *Astron. Astrophys.* 639:A70
Leon-Saval SG, Argyros A, Bland-Hawthorn J. 2010. *Opt. Express* 18(8):8430–39
Leon-Saval SG, Argyros A, Bland-Hawthorn J. 2013. *Nanophotonics* 2:429–40
Levraud N, Chambouleyron V, Sauvage JF, et al. 2024. *Astron. Astrophys.* 682:A84
Lyot B. 1933. *J. R. Astron. Soc. Can.* 27:265–80
Lyot B. 1939. *MNRAS* 99:580–94
Malbet F, Yu J, Shao M. 1995. *Publ. Astron. Soc. Pac.* 107(710):386–98
Males JR, Close LM, Haffert SY, et al. 2024. *Proc. SPIE* 13096:130960Y
Marois C, Doyon R, Nadeau D, et al. 2005. *Publ. Astron. Soc. Pac.* 117:745–56
Marois C, Lafrenière D, Doyon R, Macintosh B, Nadeau D. 2006. *Astrophys. J.* 641:556–64
Martinache F. 2013. *Publ. Astron. Soc. Pac.* 125(926):422–30
Martinache F, Guyon O, Clergeon C, Blain C. 2012. *Publ. Astron. Soc. Pac.* 124(922):1288–94
Martinache F, Guyon O, Jovanovic N, et al. 2014. *Publ. Astron. Soc. Pac.* 126(940):565–72
Martinez P. 2019. *Astron. Astrophys.* 629:L10
Martinez P, Loose C, Aller Carpentier E, Kasper M. 2012. *Astron. Astrophys.* 541:A136
Marx D, Belikov R, Sirbu D, et al. 2021. *Proc. SPIE* 11823:118230O
Mawet D, Pueyo L, Carlotti A, et al. 2013. *Astrophys. J. Suppl.* 209:7
Mawet D, Riaud P, Absil O, Surdej J. 2005. *Astrophys. J.* 633:1191–200
Mawet D, Ruane G, Xuan W, et al. 2017. *Astrophys. J.* 838(2):92
Mawet D, Serabyn E, Moody D, et al. 2011. *Proc. SPIE* 8151:81511D
Mayor M, Queloz D. 1995. *Nature* 378(6555):355–59
McLeod RR, Wagner KH. 2014. *Adv. Opt. Photonics* 6(4):368
Meimon S, Fusco T, Mugnier LM. 2010. *Opt. Lett.* 35(18):3036–38
Mendillo CB, Hewawasam K, Martel J, et al. 2021. *Proc. SPIE* 11823:118231T
Mennesson B, Belikov R, Por E, et al. 2024. *J. Astron. Telesc. Instrum. Syst.* 10:035004
Millar-Blanchaer MA, Anche RM, Nguyen MM, Maire J. 2022. *Proc. SPIE* 12184:121843X
Miller K, Guyon O, Males J. 2017. *J. Astron. Telesc. Instrum. Syst.* 3:049002
Miller K, Males JR, Guyon O, et al. 2019. *J. Astron. Telesc. Instrum. Syst.* 5:049004
Miller S, Jiang L, Pau S. 2022. *Opt. Express* 30(12):20629–46
Milli J, Kasper M, Bourget P, et al. 2018. *Proc. SPIE* 10703:107032A
Minardi S, Harris RJ, Labadie L. 2021. *Astron. Astrophys. Rev.* 29(1):6
Miskiewicz MN, Escuti MJ. 2014. *Opt. Express* 22(10):12691–706
Nickson BF, Por EH, Nguyen MM, et al. 2022. *Proc. SPIE* 12180:121805K
Norris B, Betters C, Wei J, et al. 2022. *Opt. Express* 30(19):34908
Orban de Xivry G, Quesnel M, Vanberg P, Absil O, Louppe G. 2021. *MNRAS* 505(4):5702–13
Otten GPPL, Snik F, Kenworthy MA, et al. 2017. *Astrophys. J.* 834:175

Otten GPPL, Snik F, Kenworthy MA, Miskiewicz MN, Escuti MJ. 2014. *Opt. Express* 22:30287

Pancharatnam S. 1956. *Proc. Indian Acad. Sci. A* 44(5):247–62

Paul B, Mugnier LM, Sauvage JF, Ferrari M, Dohlen K. 2013. *Opt. Express* 21(26):31751

Pfrommer T, Lewis S, Kosmalski J, et al. 2018. *Proc. SPIE* 10700:107003F

Pogorelyuk L, Kasdin NJ. 2019. *Astrophys. J.* 873(1):95

Por EH. 2017. *Proc. SPIE* 10400:104000V

Por EH. 2020. *Astrophys. J.* 888(2):127

Por EH, Haffert SY. 2020. *Astron. Astrophys.* 635:A55

Potier A, Galicher R, Baudoz P, et al. 2020. *Astron. Astrophys.* 638:A117

Potier A, Mazoyer J, Wahhaj Z, et al. 2022. *Astron. Astrophys.* 665:A136

Pourcelot R, N'Diaye M, Por EH, et al. 2022. *Astron. Astrophys.* 663:A49

Pourcelot R, Por EH, N'Diaye M, et al. 2023. *Astron. Astrophys.* 672:A73

Pourré N, Le Bouquin JB, Milli J, et al. 2022. *Astron. Astrophys.* 665:A158

Pueyo L, Kay J, Kasdin NJ, et al. 2009. *Appl. Opt.* 48(32):6296–312

Quanz SP, Amara A, Meyer MR, et al. 2013. *Astrophys. J. Lett.* 766(1):L1

Quanz SP, Meyer MR, Kenworthy MA, et al. 2010. *Astrophys. J. Lett.* 722(1):L49–53

Quesnel M, Orban de Xivry G, Louppe G, Absil O. 2022. *Astron. Astrophys.* 668:A36

Quirós-Pacheco F, Bouchez A, Plantet C, et al. 2024. *Proc. SPIE* 13097:1309725

Quirós-Pacheco F, van Dam M, Bouchez AH, et al. 2022. *Proc. SPIE* 12185:1218517

Redmond SM, Kasdin NJ, Pogorelyuk L, et al. 2020. *Proc. SPIE* 11443:114432K

Rodack AT, Frazin RA, Males JR, Guyon O. 2021. *J. Opt. Soc. Am. A* 38(10):1541–56

Roddier F, Roddier C. 1997. *Publ. Astron. Soc. Pac.* 109(737):815–20

Rouan D, Riaud P, Boccaletti A, Clénet Y, Labeyrie A. 2000. *Publ. Astron. Soc. Pac.* 112:1479–86

Ruane G, Echeverri D, Jovanovic N, et al. 2019. *Proc. SPIE* 11117:1111716

Ruane G, Riggs A, Mazoyer J, et al. 2018. *Proc. SPIE* 10698:106982S

Ruane G, Wallace JK, Riggs AJE, et al. 2023. *Proc. SPIE* 12680:1268018

Ruane G, Wallace JK, Steeves J, et al. 2020. *J. Astron. Telesc. Instrum. Syst.* 6:045005

Ruane G, Wang J, Mawet D, et al. 2018. *Astrophys. J.* 867(2):143

Salama M, Guthery C, Chambouleyron V, et al. 2024. *Astrophys. J.* 967(2):171

Sauvage JF, Fusco T, Lamb M, et al. 2016. *Proc. SPIE* 9909:990916

Sauvage JF, Fusco T, Petit C, et al. 2012. *Proc. SPIE* 8447:844715

Schmid HM, Bazzon A, Roelfsema R, et al. 2018. *Astron. Astrophys.* 619:A9

Schneider G, Smith BA, Becklin EE, et al. 1999. *Astrophys. J. Lett.* 513:L127–30

Schneider G, Thompson RI, Smith BA, Terrile RJ. 1998. *Proc. SPIE* 3356. **https://doi.org/10.1117/12.324461**

Schwieterman EW, Kiang NY, Parenteau MN, et al. 2018. *Astrobiology* 18(6):663–708

Seager S, Bains W, Petkowski JJ. 2016. *Astrobiology* 16(6):465–85

Seo BJ, Patterson K, Balasubramanian K, et al. 2019. *Proc. SPIE* 11117:111171V

Serabyn E, Mennesson B. 2006. *Proc. IAU* 200:379–84

Shaklan S, Roddier F. 1988. *Appl. Opt.* 27(11):2334–38

Singh G, Lozi J, Guyon O, et al. 2015. *Publ. Astron. Soc. Pac.* 127(955):857–69

Singh G, Martinache F, Baudoz P, et al. 2014. *Publ. Astron. Soc. Pac.* 126(940):586–94

Skemer AJ, Hinz PM, Hoffmann WF, et al. 2009. *Publ. Astron. Soc. Pac.* 121(882):897–904

Smith BA, Terrile RJ. 1984. *Science* 226(4681):1421–24

Snellen I, de Kok R, Birkby JL, et al. 2015. *Astron. Astrophys.* 576:A59

Snellen IAG, de Kok RJ, de Mooij EJW, Albrecht S. 2010. *Nature* 465(6928):1049–51

Snik F, Otten G, Kenworthy M, et al. 2012. *Proc. SPIE* 8450:84500M

Snik F, Otten G, Kenworthy M, Mawet D, Escuti M. 2014. *Proc. SPIE* 9147:91477U

Soummer R, Aimé C, Falloon P. 2003a. *Astron. Astrophys.* 397(3):1161–72

Soummer R, Dohlen K, Aimé C. 2003b. *Astron. Astrophys.* 403(1):369–81

Soummer R, Por EH, Pourcelot R, et al. 2022. *Proc. SPIE* 12180:1218026

Sparks WB, Ford HC. 2002. *Astrophys. J.* 578(1):543–64

Stark CC, Mennesson B, Bryson S, et al. 2024. *J. Astron. Telesc. Instrum. Syst.* 10:034006
Steiger S, Currie T, Brandt TD, et al. 2021. *Astron. J.* 162(2):44
Sturm E, Davies R, Alves J, et al. 2024. *Proc. SPIE* 13096:1309611
Sutlieff BJ, Birkby JL, Stone JM, et al. 2023. *MNRAS* 520(3):4235–57
Sutlieff BJ, Birkby JL, Stone JM, et al. 2024. *MNRAS* 531(1):2168–89
Sutlieff BJ, Bohn AJ, Birkby JL, et al. 2021. *MNRAS* 506(3):3224–38
Thatte N, Abuter R, Tecza M, et al. 2007. *MNRAS* 378:1229–36
Thatte NA, Melotte D, Neichel B, et al. 2022. *Proc. SPIE* 12184:1218420
Tsang M, Nair R, Lu XM. 2016. *Phys. Rev. X* 6:031033
Turnbull MC, Traub WA, Jucks KW, et al. 2006. *Astrophys. J.* 644(1):551–59
van den Born J, Jellema W. 2020. *MNRAS* 496(4):4266–75
Van Gorkom K, Douglas ES, Ashcraft JN, et al. 2022. *Proc. SPIE* 12180:121805M
Van Gorkom K, Males JR, Close LM, et al. 2021. *J. Astron. Telesc. Instrum. Syst.* 7:039001
van Holstein RG, Keller CU, Snik F, Bos SP. 2023. *Astron. Astrophys.* 677:A150
van Kooten MAM, Ragland S, Jensen-Clem R, et al. 2022. *Astrophys. J.* 932(2):109
Vievard S, Bonnefois A, Cassaing F, Montri J, Mugnier LM. 2020. *J. Astron. Telesc. Instrum. Syst.* 6:040501
Vievard S, Bos S, Cassaing F, et al. 2019. arXiv:1912.10179 [astro-ph.IM]
Vigan A, Dohlen K, N'Diaye M, et al. 2022. *Astron. Astrophys.* 660:A140
Vilas F, Smith BA. 1987. *Appl. Opt.* 26:664–68
Wilby MJ, Keller CU, Sauvage JF, et al. 2018. *Astron. Astrophys.* 615:A34
Wilby MJ, Keller CU, Snik F, Korkiakoski V, Pietrow AGM. 2017. *Astron. Astrophys.* 597:A112
Wolszczan A, Frail DA. 1992. *Nature* 355(6356):145–47
Xin Y, Jovanovic N, Ruane G, et al. 2022. *Astrophys. J.* 938(2):140
Zhang Y, Andreas Noack M, Vagovic P, et al. 2021. *Opt. Express* 29(13):19593–604
Zimmerman NT, N'Diaye M, Laurent KES, et al. 2016. *Proc. SPIE* 9904:99041Y
Zurlo A. 2024. arXiv:2404.05797 [astro-ph.EP]

RELATED RESOURCES

To achieve the scientific results presented in this article, we made use of the `Python` programming language (see **https://www.python.org**), especially the `SciPy` (Virtanen et al. 2020), `NumPy` (Oliphant 2006), `Matplotlib` (Hunter 2007), `emcee` (Foreman-Mackey et al. 2013), and `astropy` (Astropy Collab. et al. 2013, 2018) packages. The HCI simulations were calculated with `HCIpy` (Por et al. 2018). This article is reproducible using the `showyourwork!` workflow management tool (Luger et al. 2021), and the source code used to generate each figure is available in a public GitHub repository (see **https://github.com/mkenworthy/ARAA_HCC**).

Astropy Collab., Price-Whelan AM, Sipőcz BM, et al. 2018. *Astron. J.* 156:123
Astropy Collab., Robitaille TP, Tollerud EJ, et al. 2013. *Astron. Astrophys.* 558:A33
Foreman-Mackey D, Hogg DW, Lang D, Goodman J. 2013. *Publ. Astron. Soc. Pac.* 125(925):306–12
Hunter JD. 2007. *Comput. Sci. Eng.* 9:90–95
Luger R, Bedell M, Foreman-Mackey D, et al. 2021. arXiv:2110.06271 [astro-ph.SR]
Oliphant TE. 2006. *A Guide to NumPy*. Vol. 1. tramy.us: Trelgol
Por EH, Haffert SY, Radhakrishnan VM, et al. 2018. *Proc. SPIE* 10703:1070342
Virtanen P, Gommers R, Oliphant TE, et al. 2020. *Nat. Methods* 17:261–72