On Advanced Estimation Techniques for Exoplanet Detection and Characterization Using Ground-based Coronagraphs

Peter R. Lawson[a], Lisa Poyneer[b], Harrison Barrett[c], Richard Frazin[d], Luca Caucci[c], Nicholas Devaney[e], Lars Furenlid[c], Szymon Gładysz[f], Olivier Guyon[g,h], John Krist[a], Jérôme Maire[i], Christian Marois[j], Dimitri Mawet[k], David Mouillet[l], Laurent Mugnier[m], Iain Pearson[c], Marshall Perrin[n], Laurent Pueyo[o], Dmitry Savransky[b]

[a]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA;
[b]Lawrence Livermore National Laboratory, 7000 East Ave, Livermore, CA 94550, USA;
[c]College of Optical Sciences, University of Arizona, Tucson, AZ 85721, USA;
[d]Atmospheric, Oceanic and Space Sciences, Univ. Michigan, Ann Arbor, MI 48109, USA;
[e]Applied Optics Group, School of Physics, National University of Ireland, Galway, Ireland;
[f]Fraunhofer Institute, Gutleuthausstrasse 1, 76275 Ettlingen, Germany;
[g]Steward Observatory, The University of Arizona, Tucson, AZ 85721, USA;
[h]Subaru Telescope, National Astronomical Observatory of Japan, Hilo, HI 96720, USA;
[i]David Dunlap Inst., Univ. of Toronto, 50 St George St., Toronto, ON M5S 3H4, Canada;
[j]NRC, Herzberg Institute of Astrophysics, Victoria, BC V9E 2E7, Canada;
[k]European Southern Observatory, Alonso de Córdova 3107, Vitacura, Casilla 19001, Chile;
[l]IPAG, 414 rue de la Piscine, Domaine Univ., BP 53, 38041 Grenoble Cedex 09, France;
[m]ONERA, Division Optique Theorique et Appliquée, BP 72, 92322 Chatillon Cedex, France;
[n]Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA;
[o]JHU Department of Physics and Astronomy, 3400 N. Charles St, Baltimore, MD 21218, USA

ABSTRACT

The direct imaging of planets around nearby stars is exceedingly difficult. Only about 14 exoplanets have been imaged to date that have masses less than 13 times that of Jupiter. The next generation of planet-finding coronagraphs, including VLT-SPHERE, the Gemini Planet Imager, Palomar P1640, and Subaru HiCIAO have predicted contrast performance of roughly a thousand times less than would be needed to detect Earth-like planets. In this paper we review the state of the art in exoplanet imaging, most notably the method of Locally Optimized Combination of Images (LOCI), and we investigate the potential of improving the detectability of faint exoplanets through the use of advanced statistical methods based on the concepts of the ideal observer and the Hotelling observer. We propose a formal comparison of techniques using a blind data challenge with an evaluation of performance using the Receiver Operating Characteristic (ROC) and Localization ROC (LROC) curves. We place particular emphasis on the understanding and modeling of realistic sources of measurement noise in ground-based AO-corrected coronagraphs. The work reported in this paper is the result of interactions between the co-authors during a week-long workshop on exoplanet imaging that was held in Squaw Valley, California, in March of 2012.

Keywords: Exoplanets, coronagraphs, adaptive optics, instrumentation, theory

Further author information: (Send correspondence to P.R.L.)
P.R.L.: E-mail: Peter.R.Lawson@jpl.nasa.gov, Telephone: +1 (818) 354-0747

1. INTRODUCTION

Several of the co-authors of this paper took part in the conference *Seeing the Future with Imaging Science*, organized by the US National Academies Keck Futures Initiative (NAKFI). The format of that workshop was to bring together scientists and engineers from diverse backgrounds, including radiologists, astronomers, and physicists, to see if they could learn from each other's experience and envisage new approaches to problems of detection and characterization in imaging. One of the study groups at the workshop discussed possible new approaches to exoplanet imaging, and in particular adapting methods used in medical imaging to the post-processing of extreme adaptive-optics corrected coronagraphic images. Of particular interest was the concept of the ideal observer and the Hotelling observer[1–5] and how these approaches might be applied to exoplanet detection. Although these methods have been examined previously, to the authors' knowledge they have not yet been applied to real astronomical data.

Through a seed grant from the Keck Foundation, a follow-on workshop was organized to bring together experienced astronomers as well as experts in imaging theory. Of particular interest was to formally assess and compare the performance of techniques based on the Hotelling observer with the state of the art, represented by angular or spectral differential imaging using the Locally Optimized Combination of Images (LOCI) and its variations.[6–9] It was hoped that in so doing we might enable the detection of exoplanets at least an order of magnitude fainter than what is currently possible.

Direct imaging of a planet around another star is exceedingly challenging. For even the closest stars observed with the largest ground-based telescopes, the angular separation between star and planet will be near the classical diffraction limit. Moreover, a typical star will be about a billion times brighter than the planet to be imaged. The planetary image is also buried in "speckle noise," which is the result of uncorrected wavefront errors that propagate through the atmosphere and optical system. This speckle noise has a complex spatio-temporal-spectral structure, which is different from a planetary signal. While algorithms now exist that exploit some properties of the signal and noise, there has been little effort to address the full problem in a rigorous and comprehensive way. The aim of our work is to begin to address this need.

For this effort, a balance was sought between theorists and astronomers, and invitations were extended to and accepted by participants representing all the major near-term ground-based coronagraph projects, namely P1640,[10] the Gemini Planet Imager[11] (GPI), the SPHERE instrument of the Very Large Telescope, and the Subaru SCExAO project.

Our exoplanet imaging workshop took place at Squaw Valley, CA, during the week of 25-30 March 2012. Each day of the workshop two different broad topics related to imaging were discussed as the subject of possible future research. These included the topics of detection, astrometry, spectroscopy, future instrumentation, lessons learned, and an optimal observing campaign design. The sessions were run on the same model as the NAKFI sessions. The participants were split into two groups and each group decided what question they would tackle, and each reported separately at the end of each session. This way, the most compelling research topics were brought to the fore and several collaborations were formed for future work and study. An informal count of the proposed work suggests that approximately a half-dozen possible future publications will be written as a direct result of the workshop. Further details of the workshop are described in the review paper by Mawet et al.[12]

The participants readily agreed to undertake an exoplanet imaging challenge, for which simulated blind data sets would be prepared and distributed for the testing of new algorithms. This was seen as a longer-term project, extending beyond the workshop, that would form the basis for a future publication. It was also agreed to provide a preliminary report on the imaging study at this SPIE Conference, as described in the following sections.

2. CURRENT PRACTICE

Figure 1 summarizes the current state of the art in coronagraphic imaging and data post-processing. Shown in the figure are the 5-σ contrast limits versus apparent angular separation for observations of 1 hour after post-processing of data and using various coronagraphs. The current generation of instruments is represented by 1) the NIRC2 instrument on the Keck II telescope, 2) the Palomar Well-Corrected Subaperture (WCS), and 3) the NACO instrument at the European Southern Observatory's Very Large Telescope (VLT), which have been used for example to detect the planets around HR 8799[13–15] and the planet around β Pictoris,[16,17] whose K-band delta

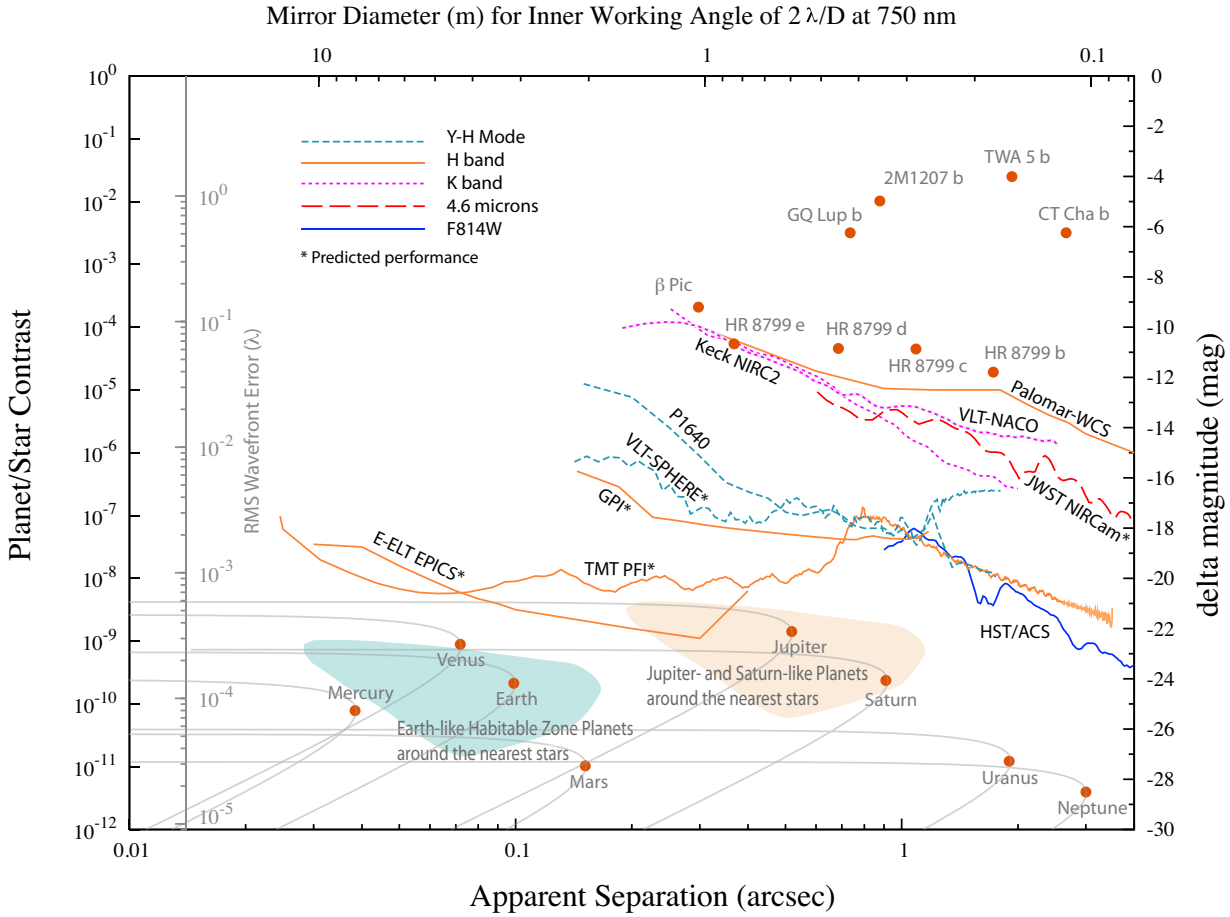


Figure 1:
The 5-σ contrast limits after post-processing of 1 hour's worth of data for various coronagraph instruments. As can be seen in the plot, there are roughly three groupings of curves: 1) the current state of the art represented by Keck NIRC2, the Palomar Well-Corrected Subaperture, and VLT-NACO; 2) the next generation of instruments, represented by P1640, GPI, and VLTI-SPHERE; and 3) future extremely large telescopes, represented by TMT PFI and E-ELT EPICS. The contrast curves for JWST NIRCam and HST/ACS are shown for reference. On the top right of the figure are plotted the K-band contrasts of 9 of the approximately 14 exoplanets imaged to date. In the lower part of the figure are plotted our solar system planets as they would appear in reflected light around a Sun-like star at a distance of 10 pc. On the left-hand of the plot is shown the corresponding RMS wavefront error for a coronagraph using a 64 $\times$ 64 element deformable mirror. The shaded regions represent where 1) Earth-like and 2) Jupiter- and Saturn-like planets would be detectable around the nearest stars.

magnitudes are also shown in the figure.[18] The upcoming generation of instruments, coming online in 2012–2013 is represented by the curves for 1) Project 1640, 2) VLT-SPHERE, and 3) the Gemini Planet Imager (GPI). Possible future instruments on telescopes larger than 10-m are represented by the curve for the Thirty Meter Telescope's Planet Finding Instrument, and the European Extremely Large Telescope's EPICS instrument. The data for these curves were obtained from the literature and the individual authors themselves and rescaled for 5-σ and 1 hour observations. (The sources for the curves are further detailed in the acknowledgments of this paper.) Also shown in the figure are data for solar system planets, as they would appear in reflected light around a Sun-like star at a distance of 10 pc. We would hope through the work discussed here to extend the performance

limits of all these instruments.

3. STATISTICAL DECISION THEORY FOR DETECTION AND CHARACTERIZATION OF EXOPLANETS

3.1 Tasks, observers and ROC curves

The tasks of interest in exoplanet studies include:

- Pure detection: Is there a planet present?
- Astrometry: Where is the planet located?
- Photometry: How bright is the planet?
- Estimation of orbital parameters: Is it really a planet and not a background star?
- Spectrometry: What are the spectral characteristics of the light from the planet?
- Detection of life: Does the spectrum contain components consistent with living organisms?

Performance on a pure detection task is quantified by the Receiver Operating Characteristic (ROC) curve, illustrated in Fig. 2. A common figure of merit for a detection task is the area under the ROC curve, denoted AUC, which ranges from 0.5 for an imaging system that gives no information at all about the presence of the signal, to 1.0 for a system that permits perfect detection. AUC is a figure of merit not only for the system that acquires the image data but also for the observer (human or algorithm) that makes a decision on the basis of the data.

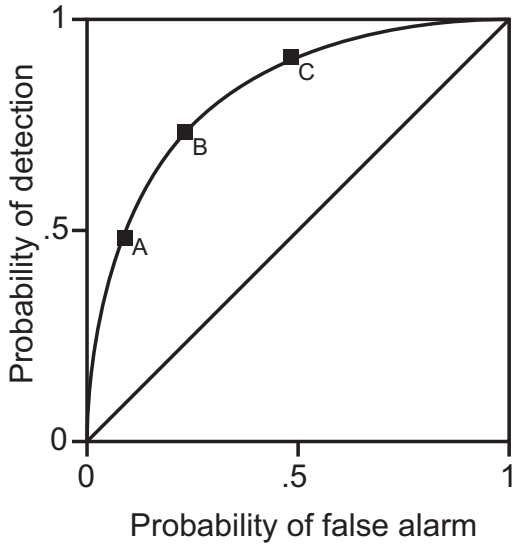


Figure 2:
Illustration of an ROC curve, which is a plot of probability of detection (also known as true-positive fraction) vs. probability of a false alarm (false-positive fraction). Points along the curve are generated by choosing different decision thresholds. Point A, for example, corresponds to a strict threshold which yields a low probability of detection but also a low false-positive fraction. Points B and C correspond to less strict thresholds.

For a combined detection-localization task, the results may be depicted by a localization-ROC (LROC), which is a plot of probability of detection of a signal and correct localization (within some tolerance) vs. probability of a false alarm (false-positive fraction) as shown in Fig. 3.

Both the ROC and LROC curves assume either zero or one signal per image, but the theory can be generalized to allow an arbirtary (but random and unknown) number of signals per image by means of the free-response ROC (FROC) curve, which is a plot of the probability of detection and correct localization vs. number of false positives per image.

A further generalization that encompasses the detection-estimation tasks enumerated above is the estimation-ROC (EROC[19]), which is a plot of expected utility of a correct detection vs. false-positive fraction.

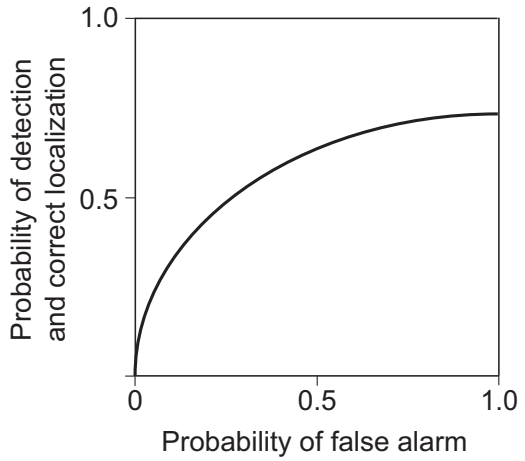


Figure 3:
Illustration of an LROC curve, Points along the curve are again generated by choosing different decision thresholds, and the location is then estimated for all candidate signals that exceed the threshold.

Whichever task and ROC is chosen, it is interesting to determine the *observer* that maximizes the task performance. For a pure detection task, the *ideal observer* is defined as one that maximizes the AUC, or equivalently maximizes the probability of detection for any specified false-alarm rate. Similar ideal observers have been derived for LROC and EROC curves.

All of these ideal observers require knowledge of a huge multivariate probability density function of the image data, which is usually impossible to calculate. A more practical alternative, in many problems in imaging, is the Hotelling observer, which performs only linear operations on the image data. The simple Hotelling observer is optimal among all linear observers for the pure detection task, in the sense of maximizing a certain detectability index derived from the ROC curve. The scanning Hotelling observer is similarly optimal for the localization-detection task. Both these observers are in fact completely optimal for their respective tasks, in an AUC sense, if the data statistics are multivariate normal.[20]

To construct and implement the Hotelling and scanning Hotelling observers, we need to determine the ensemble means of the image data and the data covariance matrices for both signal-present and signal-absent hypotheses. The basic principles leading to these statistical characterizations are summarized briefly below, and a more complete treatment will be published separately.

3.2 Data statistics

The analysis in this subsection follows Barrett et al.[2] except that it is extended to include quasistatic speckle.

The image data is denoted abstractly as $\mathbf{G}$, which indicates some sequence of two-dimensional (2D) image frames indexed by time and wavelength. A collection of frames for multiple wavelengths at one tme is known as a data cube, and the set of all data cubes will be referred to as the data set, or simply the data.

The data set is a random vector, and we consider here four separate random effects that influence the statistics of $\mathbf{G}$:

- Residual atmospheric speckle, caused by noise in the wavefront sensor and null functions in the deformable mirror. These random effects are denoted $\mathbf{R}^{atm}$.
- Quasistatic speckle, caused by random phase aberrations and log-amplitude perturbations in various optical elements outside the common path defined by the control loop for atmospheric corretion. These effects are denoted collectively as $\mathbf{R}^{qs}$.
- Noise in the science camera, $\mathbf{N}^{sci}$.
- Randomness in the object $\mathbf{F}$ being viewed (e.g., diffuse sky background, atmospheric absorption, background galaxies).

We consider two hypotheses. The null hypothesis H_0 is that there is no planet present in the scene, and the alternative hypothesis H_1 is that there is a planet. For localization or estimation tasks, H_1 can be a composite hypothesis, $H_1(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a parameter denoting planet location, brightness or other quantities of interest.

The overall mean of the data set $\mathbf{G}$ under hypothesis H_j is denoted by

$$\overline{\overline{\overline{\mathbf{G}}}}_{H_j} \equiv \left\langle \mathbf{G} \right\rangle_{\mathbf{G}|H_j} , \tag{1}$$

where the angle brackets indicate ensemble averages over the four sources of randomness identified above. Explicitly, the overall mean of the data set $\mathbf{G}$ under hypothesis H_j $(j = 0, 1)$ is given by

$$\overline{\overline{\overline{\mathbf{G}}}}_{H_j} = \left\langle \left\langle \left\langle \left\langle \mathbf{G} \right\rangle_{\mathbf{N}^{sci}|\mathbf{R}^{qs},\mathbf{R}^{atm},\mathbf{F}} \right\rangle_{\mathbf{R}^{qs}|\mathbf{R}^{atm},\mathbf{F}} \right\rangle_{\mathbf{R}^{atm}|\mathbf{F}} \right\rangle_{\mathbf{F}|H_j} . \tag{2}$$

It is useful to define three partial averages as follows:

$$\overline{\mathbf{G}}(\mathbf{R}^{qs}, \mathbf{R}^{atm}, \mathbf{F}) \equiv \left\langle \mathbf{G} \right\rangle_{\mathbf{N}^{sci}|\mathbf{R}^{qs},\mathbf{R}^{atm},\mathbf{F}} ; \tag{3}$$

$$\overline{\overline{\mathbf{G}}}(\mathbf{R}^{atm}, \mathbf{F}) \equiv \left\langle \overline{\mathbf{G}}(\mathbf{R}^{qs}, \mathbf{R}^{atm}, \mathbf{F}) \right\rangle_{\mathbf{R}^{qs}|\mathbf{R}^{atm},\mathbf{F}} ; \tag{4}$$

$$\overline{\overline{\overline{\mathbf{G}}}}(\mathbf{F}) \equiv \left\langle \overline{\overline{\mathbf{G}}}(\mathbf{R}^{atm}, \mathbf{F}) \right\rangle_{\mathbf{R}^{atm}|\mathbf{F}} . \tag{5}$$

All of the averages defined in (2) – (5) are vectors of the same dimensionality as the data vector $\mathbf{G}$ and all are indexed by spatial position and wavelength in each data cube and the time of acquisition of that cube. All of the averages are over an ensemble of exposures, not over time or space.

The subscript H_j is not needed on any of the partial averages in (3) – (5) because they are conditional on the astronomical object being imaged, $\mathbf{F}$; when we specify $\mathbf{F}$ we do not need to further specify the class from which it was drawn.

Note that the partial averages are always performed in a particular sequence, starting from the science camera and working backwards through the telescope. The lowest overbar indicates an average over noise in the science camera, the second lowest indicates an average over random effects that pertain to the quasistatic speckle, and the third is over effects that pertain to atmospheric speckle. The fourth and final overbar, as in (2), denotes an average over all objects in a class described by hypothesis H_j. Henceforth we will omit the arguments on these partial averages, for example, writing $\overline{\overline{\mathbf{G}}}$ instead of $\overline{\overline{\mathbf{G}}}(\mathbf{R}^{atm}, \mathbf{F})$.

The overall covariance matrix of the random vector $\mathbf{G}$ under hypothesis H_j is defined in outer-product form as

$$\mathbf{K}_{\mathbf{G}|H_j} \equiv \left\langle \left[\mathbf{G} - \overline{\overline{\overline{\mathbf{G}}}}_{H_j} \right] \left[\mathbf{G} - \overline{\overline{\overline{\mathbf{G}}}}_{H_j} \right]^t \right\rangle_{\mathbf{G}|H_j} , \tag{6}$$

where the superscript t denotes a vector transpose. To see the precise meaning of this transpose, it is useful to define a vector index $\mathbf{m} \equiv (m_x, m_y, m_t, m_\lambda)$, where the four components are integer indices specifying x and y

pixel locations, frame time and wavelength, respectively. Then a single component of $\mathbf{G}$ is denoted either $G_{\mathbf{m}}$ or as $[\mathbf{G}]_{\mathbf{m}}$. The overall dimension of $\mathbf{G}$ is denoted M. The covariance matrix carries two of these vector indices and is written, $[\mathbf{K}_{\mathbf{G}|H_j}]_{\mathbf{m}\mathbf{m}'}$. Explicitly,

$$[\mathbf{K}_{\mathbf{G}|H_j}]_{\mathbf{m}\mathbf{m}'} = \left\langle \left[\mathbf{G} - \overline{\overline{\overline{\mathbf{G}}}}_{H_j}\right]_{\mathbf{m}} \left[\mathbf{G} - \overline{\overline{\overline{\mathbf{G}}}}_{H_j}\right]_{\mathbf{m}'} \right\rangle_{\mathbf{G}|H_j} . \tag{7}$$

Thus $\mathbf{G}$ can be manipulated as a column vector, even though it has four indices, and $\mathbf{G}^t$ is the corresponding row vector.

By a simple mathematical tautology, (6) can be written as

$$\mathbf{K}_{\mathbf{G}|H_j} \equiv \left\langle \left[\mathbf{G} - \overline{\mathbf{G}} + \overline{\mathbf{G}} - \overline{\overline{\mathbf{G}}} + \overline{\overline{\mathbf{G}}} - \overline{\overline{\overline{\mathbf{G}}}} + \overline{\overline{\overline{\mathbf{G}}}} - \overline{\overline{\overline{\mathbf{G}}}}_{H_j}\right] \left[\mathbf{G} - \overline{\mathbf{G}} + \overline{\mathbf{G}} - \overline{\overline{\mathbf{G}}} + \overline{\overline{\mathbf{G}}} - \overline{\overline{\overline{\mathbf{G}}}} + \overline{\overline{\overline{\mathbf{G}}}} - \overline{\overline{\overline{\mathbf{G}}}}_{H_j}\right]^t \right\rangle_{\mathbf{G}|H_j} . \tag{8}$$

Remarkably, when we write out this outer product and perform the averages, all cross terms vanish identically,[2] without any assumptions of statistical independence, and we find

$$\begin{aligned} \mathbf{K}_{\mathbf{G}|H_j} = &\left\langle \left[\mathbf{G} - \overline{\mathbf{G}}\right] \left[\mathbf{G} - \overline{\mathbf{G}}\right]^t \right\rangle_{\mathbf{G}|H_j} + \left\langle \left[\overline{\mathbf{G}} - \overline{\overline{\mathbf{G}}}\right] \left[\overline{\mathbf{G}} - \overline{\overline{\mathbf{G}}}\right]^t \right\rangle_{\mathbf{G}|H_j} \\ &+ \left\langle \left[\overline{\overline{\mathbf{G}}} - \overline{\overline{\overline{\mathbf{G}}}}\right] \left[\overline{\overline{\mathbf{G}}} - \overline{\overline{\overline{\mathbf{G}}}}\right]^t \right\rangle_{\mathbf{G}|H_j} + \left\langle \left[\overline{\overline{\mathbf{G}}} - \overline{\overline{\overline{\mathbf{G}}}}_{H_j}\right] \left[\overline{\overline{\mathbf{G}}} - \overline{\overline{\overline{\mathbf{G}}}}_{H_j}\right]^t \right\rangle_{\mathbf{G}|H_j} . \end{aligned} \tag{9}$$

Now we are in a position to rewrite the overall covariance matrix as

$$\mathbf{K}_{\mathbf{G}|H_j} = \overline{\overline{\overline{\mathbf{K}}}}^{sci}_{\mathbf{G}|H_j} + \overline{\overline{\mathbf{K}}}^{qs}_{\overline{\mathbf{G}}|H_j} + \overline{\mathbf{K}}^{atm}_{\overline{\overline{\mathbf{G}}}|H_j} + \mathbf{K}^{obj}_{\overline{\overline{\overline{\mathbf{G}}}}|H_j} , \tag{12}$$

where

$$\overline{\overline{\overline{\mathbf{K}}}}^{sci}_{\mathbf{G}|H_j} = \left\langle \left\langle \left\langle \left\langle \left[\mathbf{G} - \overline{\mathbf{G}}\right] \left[\mathbf{G} - \overline{\mathbf{G}}\right]^t \right\rangle_{\mathbf{N}^{sci}|\mathbf{R}^{qs},\mathbf{R}^{atm},\mathbf{F}} \right\rangle_{\mathbf{R}^{qs}|\mathbf{R}^{atm},\mathbf{F}} \right\rangle_{\mathbf{R}^{atm}|\mathbf{F}} \right\rangle_{\mathbf{F}|H_j} , \tag{13}$$

$$\overline{\overline{\mathbf{K}}}^{qs}_{\overline{\mathbf{G}}|H_j} = \left\langle \left\langle \left\langle \left[\overline{\mathbf{G}} - \overline{\overline{\mathbf{G}}}\right] \left[\overline{\mathbf{G}} - \overline{\overline{\mathbf{G}}}\right]^t \right\rangle_{\mathbf{R}^{qs}|\mathbf{R}^{atm},\mathbf{F}} \right\rangle_{\mathbf{R}^{atm}|\mathbf{F}} \right\rangle_{\mathbf{F}|H_j} , \tag{14}$$

$$\overline{\mathbf{K}}^{atm}_{\overline{\overline{\mathbf{G}}}|H_j} = \left\langle \left\langle \left[\overline{\overline{\mathbf{G}}} - \overline{\overline{\overline{\mathbf{G}}}}\right] \left[\overline{\overline{\mathbf{G}}} - \overline{\overline{\overline{\mathbf{G}}}}\right]^t \right\rangle_{\mathbf{R}^{atm}|\mathbf{F}} \right\rangle_{\mathbf{F}|H_j} , \tag{15}$$

$$\mathbf{K}^{obj}_{\overline{\overline{\overline{\mathbf{G}}}}|H_j} = \left\langle \left[\overline{\overline{\overline{\mathbf{G}}}} - \overline{\overline{\overline{\mathbf{G}}}}\right] \left[\overline{\overline{\overline{\mathbf{G}}}} - \overline{\overline{\overline{\mathbf{G}}}}\right]^t \right\rangle_{\mathbf{F}|H_j} . \tag{16}$$

To understand the notational conventions here, we note that the innermost expectation in (13) is the covariance matrix of the science-camera noise, $\mathbf{N}^{sci}$, conditional on $\mathbf{R}^{qs}$, $\mathbf{R}^{atm}$ and $\mathbf{F}$. The three outer expectations then correspond to averaging the covariance on $\mathbf{N}^{sci}$ over the three remaining sources of randomness, hence the three overbars. Similarly, the innermost expectation in (14) is the covariance matrix of $\overline{\mathbf{G}}$, conditional on $\mathbf{R}^{atm}$ and $\mathbf{F}$, so we use $\overline{\mathbf{G}}$ as the subscript on the covariance matrix and add two overbars.

The number of overbars on the subscript of the covariance matrix plus the number on the covariance matrix itself always equals three, and a covariance matrix is also an average, so each of the four expressions in (13) – (16) involves four averages. The differences are in how many averages are applied to $\mathbf{G}$ before forming the covariance matrix and how many to the covariance matrix itself.

We emphasize that these results are exact and require no assumptions of statistical independence except those dictated by the position of the random effect along the optical chain (e.g., the science camera does not influence the atmospheric wavefront sensor or the astronomical object).

3.3 Structure of ideal and Hotelling observers

The exoplanet problem is all about detecting very weak objects and measuring their parameters. One well-justified approximation is to say that the planets, if present, are so faint that they have no measurable effect on any of the terms in the covariance expansion. This does not mean that they are undetectable, but rather that the effect of a planet is to change the mean signal rather than the covariance.

For a pure detection task and multivariate Gaussian statistics, the ideal observer uses a discriminant function (logarithm of the likelihood ratio) of the form

$$t_{ideal}(\mathbf{G}) = -\tfrac{1}{2}\left[\mathbf{G} - \overline{\overline{\overline{\mathbf{G}}}}_{H_1}\right]^t \mathbf{K}^{-1}_{\mathbf{G}|H_1}\left[\mathbf{G} - \overline{\overline{\overline{\mathbf{G}}}}_{H_1}\right] + \tfrac{1}{2}\left[\mathbf{G} - \overline{\overline{\overline{\mathbf{G}}}}_{H_0}\right]^t \mathbf{K}^{-1}_{\mathbf{G}|H_0}\left[\mathbf{G} - \overline{\overline{\overline{\mathbf{G}}}}_{H_0}\right], \tag{17}$$

The weak-signal approximation is to say that $\mathbf{K}_{\mathbf{G}|H_1} \approx \mathbf{K}_{\mathbf{G}|H_0} \equiv \mathbf{K}_{\mathbf{G}}$ and also to neglect terms quadratic in the difference in the means. The result is

$$t_{Hot}(\mathbf{G}) = \left[\overline{\overline{\overline{\mathbf{G}}}}_{H_1} - \overline{\overline{\overline{\mathbf{G}}}}_{H_0}\right]^t \mathbf{K}^{-1}_{\mathbf{G}}\left[\mathbf{G} - \overline{\overline{\overline{\mathbf{G}}}}_{H_0}\right]. \tag{18}$$

If the task is to detect the planet and measure a set of parameters $\boldsymbol{\theta}$, $\overline{\overline{\overline{\mathbf{G}}}}_{H_1}$ must be replaced by $\overline{\overline{\overline{\mathbf{G}}}}_{H_1|\boldsymbol{\theta}}$ in (18) to get the weak-signal approximation.

To interpret (18), recall that $\mathbf{G}$ can be treated as a high-dimensional column vector, so an expression of the form $\mathbf{W}^t\mathbf{G}$, where $\mathbf{W}$ is another column vector the same size as $\mathbf{G}$, indicates a scalar product. The Hotelling test statistic $t_{Hot}(\mathbf{G})$ is thus formed by first subtracting the mean vector $\overline{\overline{\overline{\mathbf{G}}}}_{H_0}$, then taking a scalar product with a template given by

$$\mathbf{W}_{Hot} = \mathbf{K}^{-1}_{\mathbf{G}}\left[\overline{\overline{\overline{\mathbf{G}}}}_{H_1} - \overline{\overline{\overline{\mathbf{G}}}}_{H_0}\right]^t. \tag{19}$$

For the scanning Hotelling observer, this template depends on $\boldsymbol{\theta}$, even in the weak-signal approximation.

A useful alternative interpretation of (18) is obtained by recognizing that the covariance matrix is symmetric and positive definite (more on this point below), so we can define the square root of the inverse of the covariance matrix and write

$$t_{Hot}(\mathbf{G}) = \left[\mathbf{K}^{-1/2}_{\mathbf{G}}\left(\overline{\overline{\overline{\mathbf{G}}}}_{H_1} - \overline{\overline{\overline{\mathbf{G}}}}_{H_0}\right)\right]^t \left[\mathbf{K}^{-1/2}_{\mathbf{G}}\left(\mathbf{G} - \overline{\overline{\overline{\mathbf{G}}}}_{H_0}\right)\right]. \tag{20}$$

When it acts on a data vector $\mathbf{G}$, the operator $\mathbf{K}^{-1/2}_{\mathbf{G}}$ converts whatever noise is present in $\mathbf{G}$ into uncorrelated noise of unit variance. In electrical engineering parlance, stationary uncorrelated noise is called *white noise*, so by analogy $\mathbf{K}^{-1/2}_{\mathbf{G}}$ is called a *prewhitening* operator. The second factor in square brackets in (20) is the prewhitened data vector, with zero mean and a covariance matrix given by the unit matrix $\mathbf{I}$ in data space. Similarly, the first factor is the prewhitened mean difference signal. Thus the Hotelling test statistic can be formed by subtracting the mean background, prewhitening the the data vector and then taking a scalar product with the prewhitened difference signal. For a scanning Hotelling observer, the prewhitening of the data needs to be performed only once.

3.4 Computational methods

The major problem in applying the Hotelling formalism in imaging is that the covariance matrix is enormous. For the challenge data described below, each data set consists of 120 data cubes, each of which consists of 277 $\times$ 277 spatial pixels for each of 5 wavelengths. Thus $\mathbf{G}$ has approximately 4.6×10^7 elements, and its covariance matrix is $(4.6 \times 10^7) \times (4.6 \times 10^7)$, for a total of 2.1×10^{15} elements. It appears quite impossible even to create and store such a matrix, much less to find its inverse or inverse square root.

The key to tractable Hotelling algorithms is the covariance decomposition of (12), which we can rewrite here as

$$\mathbf{K}_{\mathbf{G}} = \mathbf{K}^{n}_{\mathbf{G}} + \mathbf{K}^{nf}_{\mathbf{G}}, \tag{21}$$

where superscript n indicates *noise* (specifically, this term describes the noise from the science camera), *nf* indicates *noise-free* and

$$\mathbf{K}_{\mathbf{G}}^{n} \equiv \overline{\overline{\overline{\mathbf{K}}}}_{\mathbf{G}}^{sci} \; ; \tag{22}$$

$$\mathbf{K}_{\mathbf{G}}^{nf} \equiv \overline{\overline{\overline{\mathbf{K}}}}_{\overline{\mathbf{G}}}^{qs} + \overline{\mathbf{K}}_{\overline{\overline{\mathbf{G}}}}^{atm} + \mathbf{K}_{\overline{\overline{\overline{\mathbf{G}}}}}^{obj} \, . \tag{23}$$

The advantage of (21) is that $\mathbf{K}_{\mathbf{G}}^{n}$ can often be considered to be known and full rank, while (23) often has to be estimated from a set of training samples. In simulation, the noise-free training samples are readily obtained just by omitting the noise in the science camera, and we can write

$$\mathbf{K}_{\mathbf{G}} \approx \mathbf{K}_{\mathbf{G}}^{n} + \hat{\mathbf{K}}_{\mathbf{G}}^{nf} \, , \tag{24}$$

where the hat indicates a sample covariance matrix.

With real data, we can get a good estimate of $\mathbf{K}_{\mathbf{G}}^{nf}$ by forming the overall sample covariance matrix and filtering out the diagonal or near-diagonal terms that arise from noise in the science camera.

If we have obtained a set of N_s noise-free training images either by simulating them without science-camera noise or by stripping off the noise contribution to the sample covariance matrix, and we denote the k^{th} such image as $\mathbf{G}_k$, then we can form the $\mathbf{M} \times N_s$ matrix $\mathbf{R}$ given by

$$\mathbf{R} = \frac{1}{\sqrt{N_s - 1}} \left[\delta\mathbf{G}_1, \, \delta\mathbf{G}_2, \, ..., \delta\mathbf{G}_{N_s} \right] , \tag{25}$$

where

$$\delta\mathbf{G}_k \equiv \mathbf{G}_k - \frac{1}{N_s} \sum_{j=1}^{N_s} \mathbf{G}_j \, . \tag{26}$$

Thus each column of $\mathbf{R}$ is $1/\sqrt{N_s - 1}$ times a sample image minus the arithmetic average of all of the sample images.

With these definitions,

$$\hat{\mathbf{K}}_{\mathbf{G}}^{nf} = \mathbf{R}\mathbf{R}^t \, , \tag{27}$$

and (24) becomes

$$\mathbf{K}_{\mathbf{G}} \approx \mathbf{K}_{\mathbf{G}}^{n} + \mathbf{R}\mathbf{R}^t \, . \tag{28}$$

The Woodbury Matrix inversion lemma allows us to write

$$\left[\mathbf{K}_{\mathbf{G}}^{n} + \mathbf{R}\mathbf{R}^t \right]^{-1} = \left[\mathbf{K}_{\mathbf{G}}^{n} \right]^{-1} - \left[\mathbf{K}_{\mathbf{G}}^{n} \right]^{-1} \mathbf{R} \left\{ \mathbf{I} + \mathbf{R}^t \left[\mathbf{K}_{\mathbf{G}}^{n} \right]^{-1} \mathbf{R} \right\}^{-1} \mathbf{R}^t \left[\mathbf{K}_{\mathbf{G}}^{n} \right]^{-1} \, . \tag{29}$$

Thus

$$t_{Hot}(\mathbf{G}) \approx \left[\Delta\overline{\overline{\overline{G}}} \right]^t \left[\mathbf{K}_{\mathbf{G}}^{n} \right]^{-1} \left[\mathbf{G} - \overline{\overline{\overline{\mathbf{G}}}}_{H_0} \right] - \left[\Delta\overline{\overline{\overline{G}}} \right]^t \left[\mathbf{K}_{\mathbf{G}}^{n} \right]^{-1} \mathbf{R} \left\{ \mathbf{I} + \mathbf{R}^t \left[\mathbf{K}_{\mathbf{G}}^{n} \right]^{-1} \mathbf{R} \right\}^{-1} \mathbf{R}^t \left[\mathbf{K}_{\mathbf{G}}^{n} \right]^{-1} \left[\mathbf{G} - \overline{\overline{\overline{\mathbf{G}}}}_{H_0} \right] \, . \tag{30}$$

The advantage of this form is that $\mathbf{I} + \mathbf{R}^t \left[\mathbf{K}_{\mathbf{G}}^{n} \right]^{-1} \mathbf{R}$ is an $N_s \times N_s$ matrix, readily inverted numerically for N_s up to 10^4 or so. The matrix $\mathbf{K}_{\mathbf{G}}^{n}$ is still huge, but it is full rank and usually diagonal.

The requisite $N_s \times N_s$ inverse can be performed by singular-value decomposition, and this same formalism allows easy construction of the prewhitening operator. Details will be published separately.

4. EXOPLANET IMAGING CHALLENGE DATA

4.1 Description of Simulated Images

Much of the early discussion amongst the participants centered around the organization of an exoplanet imaging challenge that would allow a fair comparison of the performance of different algorithms, specifically including LOCI and methods using the Hotelling observer. The data sets were prepared by Lisa Poyneer with the help of Marshall Perrin and Jérôme Maire. This section is Lisa's overview of the data sets. The data themselves can be downloaded at http://olbin.jpl.nasa.gov/nakfi/.

The challenge data were produced using a suite of simulation tools developed for the Gemini Planet Imager (GPI) project. However, the data *do not represent a prediction of actual GPI performance, and should not be used to draw conclusions about GPI observational quality.*

The goal for this data challenge is to provide a large data set containing many (but not all) important error sources in a way that will allow testing and comparison of alternate planet finding algorithms. For this initial challenge set the group decided to consider bright stars (so as to not be photon-limited), to observe with field rotation (so that ADI could be used) and to have typical atmosphere and quasi-static speckles, though with quasi-static speckles due only to phase errors, not amplitude errors. These quasi-static errors were generated to be correlated with pointing, so that a phenomenon that may occur in a real system (correlation of quasi-static errors across different observations) could be explored.

Finally, GPI simulation tools were used to directly produce a noisy Integral Field Spectrograph (IFS) datacube, as opposed to dispersing the image onto a detector and then re-assembling a datacube via a wavelength solution.[21] This provides a simpler first case where the noise in the data is white; in a later data challenge we could address colored noise due to the dispersion and reassembly. In the following paragraphs we provide more description and explanation of various aspects.

The data challenge scenario contains 100 unique observations, drawn from a simulated GPI observing campaign. The GPI campaign scheduler (Savransky, in preparation) was used with a fake star catalog to produce observation parameters (ra, dec, ha) for each target as observed from Cerro Pachon. For this data set we wanted to limit photon noise so all targets are post-facto assigned a stellar magnitude of 5 (bright), even though most were fainter. Each star retains a specific stellar spectral type.

Along with a star and observation parameters, the campaign scheduler software also produces a random assortment of planets around that star, drawn from our models for planet mass, orbit, etc. Some stars have no planets, some have several. Each planet is given a specific angular separation, position angle in the field of view, and brightness. All planets are assumed to have the spectral type L8. In order to probe contrasts from $10^{-4.5}$ to $10^{-6.5}$, the true planet magnitudes are made brighter by 2 to adjust to the artificially bright stars.

A modified version of the GPI Adaptive Optics (AO) simulator[22] was used to produce the instrument Point Spread Function (PSF) at five wavelengths sampled in H-band. This simulator was coupled with a vastly simplified model of the GPI calibration system[23] and instrument optical design to provide quasi-static errors. The GPI AO simulator handles both common-path phase aberrations (usually atmospheric, which the AO system can correct) and non-common-path (NCP) aberrations (which the AO system does not correct, but which should be seen by the Calibration system).

Before the simulation begins, the observation pointing is used to produce a unique time-varying set of NCP errors which produce quasi-static speckle. It is far beyond the scope of this effort to completely and time-efficiently model all of the optics in GPI, their phase and amplitude variations,[24] the gravity-induced shearing that produces NCP errors, and the Calibration system's measurement and correction. Instead, we used a simple model that, while not capturing the true behavior of the system, does provide reasonable levels of NCP phase errors, temporal variations and spatial distributions of speckles.

In short, the pointing information (ra, dec, ha) for the observation is converted to a gravity vector. This gravity vector acts on a very simple cantilever model for the GPI instrument and produces a time-varying and gravity-dependent shearing between optical surfaces during the observation. Two out-of-plane optics were chosen, since these produce chromatic phase and amplitude errors, like those that GPI will actually see. (As noted above, amplitude errors for quasi-static speckles were ignored.) The instrument is assumed to start with a

very low level of quasi-static error (due to speckle nulling or the like). As the observation progresses, the optics shear. The phase errors on each optic are Talbot propagated to the pupil plane (resulting amplitude errors are ignored). At the pupil, a simple model for Calibration performance is assumed, and the corrections are applied in a slow closed loop every 1 minute. This produces a sequence of 3600 files sampled at 1 second, each representing the continually changing NCP error as compensated by a noisy Calibration system at 1-minute intervals.

The main computational stumbling block for the data challenge is the time it takes to fully simulate the AO system. GPI runs at 1 kHz and the full GPI AO simulator tracks AO performance and PSFs generation at each time step. A frozen-flow model of the atmosphere is assumed; the phase screen(s) are translated following wind velocity vectors. Generating PSFs every 1 millisecond is computationally unfeasible for a data challenge of this size. Instead, we chose to use a single short exposure (1 ms) image to approximate a 1-second exposure.

Since the typical wind speed is 10 m/s (more on that below), doing a single short exposure per second approximates the clearing time of phase across the 8-meter aperture. Though not fully accurate (the speckles will be sharper and more variable), this is reasonable. Previous work by ourselves with the AO simulator[21] and by others for non-AO applications[25] show that significant speckle averaging occurs only after a characteristic period dependent on the clearing time (pupil diameter divided by wind velocity).

So to generate a 60-minute observations, we use 3600 individual runs of the AO simulator, each producing a 1-ms short exposure PSF that approximates a 1-second PSF. Each short exposure uses a phase screen generated with a unique random seed.

The wind is the major source of variability for the AO run. For each observation, a unique time-varying atmospheric profile is generated, comprising the turbulence strength r_0 and wind velocity vector. The mean r_0 is 14 cm; the mean wind speed is 10 m/s. Simple models (e.g. smoother random noise, slow random walks) were used to produce plausible evolution of atmospheric statistics over the one hour observation time. Furthermore, since the instrument observed with field rotation to enable ADI processing, the direction of the wind in the sky changes due to the parallactic angle. This is calculated and tracked and fed into the AO simulator such that the atmospheric errors due to wind velocity (e.g. a butterfly pattern in the dark hole) will move correctly through the observation as the parallactic angle changes.

Note that this approach of using 1-ms AO short exposure has huge computational savings, but will produce images where the atmospheric speckle noise (non-smoothness) is higher than a full 1 kHz simulation.

Also note that this approach has both the AO residuals and the quasi-static NCP errors in the complex field, so that the generated PSFs fully capture the non-linear interactions between the two types of errors. (A much faster, but less accurate, method would be to use power spectral density descriptions of the errors to produce random PSFs reflecting atmospheric and quasi-static PSFs, and then add the terms.)

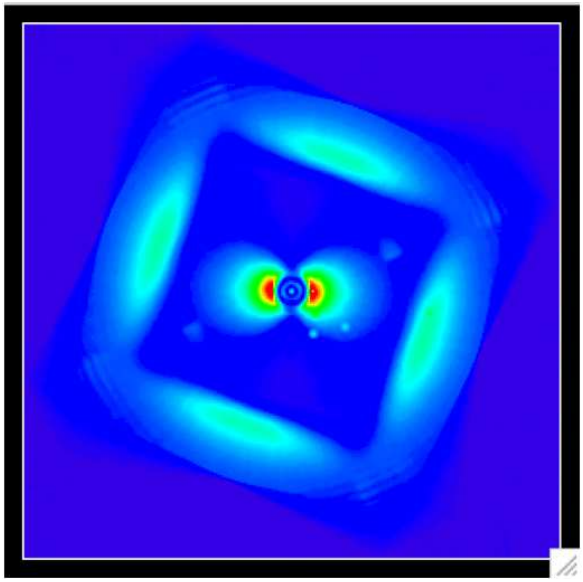
Note that for the purposes of this initial work, we ignore chromatic aberrations that would be introduced by the apodized pupil Lyot coronagraph, microlenses, or prisms. The AO simulation uses a wavelength scaling to estimate the wavelength-dependent atmospheric residual: the phase at at nominal wavelength λ_0 is multiplied by a scaling factor λ/λ_0, where λ is the wavelength of interest. For now we are therefore neglecting an important set of limiting errors.

After the 3600 short AO (with NCP) error simulations are run and produce 3600 5-wavelength PSFs, the images are composited via addition into 30-second exposures. These 120 images form the basis of each observation for the data challenge.*

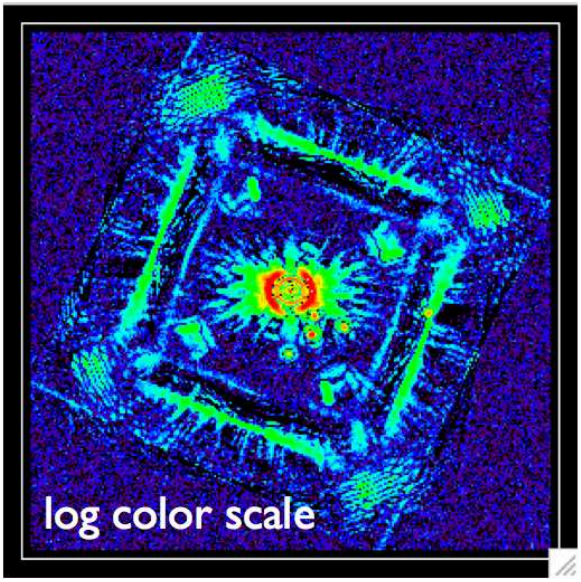
For the final step of converting to an idealized IFS data cube of the full astronomical scene, we used the GPI Data Simulation Tool[21] (DST). First, each AO output file is converted to the proper input format for the DST. Next, a script file is produced using the observation information (star brightness and spectrum, observations ra, dec, and ha, planet parameters such as magnitude, separation and position angle in the field). The data simulator uses the occulted star and un-occulted planet PSFs (produced by the AO simulator) to construct the scene for each 30-second IFS exposure. The details of how this is done, and its verification, are described elsewhere.

In this case we use only two of the many noise sources (photon noise and detector read noise). Since units are properly carried in the code, we apply photon noise directly in the ideal IFS cube (x-y-λ). In a more faithful

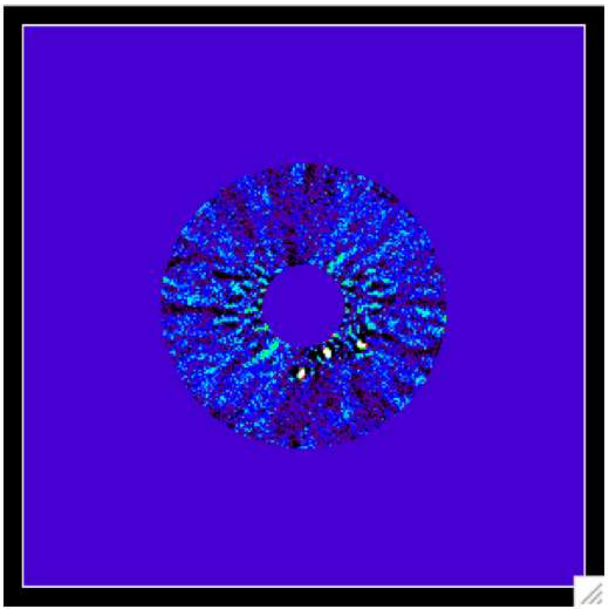
*The datasets are available at http://olbin.jpl.nasa.gov/nakfi/



(a) Rotated Sum



(b) Median subtracted



(c) LOCI

Figure 4: Preliminary results from Lawrence Livermore National Laboratory.

simulation, there would be correlations in the read-noise between pixels. Here we artificially make the read-noise in the data cube spatially white. This was done to match the expected level of read noise in a dispersed and re-assembled data cube. This works out to be 1.75 DN rms per ideal IFS pixel in a 30-second exposure.

During this process the keywords in the FITS files describing appropriate knowledge for data challenge participants (parallactic angle, star brightness, etc) are preserved. The final outputs are saved in a compressed format and collected in an archive.

All 100 observations, each with 120 30-second frames, were provided to all data challenge participants.

4.2 Preliminary Results: Poyneer and Savransky (Lawrence Livermore National Laboratory)

Our implementation of the Locally Optimized Combination of Images (LOCI) algorithm is based on the original description presented by Lafrenière[6] wherein we seek to minimize the residuals between a target image and a linear combination of a set of reference images. Using the mathematical formalism of Pueyo,[9] we wish to find the set of coefficients $\{c_k\}$ given by

$$\arg\min_{c_k} \left\| T - \sum_k^N c_k R_k \right\|_{\mathcal{O}} , \tag{31}$$

where T is the target image, $\{R_k\}_{k=1}^N$ is the set of reference images, and $\mathcal{O}$ is the optimization zone—a (possibly proper) subset of the target and reference image pixels. The solution to this problem is equivalent to the solution of

$$\mathbf{M}_{\mathcal{O}}\mathbf{c} - \mathbf{V}_{\mathcal{O}} = \mathbf{0} \tag{32}$$

where

$$\mathbf{M}_{\mathcal{O}}[i,j] \triangleq = \int_{\mathcal{O}} R_i(x,y) R_j(x,y) \mathrm{d}x\mathrm{d}y \,, \tag{33}$$

$$\mathbf{V}_{\mathcal{O}}[i] \triangleq = \int_{\mathcal{O}} R_i(x,y) T(x,y) \mathrm{d}x\mathrm{d}y \,, \tag{34}$$

and $\mathbf{c}$ is the column vector containing the values of $\{c_k\}$. The reference PSF constructed from this linear combination is subtracted from a subtraction zone $\mathcal{S}$ of the target image that is smaller than the optimization zone, to minimize the risk of subtracting the planet signal along with the background.

While schemes exist for simultaneously using both angular and spectral diversity images in the reference set,[10] the small number of wavelength bands in our data make the inclusion of spectral diversity in a given set of references marginally useful. Instead, we apply the LOCI algorithm to each wavelength independently, using only rotated images in the same wavelength for references, and then sum the final output images to increase signal-to-noise. For the optimization zones, we use radial annuli with height of 5 pixels and angular extent of 30°, with subtraction zones of one third the height and extent, centered in each subtraction zone. The subtraction zones are defined such that they cover the entire space between the inner and outer working angles of each target image, and overlap by 2% of their annular extents so that no uncorrected areas remain in the final subtracted images (areas where subtraction zones overlap are mean combined). The solution to the linear least-squares (LLS) problem is calculated using standard LLS LAPACK routines. It is important to note that the solution is exact only when matrix $\mathbf{M}_{\mathcal{O}}$ is invertible, which only occurs in the case where the set of references forms a complete basis set of the optimization zone, which is not necessarily guaranteed, depending on the specific realization of the speckle field in the reference images. For our implementation and data set, we find that $\mathbf{M}_{\mathcal{O}}$ is well conditioned for the majority of optimization zones, with singularities occurring only near the inner working angle.

This implementation of LOCI was used in the following manner. Ninety-five of the datasets were analyzed, producing a single processed 'image' for each. The detection threshold in the image was changed to different levels and standard image processing techniques were used to identify as candidate detections features that were of the right size and shape (e.g. do not select as a detection a single pixel, do not select as a detection a region which is a long streak). This was all done in an automated and fully blind mode. Then the results were compared to the known planet locations to determine if the detections were true positives or false positives.

Figure 4 shows a partial data reduction for the same dataset as illustrated by the University of Arizona in Figure 6. Figure 5 shows a more detailed analysis of our data reduction for all sets of data, with the planet separation from star plotted along the x-axis, and planet contrast plotted on the y axis, expressed on a logarithmic scale. A magnitude 20 planet in the challenge data has a contrast of 10^{-6}. Each black dot is a planet that is hidden in the data. Each colored circle represents a true detection. Bare black dots are either misses, or lie in areas that were not searched (grey regions). The size of the circle indicates the maximum threshold in the LOCI image at which the planet was detected: bright planets at the bottom have higher thresholds. The color of the

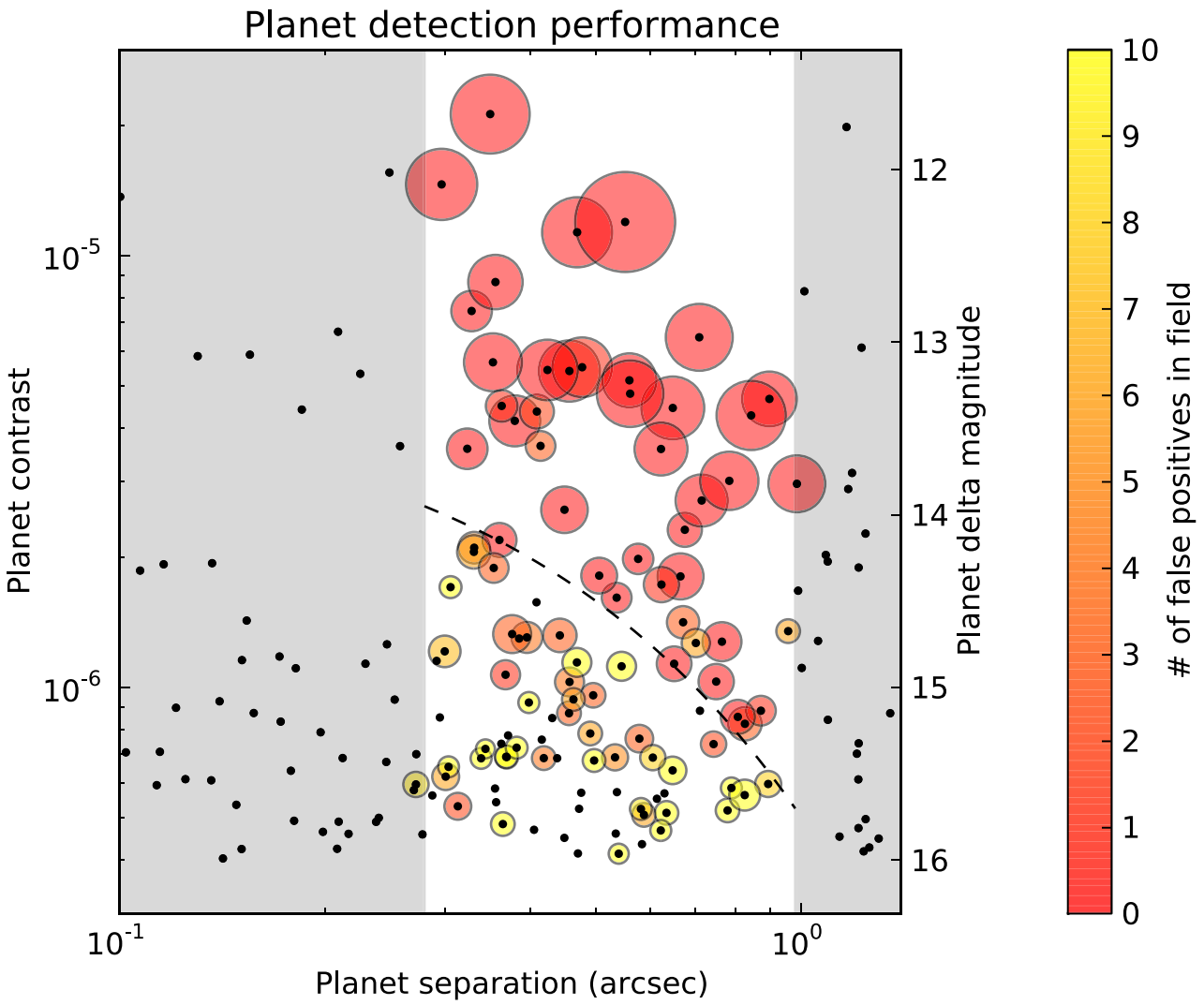


Figure 5:
The planet separation from star is plotted along the x-axis. The planet contrast is plotted on the y axis, expressed as powers of 10. A magnitude 20 planet in the challenge data has a contrast of 10^6. Each black dot is a planet that is hidden in the data. Each colored circle represents a true detection. Bare black dots are either misses, or lie in areas that were not searched (grey regions). The size of the circle indicates the maximum threshold in the LOCI image at which the planet was detected: bright planets at the top have higher thresholds. The color of the circle indicates the number of false positive detections in the entire observation for that threshold. Dark red in this case is no false positives. Orange is 5 false positives also detected at that threshold. Yellow is 10 false positives; more than this are also colored yellow.

circle indicates the number of false positive detections in the entire observation for that threshold. Dark red in this case is no false positives. Orange is 5 false positives also detected at that threshold. Yellow is 10 false positives, and aqua is 20 false positives, both of which are quite high levels.

There are several ways to investigate this further. Particularly interesting would be:

1. Searching closer in than 0.29 arcsec (our default IWA) and farther out than 0.98 arcsec (our default OWA, which is a circle inscribed in the "dark hole").

2. Improving our sensitivity limit to making true detections with fewer false alarms on fainter planets. This is roughly indicated by the dash diagonal line, which we want to move up by turning the orange and yellow circles above to dark red.

3. Presenting our results in an FROC formalism.

4. Quantifying "Prob Detection" and "Prob False Alarm" as a function of planet location and brightness and chosen threshold.

4.3 Preliminary Results: Pearson, Caucci, and Barrett (University of Arizona)

The University of Arizona's approach to the challenge is the Hotelling observer described in Section 3.3. All 100 data sets were used, each consisting of 120 temporal data cubes with 5 wavelengths for a total of 600 frames per data set. For each data set, the wind angle of the first frame in the set was estimated and all frames in the set were rotated by this angle, allowing a more accurate calculation of the mean background. As described in Section 3.2, the data in all 600 frames for each data set were placed into a column vector $\mathbf{G}$, where each element of $\mathbf{G}$ corresponds to one spatial (x, y) pixel at one time and one wavelength. Thus we have a set of column vectors $\{\mathbf{G}_j; j = 1, ..., 100\}$ to represent the 100 data sets. The 100 column vectors were averaged element-wise to give a vector $\overline{\overline{\overline{\mathbf{G}}}}$ representing the background, with the assumption that the planet signals are weak enough that they do not significantly affect the mean.

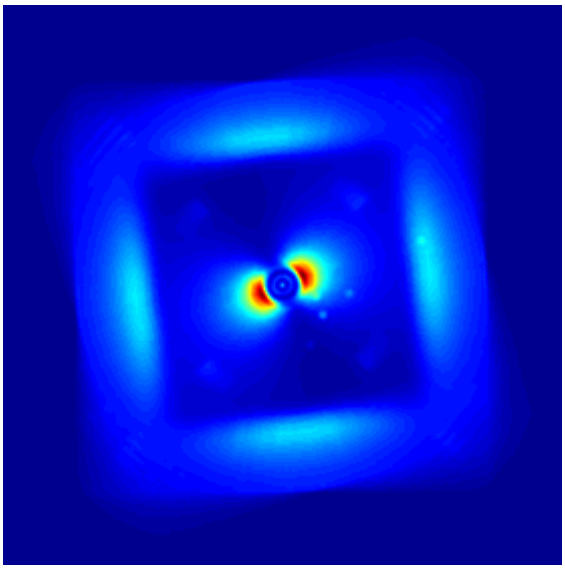
Computation of the covariance matrix was performed using the covariance decomposition described in Section 3.4. The diagonal matrix $\mathbf{K}_{\mathbf{G}}^{n}$ corresponds to the science-camera noise, which was estimated as the sum of the variances of the electronic noise and the Poisson noise from the optically induced photoelectrons. The variance of the electronic noise was provided, and the variance of the Poisson noise for each diagonal element is the corresponding element of the mean background vector $\overline{\overline{\overline{\mathbf{G}}}}$. The inverse of the overall covariance matrix can be easily calculated using the Woodbury Matrix inversion lemma and a 100×100 singular-value decomposition. For the preliminary results presented here, a similar process was used to calculate the prewhitening operator, or the inverse square-root of the covariance matrix.

Preliminary results are presented in Figure 6. Figure 6a is simply the sum of all 600 frames of one data set, each rotated by the parallactic angle specified for that frame. Figure 6b shows the same rotation and sum applied to the background-subtracted data $\mathbf{G} - \overline{\overline{\overline{\mathbf{G}}}}$, and Figure 6c shows the rotation and sum of the prewhitened data $\mathbf{K}_{\mathbf{G}}^{-1/2}\left(\mathbf{G} - \overline{\overline{\overline{\mathbf{G}}}}\right)$. An important note is that these preliminary results consist only of prewhitened images, and the final steps of the scanning Hotelling observer and thresholding the result have not yet been implemented. It is anticipated that further suppression of the star background will result when the scanning Hotelling step is performed because the prewhitened signal will have positive and negative values and act as a bandpass filter selecting the dominant spatial frequencies in the planet signal. Nevertheless, all five of the planets actually present are readily visualized in the prewhitened image, including one relatively close to the star. The angular separations in arc seconds of the five planets from the star are, in order of decreasing brightness: 0.3496, 0.2487, 0.4690, 1.0107, 0.4569.

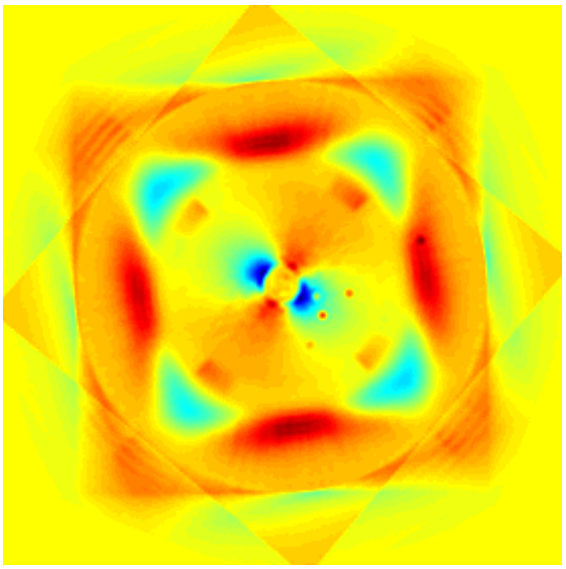
5. FUTURE DIRECTIONS AND COLLABORATIONS

5.1 Derive Practical Methods for Computation of Covariance Matrices: Barrett, Caucci, Furenlid, Pearson

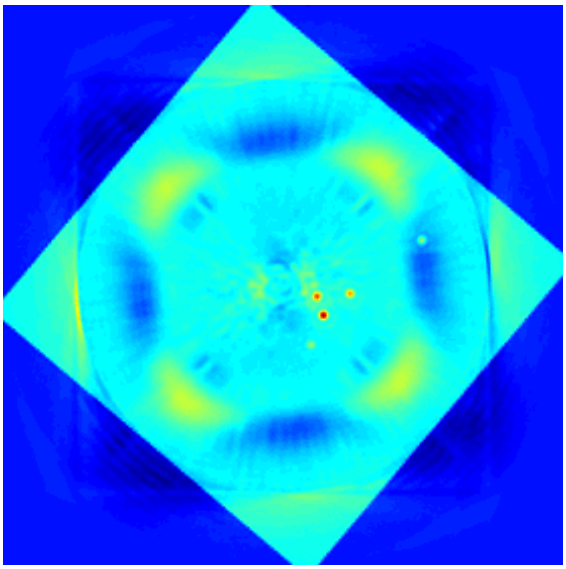
The activity we identified as Task # 1 at the conclusion of the workshop was to develop a practical and robust methodology for implementing the calculation of the data covariance matrix. This activity includes computing covariance matrices from simulation or experimental data, their inverses, and the square root of the inverses that comprises the prewhitening filter. The major challenge for the activity just identified is that data structures are very large, hence attention is required to storage, usage of efficient low-dimensional representation, recursive updates, and other mathematical tools that we will develop as we progress in our research. The preliminary



(a) Rotated Sum



(b) Mean Subtracted



(c) Prewhitened

Figure 6:
Preliminary results from the University of Arizona. Five planets are present, all of which are visible in the prewhitened image. The angular separations of the five planets are listed at the end of Section 4.3.

results reported in Section 4.3 show that the matrix inversion lemma and the corresponding prewhitening formula offer a promising avenue toward solving these problems.

Knowledge of the data covariance matrix directly and indirectly leads to many benefits, such as the optimal linear observer (Hotelling observer), the prewhitening of the data to enhance the performance of other algorithms (such as the matched filter), the simple calculation of ROC/LROC/FROC curves that characterize the performance of the algorithm and the imaging system combination. By bringing statistical rigor to data analysis and enabling the quick calculation of scalar metrics of performance (detection SNR, AUC, AULROC, AUFROC), a tool is created for algorithm, acquisition strategy, and hardware optimization.

Our objectives are to develop and validate a complete theory for the decomposition and manipulation of the data covariance matrices that can be practically realized in state-of-the-art computing hardware. We will take advantage of the latest developments in CPU/GPU parallel architectures, inherent symmetries in the covariances, sparsity, the matrix-inversion lemma, the Schur complement, tools for simultaneous diagonalization, and recursive methods.

5.2 Gain Experience Analyzing Existing Data: Gładysz

It has been demonstrated that the Hotelling observer performs better, in terms of AUC, than the commonly used simple detection algorithms.[2,3] Moreover, it has been shown that the so-called spatio-temporal Hotelling observer significantly outperforms not only the current long-exposure detection algorithms, but it is also a much better approach, in terms of AUC, than the purely spatial Hotelling observer.[5] On the other hand, the assumptions that went into these simulated tests, such as the availability of noise-free training data might be optimistic. The other approach is to try to extract the covariance matrix directly from the noisy data, available, for example, from the archive. Unfortunately there are serious obstacles to this approach, most notably strong requirements on the number of the available noisy images (or sequences in the case of the spatio-temporal Hotelling observer). This number can reach millions to tens of millions of image sequences to process one image sequence of moderate size (64 by 64 pixels by 10-50 frames).

Still, as a proof-of-concept, we will attempt that latter, brute-force approach. Four algorithms would be used: spatial Hotelling observer on long exposures,[3] spatio-temporal Hotelling observer on the sequences of short exposures,[5] statistical speckle discrimination on sequences of frames,[26] a combination of statistical speckle discrimination (as a pre-processing, object-amplifying step), and a subsequent application of the spatial Hotelling observer. The ROC curves would be constructed for each of these methods. We hope to show that the exploitation of the temporal information is beneficial, and that advanced image processing methods, like the spatio-temporal Hotelling observer can actually be used on real data and produce a benefit in terms of detectability. The crucial question is if we are given the error in the sample-based estimate of the covariance matrix do we still gain by employing the Hotelling observer? At what level of error (or sample size) does the benefit, in terms of AUC, disappear?

5.3 Compare how different algorithms make use of information in the data: Mouillet and Mugnier

The exoplanet imaging data challenge will make possible a cross-comparison of the performance of various algorithms under well-controlled conditions through the use of the simulated data sets described in Section 4.1. We propose to study these various algorithms based on their principle and their performance in the data challenge, in terms of their redundancies in the use of information as well as the a priori known properties of the data. We will identify what we think is the most useful information that helps discriminate a planet from speckles—information about the planet itself and the speckle halo properties—and identify how each approach takes advantage of it.

This analysis will also help us propose improvements to the existing algorithms in the following way:

1. Better understand the behavior of each algorithm, whose performance may depend on the observing conditions. Under different conditions the dominant limitations may vary, and the importance of any recorded information may be weighted differently.

2. Through the above understanding, propose ways of merging the best properties of each algorithm.

Each algorithm may estimate differently the prior information that it uses for detection, such as the different parts of the data covariance matrix. This may be either calibrated prior to the observation, or from the telemetry if available, or most often estimated from the data.

5.4 Derive more accurate models of the point-spread functions: Devaney

An algorithm has been developed which fits spline functions to multi-wavelength images in order to build an estimate of the PSF.[27] The fitting is inherently local and so could be equivalent to a multi-spectral LOCI. The difference is that a *model* of the PSF is built, rather than using a combination of the actual data. This model can then be scaled to any required wavelength. We believe that using the model may bring a gain in contrast ratio. The algorithm is incorporated in a matching pursuit approach whereby the PSF parameters and the planet parameters (positions and spectral energy distributions) are iteratively optimized. Alternatively, the model PSF could be used in a Hotelling approach. The algorithm has been developed for non-coronagraphic data for which the planets can be assumed to have the same PSF as the central star. The multi-wavelength model fitting may still be applicable to reducing residual speckle in coronagraphic images, and we will use the data challenge to investigate this.

5.5 Make Optimal Use of Auxiliary Data Measurable by GPI and SPHERE: Savransky

The actions on this task flow directly from the development of the Hotelling observer formalism for the specific problem of ground-based planet-finding and the experience gained from applying these tools to real and simulated data. An initial priority must be to agree upon conventions between the two (or more) groups, especially on data formatting (and the inclusion of synchronized telemetry) and on a detection criterion (i.e., using ROC/LROC as the primary metric).

Representatives from the different coronagraph projects have agreed to independently create lists of all auxiliary data that they currently plan to collect, and identify any telemetry that should be added (in the case of GPI, this includes either raw or processed outputs from the CAL system). The teams will then exchange lists and thereby begin to populate a global list of useful auxiliary data for planet-finding that will be shared with the community in general.

Along with the identification of all auxiliary data, we must decide whether each particular telemetry stream can be stored entirely, or whether we will only store a statistical description in the form of a recursively updated first and second moment estimate. The latter approach may also include using the Schur complement to calculate the conditional statistics of the science and auxiliary data.

During the first year of science operations, the teams will build up experience on applying the techniques and algorithms from Task #1 (Section 5.1) on real science data. It would be useful for the teams to communicate during this period to compare experiences and any refinements they have made to their methodologies.

At the same time, the groups must begin to prepare for the inclusion of the information provided by the Task #1 tools into their campaign scheduling; for example, incorporating AUROC statistics into dynamic campaign rescheduling after the initial K_G matrices have been constructed from initial science data. As part of this preparation, it may be useful to construct simple campaign simulations to evaluate the effects of this type of information.

Finally, Laurent Pueyo has a simulated Calibration data set which can be used to evaluate maximum likelihood wavefront reconstruction. While the data is GPI Calibration specific, it can act as a general test for the efficacy of a Maximum Likelihood Estimator in wavefront reconstruction. Furthermore, these data can be folded into an extension of the data challenge as an auxiliary data set.

Figure 7: Participants at the Exoplanet Imaging Workshop. Left to right: Harrison Barrett, David Mouillet, Szymon Gładysz, Luca Caucci, Lars Furenlid, Laurent Mugnier, Dmitry Savransky, Lisa Poyneer, Laurent Pueyo, Peter Lawson, Dimitri Mawet, Nicholas Devaney, and Richard Frazin. Not shown: Olivier Guyon, John Krist, Christian Marois, and Rémi Soummer.

6. CONCLUSION

At the time of writing the exoplanet imaging challenge is still in progress. The simulated datasets have been prepared and distributed for the first round of analysis in the exoplanet imaging challenge. The challenge itself will take place in several phases, with each phase introducing added complexity to the data. A preview of results is presented here. The final results will be published in the astronomical literature. Work with this group is actively being planned through 2012–2013.

Through support from the National Academies Keck Futures Initiative we have initiated several collaborations. This is a field that is still in its infancy. The formal approach that we advocate promises to yield a more profound understanding of correlations in noise in exoplanet data and how they may be used to our advantage. This work promises to help not only ground-based astronomy, but by extension space-based astronomy as well.

ACKNOWLEDGMENTS

This work was undertaken with support of a seed grant from the National Academies Keck Futures Initiative.

Work at the University of Arizona was supported in part by the National Institutes of Health under grant no. 2R01EB000803-21.

The authors are extremely grateful for data for Figure 1 provided by the following individuals: Markus Kasper (ESO)for ELT-EPICS; John Krist (JPL/Caltech) for JWST NIRCam and HST/ACS curves; Bruce Macintosh (UC Santa Cruz) for TMT PFI data; Christian Marois (HIA, NRC) for the K-band data for Keck NIRC2; Dimitri Mawet (ESO) for the K-band data for VLT-NACO, and the H-band curve for the Palomar Well-Corrected Subaperture; GPI data downloaded from planetimager.org, corresponding to predictions by McBride;[11] Dino Mesa

(INAF) for the data for the Integral Field Spectrograph of VLT-SPHERE;[28] and Ben Oppenheimer (AMNH) for the curve for Project 1640. Figure 1 was partly inspired by a previous illustration by Wesley Traub and John Trauger (JPL).

This work was partly conducted at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration. Reference in this paper to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement by the United States Government or the Jet Propulsion Laboratory, California Institute of Technology.

REFERENCES

[1] Barrett, H. H., Myers, K. J., Devaney, N., Dainty, J. C., and Caucci, L., "Task performance in astronomical adaptive optics," in [*Advances in Adaptive Optics II*], *Proc. SPIE* **6272** (July 2006).

[2] Barrett, H. H., Myers, K. J., Devaney, N., and Dainty, C., "Objective assessment of image quality. IV. Application to adaptive optics," *Journal of the Optical Society of America A* **23**, 3080–3105 (Dec. 2006).

[3] Caucci, L., Barrett, H. H., Devaney, N., and Rodríguez, J. J., "Application of the Hotelling and ideal observers to detection and localization of exoplanets," *Journal of the Optical Society of America A* **24**, 13 (2007).

[4] Burke, D., Devaney, N., Gladysz, S., Barrett, H. H., Whitaker, M. K., and Caucci, L., "Optimal linear estimation of binary star parameters," in [*Adaptive Optics Systems*], *Proc. SPIE* **7015** (July 2008).

[5] Caucci, L., Barrett, H. H., and Rodriguez, J. J., "Spatio-temporal Hotelling observer for signal detection from image sequences," *Optics Express* **17**, 10946 (June 2009).

[6] Lafrenière, D., Marois, C., Doyon, R., Nadeau, D., and Artigau, É., "A New Algorithm for Point-Spread Function Subtraction in High-Contrast Imaging: A Demonstration with Angular Differential Imaging," *Astrophys. J.* **660**, 770–780 (May 2007).

[7] Marois, C., Macintosh, B., and Véran, J.-P., "Exoplanet imaging with LOCI processing: photometry and astrometry with the new SOSIE pipeline," in [*Adaptive Optics Systems II*], *Proc. SPIE* **7736** (July 2010).

[8] Galicher, R., Marois, C., Macintosh, B., Barman, T., and Konopacky, Q., "M-band Imaging of the HR 8799 Planetary System Using an Innovative LOCI-based Background Subtraction Technique," *Astrophys. J.l* **739**, L41 (Oct. 2011).

[9] Pueyo, L., Crepp, J. R., Vasisht, G., Brenner, D., Oppenheimer, B. R., Zimmerman, N., Hinkley, S., Parry, I., Beichman, C., Hillenbrand, L., Roberts, L. C., Dekany, R., Shao, M., Burruss, R., Bouchez, A., Roberts, J., and Soummer, R., "Application of a Damped Locally Optimized Combination of Images Method to the Spectral Characterization of Faint Companions Using an Integral Field Spectrograph," *Astrophys. J. Supp. Ser.* **199**, 6 (Mar. 2012).

[10] Crepp, J. R., Pueyo, L., Brenner, D., Oppenheimer, B. R., Zimmerman, N., Hinkley, S., Parry, I., King, D., Vasisht, G., Beichman, C., Hillenbrand, L., Dekany, R., Shao, M., Burruss, R., Roberts, L. C., Bouchez, A., Roberts, J., and Soummer, R., "Speckle Suppression with the Project 1640 Integral Field Spectrograph," *Astrophys. J.* **729**, 132 (Mar. 2011).

[11] McBride, J., Graham, J. R., Macintosh, B., Beckwith, S. V. W., Marois, C., Poyneer, L. A., and Wiktorowicz, S. J., "Experimental Design for the Gemini Planet Imager," *Pub. Astron. Soc. Pac.* **123**, 692–708 (June 2011).

[12] Mawet, D., Baudoz, P., Beuzit, J.-L., and et al., "Review of small angle coronagraphic techniques in the wake of second-generation adaptive optics systems: choice of coronagraph, optimized wavefront control, observing strategy, and post-processing methods," in [*Space Telescopes and Instrumentation 2012: Optical, Infrared, and Millimeter Wave*], *Proc. SPIE*, in press (2012).

[13] Marois, C., Macintosh, B., Barman, T., Zuckerman, B., Song, I., Patience, J., Lafrenière, D., and Doyon, R., "Direct Imaging of Multiple Planets Orbiting the Star HR 8799," *Science* **322**, 1348– (Nov. 2008).

[14] Marois, C., Zuckerman, B., Konopacky, Q. M., Macintosh, B., and Barman, T., "Images of a fourth planet orbiting HR 8799," *Nature* **468**, 1080–1083 (Dec. 2010).

[15] Serabyn, E., Mawet, D., and Burruss, R., "An image of an exoplanet separated by two diffraction beamwidths from a star," *Nature* **464**, 1018–1020 (Apr. 2010).

[16] Lagrange, A.-M., Bonnefoy, M., Chauvin, G., Apai, D., Ehrenreich, D., Boccaletti, A., Gratadour, D., Rouan, D., Mouillet, D., Lacour, S., and Kasper, M., "A Giant Planet Imaged in the Disk of the Young Star β Pictoris," *Science* **329**, 57– (July 2010).

[17] Bonnefoy, M., Lagrange, A.-M., Boccaletti, A., Chauvin, G., Apai, D., Allard, F., Ehrenreich, D., Girard, J. H. V., Mouillet, D., Rouan, D., Gratadour, D., and Kasper, M., "High angular resolution detection of β Pictoris b at 2.18 μm," *Astron. Astrophys.* **528**, L15 (Apr. 2011).

[18] Patience, J., King, R. R., De Rosa, R. J., Vigan, A., Witte, S., Rice, E., Helling, C., and Hauschildt, P., "Spectroscopy across the brown dwarf/planetary mass boundary. I. Near-infrared JHK spectra," *Astron. Astrophys.* **540**, A85 (Apr. 2012).

[19] Clarkson, E., "The estimation receiver operating characteristic curve and ideal observers for combined detection/estimation tasks," *J. Opt. Soc. Am. A* **24**, B91–B98 (2007).

[20] Barrett, H. H. and Myers, K. J., [*Foundations of Imaging Science*], Wiley-Interscience (2004).

[21] Maire, J., Perrin, M. D., Doyon, R., Artigau, E., Dunn, J., Gavel, D. T., Graham, J. R., Lafrenière, D., Larkin, J. E., Lavigne, J.-F., Macintosh, B. A., Marois, C., Oppenheimer, B., Palmer, D. W., Poyneer, L. A., Thibault, S., and Véran, J.-P., "Data reduction pipeline for the Gemini Planet Imager," in [*Ground-based and Airborne Instrumentation for Astronomy III*], *Proc. SPIE* **7735** (July 2010).

[22] Poyneer, L. A. and Macintosh, B. A., "Optimal fourier control performance and speckle behavior in high-contrast imaging with adaptive optics," *Opt. Exp.* **14**, 7499–7514 (2006).

[23] Wallace, J. K., Angione, J., Bartos, R., Best, P., Burruss, R., Fregoso, F., Levine, B. M., Nemati, B., Shao, M., and Shelton, C., "Post-coronagraph wavefront sensor for Gemini Planet Imager," in [*Techniques and Instrumentation for Detection of Exoplanets IV*], *Proc. SPIE* **7015** (July 2008).

[24] Marois, C., Macintosh, B., Soummer, R., Poyneer, L., and Bauman, B., "An end-to-end polychromatic Fresnel propagation model of GPI," in [*Adaptive Optics Systems*], *Proc. SPIE* **7015** (July 2008).

[25] de Vries, W. H., Olivier, S. S., Asztalos, S. J., Rosenberg, L. J., and Baker, K. L., "Image Ellipticity from Atmospheric Aberrations," *Astrophys. J.* **662**, 744–749 (June 2007).

[26] Gladysz, S., Yaitskova, N., and Christou, J. C., "Statistics of intensity in adaptive-optics images and their usefulness for detection and photometry of exoplanets," *Journal of the Optical Society of America A* **27**, A260000–A75 (Nov. 2010).

[27] Devaney, N. and Thièbaut, E., "Inverse problem approach to the detection of exoplanets in multi-wavelength data," in [*Adaptive Optics: Methods, Analysis and Applications*], *Optical Society of America Technical Digest (CD)*, APDP1 (2011).

[28] Mesa, D., Gratton, R., Berton, A., Antichi, J., Verinaud, C., Boccaletti, A., Kasper, M., Claudi, R. U., Desidera, S., Giro, E., Beuzit, J.-L., Dohlen, K., Feldt, M., Mouillet, D., Chauvin, G., and Vigan, A., "Simulation of planet detection with the SPHERE integral field spectrograph," *Astron. Astrophys.* **529**, A131 (May 2011).