

# Laboratory demonstration of real-time focal plane wavefront control of residual atmospheric speckles

Benjamin L. Gerard<sup>a,b,\*</sup>, Daren Dillon<sup>a,b</sup>, Sylvain Cetre<sup>b,c,d</sup>  
and Rebecca Jensen-Clem<sup>a,b</sup>

<sup>a</sup>University of California Observatories, Santa Cruz, California, United States

<sup>b</sup>University of California Santa Cruz, Santa Cruz, California, United States

<sup>c</sup>W.M. Keck Observatory, Waimea, Hawaii, United States

<sup>d</sup>Durham University, Durham, United Kingdom

**Abstract.** Current and future high contrast imaging instruments aim to detect exoplanets at closer orbital separations, lower masses, and/or older ages than their predecessors. However, continually evolving speckles in the coronagraphic science image limit contrasts of state-of-the-art ground-based exoplanet imaging instruments. For ground-based adaptive optics (AO) instruments, it remains challenging for most speckle suppression techniques to attenuate both the dynamic atmospheric as well as quasistatic instrumental speckles on-sky. We have proposed a focal plane wavefront sensing and control algorithm to address this challenge, called the fast atmospheric selfcoherent camera (SCC) technique (FAST), which in theory enables the SCC to operate down to millisecond timescales even when only a few photons are detected per speckle. Here, we present the first experimental results of FAST on the Santa Cruz Extreme AO Laboratory (SEAL) testbed. In particular, we illustrate the benefit of “second stage” AO-based focal plane wavefront control, demonstrating up to 5× contrast improvement with FAST closed-loop compensation of evolving residual atmospheric turbulence—both for low and high order spatial modes—down to 20-ms timescales. © The Authors. Published by SPIE under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JATIS.8.3.039001](https://doi.org/10.1117/1.JATIS.8.3.039001)]

**Keywords:** wavefront sensing; wavefront control; coronagraphy; adaptive optics.

Paper 21176G received Dec. 21, 2021; accepted for publication Jun. 17, 2022; published online Jul. 4, 2022.

## 1 Introduction

Current state-of-the-art instrumentation enables ground-based exoplanet imaging instruments to be sensitive to detecting self-luminous giant exoplanets down to a few Jupiter masses at separations beyond around 10 au within star systems that are within around 100 pc and are younger than a few hundred Myrs.<sup>1</sup> Despite exciting new detections enabled by these instruments (e.g., Ref. 2), large surveys have both (a) shown that the giant exoplanets these facilities are sensitive to detecting are rare—as low as 1% of stars host such planets<sup>3</sup>—and as a result (b) illustrated that future instruments will need to improve detection and characterization sensitivity to lower mass, closer-in, and older systems. The main technological factor limiting such improved sensitivity for near-infrared imaging is uncorrected/subtracted speckle noise.<sup>4</sup> Speckle noise is a result of leftover diffracted starlight in the coronagraphic science image that prevents detecting exoplanets below a flux ratio (i.e., planet flux normalized to the host star flux), of about  $10^{-6}$  at radial separations at  $10\lambda/D$  from the star.<sup>2</sup> It is therefore crucial to improve on these speckle subtraction and correction techniques to enable both the current and next generation of exoplanet imagers to detect and characterize new exoplanetary systems, shedding light on the processes of planet formation, evolution, and ultimately the prevalence of life beyond the solar system.<sup>5</sup>

Focal plane wavefront sensing and control<sup>6</sup> for ground-based 8 m-class telescopes—using either an active feedback loop between the coronagraphic science image and adaptive optics

---

\*Address all correspondence to Benjamin L. Gerard, [blgerard@ucsc.edu](mailto:blgerard@ucsc.edu)

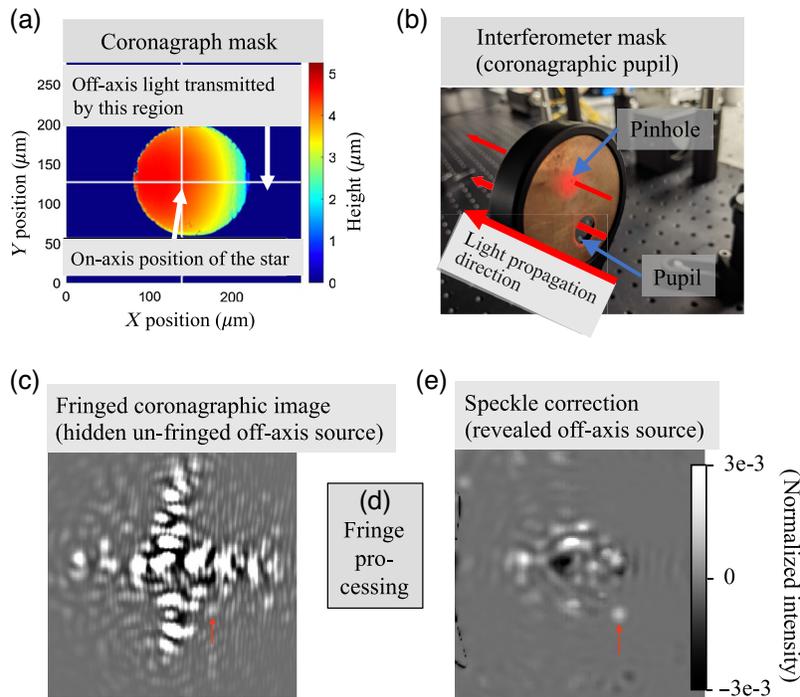
(AO) system’s deformable mirror (DM) and/or via postprocessing methods (which in this paper we will refer to as coherent differential imaging, or CDI)—is one such speckle correction technology that could enable the aforementioned sensitivity/contrast gains needed for fainter exoplanet detection and characterization. However, minimal on-sky gains with this approach have been demonstrated thus far,<sup>7,8</sup> largely due to speckle evolution occurring on timescales faster than measurement and correction algorithms and/or system hardware can enable.<sup>4</sup> However, a number of projects are now being pursued to further develop the promise of this technology on-sky, including ongoing efforts at Keck/NIRC2,<sup>9</sup> Subaru/SCEXAO,<sup>8,10</sup> Magellan/MagAO-X,<sup>11</sup> ESO/SPHERE,<sup>12</sup> and Gemini/GPI.<sup>13</sup> In this paper, we will discuss related laboratory developments of one such technology, based on the self-coherent camera (SCC),<sup>14</sup> called the fast atmospheric SCC technique (FAST),<sup>4,15</sup> using the Santa Cruz Extreme AO Laboratory (SEAL).<sup>16</sup> This FAST concept was initially developed at the high contrast testbed at the National Research Council of Canada Herzberg Astronomy and Astrophysics research center, which has published a laboratory optical design and preliminary FAST testing results in Ref. 17. A conference proceeding publication of preliminary SEAL testing and development that led to this paper was published initially in Ref. 18.

We summarize the FAST concept in Sec. 2 (with additional background given in [Appendix A](#)), provide an overview of our FAST SEAL setup in Sec. 3, and describe various FAST calibration wavefront reconstruction procedures in Sec. 4. We present results and analysis of closed-loop FAST correction in Sec. 5, including for static dark hole (DH) generation (Sec. 5.1), quasi-static on-air aberrations (Sec. 5.2), and AO residual turbulence (Sec. 5.3), clearly demonstrating the gain that real-time focal plane wavefront control enables. We then provide further discussion in Sec. 6 and conclude in Sec. 7.

## 2 FAST Summary

FAST, illustrated and described in Fig. 1, relies on a technique in which coherent starlight is interfered with itself to enable a measurement and correction of the complex speckle electric field (i.e., phase and amplitude) in the coronagraphic image. See [Appendix A](#) for a detailed description of the SCC,<sup>14</sup> upon which FAST is based. Speckle attenuation/subtraction with this approach can be accomplished using a DM (i.e., focal plane wavefront control)<sup>19</sup> and/or by postprocessing (i.e., CDI).<sup>15,18</sup> In this paper, we will only focus on FAST DM control. As shown in Fig. 1, FAST produces stellar interference fringes by way of two specialized coronagraphic masks: a setup-specific focal plane mask (FPM) [Fig. 1(a), illustrating a tip + Gaussian component within  $3\lambda/D$  and a piston component elsewhere] modulates starlight but not exoplanet light such that the downstream pupil plane mask [Fig. 1(b)] transmits remaining starlight (i.e., including residual aberrations from the atmosphere, telescope and instrument optics, and coronagraph mask diffraction effects) and exoplanet light through a traditional Lyot stop, whereas much of the starlight (but not exoplanet light) passes through a separate pinhole in the mask. In the final focal plane (the coronagraphic image), the residual stellar speckles interfere with the starlight that passed through the pinhole to produce fringes [Fig. 1(c)], in principle even when only a few photoelectrons ( $\gtrsim 10$ ) are detected per pixel,<sup>15</sup> while any off-axis sources remain unfringed). FAST simulations have shown that Fourier-based processing of the fringes in this final image (see [Appendix A](#) for details of this process) enables estimating and attenuating (via DM control and/or CDI-based postprocessing) remaining stellar speckles from a single, millisecond-exposure image, with no need to create additional phase diversity by defocusing and/or probing individual speckles with the DM [Fig. 1(e)].

FAST’s key innovation compared with previous focal plane wavefront control techniques is the coronagraphic FPM [Fig. 1(a)], which provides a high enough fringe visibility (i.e., fringed versus unfringed components) in the final image such that fringes can in theory be detected on millisecond timescales, even when only a few photons are detected per pixel. The classical SCC also generates fringes in the coronagraphic image to enable focal plane wavefront control and CDI<sup>14</sup> but with fringe visibilities about  $10^6$  times lower than Fig. 1(c) [e.g., see Fig. 10(b) in [Appendix Sec. 8.1](#)], removing any possibility of AO-based wavefront control on millisecond timescales. Also due to the nature of the FAST FPM, which sends the core of the point-spread-function (PSF) (i.e., low order modes of the entrance pupil wavefront) into the off-axis Lyot stop



**Fig. 1** Overview of FAST (from Ref. 18) using images from our Santa Cruz Extreme AO Laboratory (SEAL) high contrast testbed<sup>16</sup> (these FAST/SEAL laboratory results will be presented throughout this paper and are shown here to first illustrate the FAST concept). A setup-specific coronagraph mask (a) is designed to both (i) attenuate starlight while transmitting exoplanet light and (ii) enhance the stellar fringe visibility using an interferometer mask in the downstream coronagraphic pupil (b). The FPM shape, as shown in (a), includes a Gaussian and tilt component, designed to redistribute starlight into the pinhole aperture in (b) that is within  $3\lambda/D$  of the optical axis in (a), and a flat/piston component that sends star and exoplanet light beyond  $3\lambda/D$  in (a) into the pupil aperture in (b). This design enables the detection of an interference pattern on top of the leftover stellar speckles in the coronagraphic image (c), enabling coronagraphic science imaging and AO wavefront sensing. The fringed stellar speckles are then processed (d) and further attenuated using both a DM and via postprocessing (e), revealing an unfringed off-axis source. See Secs. 4 and 5 for more details on (d) and (e).

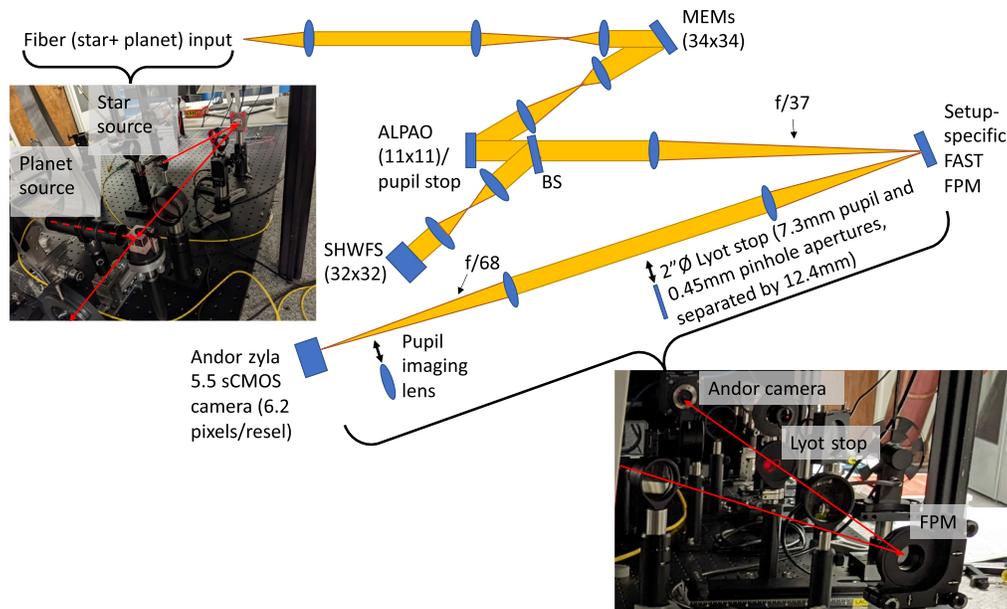
pinhole, in turn transmitted through to the SCC image, on-axis FAST images are highly sensitive to detecting low order modes of the entrance pupil wavefront;<sup>20</sup> typically, other coronagraphs usually used with the SCC are comparatively not as sensitive to such modes (see Sec. 4.2).

### 3 Seal Laboratory Setup and Alignment

#### 3.1 Setup

In this section, we will describe our refractive setup at UCSC's Laboratory for AO facilities developed to test FAST, which is the same setup from which subsequent results in this paper are obtained. Although we will refer to this setup as the Santa Cruz Extreme AO Laboratory (SEAL), SEAL will ultimately be a mostly reflective setup designed for multipurpose AO-based wavefront sensing and control techniques, beyond the scope of just testing FAST; the higher-level setup of the full SEAL testbed, including the optical design and both current and ongoing projects (including FAST), are presented and described in Ref. 16. Note that the main facilities used here are the same as previously described in Ref. 21, including a highly stabilized granite testbed and custom-made enclosure.

Figure 2 shows our SEAL testbed setup and outlines the main hardware components relevant to FAST testing. The visible light star and planet light sources are commercial Thorlabs lasers



**Fig. 2** An outline of our FAST SEAL setup, showing a conceptual diagram of the full setup and an inset of testbed images in two areas. In the inset, red arrows indicate light propagation (with different line styles combined/split from/into different paths). Double-sided black arrows indicate a deployable mechanism. Starting in the upper left, a planet and star light source, both aligned to the same focal plane, are collimated and then combined, with the planet source positioned slightly off-axis a few diffraction bandwidths from the star. Re-imaging optics place the pupil first on the  $29 \times 29$  (across the beam) MEMS DM, and then on the  $11 \times 11$  ALPAO DM, the latter defining the system pupil stop. A plate beam splitter (labeled as “BS”) reflects light to an SHWFS with 32 subapertures across the beam diameter and transmits light to the FAST arm. The FAST beam is then focused on our setup-specific FPM, then recollimated to a downstream coronagraphic pupil, where a deployable Lyot stop can modulate the beam (the central and off-axis coronagraphic pupil are visible on the Lyot stop in the lower left inset). Finally, downstream imaging optics enable coronagraphic focal plane (i.e., FAST) or pupil plane imaging on an Andor camera.

KLS635 and HNLS008L, respectively, centered at 633 nm. The adjustable star light source is set to 0.15 mW unless otherwise noted, and the planet light source is fixed at 0.8 mW. We place a neutral density (ND) filter with an optical density (OD) of 2 on the planet source. The beamcube combining the two light sources further attenuates the planet by a factor of 9, theoretically setting the planet at a flux ratio of  $6 \times 10^{-3}$  (see Sec. 4 for a measurement of the star-to-planet flux ratio). Although for most systems this flux ratio is unlikely to yield planet -mass sensitivity (i.e., more realistically in the brown dwarf mass sensitivity range), for the purposes of this paper (i.e., demonstrating real-time correction of AO residuals, not demonstrating long-exposure high-contrast capabilities), we will subsequently refer to this off-axis source as the planet light source. Additional higher OD ND filters (and/or increasing the star light source power) can further decrease this flux ratio for future high contrast experiments, but for the purposes of our main focus in this paper (i.e., real-time FAST correction of AO residuals) the above-described planet flux ratio is sufficient. Moving downstream, a set of relay lenses then images the system pupil (note that the system pupil stop is defined further downstream; see below) onto our  $34 \times 34$  actuator microelectrical-mechanical system DM (MEMS; 952 actuators forming a circular pupil), illuminating 29 actuators across the system beam diameter (but see the Lyot stop sizing implications of the MEMS footprint two paragraphs below). Reimaging optics then relays a pupil plane to an  $11 \times 11$  actuator ALPAO DM (97 actuators forming a circular pupil), also serving as the primary system pupil stop (i.e., a circular unobscured pupil geometry; back-propagation determines the relevant illuminated pupil on the MEMS). A plate beam splitter then reflects 50% of the beam through a pair of relay lenses to a commercial Thorlabs Shack

Hartmann wavefront sensor (SHWFS; model WFS-20), sampling 32 lenslets across the beam diameter. The transmitted light is sent through a 500-mm focal length lens to generate an  $f/37$  beam on the FPM, which is designed and fabricated specifically for that  $f$ -number and the 633-nm light source wavelength.

Figures 1(a) and 2 already show the fully reflective setup-specific FAST tip/tilt Gaussian (TG) FPM fabricated and characterized via Zygo interferometry measurements by University of Alberta's nanoFAB laboratory, made from a nanoscribe three-dimensional (3D) printing machine and subsequently aluminum coated. This mask design only deviates from a flat surface over the central region (a  $3\lambda/D$  radius from the optical axis) that redistributes the core starlight, and with this 3D printing approach, the full tilt can be made without phase wrapping (i.e., avoiding additional related chromatic effects). However, the design is also essentially a Lyot coronagraph, with no component optimized for diffracted starlight suppression, and so it is not ideal for reaching high contrasts. However, such a mask is still sufficient for the focus of this paper (i.e., ground-based observations where coronagraphic images are dominated by unpinned speckle noise). Further characterization of this mask is presented in Ref. 18, which shows that it is fabricated at sufficient quality to enable high fringe visibility SCC images to enable real-time wavefront control, which is the main focus of this paper.

An  $f = 200$  mm, 2"Ø lens then collimates the post-FPM beam, oversizing the clear aperture relative to the noncoronagraphic pupil footprint by a factor of 6.2, sufficient for the "classical" SCC.<sup>22</sup> A Lyot stop then transmits light through the central 7.3-mm coronagraphic pupil (90% undersized) and off-axis 0.45 mm pinhole (which is the maximum pinhole diameter required to enable the first pinhole PSF Airy minimum to lie outside the DM control region). The 90% undersized Lyot stop pupil means that the 29 actuators sampled across the noncoronagraphic pupil become only 26 actuators sampled across the coronagraphic pupil. Avoiding edge effects, we found that using 24 actuators across the MEMS pupil diameter enabled sufficient performance in the FAST coronagraphic image. Although the FAST FPM prescription is designed for a theoretical pinhole–pupil separation (center-to-center) of 1.6 pupil diameters, the separation is instead empirically determined to be 1.524 pupil diameters (see Sec. 3.2). An  $f = 500$  mm 2"Ø lens lastly focuses the post-Lyot stop beam onto our Andor Zyla 5.5 sCMOS camera, enabling a theoretical plate scale of 6 pixels/resel (see Sec. 9.3 for a corresponding measurement), sufficiently oversampled for the SCC fringes.<sup>22</sup> An  $f = 50$  mm lens is placed 50 mm from the Andor detector on a flip mount to reimage the coronagraphic pupil on the same detector when deployed to be used for calibration purposes. The Lyot stop is also on a flip mount to enable coronagraphic pupil imaging both with and without the Lyot stop.

Our SEAL software infrastructure is based on the Keck pyramid wavefront sensor real-time control (RTC) architecture, described in Ref. 23. Reference 16 describes more specifically the hardware and infrastructure for SEAL software control and development, which includes a GPU-based clone of the Keck RTC in Linux (which we will refer to as the SEAL RTC) and a Windows machine for some off-the-shelf components without available Linux drivers. Low-level drivers for the MEMS and Andor camera (i.e., to get and send DM commands and to get images/subarrays) are written on the SEAL RTC in C and interfaced with Python to enable high-level FAST software development. Unless otherwise mentioned, for the Andor camera we use a  $320 \times 320$  subarray updating at 100 Hz (although in future work we will plan to demonstrate the higher-speed FAST closed-loop control: Andor benchmarks the Zyla 5.5 with  $<1$  e- read noise at frame rates up to 1 kHz; Andor, private communication). Running in serial mode in Python (i.e., not parallelized), a 20-ms pause between grabbing Andor frames and applying DM commands is needed to enable a given image to be used to compute the corresponding DM commands. However, benchmarking tests for other projects on the SEAL RTC have already demonstrated the ability to reach  $\sim 5$  ms latency in Python by multithreading and synchronizing imaging-grabbing and DM-command-sending threads.<sup>24</sup> Our windows machine is used to operate the adjustable star light source, deployable pupil imaging lens, and SHWFS. For the SHWFS, although Thorlabs does not provide Linux-based drivers for the commercial WFS-20, we use the Windows serial interface in Python and an ethernet connection to the SEAL RTC to use SHWFS frames in our SEAL RTC Python interface for high-level AO software development (as with the Andor and MEMS above); current frames are limited to updating at 50 Hz (i.e., limited by the above-mentioned 20 ms pause) to interface with Python. However, separate

benchmarking with this SHWFS setup shows that we can readout slopes with 23 subapertures across the beam (i.e., slightly undersampled relative to the 29 DM actuators across the beam) at 460 Hz for future “first stage AO + FAST” tests at higher speeds (see Sec. 7).

### 3.2 Alignment

Off-the-shelf lenses are aligned with standard techniques, using a shear plate to collimate the beam and a knife-edge test to calibrate focal lengths but will not be discussed in detail further in this paper. The FAST FPM is first aligned manually using a series of stage adjustments and then algorithmically using the ALPAO DM. Further details of this FPM alignment process are given in Appendix Sec. 9.1. Appendix Sec. 9.2 additionally describes the steps taken after FPM alignment to empirically define the Lyot stop prescription from recorded coronagraphic pupil images.

## 4 Calibration Procedure

In this section, we describe the process and measurements for high and low order reconstruction (Secs. 4.1 and 4.2, respectively). This analysis relies on a number of initial calibration steps, including ALPAO DM flattening, DM-to-wavefront conversion, and more, which are described in detail in Appendix Sec. 9.3.

### 4.1 High Order Reconstruction

Our high order wavefront calibration and reconstruction procedure largely follows the standard methods developed for calibrating the SCC as outlined in Appendix Sec. 8.2. In addition to the overview given in Sec. 8.2, we briefly summarize these steps here, specifically noting additional setup-specific subtleties we have implemented. Low order WFS (LOWFS) calibration and linearity will be discussed next in Sec. 4.2.

To generate the high order interaction matrix (IM), sine and cosine waves are applied to the MEMS for every spatial frequency greater than five cycles per pupil ( $c/p$ ) within the DM control region at intervals of 1  $c/p$ . Applying lower order Fourier modes (i.e., those that are close to or within the edge of the FAST FPM at  $3 \lambda/D$ ) in the same IM as these mid to high order modes would increase nonlinearities by modulating the pinhole PSF throughput for only those modes. To minimize aliasing effects, we limit the highest MEMS spatial frequency along the  $x$  and  $y$  directions to 12  $c/p$ , below the theoretical 13  $c/p$  maximum.<sup>19,20</sup> For each DM Fourier mode, we compute  $I_-$  [i.e., the Fourier-filtered complex-valued fringe component of the SCC image, as described in Sec. 8.1, Eq. (2), and Fig. 11] pixel values (real and imaginary) within a user-defined precomputed (DH) region (i.e., which sets the 12  $c/p$  maximum spatial frequency limit along the  $x$  direction described above) for a differential image (i.e., as in Fig. 12). As in Ref. 20, we implement an additional binary mask before storing the computed  $I_-$  for a given mode in the IM, which isolates the given  $I_-$ -plane sine spot within a  $1.3 \lambda/D$  radius of its central location and sets all other pixel values within the DM control region equal to zero, minimizing cross-talk between modes. Using the same notation as Sec. 8.2, looping through all Fourier modes, these  $I_-$  pixel values are then stored in a reference vector,  $\vec{A}$ , with dimensions  $n \times m = 456 \times 27,936$ . The  $456 \times 27,936$  command matrix (CM),  $\vec{C}$ , and reconstructed high order DM commands,  $\vec{D}_n$ , are then generated via Eqs. (3) and (4), respectively.

We compute the pseudoinverse via singular value decomposition (SVD), choosing the lowest SVD cutoff value (i.e., the lowest number of IM Eigenmodes conserved in the CM) that sufficiently reconstructs individual Fourier modes applied on the DM without cross-talk from other modes. Note that although with a more conventional zonal-based pupil plane WFS where the SVD cutoff is used to minimize the propagation of waffle modes onto the DM,<sup>25</sup> this waffle mode propagation is less relevant for a focal plane WFS. A binary DH mask is spatially filtering modes that the DM controls, inherently removing waffle modes at the DM Nyquist frequency.<sup>20</sup>

## 4.2 LOWFS Reconstruction and Linearity

Using FAST, the SCC image is also an LOWFS, as demonstrated in simulations in Ref. 20. Here, we demonstrate laboratory validation of this concept with SEAL. FAST LOWFS capabilities are enabled by the setup-specific FAST FPM, which sends the PSF core (i.e., low order pupil plane wavefront aberrations) through the Lyot stop pinhole aperture, forming fringes in the SCC image that track the amplitude and phase of low order modes (including for even modes such as focus, whose sign ambiguity is resolved from the fringes).

We apply a similar matrix vector multiplication (MVM)-based approach to generating an IM and CM for low order modes as described in Sec. 4.1 for high order modes. During the IM generation, instead of applying binary masks around individual sine spots, a binary mask in the  $I_{\perp}$  plane uses pixel values only within a  $5\lambda/D$  radius—the limiting DH inner working angle (IWA) for this system<sup>19</sup>—of the optical axis. After testing and development, we resolved to implement closed-loop LOWFS performance for three different modal groups: (1) tip/tilt/focus (TTF), (2) the next five Zernike modes, and (3) low order Fourier modes between 2 and 4 c/p with modal groups (1) and (2) deprojected from this basis. The linearity curves, showing reconstructed output wavefront error (WFE) versus input WFE amplitude for all modes applied and measured, are in Figs. 3–5.

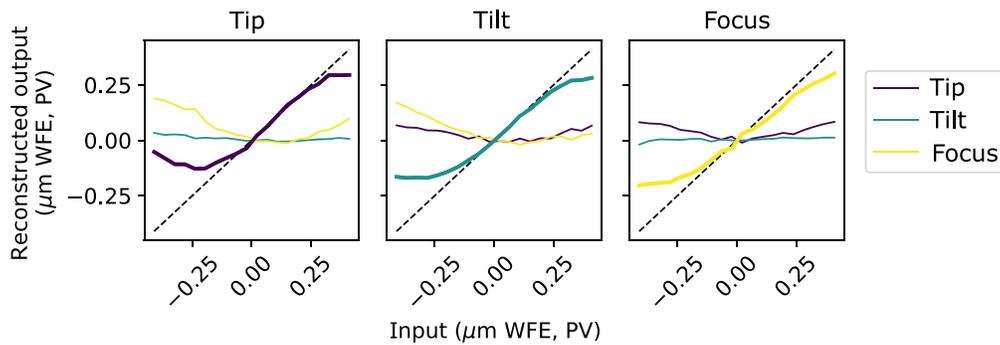


Fig. 3 LOWFS linearity for tip/tilt and focus.

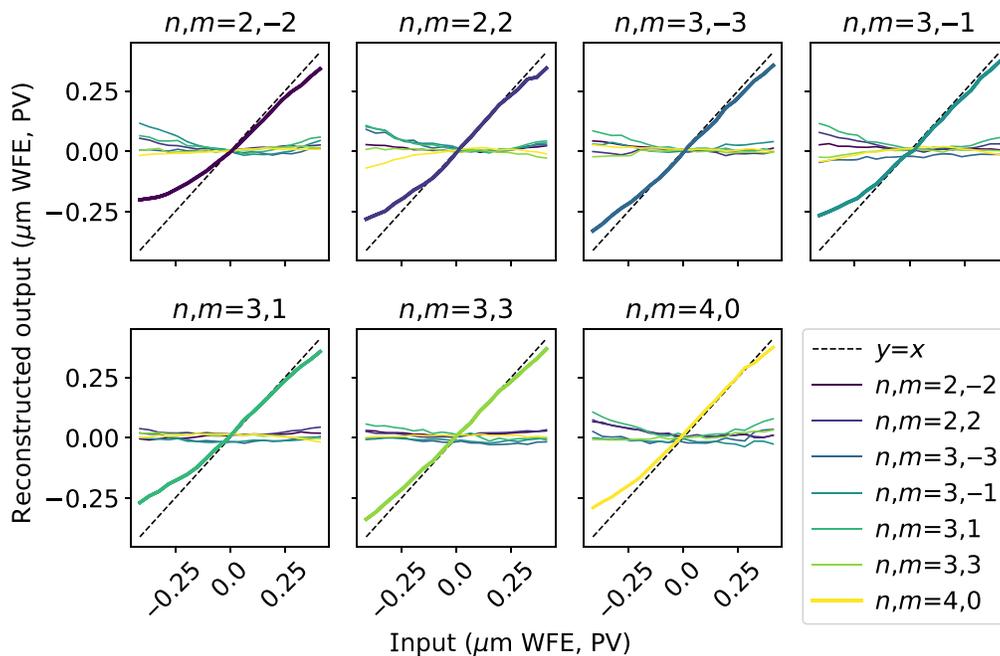
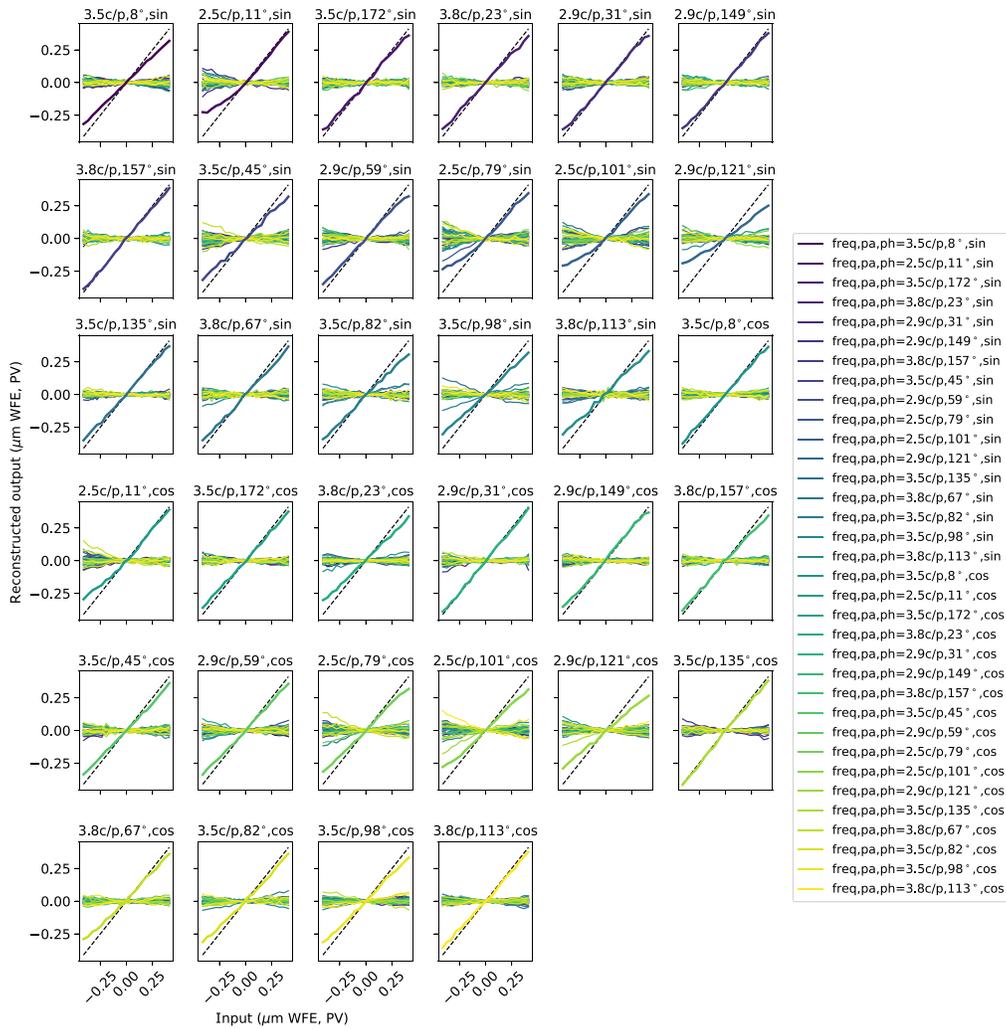


Fig. 4 LOWFS linearity for Zernike modes beyond TTF, labeled by their  $n$  and  $m$  indices.



**Fig. 5** LOWFS linearity for Fourier modes between 2 and 4 c/p, indicated by the amplitude, position angle, and relative phase of their spatial frequency (labeled “freq,pa,ph,” respectively). The relative phase is reconstructed from a pair of sine and cosine modes for a given spatial frequency, as labeled. All panels show the same x and y scale of  $\pm 0.45 \mu\text{m}$  WFE.

As shown, these modal groups form an excellent basis set to enable low order control, with a clear linear regime defined for each mode and minimal cross-talk between other modes in this regime. Figure 3 shows a tip/tilt capture range of  $\sim 0.2 \mu\text{m}$  PV corresponds to  $\sim 0.3 \lambda/D \approx 10$  mas for a 8-m telescope at  $\lambda = 1.6 \mu\text{m}$ , consistent with simulations from Ref. 20 and the need for a diffraction-limited AO residual wavefront (i.e., Strehl ratio  $\gtrsim 20\%$ ) to close the FAST loop. For each modal group, the input “poke” amplitude used to generate the IM and SVD cutoff used to generate the CM are separately optimized to produce optimal linearity and minimal cross-talk. It is for this reason that we placed focus in the same modal group as tip/tilt rather than with the other Zernike modes: a single SVD cutoff value produced good linearity for all Zernikes other than focus, whereas the same tip/tilt SVD cutoff preserved good linearity for focus. We also confirmed that if all modes were combined in the same IM and CM that cross-talk between any mode and all the other modes was still negligible within the linear range of that mode (although with a single SVD cutoff for all modal groups, linearity appeared qualitatively worse than as shown in Figs. 3–5, which is why we chose to separate these groups into different IMs and CMs). The robustness of our choice to use separate SVD cutoffs and IM amplitudes for each modal group is also further supported by closed-loop results in Sec. 5 (i.e., if cross-talk between modal groups remained significant, the loops would not close and/or go unstable).

Interestingly, Fig. 3 shows some conceptual discrepancies from TTF linearity simulations in Ref. 20, including a smaller capture range for tip and for tilt and cross talk in focus only from tip but not from tilt. As we have not observed this behavior in simulation and we have already optically aligned the system for these modes, we conclude (by process of elimination) that these behaviors are due to FPM fabrication defects. Reference 20 did not analyze SCC/LOWFS modes beyond TTF, so we are not able to compare simulations to Figs. 4 and 5, but in general the strong linearity and low cross talk for each mode indicates the potential for efficient closed-loop control, which is the main focus of this paper.

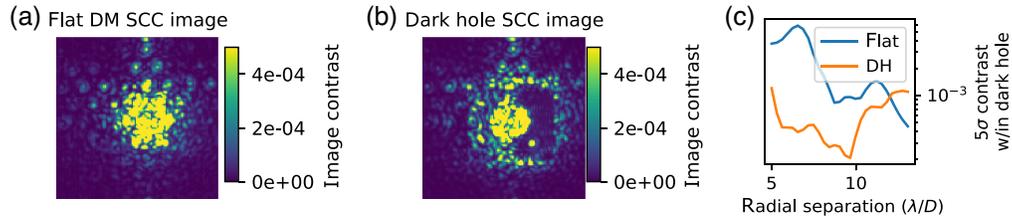
## 5 Closed-Loop Operations

This section is split into three parts: first, presenting results and analysis from generating a static DH (Sec. 5.1); second, from closing the FAST LOWFS and high order WFS (HOWFS) loops for on-air quasistatic errors within the SEAL testbed enclosure (Sec. 5.2); and last in closing, the LOWFS and HOWFS loops on simulated AO residual turbulence (Sec. 5.3). The latter is an important milestone for focal plane wavefront sensing technologies, demonstrating the potential of FAST to be implemented as a second stage AO WFS. In particular, the ability to enable low and high order control is critical, as the latter is dependent on the former to enable optimal performance gains.<sup>26</sup> Common to all subsections, our current Python-based real-time code runs at  $\sim 50$  Hz with 1 frame of system latency, set by a user-defined 20 ms pause at each iteration to enable FAST images (at the end of the iteration) to see the effect of DM commands applied (at the beginning of the iteration); we found that pauses smaller than this amount could prevent this sequential dependence from functioning properly, although we do not expect this to remain an issue after a planned future migration to a parallelized framework (see Sec. 3.1).

Common to Secs. 5.2 and 5.3, we record HOWFS and LOWFS IMs and CMs (as described in Secs. 4.1 and 4.2, respectively) immediately before running open- and closed-loop sequences, minimizing the impact of time-dependent control errors. LOWFS corrections are calibrated and applied on the  $11 \times 11$  ALPAO DM (hereafter, the low order DM), whereas HOWFS corrections use the  $34 \times 34$  MEMS (hereafter, the high order DM, or HODM). Closed-loop target SCC images use differential images (with respect to a static DH image taken just before the sequence begins) dotted with the CM [i.e., Eq. (4)] for each modal group to generate DM commands (i.e., instead of absolute speckle correction, preventing achievable closed-loop contrasts deeper than the static DH but enabling faster convergence to closed-loop residual levels). For all sections, we use a leaky integrator controller with the leaks and gains tuned separately by hand (and separately for each section) for each modal group (but with a fixed value for each parameter across a given modal group), balancing rejection of temporal disturbances (which prefers higher gains and leaks) with noise propagation and loop stability (which prefers lower gains and leaks).<sup>27</sup>

### 5.1 Static Errors

Figure 6 shows images and contrast curves of speckle correction by DM control of static aberrations within the SEAL system. The off-axis planet source (to the lower right of the star) is visible in the middle “DH SCC image” panel. Due to the static nature of this initial DH, two control approaches produce similar results: (1) using a low gain integral controller (typically below 0.01) to ensure loop convergence over a large number of iterations (converging typically after a few hundred frames) or (2) using a higher gain but limiting the number of iterations, beyond which the loop diverges due to detector and photon noise propagation into the reconstructed DM commands (e.g., a gain of 0.1 fixed at 100 iterations typically produces similar results as option 1). Contrast curves in Fig. 6(c), calibrated to the peak off-axis stellar flux as described in Sec. 9.3, compute five times the standard deviation within a given radial annulus of the SCC image (masking the planet signal to prevent bias in computing contrast gains) that is also within the half DH region on the right-hand side of the image, showing up to a  $10\times$  contrast gain depending on separation for the shown 20 ms exposure time. In Fig. 6(c), it is also apparent that although we defined the DH control region out to  $12 \lambda/D$  along the  $x$  direction (Sec. 4.1), contrast gains are negligible beyond  $\sim 10 \lambda/D$  due to a trade-off we found between SVD cutoff and MEMS stroke

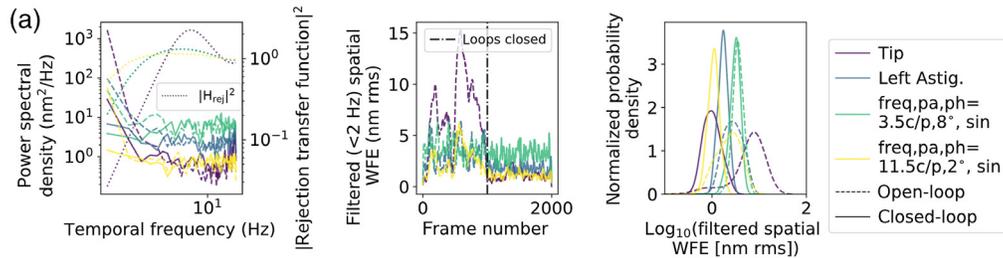


**Fig. 6** Speckle correction/subtraction results of quasistatic aberrations on SEAL. (a) FAST image before DM correction. (b) The same image after DM correction, generating a DH on the right side of the star with the planet visible to the lower right of the star. (c)  $5\sigma$  contrast curves of the SCC image within the DH region (masking the planet signal) for before (flat) and after (DH) static correction, showing up to a 10 $\times$  improvement.

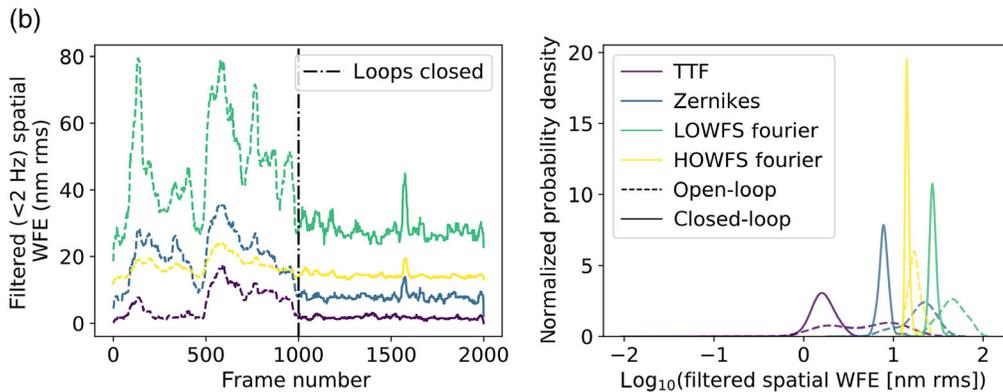
limitations: a lower SVD cutoff (i.e., including more eigen modes in the HOWFS CM) produced a deeper DH between 10 and 12  $\lambda/D$  but also required  $\sim 2\times$  more stroke (which cannot be compensated by the higher-stroke ALPAO device due to the spatial frequencies being considered) than a higher SVD cutoff, which produced a DH as deep out to 10  $\lambda/D$  but not as deep between 10 and 12  $\lambda/D$ . Given the subsequent need to use additional MEMS stroke to apply and then correct AO residual turbulence (Sec. 5.3), we chose to proceed with the higher SVD cutoff option.

## 5.2 Quasistatic Errors

Figure 7 shows the temporal effects of closing the loop on air, stabilizing any quasistatic effects such as air flow through the enclosure. The histograms in Fig. 7 clearly show a gain in WFE for



On-air real-time correction of four individual modes.



On-air real-time correction of all modes for all four modal groups.

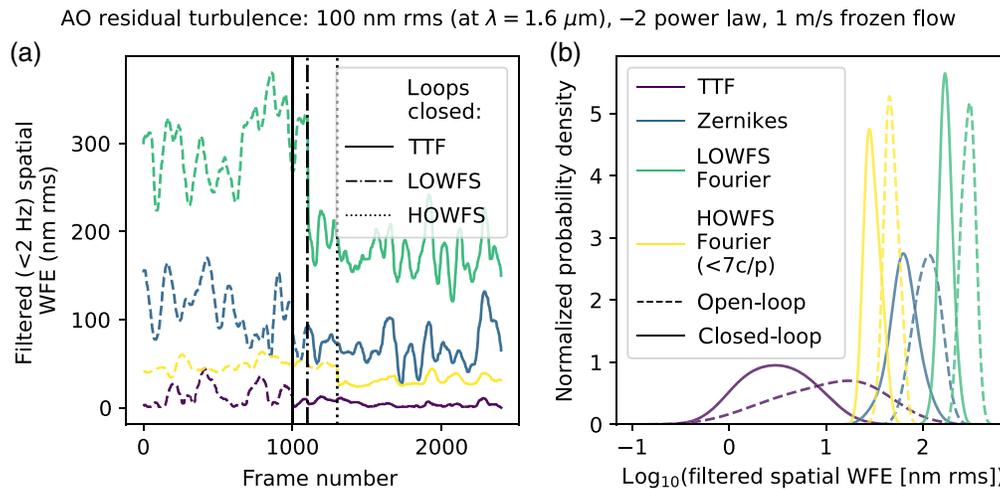
**Fig. 7** SEAL temporal stability before and after closing the loop on only quasistatic turbulence. (a) The statistics for a single modal within each of the four modal groups (with each group indicated by a different color; the color scheme is consistent across all panels in this figure), (b) the cumulative statistics for all modes within a given modal group. Temporal PSD plots on the left panel side of panel (a) motivate the subsequent display of low-pass filtered wavefront statistics, where the choice of temporal frequency cutoff (2 Hz) is the 0 dB bandwidth for tip. Kernel density distributions are computed using the seaborn.kdeplot package.<sup>28</sup> All panels clearly illustrate a WFE reduction and stabilization for all modes.

closed- versus open-loop control for all modal groups. The median TTF open- and closed-loop WFE are 6.2 and 1.6 nm rms, respectively, which for the latter corresponds to  $0.003 \lambda/D$  for tip/tilt, illustrating the highly stabilized environment the SEAL granite testbed and enclosure enables. Although this nm level of on-air stability is perhaps unrealistic for an on-sky ground-based environment due to realistic quasistatic effects (e.g., mechanical effects from telescope tracking, changing thermal environment, instrument flexure for Cassegrain instruments), it sets the stage for similar gains to be made on AO residual turbulence correction, discussed next.

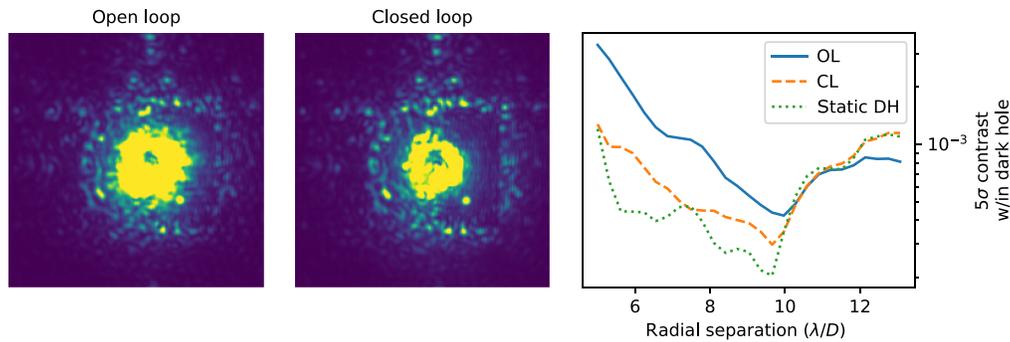
### 5.3 Dynamic AO Residual Errors

In this section, we present an important milestone for focal plane wavefront control technologies, demonstrating FAST correction of AO residual speckles in quasi-real-time for both low- and high-order modes. An evolving full power law AO residual wavefront (i.e., including low order modes) is applied on the MEMS in open loop and then corrected by the FAST LOWFS and HOWFS in closed-loop. AO residual WFE is normalized to 100 nm rms between 0 and 16 c/p assuming  $\lambda = 1.6 \mu\text{m}$  with a  $-2$  spatial power spectral density (PSD) power law. A 10-m diameter telescope, 1 m/s ground layer “frozen flow” wind speed is assumed. As discussed at the beginning of Sec. 5, our Python real-time code limits operational frame rate to 50 Hz with a 1-frame delay. This limited loop speed is the main reason why a 1-m/s wind speed is used, although a more realistic 1 ms atmospheric lag, as motivated in Ref. 19, is applied at the time of DM correction. Regardless, future more-optimized real-time code running at kHz speeds will enable closed-loop correction on more typical wind speeds of order 10 m/s. With this setup implemented, Fig. 8 shows the time series and WFE distributions and Fig. 9 shows long exposure SCC images and contrast curves (computed analogously to Fig. 6 as described in Sec. 5.1) for open- and closed-loop FAST correction of evolving AO residuals.

Figure 9 shows up to a  $5\times$  contrast gain comparing open- and closed-loop correction of AO residual turbulence, illustrating the benefit of closed-loop operations at all spatial frequencies. Figure 8 shows a  $1.5\times$  HOWFS median WFE gain (46 versus 30 nm rms for open versus closed-loop HOWFS WFE, respectively, for spatial frequencies between 5.5 and 7 c/p), which is



**Fig. 8** (a) Time series and (b) corresponding kernel density distributions of open- and closed-loop operations of AO residual turbulence, using the same temporal filter based on the analysis in Fig. 7. The same colors in the right panel legend also apply to the left panel, noting that high order Fourier modes are only shown for spatial frequencies between 5 and 7 c/p. Atmospheric parameters are shown in the title; the loop speed is  $\sim 50$  Hz with 1 frame delay. Gains and leaks for a leaky integrator controller are separately tuned for each modal group. Loop closures occur at three different times for TTF, other LOWFS modes (including Zernike and LOWFS Fourier modes), and HOWFS modes, indicated by the three different vertical lines on the left panel. Open- and closed-loop distributions in the right panel are computed with data in the left panel to the left of the solid line (TTF loops closed) and to the right of the dotted line (HOWFS loops closed), respectively.



**Fig. 9** Stacked FAST images (on the same relative contrast scale) and corresponding contrast curves for open- and closed-loop (OL and CL, respectively) control of simulated AO residual turbulence (using the same frame indices to define the OL and CL regimes as in the right panel of Fig. 8), also showing the same “static DH” contrast curve from Fig. 6 for reference. The same planet source is present in the lower right side of the DH images. Closed-loop correction of AO residual turbulence is demonstrated to improve contrast by up to a factor 5 relative to no residual atmospheric correction (OL).

consistent with the  $2.3\times$  median OL versus CL contrast gains between  $5.5$  and  $7\lambda/D$  in Fig. 9 (i.e., contrast gains should scale as the square of the WFE gains). Figure 8 also illustrates the sequential process of closing loops: in contrast to on-air control in Fig. 7, here we found that all loops cannot be closed at once, requiring initial TTF loop closure, then LOWFS (including Zernike and Fourier modes simultaneously), and finally HOWFS Fourier modes. When we tried to close all modes simultaneously, or, e.g., HOWFS modes first, the loop would go unstable shortly thereafter. This is to be expected: LOWFS modes must be closed and stabilized first (i.e., producing a stable stellar core on the FAST FPM, which is needed for HOWFS measurements), as we predicted in Ref. 20. The number of frames needed between closing the TTF, LOWFS, and HOWFS loops for the different modal groups was determined by applying AO residuals on the MEMS with only the modes from a given group applied and then determining by eye how long it took for the loop to converge; we did not further explore flexibility in this parameter space. Contrast improvement in Fig. 9 decreases with separation, however, suggesting that loop control is not yet optimized. In Fig. 8, we only plot HOWFS telemetry for spatial frequencies below  $7$  c/p, as including these higher order modes “dilutes” the relative gain from these plotted HOWFS modes. By applying a single gain and leak ( $0.3$  and  $0.8$ , respectively) to all HOWFS modes (i.e., as described in Sec. 8.2), there is an inherent nonoptimal trade-off between temporal rejection at lower spatial frequencies versus noise amplification at higher spatial frequencies;<sup>27</sup> our chosen scalar gain and leak values gave better rejection at lower spatial frequencies and more noise amplification at higher ones, consistent with the smaller contrast gains at larger separations in Fig. 9. This clearly motivates the need to implement more advanced control, such as modal gain optimization.<sup>19</sup> Further comparing the closed-loop performance to static DH limits (dashed orange vs. dotted green curves), it is clear that across most spatial frequencies there are additional gains to be made with real-time correction of AO residuals. We discuss such further areas for contrast improvement in Sec. 6.

Regardless of additional gains to be made, Fig. 8 and particularly Fig. 9 represent an important milestone in using real-time focal plane wavefront control to boost the achievable final contrasts with exoplanet imaging instruments. To be clear, this is not the first demonstration of real-time focal plane wavefront control; Ref. 8 demonstrated on-sky linear dark field control at  $20$  Hz but with negligible contrast gains due to a noisy detector and insufficient bandwidth for control of residual atmospheric speckles. In this paper, however, we have demonstrated a clear raw contrast gain of up to  $5\times$  for simulated AO residuals in real-time. The implications of this result scales to an up to  $5\times$  final contrast gain, supported by the work from Ref. 29 (correlating AO performance to achievable raw contrast and raw contrast to final contrast). In other words, this demonstration of  $5\times$  raw contrast gain with FAST translates to boosting postprocessed contrast by  $5\times$  with angular differential imaging<sup>30</sup> (even in second generation high contrast imaging

instruments, spectral differential imaging<sup>31</sup> yields negligible contrast gains to due significant instrument chromaticity<sup>4</sup>).

## 6 Key Areas for Contrast Improvement

Perhaps the most significant area for further improvement is the optimal choice of LOWFS modal bases to reduce the impact of nonlinearities on the achievable closed-loop WFE. The choice of Fourier modes for LOWFS control is perhaps a nonintuitive one; applying DM Fourier modes with low enough spatial frequencies ( $<3$  c/p for this TG FPM) generates sine spots that are simply blocked by the FPM and redirected to the off-axis FAST pupil, rather than generating sine spots in the coronagraphic image. We could instead consider higher order Zernike modes; however, these increasingly become a non-natural basis with increasing radial order, as entrance pupil wavefront spatial frequencies corresponding to focal plane separations beyond the TG FPM core are not sent to the off-axis Lyot stop pupil, instead overlapping (i.e., increasing nonlinear cross-talk) with the FAST HOWFS Fourier modal group. Instead, we first tried applying a basis of higher order spatially-filtered ( $<3$  c/p, using an algorithmic Butterworth Fourier filtering mask) Zernike modes, but at the cost of higher order Zernike modes becoming increasingly nonlinear due to the loss of information by a low-pass filter. Alternatively, considering the PSFs of Zernike modes, PSF structure deviating from an Airy pattern increases in radial extent as a function of Zernike mode radial order when amplitudes are otherwise fixed, and so restricting a smaller region of the PSF core to use for wavefront reconstruction causes increased nonorthogonality versus Zernike mode order. Thus, an LOWFS Fourier modal basis is instead a natural choice, restricting diffracted light to within the FPM IWA. However, for reasons we do not yet understand, extending this Fourier modal basis to spatial frequencies below  $2$  c/p also did not produce good linearity but low-order Zernike modes instead did, which motivates our choice of both modal bases (the Zernike mode cutoff was determined based on the mode above which nonlinearities would significantly increase if spatially filtered to  $<2$  c/p). Other bases should be further explored (e.g., KL modes, disk harmonics, etc.) that can span the full range of accessible LOWFS spatial frequencies (or perhaps in combination with Fourier and/or Zernike modes, as was done in this paper) to optimally “linearize” FAST LOWFS performance. Furthermore, such additional bases should be explored specifically in the context of realistic obstructed pupil geometries for 8 m-class telescopes (e.g., using annular Zernike polynomials), as the orthogonality of bases used in this paper may be less robust in the presence of such pupil geometries.

Higher quality of fabricated FAST FPMs will also enable further gains. In Ref. 18, we modeled that our current TG FPM has around 300 nm rms of phase errors. In Ref. 19, we showed that getting such errors below 100 nm rms would enable TG fringe visibilities at  $10 \lambda/D$  up to 15% (about 10 $\times$  greater than what we are getting now). Such a higher quality mask would improve measurement sensitivity to photon noise (i.e., lowering closed-loop WFE for a fixed number of input photons) and also nonlinearities. We noticed the effects of the latter due to fabrication defects on our FPM, where some specific LOWFS Fourier modes were intentionally not controlled due their empirically measured bad linearities (while other modes at the same spatial frequency but different position angle remained sufficiently linear). We have already demonstrated that it is realistic to fabricate FAST FPMs at the nm rms level,<sup>32,33</sup> so it is reasonable to expect that future FAST FPMs can improve on this front.

Additional algorithmic optimization procedures can also enable further contrast gains. Approaches to improving HOWFS linearities include varying the Fourier modal amplitude with spatial frequency so IM spot intensities optimally match the power law of AO residual aberrations, optimizing the binary mask radius used to isolate sine spots during the HOWFS IM calibration (currently fixed at  $1.3 \lambda/D$ ; Sec. 4.1), and considering different modal groups within the HOWFS control region to control separately with different SVD cutoffs and CMs (e.g., as has already been done for LOWFS TTF, Zernike, and Fourier modes in Sec. 4.2). FAST modal gain optimization,<sup>19</sup> both for LOWFS and HOWFS control is also a key area to develop that we have already been working on with SEAL.<sup>18</sup> However, this approach requires characterizing the rejection transfer function<sup>34</sup> (i.e., which is defined for a given mode relative to its open- and

closed-loop temporal PSDs, or  $\text{PSD}_{\text{OL}}$  and  $\text{PSD}_{\text{CL}}$ , respectively, as:  $\text{RTF} \equiv \sqrt{\text{PSD}_{\text{OL}}/\text{PSD}_{\text{CL}}}$ , a typically linear assumption for which FAST nonlinearities may pose limitations to its success. FAST predictive wavefront control, which we are also beginning to implement,<sup>35</sup> is also a critical area to develop, as the inherent 1/2 frame delay only recoverable by prediction is much larger at the slower 100 Hz speeds FAST will be limited to on fifth magnitude stars with an 8-m telescope<sup>15</sup> (note that the data acquired in this paper, collecting hundreds of photons per pixel per frame, is more akin to a zeroth magnitude star); data-driven predictive control methods (e.g., Ref. 36) can also be used to “learn” FAST nonlinearities, also further improving achievable gains.

## 7 Conclusion

Exoplanet imaging speckle subtraction via focal plane wavefront control is a promising technology that has the potential to enable current and future instruments to detect and characterize lower mass, closer in, and/or older exoplanetary systems than is currently possible. However, one issue thus far in making such technology operational on-sky has been achievable temporal bandwidth, due to continual speckle evolution from atmospheric and quasistatic effects. We have developed a technique to mitigate this bandwidth issue, called the fast atmospheric self-coherent camera technique (FAST), for which in this paper we present laboratory validation on the Santa Cruz Extreme AO Laboratory (SEAL) testbed. Our main findings are as follows:

1. We tested FAST quasistatic DH generation and closed-loop temporal stability, demonstrating  $5\sigma$  contrasts within the DH of around  $5e-4$  (Fig. 6) and nm rms-level temporal stabilities on-air (Fig. 7).
2. We tested FAST correction of AO residual speckles, demonstrating a clear gain of such a real-time “second stage AO” capability using a focal plane wavefront sensor, showing up to a factor of  $5\times$  contrast gain, depending on radial separation (Fig. 9).

These testing results clearly illustrate the benefit of running real-time focal plane wavefront control technology. However, the developments presented here are by no means comprehensive, instead opening many new avenues of real-time focal plane wavefront control approaches to explore and optimize, some of which we discuss in Sec. 6. In future manuscripts, we also plan to investigate FAST SEAL testing of many related additional topics, including increasing spectral bandwidth capabilities (for which fringe smearing due to PSF magnification with wavelength can be mitigated with the multireference SCC<sup>37</sup> adapted specifically for FAST<sup>4</sup>), optimizing mid-to-high order temporal bandwidth (including predictive control), real-time integration with a first stage AO correction, and optimizing control for changing atmospheric conditions and corresponding first stage AO performance.

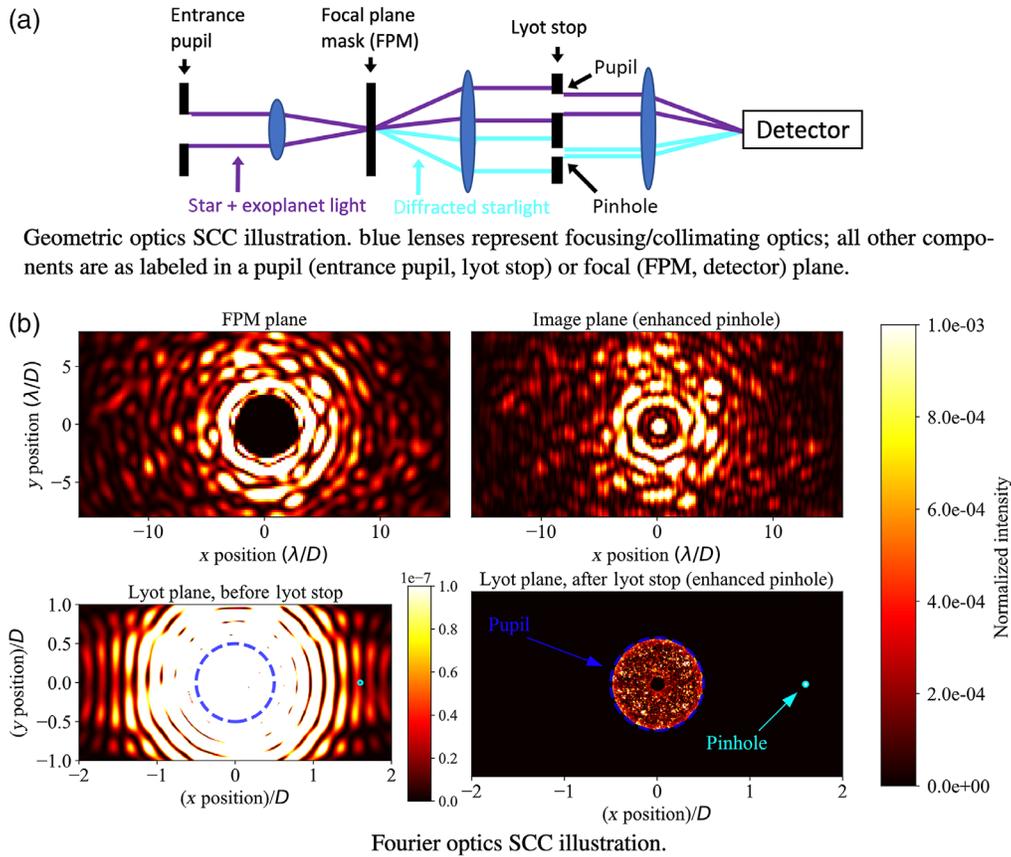
## 8 Appendix A: Summary of the Self-Coherent Camera

The SCC, developed by Refs. 14, 22, and others, is a method designed to use the coronagraphic science image as both a WFS to control an upstream DM and enable CDI-based postprocessing. We review the basics of classical (i.e., not FAST-specific) SCC image formation and processing (Appendix A.1) and DM control (Appendix A.2). SCC CDI techniques are not summarized due to the scope of this paper.

### 8.1 Image Formation and Processing

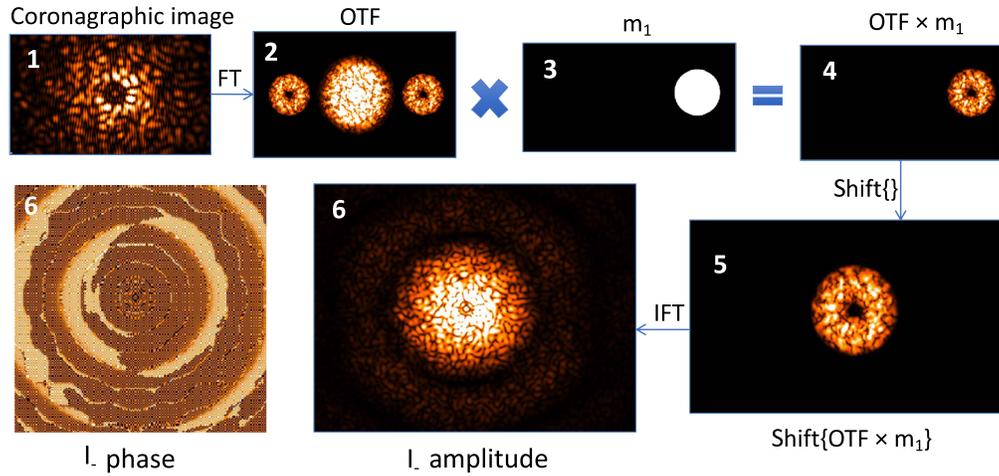
Figure 10 shows the concept of SCC image formation, both from a geometric (a) and Fourier (b) optics perspective. Reviewing a mathematical description of the SCC image, the recorded image on the focal plane detector,  $I$ , produced from noiseless (i.e., without photon/detector/sky background noise) propagation of WFE (i.e., from atmospheric and/or instrumental origin) through to the detector, can be described as<sup>14</sup>

$$I = I_S\{\theta\} + I_P\{\theta\} + I_R\{\theta\} + 2\sqrt{I_S\{\theta\}I_R\{\theta\}} \cos[2\pi(\xi_0/\lambda)\phi\{\theta\}], \quad (1)$$



**Fig. 10** An simulated illustration of SCC image formation due to (a) the geometric and (b) diffractive properties of an FPM, adapted from Ref. 15. We describe the four subpanels in (b). Upper left: in an intermediate focal plane, an FPM blocks the on-axis star while transmitting remaining diffracted starlight and off-axis exoplanet light. Lower left: in the downstream pupil plane, starlight is diffracted outside the footprint of the primary mirror if no FPM was deployed (illustrated by the dashed blue circle); such diffraction effects occur for any FPM with sub- $\lambda/D$  spatial scales, in this case encoded by the hard edge of an amplitude “dot” FPM. Lower right: a Lyot stop blocks most of this diffracted light outside the pupil footprint, except for a small off-axis pinhole (whose cyan aperture footprint is also illustrated in the lower left panel) that is transmitted downstream. Upper right: The Lyot stop pinhole and pupil apertures interfere in the coronagraphic image to form fringes on sub- $\lambda/D$  spatial scales, overlaid on the existing speckle pattern (with spatial scales larger than  $\lambda/D$ ) from the main Lyot stop pupil aperture. The electric field intensity in the pinhole aperture used to simulate the lower and upper right panels is (unphysically) enhanced by a factor of  $2.5 \times 10^5$  for the purposes of this illustration.

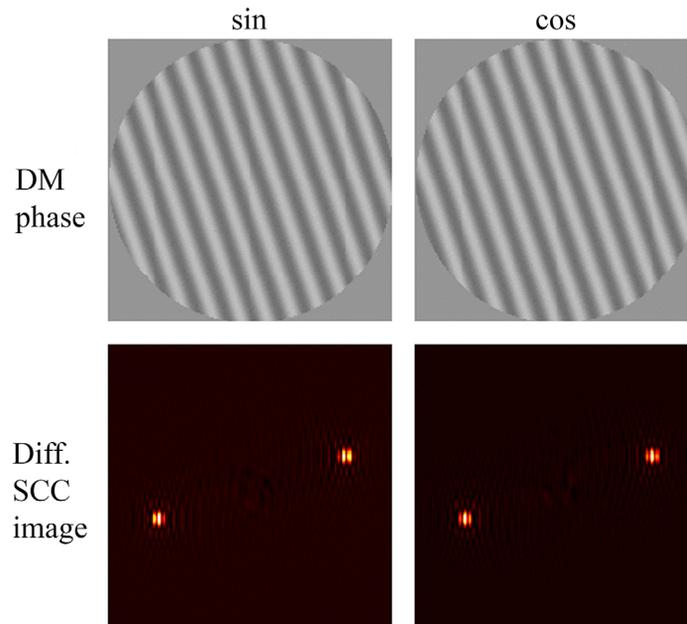
where  $\theta$  represents the two-dimensional (2D) coordinates in the image,  $I_S$  is the unfringed stellar speckle intensity component,  $I_P$  is the unfringed planet intensity component,  $I_R$  is the pinhole PSF component (i.e., the PSF recorded if the main pupil aperture was blocked), and  $\phi$  is the electric field phase difference between the  $I_S$  and  $I_R$  image components. The spatial scale of the fringe term  $2\sqrt{I_S I_R} \cos[2\pi(\xi_0/\lambda)\phi]$  is set by the Lyot stop’s pupil-to-pinhole separation ( $\xi_0$ , projected onto the entrance pupil) and observing wavelength,  $\lambda$ , enabling sub- $\lambda/D$  spatial scales (between  $\lambda/[0.5\xi_0]$  and  $\lambda/[1.5\xi_0]$ ) to be isolated in the Fourier plane of the image (i.e., the optical transfer function, or OTF) without bias from the unfringed exoplanet signal  $I_P$ . Figure 11 shows this Fourier filtering process used to isolate the fringed SCC image component. The phase of a given stellar speckle [i.e.,  $\phi\{\theta\}$  in Eq. (1)], which is not normally measurable in a single coronagraphic image (i.e., not encoded in  $I_S$  or  $I_R$ ), is encoded by the relative position of fringes on that speckle, which is measured as the phase component of  $I_-$  in Fig. 11. In equation form, the Fourier filtering process in Fig. 11 is



**Fig. 11** An illustration of the SCC wavefront sensing and reconstruction algorithm, adapted from Ref. 15. Starting from the recorded coronagraphic SCC image (step 1), a series of Fourier transforms (FT and IFT, the “I” representing “inverse”) and Fourier filtering steps (steps 2 to 5) enable a complex-valued measurement of the focal plane electric field (step 6) known as  $I_-$ . Note that  $m_1$  is a 2D array of ones and zeros (black and white representing zeros and ones, respectively; subsequently referred to as an “algorithmic” or “binary” mask), and the “x symbol” represents a multiplication of OTF (i.e., including both real and imaginary components) by  $m_1$ .

$$I_- = IFT\{\text{shift}\{FT\{I\}m_1\}\}, \quad (2)$$

where Eq. (2) and Fig. 11 use the same operators for Fourier transform( $FT\{\}$ ), inverse Fourier transform ( $IFT\{\}$ ), and OTF side-lobe centering ( $\text{shift}\{\}$ ) as well as a binary mask to isolate the OTF side-lobe fringe component ( $m_1$ ).



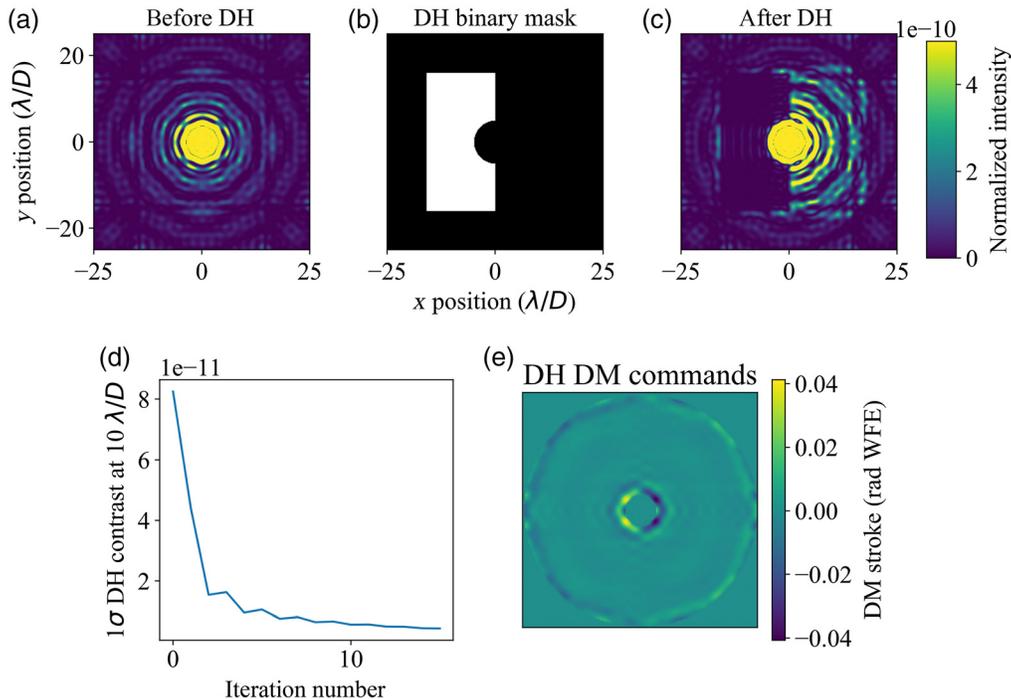
**Fig. 12** An illustration of the procedure for generating an SCC IM. The top row shows the phase applied on the DM, whereas the bottom row shows the corresponding differential SCC image (recorded image relative to a flat wavefront image). The two columns show responses for sine and cosine waves at the same spatial frequency. Such differential SCC images are recorded at the full range of controllable DM spatial frequencies to generate an IM in the complex-valued  $I_-$  plane (see Fig. 11).

## 8.2 DM Control

Recording SCC images and processing them to produce  $I_-$  can be calibrated empirically to DM commands applying sine and cosine waves over the full spatial frequency range of the DM, illustrated in Fig. 12. The Lyot stop pinhole size sets the image-plane region over which DM actuators can correct for speckles, so a higher order correction requires a smaller pinhole. A typical SCC pinhole diameter requirement is that the diagonal along the DM Nyquist region lies with the first minimum of the  $I_R$  Airy function, which for a square grid DM is given in Ref. 43 [Eq. (40)].

An SCC IM can be built by recording differential SCC images as shown in Fig. 12 for each DM Fourier mode [Eq. (1) for a given DM Fourier minus Eq. (1) for a flat DM] that are then converted to  $I_-$  [Eq. (2)] and lastly multiplied by a binary DH mask. This step, not illustrated in Fig. 12, is completed by multiplying (1) the complex-valued  $I_-$  for a differential SCC image of a given DM Fourier mode by (2) an binary DH mask (i.e., using ones to represent the region of desired image-plane speckle attenuation). Note that with only a pupil plane conjugated DM(s), speckle correction of phase and amplitude errors (including diffraction) is limited to a half DH.<sup>38</sup> A vectorized version of these  $I_-$  DH pixel values are then used to build an IM and then CM to enable a subsequent linear least-squares-based MVM linking SCC images and DM commands as follows:

1. Applying  $n$  Fourier modes (which includes sine and cosine components), recording  $mI_-$  pixel values (which includes the real and imaginary components) within the DH for each mode, an SCC reference matrix,  $\vec{A}$ , is generated with dimensions  $n \times m$ . A separate DM modal reference matrix,  $\vec{F}$ , with dimensions  $n \times p$  (where  $p$  is the number of DM actuators), is saved during this process.



**Fig. 13** A simulated half DH on a diffraction-limited (i.e., no WFE) coronagraphic image (using the apodized Lyot coronagraph from Ref. 39). (a) The raw APLC SCC image (the fringe visibility is too low to see fringes by eye). (b) The binary mask (i.e., defining the DH region with ones and elsewhere with zeros) used to generate  $mI_-$  pixel values. (c) Image after DH generation, after 15 iterations (the upper left and right panels are shown on the same scale). (d) Contrast within the DH at  $10\lambda/D$  versus iteration number, showing that the majority of contrast gains occur after just two iterations. (e) DM commands used to generate the upper right DH image, illustrating the relative minimal amount of DM stroke needed for this SCC configuration.

2. A CM,  $\vec{C}$ , is then generated via

$$\vec{C} = \vec{F}^T \cdot (((\vec{A} \cdot \vec{A}^T)^\dagger) \cdot \vec{A}), \quad (3)$$

where  $\cdot$ ,  $T$ , and  $\dagger$  represent a matrix dot product, transpose, and pseudoinverse, respectively.

3. In closed-loop, the  $p \times m$  CM is then used to convert SCC  $I_{-}$  DH pixel values for a new target image ( $\vec{T}$ , which is an  $m \times 1$  vector) into DM commands ( $\vec{D}_n$ ; a  $p \times 1$  vector, where  $\vec{D}_{n-1}$  represents the previous DM commands), via

$$\vec{D}_n = g(\vec{C} \cdot \vec{T}) + l(\vec{D}_{n-1}), \quad (4)$$

where  $g$  is a scalar gain value (i.e., the same for all modes) and  $l$  is the scalar leak value for a leaky integrator closed-loop controller.

This process to generate DM commands is then applied iteratively in closed-loop until the desired DH contrast convergence in contrast achieved. The number of iterations needed to reach such is convergence system-dependent (e.g., coronagraph type, input WFE level, desired convergent DH level, available DM stroke, temporal WFE/stability environment, and more can all influence convergence timescale and achievable contrast) but can in principle be as few as 0 (i.e., only one image). Figure 13 shows this DH generation process. Note that in the context of this section serving as a high-level summary of SCC DM control, many additional low-level details used for further improving achievable contrast are not described in this appendix section, such as optimizing/tuning the CM pseudoinverse SVD cutoff, IM Fourier modal amplitude, additional IM binary masks, the optimal basis set for low order modes, and more.

## 9 Appendix B: SEAL Alignment and Calibration

Here, we discuss SEAL-specific design and operations as it relates to the main text.

### 9.1 FPM Alignment

FAST FPM alignment is carried out in multiple steps: initial coarse adjustment with a 5-degree-of-freedom stage ( $x$ - $y$ -pitch-yaw-roll; focus is predetermined by a knife edge test as described above). With the Lyot stop in place, pitch and yaw (i.e., tip and tilt) of the FPM mount are adjusted until the noncoronagraphic pupil aperture is centered on the intended central aperture of the Lyot stop (which displays a 0.8-mm-wide ring around the aperture edge due to the 90% undersizing). Then, a pinhole aperture is placed on the focusing lens just upstream of the FPM to simulate a flat field. The FPM core in this mode is visible as a dark circular shadow on the otherwise illuminated flat field;  $x$  and  $y$  positions are then adjusted so the center of this black dot is positioned at the coordinates of the normal Airy disk's center. After removing the pinhole/flatfield, a somewhat-aligned coronagraphic pupil is visible, by eye, on the Lyot stop (as in Figs. 1 and 2); further  $x$ - $y$  adjustment can then enable visually maximizing the off-axis pupil intensity into the pinhole. The FPM roll is then adjusted to the desired position angle (i.e., defined by the Lyot stop pinhole's position angle), after which the above  $x$ - $y$  positioning steps are repeated (since the setup-specific shape on the FPM is not exactly in the center of the optic's clear aperture). Lastly, fine  $x$ - $y$  adjustment is performed in software with the DM tip/tilt: a grid search is performed in a square region around the estimated best alignment position, optimizing a combination of maximal fringe visibility and minimal raw contrast. Although a fringe visibility metric naively seems like a sufficient metric to align the FAST FPM (since once fully aligned the maximal amount of light is transmitted through the Lyot stop off-axis pinhole), fringe intensities are equally weighted between the focal plane amplitudes from the Lyot stop pinhole and pupil [see Eq. (1)], which causes fringe intensities to increase when the FPM is misaligned because the amount of starlight increase is greater than the amount of pinhole light decrease. This is mitigated as described above by requiring minimal starlight (measured in the image plane) in addition to maximal fringe visibilities (measured by integrated flux on the MTF sidelobe).

## 9.2 Lyot Stop Design

The pupil-to-pinhole separation for our Lyot stop prescription is determined empirically, as was done in Ref. 40 and as we will illustrate should be a standard process for future FAST FPMs for laboratory and observatory instrument projects, by taking coronagraphic pupil images with the Lyot stop out and fitting the centers of each pupil (where a separate system is used with DMs replaced by fold mirrors for this characterization to enable a lower WFE characterization). As to be discussed next in Sec. 9.3, we are only able to reach 40-nm rms closed-loop WFE with the ALPAO DM while in a separate system with only fold mirrors produces  $\sim$ nm rms level WFE, making the latter more optimal to characterize the quality of the fabricated FAST FPM (i.e., avoiding ambiguity issues between system WFE and FPM-specific fabrication defects). The off-axis pupil center in this setup with only fold mirrors is determined by fitting a 2D Gaussian, whereas the central coronagraphic pupil position is determined by eye (incurring  $\sim$ 1–2 pixel error  $\approx$ 2% pupil diameter error  $\approx$ 16% pinhole diameter error). Measuring this separation is important to enable a Lyot stop prescription that both (1) optimizes fringe visibility (i.e., limiting the pinhole position to deviate from the peak intensity) and (2) prevents a differential tilt between the pinhole PSF center and unfringed coronagraphic image center (i.e., preventing an offset between the center of the fringed and unfringed image components). We initially found that a Lyot stop prescription with a pinhole separation too discrepant from what was needed indeed caused a tilted pinhole PSF with respect to the unfringed coronagraphic image due to an amplitude gradient across the pinhole aperture causing a tilt. Although our initial pinhole separation prescription was for the intended FPM design, as discussed in Sec. 3.1, it turned out that this differed significantly enough from the fabricated mask to cause such a measurable discrepancy, but with an empirically measured Lyot stop prescription this was no longer a problem. A recorded coronagraphic pupil image should therefore be the main determinant for future SCC Lyot stop prescriptions, both for lab experiments and future FAST instruments, to avoid a tilted pinhole PSF.

## 9.3 Initial Calibrations

We start bench operations by flattening the ALPAO using the SHWFS, also compensating for any corresponding common path aberrations. In reality, the ALPAO surface is not flat, but instead compensates for the unflat resting MEMS and IrisAO surfaces. Although the latter produces an 8 nm rms flat as measured by our Zygo interferometer (and so likely further compensation by the ALPAO as measured by the SHWFS is negligible), we chose to keep the MEMS in a resting nonflat position while flattening the ALPAO, enabling ALPAO compensation for MEMS aberrations below 5.5 cycles/pupil (*c/p*). Although this option leaves the remaining MEMS figure errors between 5.5 and 12 *c/p* (the limiting spatial frequency on a side we correct for) initially uncorrected, we found this approach more favorable than using an optical flat command for the MEMS. We found that the latter approach—which we also generated with our Zygo interferometer—reaches diffraction-limited contrasts in the SCC image, requiring more MEMS stroke to dig a DH on diffracted light rather than on unpinned speckles, consistent with Ref. 41 (although note the differing contrast regimes between Ref. 41, at  $10^{-10}$  levels, and this paper, at  $10^{-4}$  levels). Instead, leaving the MEMS figure error initially uncorrected at these mid to high spatial frequencies significantly relaxed the subsequent MEMS DH stroke limits (and also enabling the subsequent application of AO residual turbulence without running out of stroke) while having a negligible effect on initial Strehl ratio since aberrations at these spatial frequencies are a negligible component of the overall WFE. Other ALPAO devices similar to ours have been known to drift on hours to days timescales;<sup>42</sup> we found that daily generation of a new best flat delivers sufficient performance. The flattening is done in SHWFS slope space, applying a classical zonal “push minus pull” actuator pokes to generate an IM.<sup>25</sup> The CM pseudoinversion SVD cutoff is chosen to optimize the convergent closed-loop WFE as measured by the precalibrated SHWFS wavefront reconstructor (i.e., we choose the SVD cutoff value that produces the flattest wavefront), usually converging to a residual system WFE of around 40 nm rms. There are likely a few key limitations to reaching a deeper closed-loop residual WFE: (1) the presence of a pinned (bad) MEMS actuator near the pupil center, which we do not account for in SHWFS

reconstruction (e.g., algorithmically masking or interpolating over this region of the pupil), and (2) segment gaps caused by the IrisAO DM, which we also do not algorithmically account for in the SHWFS IM. Prior to the IrisAO and MEMS installation into SEAL, the SHWFS driving the ALPAO typically converged to  $\sim 5$  nm rms. Note that this SHWFS flattening procedure is completed before running the DM-based FPM alignment procedure described in Sec. 3.2.

To calibrate contrast, with the best flat on the MEMS applied as described above, we tilt the PSF off the FAST FPM core to record an off-axis PSF, adjusting the exposure time to prevent the star from saturating. Coronagraphic images in units of normalized intensity and/or contrast curves shown later in this paper are derived from dividing the dark-subtracted coronagraphic image in ADUs by the aforementioned maximum flux of the (also dark-subtracted) off-axis stellar PSF, scaling to account for the different exposure times in the two images. We can also calibrate flux ratio of the off-axis planet with this approach: turning off the starlight source but leaving on the planet source and scaling for exposure time gives a normalized planet flux ratio of  $2.4 \times 10^{-3}$ .

To find the coronagraphic image center and empirically calibrate the plate scale, we place two sine waves simultaneously on the DM—one with a position angle offset 90 deg from the other and each at separations of  $10 \lambda/D$ —to generate “satellite spots.” Viewing aligned coronagraphic pupil images and poking MEMS actuators confirmed the 26 actuator sampling across the Lyot stop pupil described in Sec. 3.1, which is needed to ensure a  $10 c/p$  sine wave places spots at separations of  $10 \lambda/D$  in the SCC image. We also determined that the MEMS actuator grid has a 1-deg position angle offset with respect to the Andor detector pixel grid, likely due to small differences in how the two devices are mounted. We correct for this algorithmically to ensure that the MEMS Fourier modes are aligned with the Andor camera pixel grid. Averaging the four 2D Gaussian-fit positions of these spots provides the coordinates of the image center. The average separation between the sine spot pairs also provides a plate scale measurement: 6.2 pixels/resel, sufficiently oversampled for the SCC fringes.

A calibration between DM voltage units and reconstructed (reflected) wavefront units for both the ALPAO and MEMS, which we use subsequently in this paper, were obtained by measuring 4 off-axis PSFs, each at four different combinations of tip and tilt commands (repeating this for both the ALPAO and MEMS). The above-calibrated plate scale is then used in combination with the fitted positions of the four off-axis PSFs to convert differential tip and tilt commands into separations in  $\lambda/D$ , which relates to pupil peak-to-valley (PV) tip/tilt WFE via:  $\text{WFE}/\lambda = X(\lambda/D)$ , where  $X$  is the measured separation in the focal plane in  $\lambda/D$  units. Averaging the separate tip and tilt differences produces 8 and 55  $\mu\text{m}$  of PV WFE for 1 MEMS and ALPAO unit, respectively (i.e., the tip/tilt stroke limits of these devices). It is important to note that in reality these conversion factors are mode-dependent. E.g., the ALPAO DM waffle stroke limit is 4  $\mu$  PV WFE, not 55 (ALPAO, private communication). However, because the main focus of this paper is the first laboratory demonstration of FAST correction of dynamic AO residual WFEs rather than high accuracy wavefront reconstruction, for simplicity we use these scalar conversion factors for all analyses and plots in this paper.

## Acknowledgments

We gratefully acknowledge research support of the University of California Observatories for funding this research. We thank Maaïke van Kooten, Jules Fowler, Rachel Bowens-Rubin, Dominic Sanchez, Renate Kupke, and Phil Hinz for comments, suggestions, and discussions that have contributed to this manuscript. We thank the reviewers for their helpful feedback and suggestions that improved this paper.

## References

1. E. L. Nielsen et al., “The Gemini Planet Imager exoplanet survey: giant planet and brown dwarf demographics from 10 to 100 au,” *Astrophys. J.* **158**, 13 (2019).
2. B. Macintosh et al., “Discovery and spectroscopy of the young jovian planet 51 Eri b with the Gemini Planet Imager,” *Science* **350**, 64–67 (2015).

3. R. Galicher et al., “The international deep planet survey. II. The frequency of directly imaged giant exoplanets with stellar mass,” *Astron. Astrophys.* **594**, A63 (2016).
4. B. L. Gerard, “Exoplanet imaging speckle subtraction: current limitations and a path forward,” PhD thesis, University of Victoria (2020).
5. M. Fitzgerald et al., “The planetary systems imager for TMT,” *Bull. Am. Astron. Soc.* **51**, 251 (2019).
6. N. Jovanovic et al., “Review of high-contrast imaging systems for current and future ground-based and space-based telescopes: part II. Common path wavefront sensing/control and coherent differential imaging,” *Proc. SPIE* **10703**, 107031U (2018).
7. F. Martinache et al., “On-sky speckle nulling demonstration at small angular separation with SCEXAO,” *Publ. Astron. Soc. Pacific* **126**, 565 (2014).
8. S. P. Bos et al., “First on-sky demonstration of spatial linear dark field control with the vector-apodizing phase plate at Subaru/SCEXAO,” *Astron. Astrophys.* **653**, A42 (2021).
9. M. Bottom et al., “Speckle nulling wavefront control for Palomar and Keck,” *Proc. SPIE* **9909**, 990955 (2016).
10. S. Steiger et al., “SCEXAO/MEC and CHARIS discovery of a low-mass, 6 AU separation companion to HIP 109427 using stochastic speckle discrimination and high-contrast spectroscopy,” *Astrophys. J.* **162**, 44 (2021).
11. A. T. Rodack et al., “Millisecond exoplanet imaging: I method and simulation results,” *J. Opt. Soc. Am. A* **38**, 1541 (2021).
12. A. Potier et al., “Increasing the raw contrast of VLT/SPHERE with the dark hole technique. I. Simulations and validation on the internal source,” *Astron. Astrophys.* **638**, A117 (2020).
13. C. Marois et al., “Upgrading the Gemini Planet Imager calibration unit with a photon counting focal plane wavefront sensor,” *Proc. SPIE* **11448**, 1144873 (2020).
14. P. Baudoz et al., “The self-coherent camera: a new tool for planet detection,” in *IAU Colloq. 200: Direct Imaging of Exoplanets: Science & Techniques*, C. Aime and F. Vakili, Eds., pp. 553–558 (2006).
15. B. L. Gerard, C. Marois, and R. Galicher, “Fast coherent differential imaging on ground-based telescopes using the self-coherent camera,” *Astrophys. J.* **156**, 106 (2018).
16. R. Jensen-Clem et al., “The Santa Cruz Extreme AO Lab (SEAL): design and first light,” *Proc. SPIE* **11823**, 118231D (2021).
17. O. Lardière et al., “Optical design and preliminary results of NEW EARTH, first Canadian high-contrast imaging laboratory test bench,” *Proc. SPIE* **11448**, 1334–1341 (2020).
18. B. L. Gerard et al., “First experimental results of the fast atmospheric self-coherent camera technique on the Santa Cruz extreme adaptive optics laboratory testbed: demonstration of high speed focal plane wavefront control of residual atmospheric speckles,” *Proc. SPIE* **11823**, 118231H (2021).
19. B. L. Gerard et al., “Fast focal plane wavefront sensing on ground-based telescopes,” *Proc. SPIE* **10703**, 1070351 (2018).
20. B. L. Gerard et al., “Fast focal plane wavefront sensing as a second stage adaptive optics wavefront sensor,” *Proc. SPIE* **11448**, 483–499 (2021).
21. J. W. Evans et al., “Contrast analysis and stability on the ExAO testbed,” *Proc. SPIE* **7015**, 70156K (2008).
22. R. Galicher et al., “Self-coherent camera as a focal plane wavefront sensor: simulations,” *Astron. Astrophys.* **509**, A31 (2010).
23. S. Cetre et al., “A near-infrared pyramid wavefront sensor for Keck adaptive optics: real-time controller,” *Proc. SPIE* **10703**, 1070339 (2018).
24. A. Sengupta et al., “Laboratory demonstration of LQG tip-tilt control,” in *Wavefront Sensing in the VLT/ELT Era VII* (2021). <https://www.youtube.com/watch?v=Giz0E73S4C4>.
25. J. W. Hardy, *Adaptive Optics for Astronomical Telescopes*, Oxford University Press (1998).
26. O. Guyon, “Extreme adaptive optics,” *Annu. Rev. Astron. Astrophys.* **56**, 315–355 (2018).
27. E. Gendron and P. Lena, “Astronomical adaptive optics. I. Modal control optimization,” *Astron. Astrophys.* **291**, 337–347 (1994).
28. M. L. Waskom, “Seaborn: statistical data visualization,” *J. Open Source Softw.* **6**(60), 3021 (2021).

29. V. P. Bailey et al., “Status and performance of the Gemini Planet Imager adaptive optics system,” *Proc. SPIE* **9909**, 99090V (2016).
30. C. Marois et al., “Angular differential imaging: a powerful high-contrast imaging technique,” *Astrophys. J.* **641**, 556–564 (2006).
31. C. Marois et al., “Efficient speckle noise attenuation in faint companion imaging,” *Publ. Astron. Soc. Pacific* **112**, 91–96 (2000).
32. B. L. Gerard et al., “The fast atmospheric self-coherent camera technique: laboratory results and future directions,” arXiv:1910.04554 (2019).
33. Q. Fu et al., “Additive lithographic fabrication of a tilt-Gaussian-vortex mask for focal plane wavefront sensing,” *Proc. SPIE* **11889**, 162–170 (2021).
34. C. M. Correia, “Advanced control laws for the new generation of AO systems,” *Proc. SPIE* **10703**, 107031M (2018).
35. M. van Kooten et al., “Predictive control for pupil and focal plane wavefront sensing,” in *Wavefront Sensing in the VLT/ELT Era VII* (2021). <https://www.youtube.com/watch?v=XDgN212uTuQ>.
36. M. A. van Kooten et al., “Status of predictive wavefront control on Keck II adaptive optics bench: on-sky coronagraphic results,” *Proc. SPIE* **11823**, 118231F (2021).
37. J. R. Delorme et al., “Focal plane wavefront sensor achromatization: the multireference self-coherent camera,” *Astron. Astrophys.* **588**, A136 (2016).
38. P. J. Bordé and W. A. Traub, “High-contrast imaging from space: speckle nulling in a low-aberration regime,” *Astrophys. J.* **638**, 488–498 (2006).
39. R. Soummer et al., “Apodized pupil lyot coronagraphs for arbitrary apertures. III. Quasi-achromatic solutions,” *Astrophys. J.* **729**, 144 (2011).
40. O. Lardière et al., “Optical design and preliminary results of NEW EARTH, first Canadian high-contrast imaging laboratory test bench,” *Proc. SPIE* **11448**, 114486Y (2020).
41. L. Pueyo et al., “Optimal dark hole generation via two deformable mirrors with stroke minimization,” *J. Appl. Opt.* **48**, 6296 (2009).
42. U. Bitenc et al., “Assessing the stability of an ALPAO deformable mirror for feed-forward operation,” *Opt. Express* **22**, 12438 (2014).
43. J. Mazoyer et al., “Estimation and correction of wavefront aberrations using the self-coherent camera: laboratory results,” *Astron. Astrophys.* **557**(A9), 13 (2013).

**Benjamin L. Gerard** is a postdoctoral scholar at UC Santa Cruz and UC Observatories (UCO), working on exoplanet imaging and adaptive optics in the Laboratory for Adaptive Optics (LAO) Research Group. He completed his PhD and master’s at the University of Victoria, British Columbia, Canada, working primarily in the exoplanet imaging and adaptive optics groups at the National Research Council of Canada, Herzberg Astronomy and Astrophysics, also in Victoria.

**Daren Dillon** is a development engineer at UCO, working primarily in the LAO group. He has decades of experience and expertise in opto-mechanical development, deployment, and operations of AO technologies.

**Sylvain Cetre** is a software engineer, working primarily for Durham University but also working with UCO and W.M. Keck Observatory. He has decades of experience and expertise in AO software development, deployment, and operations.

**Rebecca Jensen-Clem** is an assistant professor at UCSC, working on exoplanet imaging and adaptive optics science and instrumentation. She received her PhD from Caltech and was a Miller postdoctoral fellow at UC Berkeley.