SAG19: exoplanet imaging signal detection theory and rigorous contrast metrics

Dimitri Mawet[1,2], Rebecca Jensen-Clem[3], Olivier Absil[4], Faustine Cantalloube[5,6], Elodie Choquet[6], Jean-Baptiste Ruffio [1], John Krist [2], Garreth Ruane[2], John Trauger[2], Carlos Gomez-Gonzalez[7], Marie Ygouf[2], Bijan Nemati[8], Ruslan Belikov[9], Thayne Currie[9,10], Steve Bryson[9], M. Kenworthy[11], Johan Mazoyer[2,12], Christian Marois[13,14], Tiffany Glassman[15], Maggie Turnbull[16], Angelle Tanner[17], and T.J. Rodigas[18]

[1] *Department of Astronomy, California Institute of Technology, 1200 E. California Blvd, MC 249-17, Pasadena, CA 91125, USA*
[2] *Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., Pasadena CA 91109, USA*
[*] *dmawet@astro.caltech.edu*
[3] *Astronomy Department, University of California Santa Cruz, 1156 High St., Santa Cruz, CA 95064, USA*
[4] *Space sciences, Technologies and Astrophysics Research (STAR) Institute, Université de Liège, 19c Allée du Six Août, 4000 Liège, Belgium*
[5] *Max-Planck-Institut für Astronomie (MPIA), Königstuhl 17, D-69117 Heidelberg, Germany*
[6] *Aix Marseille Univ, CNRS, CNES, LAM, Marseille, France*
[7] *Barcelona Supercomputing Center, Spain*
[8] *The Univ. of Alabama in Huntsville (United States)*
[9] *NASA-Ames Research Center, Moffett Blvd., Moffett Field, CA, USA*
[10] *Subaru Telescope, National Astronomical Observatory of Japan, 650 North A'ohōkū Place, Hilo, HI 96720, USA*
[11] *Leiden Observatory, Leiden University, P.O. Box 9513, 2300 RA Leiden, The Netherlands*
[12] *LESIA, Observatoire de Paris, Université PSL, CNRS, Sorbonne Université, Université de Paris, 5 place Jules Janssen, 92195 Meudon, France*
[13] *National Research Council of Canada Herzberg, 5071 West Saanich Road, Victoria, BC, V9E 2E7, Canada*
[14] *29 University of Victoria, 3800 Finnerty Road, Victoria, BC, V8P 5C2, Canada*
[15] *Northrop Grumman Corporation, One Space Park Drive, E1-4068, Redondo Beach, CA 90278.*
[16] *SETI Institute, Carl Sagan Center for the Study of Life in the Universe, Off-Site: 2613 Waunona Way, Madison, WI 53713, USA*
[17] *Mississippi State University, Department of Physics Astronomy, Hilbun Hall, Starkville, MS, 39762, USA*
[18] *Department of Terrestrial Magnetism, Carnegie Institution for Science, 5421 Broad Branch Road, Washington, DC 20015, USA*

ABSTRACT

As planning for the next generation of high contrast imaging instruments (e.g. Roman-CGI, HabEx, and LUVOIR, TMT-PSI, ELT-EPICS, Magellan-GMagAO-X, SPHERE+, GPI-2.0, SCExAO, and MagAO-X) matures, and first-generation extreme adaptive optics facilities (e.g. VLT-SPHERE, Gemini-GPI) finish their large main surveys, it is imperative that the performance of different designs, post-processing routines, observing strategies, observing calibration procedures, and survey results be compared in a consistent, statistically robust framework. SAG19, exoplanet imaging signal detection theory and rigorous contrast metrics, is meant to address the performance of these direct imaging instruments, strategies, and methods. In this report, we create a reference document for high contrast exoplanet detection terminology (Sections 2 and 3, drawing from work published as part of SAG19), summarize the results of three additional publications that resulted from SAG19 (Section 4, 5, and 6), and summarize the broad findings of SAG19 (Section 7).

Contents

1 The SAG 19 Charter

The scope of SAG19 is:

- To go back to the basics of Bayesian Signal Detection Theory (SDT). Bayesian SDT implies H0:signal absent / H1:signal present hypothesis testing, and invokes well-known concepts such as: the confusion/contingency matrix, false positive (type I error), false negative (type II error), true positive, and true negative fractions, and useful combinations of these quantities such as sensitivity (or completeness) and specificity.
- To rebuild a solid set of usual definitions used for or in lieu of "contrast" in different contexts, such as astrophysical contrast or ground truth, instrumental contrast used for coronagraph/instrument designs, and the measured on-sky data-driven contrast.

  Bayesian, hypothesis testing SDT will automatically force our community to be inclusive of all possible aspects of exoplanet detection, and signal-to-noise ratio (SNR) metrics, including signal-related parameters: planet-star contrast, SED, polarization, variability; instrument parameters: throughput, bandwidth, Strehl ratio/encircled energy, background (sky/thermal, or astrophysical), detector characteristics; noise characteristics as affected by the starlight suppression technique (in a very broad sense): mean intensity, RMS pixel intensities, RMS resolution element (resel, of characteristic size wavelength/telescope diameter) intensities, the probability density function (PDF) computed over pixels, the same PDF computed over resels, their nature and higher order moments, the sample zone and size, outlier management, etc.
- To identify what we can learn and apply from communities outside our field (e.g. medical imaging). A good example is the widespread use of receiver operating characteristic curve (ROC) and area under the curve (AUC). ROC plots the true positive fraction against the false positive fraction, and is useful to capture the true performance of a given high contrast imaging instrument, or post-processing/detection algorithm. Other formalisms, alternative to the ROC curve, such as the precision-recall curve will also be considered.
- To define precise contrast computation and ROC curve computation recipes, a new "industry standard". The goal is to be able to compare results from surveys, instrument and/or telescope designs on a level-playing field.
- To identify how the new metrics and recipes can be used to define confidence levels for detection (H1) and subsequently error bars for photometric, spectroscopic, astrometric characterization. Ancillary goals: better understanding what limits exoplanet characterization, not just detection. For instance, understanding the limiting precision of extracted planet spectra from algorithms that anneal the planet signal and gaining proper error assessments from spectral extraction. This is particularly important in cases where the prior on the signal/wavelength to be detected is unknown and iterative forward-modeling must be applied.
- To perform a community data challenge before and after applying our proposed set of standardized SDT rules and recipes, and apply lessons learned.

2 Definitions, contrast and yield metrics (partly adapted from Krist, 2016; Ruane et al., 2018)

We define the "contrast" as the ratio of the planet intensity to the star intensity when both objects are observed in exactly the same manner. When contrast is used to describe coronagraphic or survey performance, however, its meaning becomes more complicated: the field of interest (e.g., a coronagraphic dark hole) as measured with the detector will be filled with diffracted light from the star (mostly, but not completely, eliminated by the starlight suppression system), scattered light speckles and stray light from optical defects (perhaps reduced by wavefront control), and extended astrophysical sources (e.g., zodi or exozodi). Even with a perfect detector and post-processing capabilities, the shot noise from these features limits the system's sensitivity to faint exoplanets. We assume here that the scattered and stray light, as well as the astrophysical backgrounds, are generally independent of the suppression technique, and hence we ignore them in favor of focusing on comparative performances.

2.1 Point Spread Function (PSF)

$\mathrm{PSF}_0(\mathbf{x},\lambda)$, is the PSF of the imaging system with a perfect wavefront, flat deformable mirrors, and with all of the coronagraph masks removed (including apodizers, focal plane masks, and Lyot stops). We take $\mathbf{x}$ to be the position vector in the image plane and λ as the wavelength. In this case, the PSF is theoretically shift invariant. The PSF is normalized such that

$$\int \mathrm{PSF}_0(\mathbf{x})d\mathbf{x} = 1 \tag{1}$$

2.2 Coronagraphic, shift-variant PSF

The coronagraphic PSF, $\mathrm{PSF}_{\mathrm{coro}}(\mathbf{x},\mathbf{x_0},\lambda)$, is the PSF with the coronagraph masks in place and DMs set to minimize starlight in the desired dark hole region of the image plane. $\mathrm{PSF}_{\mathrm{coro}}$ is different from PSF_0 in that it varies as a function of source position $\mathbf{x_0}$.

2.2.1 Off-axis coronagraphic PSF and planet throughput η_p

Most starlight suppression techniques reduce the intensity of the stellar diffraction pattern in a manner that also modifies the shape and intensity of the PSF of a field sources such as a planet. Pupil apodization, Lyot stops, and intentionally-induced wavefront modulations for obscuration compensation may block light and/or diffract it out of the core and into the wings. In most cases, planets are sufficiently faint that only the light within the core of the planet's PSF will be measurable, so even a lossless suppression system may have poor net performance if the majority of the planet's flux is dispersed over large angles due to PSF reshaping. Thus, the metric that defines an effective suppression performance must include the field PSF characteristics to make comparisons among suppression methods reasonable.

Low/no loss systems such as the starshade or (for unobscured systems) PIAA or vector vortex coronagraphs have field PSFs that are very close in morphology and photometry to the non-coronagraphic ones, except near and within the inner working angle defined by the occulter or focal plane mask. Shaped pupil and hybrid Lyot coronagraphs (HLC), however, suffer from low transmission due to their pupil/Lyot masks, redistribution of light from the planet PSF's core into the wings as a result of the masks, and, in the case of the HLC, the wavefront modulations.

The signal detected in photo-electrons from the planet and star at position $\mathbf{x_0}$ is

$$S_p(\mathbf{x_0}) = \int_{\Delta\lambda} \eta_p(\mathbf{x_0},\lambda)\Phi_p(\lambda)A\Delta t\, q(\lambda)T(\lambda)d\lambda \tag{2}$$

respectively, where η_p is the fraction of collected planet light that is detected, Δt is the integration time, $\Phi_p(\lambda)$ and $\Phi_s(\lambda)$ are the flux from the planet and star (photons per unit area per unit time per unit wavelength at the primary mirror), $\Delta\lambda$ is the spectral bandwidth, A is the collecting area of the telescope, $q(\lambda)$ is the detective quantum efficiency, and $T(\lambda)$ is the transmission of the instrument describing losses that affect the star and planet equally. Specifically, η_p may be computed by

$$\eta_p(\mathbf{x_0},\lambda) = \int_{\mathrm{AP}(\mathbf{x_0})} \mathrm{PSF}_{\mathrm{coro}}(\mathbf{x},\mathbf{x_0},\lambda)d\mathbf{x} \tag{3}$$

where the signal is integrated over a finite circular aperture for photometric estimation in the image plane, $\mathrm{AP}(\mathbf{x_0})$, centered on the source position $\mathbf{x_0}$.

2.2.2 Throughput definitions

The relative throughput is normalized to the throughput of the system without the coronagraph masks: $\eta_p(\mathbf{x_0},\lambda)/\eta_{\mathrm{tel}}(\lambda)$, where

$$\eta_{\mathrm{tel}}(\lambda) = \int_{\mathrm{AP}(0)} \mathrm{PSF}_0(\mathbf{x},\lambda)d\mathbf{x}. \tag{4}$$

The total energy throughput is the integral of $\eta_p(\mathbf{x_0}, \lambda)$ with respect to $\mathbf{x_0}$ over the image plane and with respect to λ over the spectral bandwidth $\Delta\lambda$. The spectral bandwidth is most often reported as a fraction ($\Delta\lambda/\lambda$) or percentage ($\Delta\lambda/\lambda \times 100$) for broadband imaging, whereas the spectral resolution of spectrographs is typically given by $R = \lambda/\Delta\lambda$.

2.2.3 On-axis coronagraphic PSF, residual stellar intensity η_s

The signal detected in photo-electrons from the star at position $\mathbf{x_0}$ is

$$S_s(\mathbf{x_0}) = \int_{\Delta\lambda} \eta_s(\mathbf{x_0}, \lambda)\Phi_s(\lambda)A\Delta t\, q(\lambda)T(\lambda)d\lambda, \tag{5}$$

respectively, where η_s is the fraction of collected astrophysical light that is detected. Specifically, η_s may be computed by

$$\eta_s(\mathbf{x_0}, \lambda) = \int_{\mathrm{AP}(\mathbf{x_0})} \mathrm{PSF}_{\mathrm{coro}}(\mathbf{x}, 0, \lambda)d\mathbf{x}, \tag{6}$$

where the signal is integrated over a finite circular aperture for photometric estimation in the image plane, $\mathrm{AP}(\mathbf{x_0})$, centered on $\mathbf{x_0}$.

2.3 Starlight suppression

The term suppression has been used in the past, especially for interferometers including the visible nulling coronagraph concept, to describe how much the diffraction pattern of the star has been reduced by the instrument. Its use, however, is fraught with potential interpretation errors. It might reasonably mean that the wings of the stellar PSF are suppressed to the level of a planet with a given planet/star contrast, or that the entire stellar PSF has reduced, or that the surface brightness of the wings has been reduced. Without any other qualifiers, the term “suppression" does not indicate how the brightness of the suppressed stellar PSF compares to the planet intensity. As a quantifier, suppression is therefore only useful within the context of a specific system configuration and cannot be used across different instruments.

2.4 Raw (coronagraphic, background) contrast

The most widely used coronagraph performance metric is the raw contrast defined by

$$C(\mathbf{x_0}, \lambda) = \frac{\eta_p(\mathbf{x_0}, \lambda)}{\eta_s(\mathbf{x_0}, \lambda)}. \tag{7}$$

The raw contrast may also be integrated in over spatial and spectral dimensions. Subsequently in this document, we use the term C to refer to the raw contrast in Equation 7.

2.5 Normalized intensity

Another widely used coronagraph performance metric is normalized intensity defined by

$$C_I(\mathbf{x_0}, \lambda) = \frac{\eta_s(\mathbf{x_0}, \lambda)}{\eta_s(\mathbf{0}, \lambda)}. \tag{8}$$

It may also be integrated in over spatial and spectral dimensions.

2.6 Signal to noise ratio

The signal to noise ratio is one of the most relevant metric pertaining to signal detection.

2.6.1 Photon noise limited

In the presence of stellar photon noise, the signal-to-noise ratio for detection is $\mathrm{SNR} = S_p/\sqrt{S_s}$. For the sake of simplicity in this discussion, we assume η_s, η_p, and the planet-to-star flux ratio, $\varepsilon = \Phi_p/\Phi_s$, are approximately constant as a function of wavelength. The expression for the planet and star signals simplify to

$$S_p = \eta_p \varepsilon N_\star, \tag{9}$$
$$S_s = \eta_s N_\star, \tag{10}$$

where

$$N_\star = \int_{\Delta\lambda} \Phi_s(\lambda)A\Delta t\, q(\lambda)T(\lambda)d\lambda \tag{11}$$

is the unocculted stellar signal in units of photo-electrons. From this point, we assume the $\mathbf{x_0}$ argument is implicit. The stellar photon noise limited SNR becomes

$$\mathrm{SNR} = \frac{\eta_p}{\sqrt{\eta_s}}\varepsilon\sqrt{N_\star} = \varepsilon\sqrt{\frac{\eta_p N_\star}{C}}. \tag{12}$$

The detection limits of an observation are most often communicated in terms of the minimum detectable planet-to-star flux ratio $\varepsilon_{\mathrm{lim}}$. To achieve an SNR of unity,

$$\varepsilon_{\mathrm{lim}} = \sqrt{\frac{C}{\eta_p N_\star}}. \tag{13}$$

An exoplanet with $\varepsilon = 10^{-10}$ may be detected if the C is small enough to ensure that the flux ratio limit set by photon noise $\varepsilon_{\mathrm{lim}}$ is less than 10^{-10}. A system with $\eta_p = 10\%$ and $N_\star \approx 10^{12}$, which is typical for a one hour integration on a solar type star at 10 pc, requires $C \approx 10^{-10}$.

2.6.2 Speckle noise limited

The dominant noise source in high-contrast images is often spatial speckle noise. Speckles appear as blobs in the image whose spectral irradiance at a single wavelength is described by

$$I(\mathbf{x},\lambda) = |E_0(\mathbf{x},\lambda)|^2 + |E_{\mathrm{speck}}(\mathbf{x},\lambda)|^2 + 2\mathrm{Re}\{E_0(\mathbf{x},\lambda)^* E_{\mathrm{speck}}(\mathbf{x},\lambda)\}, \tag{14}$$

where $|E_0(\mathbf{x},\lambda)|^2 \approx \mathrm{PSF}_{\mathrm{coro}}(\mathbf{x},0,\lambda)$ with an unaberrated wavefront and $E_{\mathrm{speck}}(\mathbf{x},\lambda)$ represents the field in the focal plane due to a small additive wavefront error normalized by the coronagraph throughput (Racine et al., 1999; Bloemhof et al., 2001; Sivaramakrishnan et al., 2002; Perrin et al., 2003; Aime & Soummer, 2004a; Soummer et al., 2007). The additive wavefront errors that cause speckles interfere with the underlying residual starlight.

Aberrations that change slowly may be calibrated using differential imaging techniques, such as Angular Differential Imaging (ADI) (Davies, 1980; Marois et al., 2006). Those that change on small timescales (e.g. atmospheric residuals) average out into a speckle halo, which may also be subtracted from the image. In these cases, we are only left with the photon noise from the subtracted speckles. However, aberrations that vary on intermediate timescales (i.e. quasi-static) are the most problematic since they are responsible for residual spatial speckle noise after differential imaging. The remaining spatial speckle noise introduces a systematic noise floor for detecting companions through direct imaging.

The SNR when limited by spatial speckle noise is $\mathrm{SNR} = S_p/\sigma_{\mathrm{speck}}$, where (Aime & Soummer, 2004a; Soummer et al., 2007)

$$\sigma_{\mathrm{speck}}^2 \approx S_{\mathrm{speck}}^2 + 2S_{\mathrm{speck}}S_0, \tag{15}$$

where S_{speck} and S_0 are the signals that would be detected from the aberrated and unaberrated wavefronts alone. Expanding the signal terms:

$$S_{\mathrm{speck}} \approx C_{\mathrm{speck}}\eta_p N_\star, \tag{16}$$

$$S_0 \approx C_0 \eta_p N_\star, \tag{17}$$

where we have made the approximation that η_p is constant over the spectral bandwidth. The raw contrast of a speckle at a position $\mathbf{x}'$ is directly tied to the amplitude of an aberration at spatial frequency $\boldsymbol{\xi} = \mathbf{x}'/\lambda F\#$ cycles per pupil diameter, where $F\#$ is the focal ratio. For a wavefront error ω at the spatial frequency $\boldsymbol{\xi}$, the raw contrast at $\mathbf{x}$ is $C_{\mathrm{speck}} = 2(\pi\omega)^2$, where ω is in units of waves rms (Malbet et al., 1995; Ruane et al., 2018). C_0 is the minimum possible raw contrast (unaberrated case). The SNR may be written

$$\mathrm{SNR} = \frac{\varepsilon\eta_p N_\star}{\sqrt{C_{\mathrm{speck}}^2\eta_p^2 N_\star^2 + 2C_{\mathrm{speck}}C_0\eta_p^2 N_\star^2}} = \frac{\varepsilon}{\sqrt{C_{\mathrm{speck}}^2 + 2C_{\mathrm{speck}}C_0}}. \tag{18}$$

The planet-to-star flux ratio at which SNR is unity is given by

$$\varepsilon_{\mathrm{lim}} = \sqrt{C_{\mathrm{speck}}^2 + 2C_{\mathrm{speck}}C_0}. \tag{19}$$

More generally, we write the flux ratio limits as three terms added in quadrature:

$$\varepsilon_{\text{lim}} = \sqrt{\varepsilon_{\text{speck},1}^2 + \varepsilon_{\text{speck},2}^2 + \varepsilon_{\text{phot}}^2}, \tag{20}$$

where $\varepsilon_{\text{speck},1} = C_{\text{speck}}$, $\varepsilon_{\text{speck},2} = \sqrt{2C_{\text{speck}}C_0}$, and $\varepsilon_{\text{phot}} = \sqrt{C/\eta_p N_\star}$ represent the individual flux ratio limit terms.

Coronagraph instruments are often designed to minimize the limiting flux ratio. The first term, $\varepsilon_{\text{speck},1}$, is set by wavefront stability and the effectiveness of the differential imaging strategy. It is minimized by creating a dark hole in the stellar PSF with the adaptive optics system and maintaining it throughout the observation. The second term, $\varepsilon_{\text{speck},2}$, describes the interaction between wavefront errors and starlight diffracted through the coronagraph in the unaberrated case. Lastly, $\varepsilon_{\text{phot}}$ is the stellar photon noise term.

2.6.3 Integration time

Another way to define coronagraph design metrics is to predict the minimum integration time for a detection. For example, in the stellar photon noise limited regime, the integration time for SNR=1 is

$$\Delta t = \frac{C}{\eta_p} \frac{1}{\varepsilon^2 \dot{N}_\star} = \frac{\eta_s}{\eta_p^2} \frac{1}{\varepsilon^2 \dot{N}_\star}, \tag{21}$$

where $\dot{N}_\star = N_\star/\Delta t$ is the count rate. Thus, an optimal coronagraph minimizes η_s/η_p^2. Generalizing the SNR expressions above to include other noise sources,

$$\text{SNR} = \frac{S_p}{\sqrt{S_s + \sigma_{\text{det}}^2 + \sigma_{\text{bkgd}}^2}}, \tag{22}$$

where σ_{det} is the detector noise and σ_{bkgd} is the background noise (e.g. thermal background). Since the additional noise terms do not depend on the coronagraph throughput, we can characterize the relative importance of each additional noise term as constants a_n, where $\sigma_n^2 = a_n N_\star$, and write the integration time as follows:

$$\Delta t = \frac{1}{\eta_p^2 \varepsilon^2 \dot{N}_\star} \left[\eta_s + \sum_n a_n \right]. \tag{23}$$

Thus, the coronagraph is only designed to suppress starlight enough such that the stellar photon noise no longer dominates the error budget.

Finally, when limited by mid-spatial frequency aberrations that generate speckles, an optimal coronagraph minimizes their photon noise contribution by maximizing the coronagraph throughput. We write $\sigma_{\text{speck}} = bN_\star$, where b is a constant representing the strength of the spatial speckle noise. The expression for the integration time becomes

$$\Delta t = \frac{1}{\varepsilon^2 \dot{N}_\star} \left[\frac{\eta_s + \sum_n a_n}{\eta_p^2 - b^2/\varepsilon^2} \right]. \tag{24}$$

If $b > \eta_p \varepsilon$, spatial speckle noise prevents detection (i.e. integration time is negative). Therefore, an optimal coronagraph minimizes $|\Delta t|$, but maintains $\eta_p > b/\varepsilon$ to ensure planets may be detected. Furthermore, the removal of speckles through differential imaging is most efficient when the coronagraph throughput is maximized because the SNR of the speckle measurement is also maximized in each frame.

3 Signal detection theory (partly adapted from Mawet et al., 2014)

The detection problem consists of making an informed decision between two hypotheses, H_0, signal absent, and H_1, signal present. While the application of hypothesis testing for the binary classification problem of exoplanet imaging using matched filtering and Bayesian techniques has been discussed in detail elsewhere (eg Kasdin & Braems, 2006), here we focus on the background and photon noise only without any considerations for speckle noise or sample sizes.

Because most exoplanet hunters want to minimize the risk of announcing false detections or waste precious telescope time following up artifacts, high contrast imaging has mostly been concerned (so far) with choosing a detection threshold τ, defining the contrast which minimizes the false positive fraction (FPF), defined as

$$FPF = \frac{FP}{TN + FP} = \int_{\tau}^{+\infty} pr(x|H_0)dx \tag{25}$$

where x is the intensity of the residual speckles, and $pr(x|H_0)$ is the probability density function of x under the null hypothesis H_0. FP is the number of false positives and TN is the number of true negatives. Under H_0, the confidence level $CL = 1 - FPF$ is called the "specificity." However, exoplanet hunters who want to optimize their survey, and derive meaningful conclusions about null results, also wish to maximize the so-called "True Positive Fraction" (TPF), or in statistical terms the "sensitivity" (some authors refer to "completeness", see, e.g. Wahhaj et al. (2013)), which is defined as

$$TPF = \frac{TP}{TP+FN} = \int_{\tau}^{+\infty} pr(x|H_1)dx \tag{26}$$

with $pr(x|H_1)$ is the probability density function of x under the hypothesis H_1, TP is the number of true positives, and FN is the number of false negatives. For instance, a 95% sensitivity (or completeness) for a given signal μ_c, and detection threshold τ means that 95% of the objects at the intensity level μ_c will statistically be recovered from the data. Ultimately, the goal of high contrast imaging, as a signal detection application, is to maximize the TPF while minimizing the FPF. Optimizing detection thus consists of maximizing the so-called AUC, i.e. the area under the "Receiver Operating Characteristics" (ROC) curve. The ROC curve plots the TPF as a function of the FPF. The optimal linear observer, or discriminant, maximizing the AUC is called the Hotelling observer, and can be regarded as a generalization of the familiar pre-whitening matched filter (see, for instance Caucci et al. (2007), Lawson et al. (2012), Ruffio et al. (2017)).

3.1 The effect of non-Gaussian statistics

Statistical tools to assess the significance of a point source detection at large angles are most often based on the assumption that the underlying noise is Gaussian. However, it was clear early in the history of direct imaging that speckle noise in raw high contrast images is never Gaussian (Perrin et al., 2003; Aime & Soummer, 2004a; Bloemhof, 2004; Fitzgerald & Graham, 2006; Soummer et al., 2007; Hinkley et al., 2007; Marois et al., 2008). The main conclusion of this series of pioneering papers is that the probability density function (PDF) of speckles in raw images does not follow a well-behaved Gaussian distribution, but rather it is better described by a modified Rician (MR) distribution, which is a function of the local time-averaged static point-spread function (PSF) intensity I_c and random speckle noise intensities I_s:

$$p_{MR}(I, I_c, I_s) = \frac{1}{I_s} \exp\left(-\frac{I+I_c}{I_s}\right) I_o \left(\frac{2\sqrt{II_c}}{I_s}\right) \tag{27}$$

where I_0 is the modified Bessel function of the first kind, and where the mean and variance of I are $\mu_I = I_c + I_s$, and $\sigma_I^2 = I_s^2 + 2 * I_c I_s$, respectively (Soummer et al., 2007).

The bulk of past studies related to speckle statistics focused on the temporal aspects of speckle noise in the presence of atmospheric turbulence, corrected or not by adaptive optics systems. In the virtual case of an instrument with perfect optics on a ground-based telescope, the practical impact of the temporal MR PDF of speckles would only have a minor effect on detection limits by virtue of the central limit theorem (CLT). Indeed, if a large number of independent and identically distributed (i.i.d.) speckles are co-added at a specific location in the image, then the sample means will be normally distributed (Marois et al., 2008). In other words, the speckle sampling distribution will be Gaussian.

Unfortunately, optics are never perfect nor hold their imperfect shape constant over time, and so they induce slowly varying wavefront errors, creating quasi-static speckles. Marois et al. (2008) used a heuristic argument to show that quasi-static speckle noise inside annuli centered on the PSF core would follow the MR PDF Eq. 27, because it is basically produced with the same value of I_c (the unaberrated PSF). The typical lifetime of quasi-static speckles has been found to range from several minutes to hours (Hinkley et al., 2007). I_s has thus a complex spatio-temporal dependence $I_s(\theta, t)$. Slowly varying wavefront errors disturb the spatio-temporal autocorrelation of the PSF accordingly, and thus its temporal and spatial statistical properties: the samples of resolution elements used to compute noise properties are no longer independent and identically distributed.

Marois et al. (2008) showed that the net effect of the MR nature of quasi-static speckle noise is that the confidence level (CL) at a fixed detection threshold τ significantly deteriorates. Subsequently, in order to preserve CLs, or equivalently control the likelihood of type I errors (false alarm probability, or false positive fraction) the detection thresholds need to be increased significantly, e.g. up to a factor 4 (Marois et al., 2008).

Fortunately, observing strategies such as angular differential imaging (ADI, Marois et al., 2006), and data reduction techniques such as principal component analysis (PCA, Soummer et al., 2012; Meshkat et al., 2013) routinely demonstrate their "whitening" capability, i.e. the efficient removal of the correlated component of the noise. Whitening yields independent Gaussian noise samples through complementary mechanisms. First, once the correlated component has been removed (even partially), other noise sources start to dominate. The latter (background, photon noise, readout or dark current) are independent noise processes and thus Gaussian by virtue of the central limit theorem (CLT). Second, ADI and other differential imaging techniques enhance the efficiency of the first mechanism (if the removal is only partial) by introducing geometrical diversity (field rotation in the case of ADI), further consolidating the independence of noise samples.

3.2 The effect of small sample statistics

In the close separation regime (down to the diffraction limit at $1\lambda/D$), speckle noise dominates at all contrast levels, even after being controlled or nulled by active speckle correction (Malbet et al., 1995; Bordé & Traub, 2006; Give'on et al., 2007) and/or a dedicated low-order wavefront sensor (see, e.g., Guyon et al. 2009). In the case of very high contrast images ($10^9:1$ and higher), other sources of noise such as photon noise, readout or dark current might become dominant, especially at larger separations (see, e.g., Brown (2005), and Kasdin & Braems (2006) for thorough treatments of the uniform background case). At small separations, these factors are presumably less important compared to the speckle variability induced by residual low-order aberrations. The detailed error budget largely depends on the hardware available, and must therefore be studied on a case-by-case basis, which is beyond the scope of this report.

Quasi-static speckles at a given radius r are all drawn from the same parent population of mean μ and standard deviation σ (Marois et al., 2008). Assuming that the detection is performed on individual resolution elements λ/D, we must treat speckle noise on this characteristic spatial scale as well. We also note that the size of residual speckles is always $\sim\lambda/D$, even after coherent (interference) or incoherent (intensity image) linear combinations. At the radius r (in resolution element units λ/D), there are $2\pi r$ resolution elements and thus possible non-overlapping speckles, i.e. about 6 at $1\lambda/D$, 12 at $2\lambda/D$, 18 at $3\lambda/D$, and 100 at $16\lambda/D$. The empirical estimators of the mean and standard deviation, $\bar{x}$ and s, are thus calculated from a sample with a limited number of elements, or degree of freedom (DOF), shrinking with r. For samples containing less than ~ 100 elements (this number is somewhat arbitrary and varies according to practices and applications), we are in the regime of small sample statistics, which significantly affects the calculation of Eq. 25 and Eq. 26.

Mawet et al. (2014) quantified the effect of small sample statistics on the computation of the $pr(x|H_0)$ (and $pr(x|H_1)$), and its impact on the choice of the detection threshold τ (and thus contrast) using the two-sample t-test.

4 The Performance Map: A New Standard for Assessing the Performance of High Contrast Imaging Systems (a summary of Jensen-Clem et al., 2018)

The contrast curve is a means of representing the true and false positive fractions associated with a range of signals and positions in a final image. Schematically, we can define the "observer's" contrast as:

$$\text{contrast} = \left(\frac{\text{factor} \times \text{noise}}{\text{stellar aperture photometry}}\right)\left(\frac{1}{\text{throughput}}\right) \tag{28}$$

where the numerator is the detection threshold, expressed as a multiple of the noise distribution's width. Often, the width of the noise distribution (here, the "noise") is chosen to be the standard deviation of the resolution element intensities at a given separation from the star, while the multiplicative "factor" is chosen to be three or five to produce a 3σ or 5σ contrast curve. The detection threshold is then converted to a fraction of the parent star's brightness via the "stellar aperture photometry" term. Finally, the "throughput" term corrects this brightness ratio for any attenuation of the off-axis signal relative to the star's (e.g. due to field-dependent flux losses imposed by the coronagraphic system and post-processing algorithms). The final contrast is therefore the planet-to-star flux ratio of a planet whose brightness is equal to the detection threshold. Hence, the contrast curve can be interpreted as the signal for which we achieve 50% completeness given our choice of detection threshold in the numerator. The numerator also fixes the false positive fraction – for example, choosing factor$= 3$ for a white noise distribution gives FPF = 0.001. Finally, it is important to note that the contrast curve's statistics refer to planet detectability, and not to the photometric accuracy associated with any given planetary signal.

Where the Contrast Curve Falls Short: Both practical and fundamental shortcomings, however, undermine the utility of the contrast curve as a general purpose performance metric. First, the contrast is inflexible: by fixing the true positive fraction to 0.5 and the false positive fraction to a value set by the numerator, we cannot explore the (TPF, FPF, detection threshold) trade space. Fixing the TPF, FPF, and detection threshold for all separations may not be desirable for all applications – because the number of resolution elements, the PDF of the noise, and the predicted population of planets all vary as a function of separation, a particular imaging program's science goals may be better served by a detection threshold that also varies with separation. Also problematic are the calculation of the terms in Equation 28.The choice of the"noise" term as the standard deviation of resolution elements in a region of the image (whose shape and size widely varies in the literature) is valid if two conditions are met: 1) if the region includes enough statistically independent realizations of the noise to allow for an accurate measure of the distribution's standard deviation, and 2) if the underlying noise distribution is Gaussian. While there is no hard and fast rule for deciding whether the first condition is met, statisticians generally consider 30 independent samples to be the boundary between large and small sample statistics. For the case of $1\,\lambda/$D-wide annular regions, 30 samples corresponds to a separation of $\sim 5\lambda/$D. Below this threshold, the sample standard deviation is an increasingly uncertain estimate of the width of the underlying noise distribution. The mitigating strategy proposed by Mawet et al. (2014), however, also requires that condition #2 (Gaussian noise) is met. Aime & Soummer (2004b) and others show that uncorrected low-order wavefront aberrations cause the noise at small separations to follow a positively skewed modified rician distribution rather than a normal distribution. While numerous observing and post-processing strategies have been employed to whiten this skewed distribution, their success at small separations is limited by the temporal and spectral variability of the noise. The result is that the noise distribution at small angles retains an unknown skewness at small separations that increases the false positive fraction compared to a Gaussian distribution. Hence, neither condition for the use of the standard deviation as a proxy for the FPF is met at small separations.

The Performance Map: We propose two modifications to the contrast curve: 1) a detection threshold (and hence FPF) that varies with separation, and 2) the inclusion of all possible TPFs as a heatmap. When the detection threshold is held constant with separation, the radial distribution of false positives is not uniform because the number of resolution elements varies with separation. If the expected number of false positives N_{FP} is given by $N_{\text{FP}}(r) = \text{FPF} \times 2\pi r$ for separation r, then a constant detection threshold (and hence a constant FPF) allows more total false positives at wide separations than at small separations. If we instead keep the radial distribution of false positives constant, we allow the detection threshold to adapt to the changing number of resolution elements with separation. Next, we plot the astrophysical flux ratios of those planets that give any desired TPF as a function of separation. Rather than choosing a single TPF contour, we propose to show the full $0 \leq \text{TFP} \leq 1$ space as a heatmap. A representative TPF contour can be overplotted for clarity. We call this modified figure the performance map. We argue that the performance map highlights the most scientifically and programmatically relevant quantities, namely the TPFs of the signals of interest for a given number of false positives. The contrast curve, on the other hand, highlights the detection threshold, which has no intrinsic meaning beyond pointing to a false positive fraction. Our proposed performance map is one among many possible methods for visualizing the true and false positive fractions associated with a high dynamic range image. The performance map is an opportunity for displaying the results of planet search programs in a consistent and statistically correct way as well as comparing the performance of various post-processing algorithms within a well-defined statistical framework. By encouraging the scrutiny of this new metric, we hope to improve the prediction and evaluation of the performance of the next generation of high contrast imaging instruments.

5 A Bayesian Framework for Exoplanet Direct Detection and Non-Detection (a summary of Ruffio et al., 2018)

Overview: Rigorously quantifying the information in high contrast imaging data is important for informing follow-up strategies to confirm the substellar nature of a point source, constraining theoretical models of planet-disk interactions, and deriving planet occurrence rates. However, within the exoplanet direct imaging community, non-detections have almost exclusively been defined using a frequentist detection threshold (*i.e.* contrast curve) and associated completeness. This can lead to conceptual inconsistencies when included in a Bayesian framework. A Bayesian upper limit is such that the true value of a parameter lies below this limit with a certain probability. The associated probability is the integral of the posterior distribution with the upper limit as the upper bound. In summary, a frequentist upper limit is a statement about the detectability of planets while a Bayesian upper limit is a statement about the probability of a parameter to lie in an interval given the data. The latter is therefore better suited for rejecting hypotheses or theoretical models based on their predictions. Here, we emphasize that Bayesian statistics and upper limits are more easily interpreted and typically more constraining than the frequentist approach.

A companion at a known location: We first need to define a probability, hereafter cutoff probability, of the true planet flux to fall below the upper limit. The Bayesian upper limit is then defined as the value for which the cumulative distribution of the posterior is equal to the cutoff probability. For a Gaussian noise with known standard deviation σ, an unbounded uniform prior and a 97.7% cutoff probability, the upper limit is 2σ above the estimated flux. The flux can be estimated even if the planet is not formally detected as long as its position is known. In the unlucky event of a very negative noise sample (for example lower than -2σ) at the location of the planet, the estimated flux and the upper-limit could become negative. This might be unsettling as we know that a flux is strictly positive, but this will be corrected by a more informative prior, which will forbid negative values of the flux. For a given noise distribution, it is important to note that the upper-limit is a function of the data, here the estimated flux, while the detection threshold is a property of the noise; using the detection threshold in place of an upper limit is not making optimal use of the data. Assuming Gaussian noise, we find that the planet's flux upper limit $f_{\lim}$ is the value for which the cumulative distribution of the posterior is equal to a cutoff probability η (for example 97.7%),

$$f_{\lim} = \mathscr{Q}_{\tilde{f}_x,\sigma_x}\left(\eta + (1-\eta)\mathscr{C}_{\tilde{f}_x,\sigma_x}(0)\right) \tag{29}$$

where $\mathscr{Q}$ and $\mathscr{C}$ are respectively the quantile and cumulative distribution functions associated with a gaussian distribution with mean $\tilde{f}_x$ and standard deviation σ_x, where $\tilde{f}_x$ and σ_x are the estimated planet flux and flux uncertainty at the image position x.

A companion with a known orbit: Here, we assume that we know the orbit of a planet projected onto the sky plane, but that we do not know its position along it. For example, this situation arise when trying to constrain the mass of a planet in the gap of a proto-planetary disk, where the geometry of the gap defines its orbit. We define the data d, or observation, as the random vector representing the pixel values of the image. The point-source parameters are its position on the projected orbit defined as the curvilinear abscissa s and its flux f. We also define n as a Gaussian random vector with zero mean and covariance matrix Σ. In practice, the noise is assumed to be independent, in which case Σ becomes diagonal. Data, signal and noise are related through $d = fm + n$, with $m = m(s)$ the planet model in the direct imaging data, which is effectively a function of the planet position S. We assume a uniform positive prior over flux, and note that the probability of finding the companion at any position in the orbit is proportional to the time spent around that position according to Kepler's laws. We further assume that we are assessing the planet's flux from speckle-subtracted images, that planet signal is faint relative to the speckles, and that the noise is Gaussian. With these assumption, we find that the posterior distribution is given by

$$\mathscr{P}(f|\boldsymbol{d}) \propto \mathscr{H}(f)\int_{s=s_i}^{s_f} \exp\left\{-\frac{1}{2\sigma_s^2}\left(f^2 - 2f\tilde{f}_s\right)\right\}\frac{1}{Tv_{proj}}\,\mathrm{d}s. \tag{30}$$

where T is the orbital period and $v_{proj} = v_{proj}(s)$ is the projected velocity of a companion at the position s on the ellipse representing the projected orbit. $\tilde{f}_s$ is the estimated flux of the planet at the location s on its projected orbit.

Combining Radial Velocity (RV) and Direct Imaging Observations: Combining RV data with direct imaging is another promising avenue for constraining masses of non-transiting wide-orbit planets detected with Doppler measurements. An example of application of this method can be found in the case of ε Eridani in Mawet et al. (2019) for which Ruffio et al. (2018) describes the theoretical concepts. RV data only provides a lower limit on the mass of the planet due to the $M\sin(i)$ mass-inclination degeneracy. Direct imaging non detection can provide a mass upper bound and therefore reject the lower inclinations. Combining the data involves defining a joint likelihood for the inference of the orbital elements and the masses of the planet and the star. These parameters are all combined into the symbol Θ. We define d_{RV} as the time series of radial velocities and d_{DI} as the direct imaging observation. Since the data are independent, the likelihood can be decomposed as

$$\mathscr{P}(d_{DI}, d_{RV}|\Theta) = \mathscr{P}(d_{RV}|\Theta)\mathscr{P}(d_{DI}|\Theta), \qquad \log\mathscr{P}(d_{DI}|f,x) = -\frac{1}{2\sigma_x^2}\left(f^2 - 2f\tilde{f}_x\right). \tag{31}$$

6 Exoplanet Imaging Data Challenge: benchmarking the various image processing methods for exoplanet detection (a summary of Cantalloube et al., 2020)

Overview: The *Exoplanet Imaging Data Challenge* is a community-wide effort meant to offer a platform for a fair and common comparison of image processing methods designed for exoplanet direct detection. For this purpose, it gathers on a dedicated repository (Zenodo), data from several high-contrast ground-based instruments worldwide in which we injected synthetic planetary signals. The data challenge is hosted on the CodaLab competition platform, where participants can upload their results. The specifications of the data challenge are published on our website [1]. The first phase, launched on the 1st of September 2019 and closed on the 1st of October 2020, consisted in detecting point sources in two types of common data-set in the field of high-contrast imaging: data taken in pupil-tracking mode at one wavelength (subchallenge 1, also referred to as ADI) and multispectral data taken in pupil-tracking mode (subchallenge 2, also referred to as ADI+mSDI).

Based on the planetary signal injections in the data sets available for each subchallenge, we counted the number of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) at the detection threshold provided by the participants. We then repeated this procedure for a range of detection thresholds. A TP is defined as a detection within one resolution element from the position of an injected companion, while a FP is a detection at any other location. Any signal above the chosen threshold is considered as a detection. From these counts, we compute different metrics:

1. the true positive rate (also called sensitivity or recall): $TPR = TP/(TP+FN)$,
2. the false discovery rate (also called precision): $FDR = FP/(FP+TP)$,
3. the F1-score (harmonic mean of precision and sensitivity): F1-score $= 2 \cdot TP/(2 \cdot TP + FP + FN)$.

As a result, we chose to use three figures of merit: (1) the F1-score computed at the submitted threshold, (2) the AUC of the TPR (must be close to 1), and (3) the AUC of the FDR (must be close to 0). The three scores have values between 0 and 1. If no synthetic planetary signals have been injected into the data set, these three scores are undefined. We computed these scores for each data set separately, and averaged them for each instrument, then finally averaged them for all the data sets of a given subchallenge.

We received 22 valid submissions to the ADI subchallenge and 6 valid submissions to the ADI+mSDI subchallenge. We split these submissions into three main families: (1) speckle subtraction techniques, such as PCA and LOCI, (2) inverse problem approaches, such as forward modeled matched filter, and (3) supervised machine learning techniques such as SODINN. Based on the three scores defined above, we compute the F1-score, the AUC of the TPR and AUC of the FDR for all data sets.

We highlight four conclusions from the ADI subchallenge: (1) The three rankings based on each of the three scores give consistent results; however some algorithms are more robust to false alarms (high AUC of the FDR) but are not very sensitive to faint signals (low AUC of the TPR). Depending on the science case (such as deep search for a point source or homogeneous processing of a large sample of data) one may use the most relevant metric to choose some methods over the others. (2) As expected, we observe that globally the speckle subtraction based techniques do not perform as well as more recent techniques, with the exception of supervised machine learning techniques for which we observe numerous false positives. (3) Any technique's performance is dependent on the type of instrument (e.g. coronagraphic versus non-coronagraphic). (4) The latest algorithms, such as RSM and SLIMask for the speckle subtraction techniques, and FMMF, PACO and TRAP for the inverse problem approaches, are providing with the best performance along with the three scores used here.

We highlight four conclusions from the ADI+mSDI subchallenge: (1) Properly exploiting the spectral information helps to detect fainter sources. (2) The spectral template used as a prior for the planetary signal in some methods plays an important role. (3) As a general rule, the more advanced/recent techniques perform better at detecting planetary signals but the results are not homogeneous. This shows once again that applying various algorithms based on different concept is the most appropriate way of validating or invalidating a candidate based on a single data set. (4) As for the ADI subchallenge, inverse problem approaches are the only methods providing directly with the contrast of the planetary signal candidates. This allows to additionally help disentangling the candidate from a stellar residual, by comparing the extracted spectrum of the candidate with the extracted spectrum of a residual speckle.

In the future, we intend to increase the number of data sets available. For instance, we put a link towards the NEAR campaign data on the EIDC website, to allow anyone to test advanced algorithms on this challenging observation sequence. On top of using comparison metrics that are more adapted, we also plan to launch new phases of the data challenge, adding, for example, the characterisation of the detections (position and contrast with their uncertainties) or adding extended sources to be detected. We also plan to offer alternative data set enabling the direct detection of exoplanets, such as high spectral resolution images. We aim to regularly publish the results of the EIDC at every future SPIE astronomical telescopes+instrumentation conference.

[1] https://exoplanet-imaging-challenge.github.io/

7 Summary of Findings

- The contrast curve represents a single choice of a detection threshold, false positive fraction, and true positive fraction (e.g. a 3σ contrast curve indicates a detection threshold of 3σ, a false positive fraction of 0.001 assuming the noise follows a Gaussian distribution, and a true positive fraction of 0.5). Fixing these values with respect to separation from the central star may not be optimal for all applications.

- A contrast curve can be invoked as a frequentist upper limit on the detectability of a planet with a certain separation and brightness. However, a Bayesian upper limit is better suited to rejecting hypotheses regarding a particular planet's detection because it is a statement about the probability of a parameter (e. g. a planet's brightness) to lie within a given interval in the data.

- The contrast curve includes the implicit assumption that the noise in the science image follows a Gaussian distribution. However, the presence of speckles near the coronagraph's inner working angle in the final post-processed science image indicate that the noise is not fully whitened.

- Modern techniques for speckle subtraction (e. g. the Regime-Switching Model detection map; Dahlqvist et al., 2020), generally outperform first-generation techniques (e. g. LOCI; Lafrenière et al., 2007) when applied to the same datasets. However, machine learning techniques (e. g. SODINN; Gomez Gonzalez et al., 2018), do not yet outperform first-generation techniques.

References

Aime, C., & Soummer, R. 2004a, ApJ, 612, L85
—. 2004b, ApJ, 612, L85
Bloemhof, E. E. 2004, Optics Letters, 29, 159
Bloemhof, E. E., Dekany, R. G., Troy, M., & Oppenheimer, B. R. 2001, Astrophys. J. Lett., 558, L71
Bordé, P. J., & Traub, W. A. 2006, ApJ, 638, 488
Brown, R. A. 2005, ApJ, 624, 1010
Cantalloube, F., Gomez-Gonzalez, C., Absil, O., et al. 2020, in Adaptive Optics Systems VII, ed. L. Schreiber, D. Schmidt, & E. Vernet, Vol. 11448, International Society for Optics and Photonics (SPIE), 1027 – 1062. https://doi.org/10.1117/12.2574803
Caucci, L., Barrett, H. H., Devaney, N., & Rodríguez, J. J. 2007, Journal of the Optical Society of America A, 24, 13
Dahlqvist, C. H., Cantalloube, F., & Absil, O. 2020, A&A, 633, A95
Davies, D. 1980, Icarus, 42, 145
Fitzgerald, M. P., & Graham, J. R. 2006, ApJ, 637, 541
Give'on, A., Kern, B., Shaklan, S., Moody, D. C., & Pueyo, L. 2007, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 6691, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series
Gomez Gonzalez, C. A., Absil, O., & Van Droogenbroeck, M. 2018, A&A, 613, A71
Guyon, O., Matsuo, T., & Angel, R. 2009, ApJ, 693, 75
Hinkley, S., Oppenheimer, B. R., Soummer, R., et al. 2007, ApJ, 654, 633
Jensen-Clem, R., Mawet, D., Gomez Gonzalez, C. A., et al. 2018, AJ, 155, 19
Kasdin, N. J., & Braems, I. 2006, ApJ, 646, 1260
Krist, J. 2016, JPL Internal Memo
Lafrenière, D., Marois, C., Doyon, R., Nadeau, D., & Artigau, É. 2007, ApJ, 660, 770
Lawson, P. R., Poyneer, L., Barrett, H., et al. 2012, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 8447, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series
Malbet, F., Yu, J. W., & Shao, M. 1995, PASP, 107, 386
Marois, C., Lafrenière, D., Doyon, R., Macintosh, B., & Nadeau, D. 2006, ApJ, 641, 556
Marois, C., Lafrenière, D., Macintosh, B., & Doyon, R. 2008, ApJ, 673, 647
Mawet, D., Milli, J., Wahhaj, Z., et al. 2014, ApJ, 792, 97
Mawet, D., Hirsch, L., Lee, E. J., et al. 2019, AJ, 157, 33
Meshkat, T., Bailey, V., Rameau, J., et al. 2013, ApJ, 775, L40
Perrin, M. D., Sivaramakrishnan, A., Makidon, R. B., Oppenheimer, B. R., & Graham, J. R. 2003, ApJ, 596, 702
Racine, R., Walker, G. A. H., Nadeau, D., Doyon, R., & Marois, C. 1999, Publ. Astron. Soc. Pac., 111, 587
Ruane, G., Mawet, D., Mennesson, B., Jewell, J., & Shaklan, S. 2018, J. Astron. Telesc. Instrum. Syst., 4, 015004
Ruane, G., Riggs, A., Mazoyer, J., et al. 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10698, Space Telescopes and Instrumentation 2018: Optical, Infrared, and Millimeter Wave, ed. M. Lystrup, H. A. MacEwen, G. G. Fazio, N. Batalha, N. Siegler, & E. C. Tong, 106982S
Ruffio, J.-B., Macintosh, B., Wang, J. J., et al. 2017, ApJ, 842, 14
Ruffio, J.-B., Mawet, D., Czekala, I., et al. 2018, AJ, 156, 196
Sivaramakrishnan, A., Lloyd, J. P., Hodge, P. E., & Macintosh, B. A. 2002, Astrophys. J. Lett., 581, L59. http://stacks.iop.org/1538-4357/581/i=1/a=L59
Soummer, R., Ferrari, A., Aime, C., & Jolissaint, L. 2007, ApJ, 669, 642
Soummer, R., Pueyo, L., & Larkin, J. 2012, ApJ, 755, L28
Wahhaj, Z., Liu, M. C., Biller, B. A., et al. 2013, ApJ, 779, 80