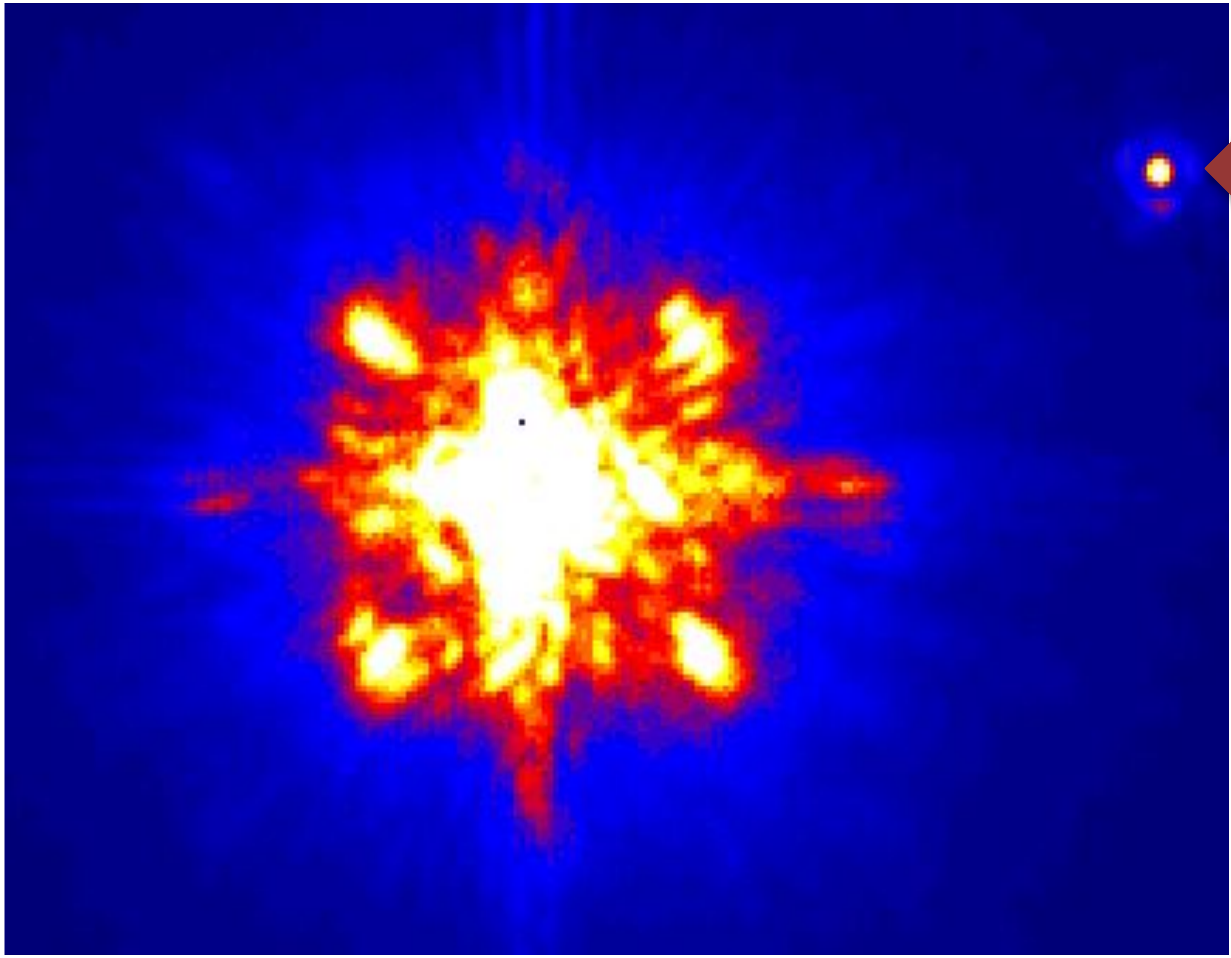
Direct Imaging of Exoplanets *sans* Background Subtraction

Richard A. Frazin
University of Michigan

Direct Imaging Challenges

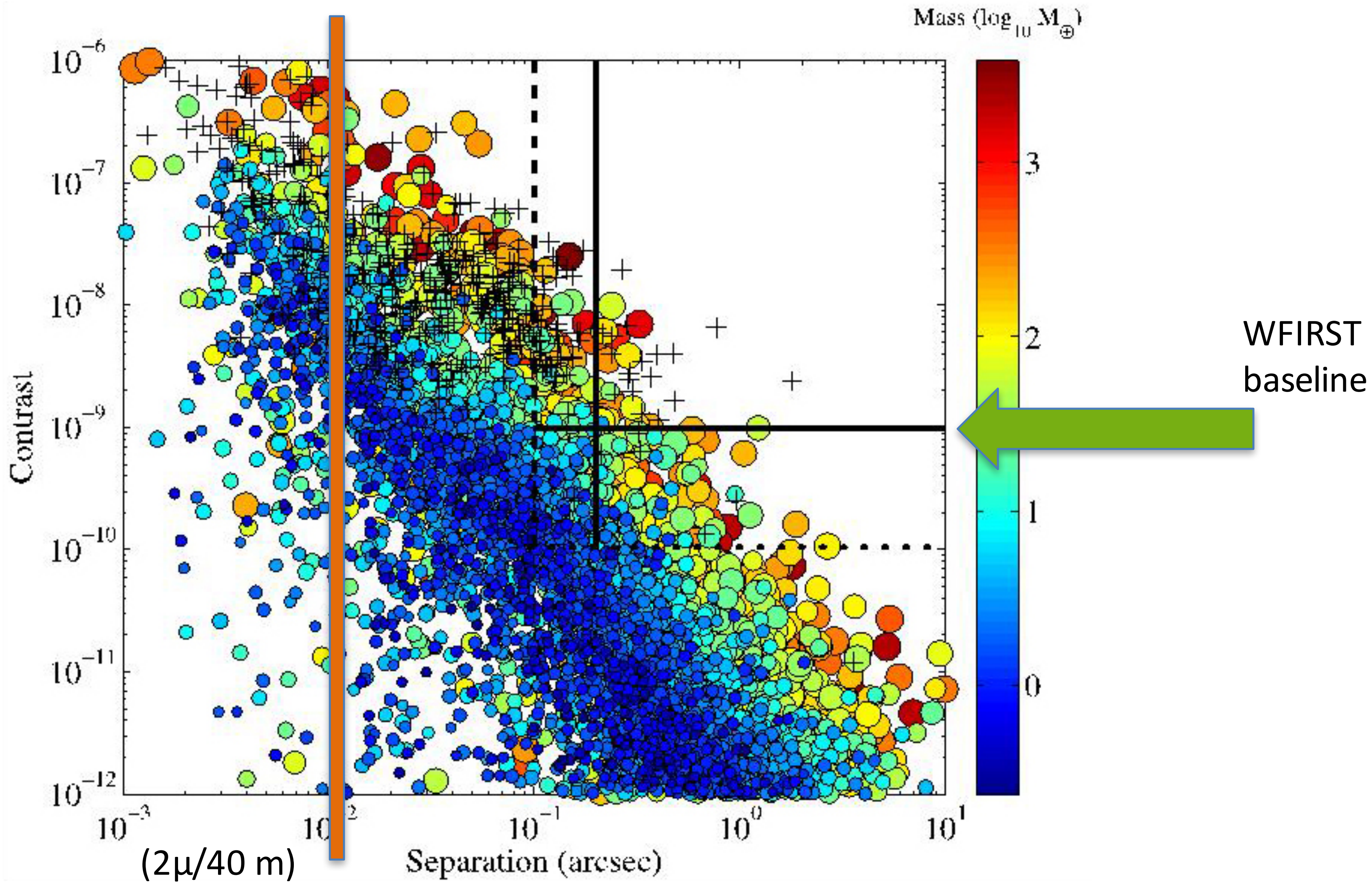
"Pale Blue Dot" photo by Voyager 1, 1990
distance = 40.5 AU = 0.0002 pc

Direct Imaging Challenges



stellar companion, 30 AU. Earth-like planet would be ~10^7 fainter.

IR Hale Telescope (AO) image of Gliese 105 (~ 9 pc).

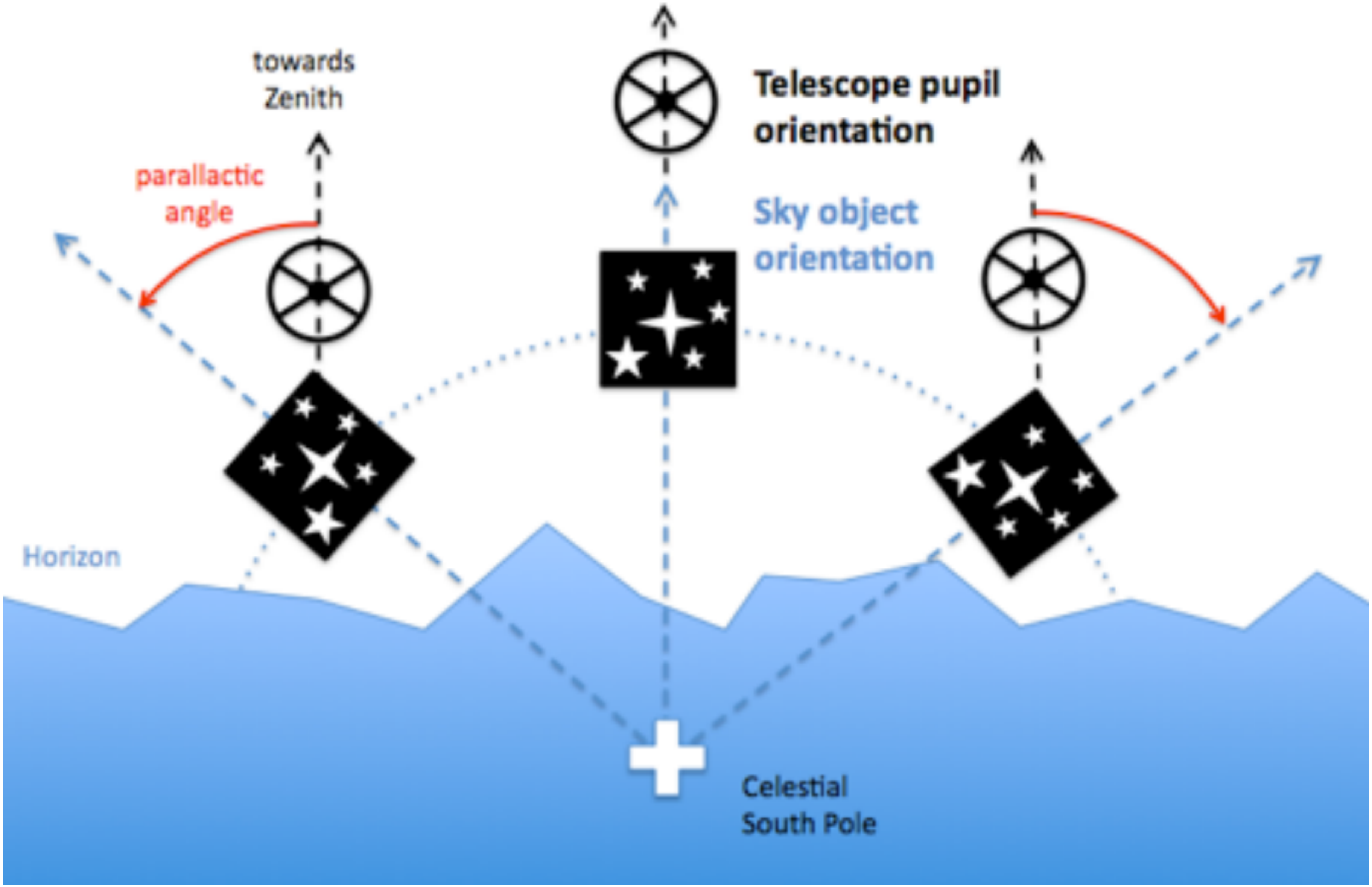


Contrast vs. Separation. Colored circles show a simulation of model planets, ranging in size from Mars-like to several times the radius of Jupiter, placed in orbit around ~200 of the nearest stars within 30 pc. The model assumes roughly four planets per star with a mixture of gas giants, ice giants, and rocky planets, and a size and radius distribution consistent with Kepler results. Color indicates planet mass while size indicates planet radius. Crosses represent known radial velocity planets at their maximum possible contrast values. (WFIRST website)

Differential Imaging, ADI

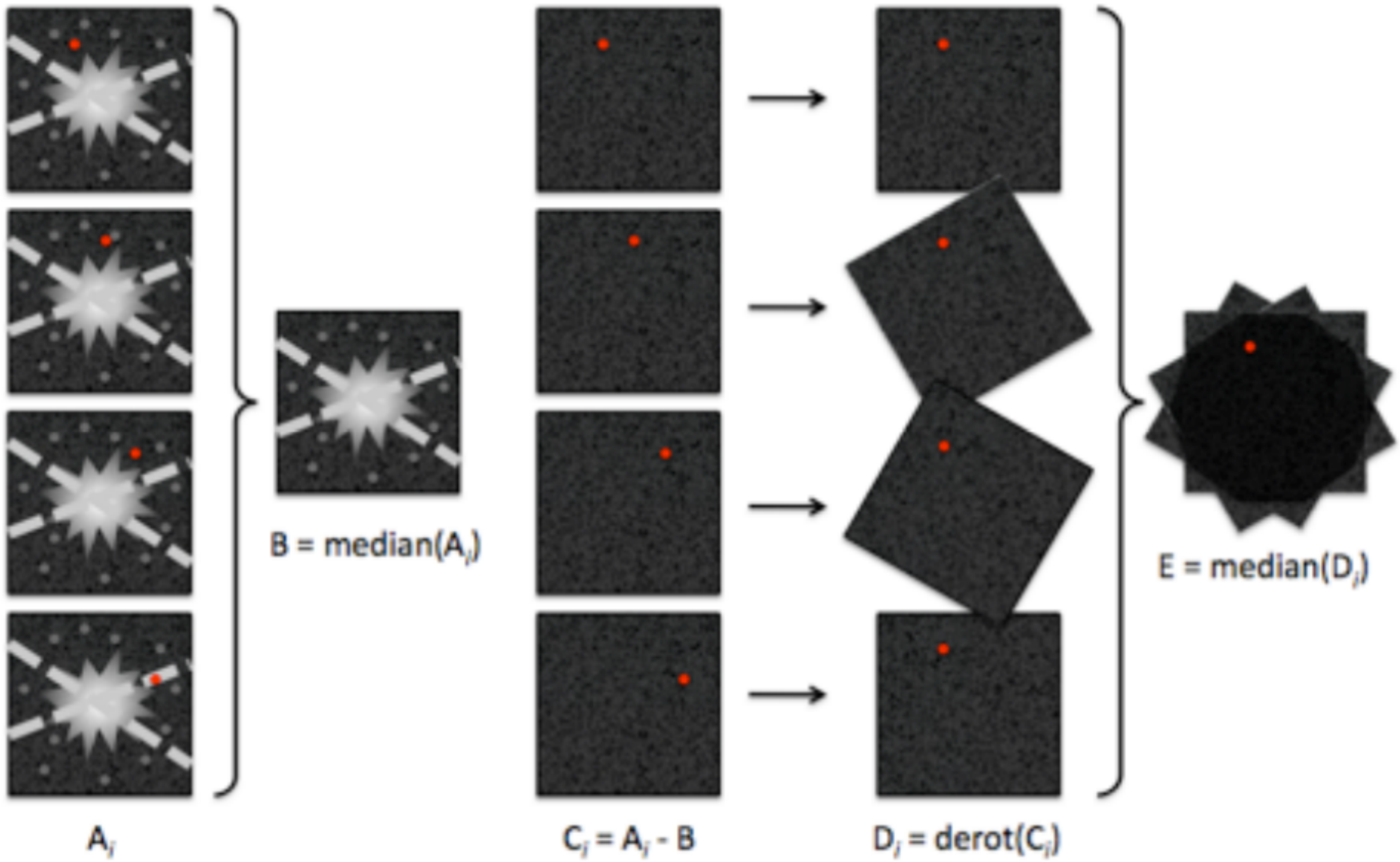
- In order to try to separate the planets from the speckles, the community has employed a number of differential imaging techniques, which attempt to subtract the telescope PSF, hopefully, leaving only planetary light behind.
- The most important of these is called ADI, which relies on the diurnal rotation of the sky to move the planet with respect to the PSF.

ADI principles



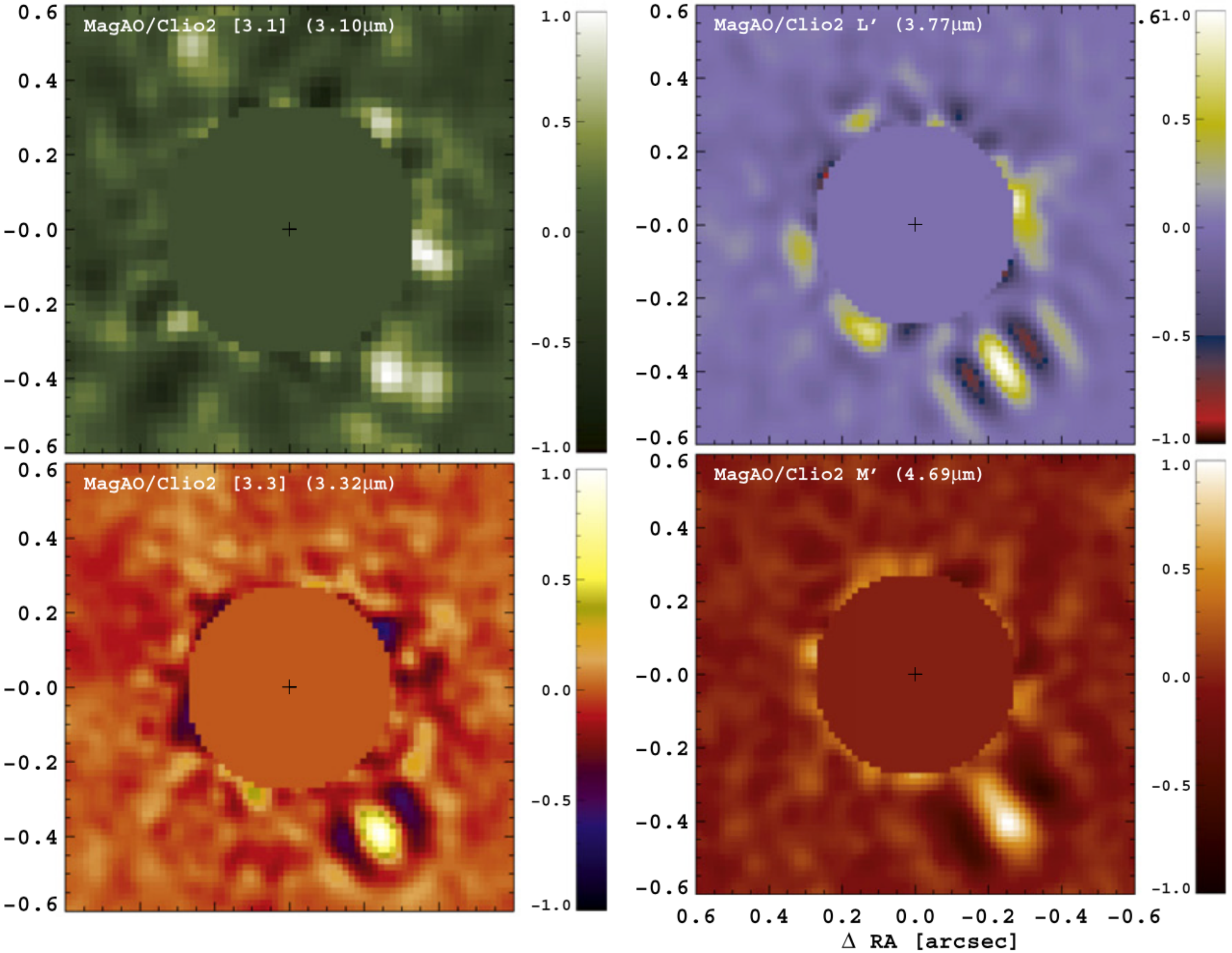
R. Claudi, (http://www.iiassvietri.it/down/ases_2015/Lectures_notes/DI_5.pdf)

ADI principles



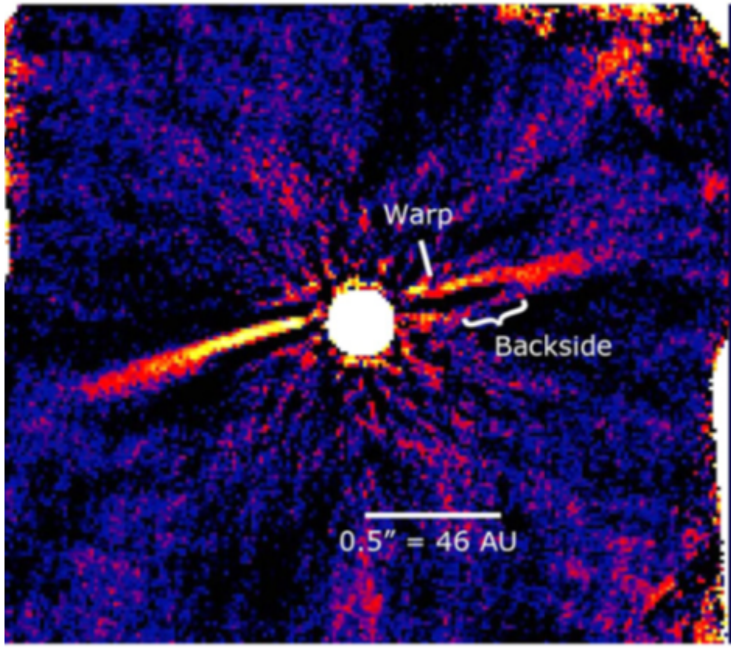
R. Claudi, (http://www.iiassvietri.it/down/ases_2015/Lectures_notes/DI_5.pdf)

β pic results from MagAO Clio, ADI (KLIP) processing. (Morzinski et al. 2015)



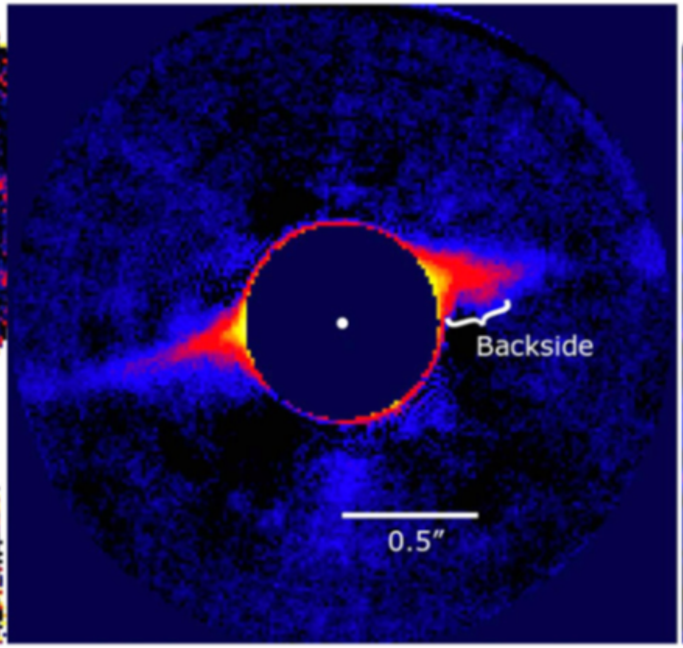
HD106906 pic results from GPI, ADI (KLIP) processing.
Kalas et al. 2015)

standard KLIP
Stokes I



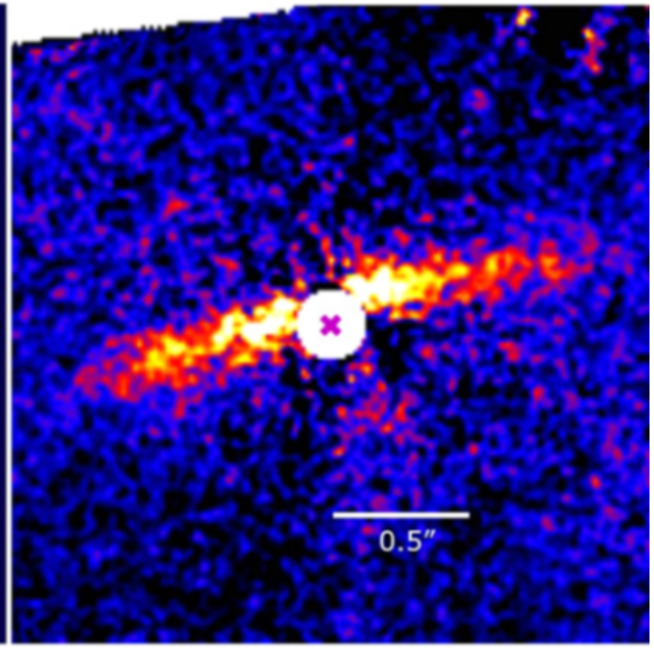
self-subtraction
evident

interpolated KLIP
Stokes I



self-subtraction
mitigated

standard KLIP
Stokes Qr

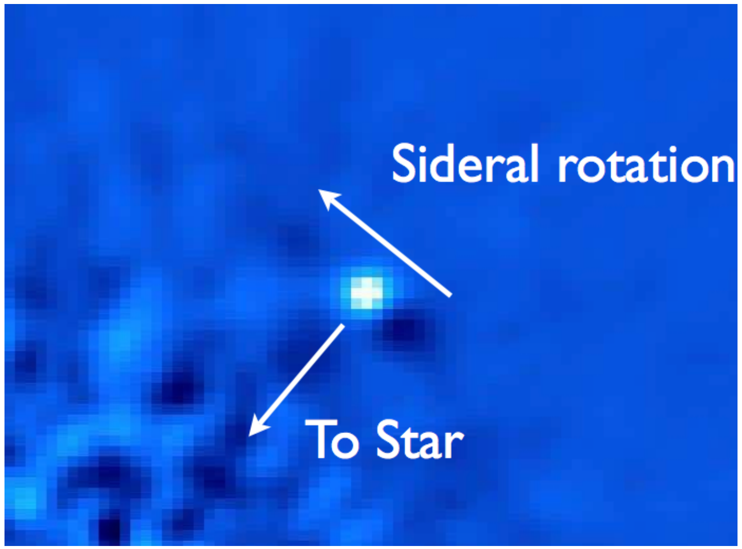


polarized speckle
contamination?

So, why are ADI images problematic in practice? – Big Reasons

- ADI implicitly assumes that the aberrations are not evolving during the course of the observing period (hours, days, or more). But, due to varying mechanical and thermal stress, they are.
- ADI will remove any feature with circular symmetry, whether or not it part of the image. Thus, it is not true imaging.
- Self-subtraction is very problematic since the most informative images are the closest in time and have the least diurnal rotation.

(Marois, SPIE 2010)



So, why are ADI images problematic in practice? – Little Reasons

- Wind characteristics change over time, resulting in a turbulent PSF that changes
- Time variable rotation rate (fastest at zenith)
- Any pointing jitter changes PSF with coronagraph
- Complicated polarization effects in telescope optics (esp. for slightly polarized host stars)
- Statistical penalty increases towards center
- Speckle cancellation/“dark hole” methods won’t work nearly as well in space

A&A 581, A80 (2015)
DOI: 10.1051/0004-6361/201525879

Astronomy & Astrophysics

Detection limits with spectral differential imaging data⋆⋆⋆

J. Rameau[1,2], G. Chauvin[1], A.-M. Lagrange[1], A.-L. Maire[3], A. Boccaletti[4], and M. Bonnefoy[1]

[1] Univ. Grenoble-Alpes/CNRS, IPAG, 38000 Grenoble, France
[2] Institut de Recherche sur les Exoplanètes (iREx), Université de Montréal, Département de physique, CP 6128 Succ. Centre-ville, Montréal, QC, H3C 3J7, Canada
e-mail: julien.rameau@astro.umontreal.ca
[3] INAF, Osservatorio Astronomico di Padova, Vicolo dell'Osservatorio 5, 35122 Padova, Italy
[4] LESIA, Observatoire de Paris, CNRS, Université Pierre et Marie Curie Paris 6 et Université Denis Diderot Paris 7, 5 place Jules Janssen, 92195 Meudon, France

Received 12 February 2015 / Accepted 5 June 2015

ABSTRACT

Context. Direct imaging of exoplanets is polluted by speckle noise that severely limits the achievable contrast. Angular and spectral differential imaging have been proposed to make use of the temporal and chromatic properties of the speckles. Both modes, associated with extreme adaptive-optics and coronagraphy, are at the core of the new generation of planet imagers SPHERE and GPI.
Aims. We aim to illustrate and characterize the impact of the SDI and SDI+ADI (ASDI) data reduction on the detection of giant planets. We also propose an unbiased method to derive the detection limits from SDI/ASDI data.
Methods. Observations of AB Dor B and β Pictoris made with VLT/NaCo were used to simulate and quantify the effects of SDI and ASDI. The novel method is compared to the traditional injection of artificial point sources.
Results. The SDI reduction process creates a typical radial positive-negative pattern of any point-source. Its characteristics and its self-subtraction depend on the separation, but also on the spectral properties of the object. This work demonstrates that the self-subtraction cannot be reduced to a simple geometric effect. As a consequence, the detection performances of SDI observations cannot be expressed as a contrast in magnitude with the central star without the knowledge of the spectral properties of detectable companions. In addition, the residual noise cannot be converted into contrast and physical characteristics (mass, temperature) by standard calibration of flux losses. The proposed method takes the SDI bias into account to derive detection limits without the cost of massively injecting artificial sources into the data. Finally, the sensitivity of ASDI observations can be measured only with a control parameter on the algorithms that controls the minimum rotation that is necessary to build the reference image.

This shows the IDEALIZED (no differential aberration) SDI response pattern for 4 different on/off methane band flux ratios. Difficult to calibrate. Not helpful near λ/D (0.05” for VLT).
(Rameau et al. 2015)

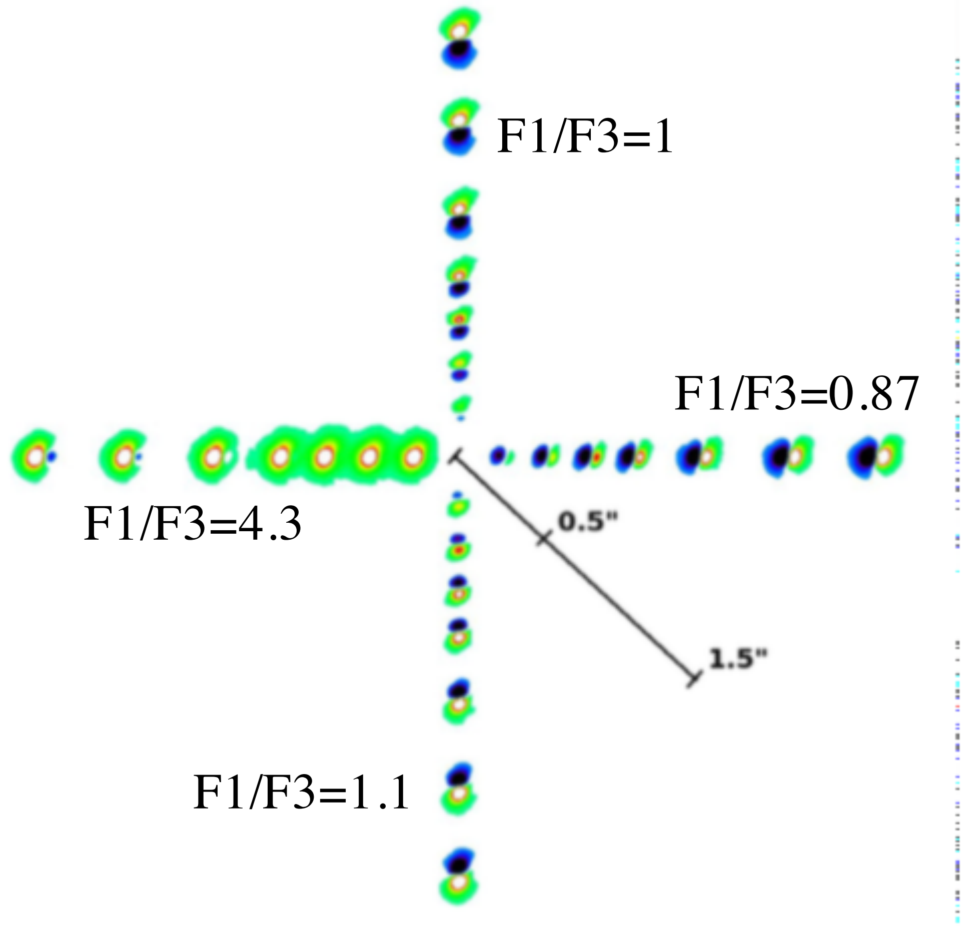


Fig. 3. Effects of SDI processing as a function of the separation and flux ratio (F1/F3) in noise-free images. Upward residuals have a flux ratio of 1 as an example; leftward residuals of 4.3, which is typical of a T7 dwarf; downward residuals have a ratio of 1.1, typical of a mid-L to early-T dwarf; rightward residuals have a ratio of 0.87, which is like that of 2M 0122 B. SDI residuals suffer from a strong self-subtraction (except for a highly methaned companion) and are thus difficult to characterize at small radii, i.e., in the region where the speckle pattern will be subtracted. The bifurcation point is at 1.9″ with these PSFs.

This shows the IDEALIZED (no differential aberration) for combined SDI/ADI response pattern for 4 different on/off methane band flux ratios. Difficult to calibrate. Should be better near λ/D (0.05”). (Rameau et al. 2015)

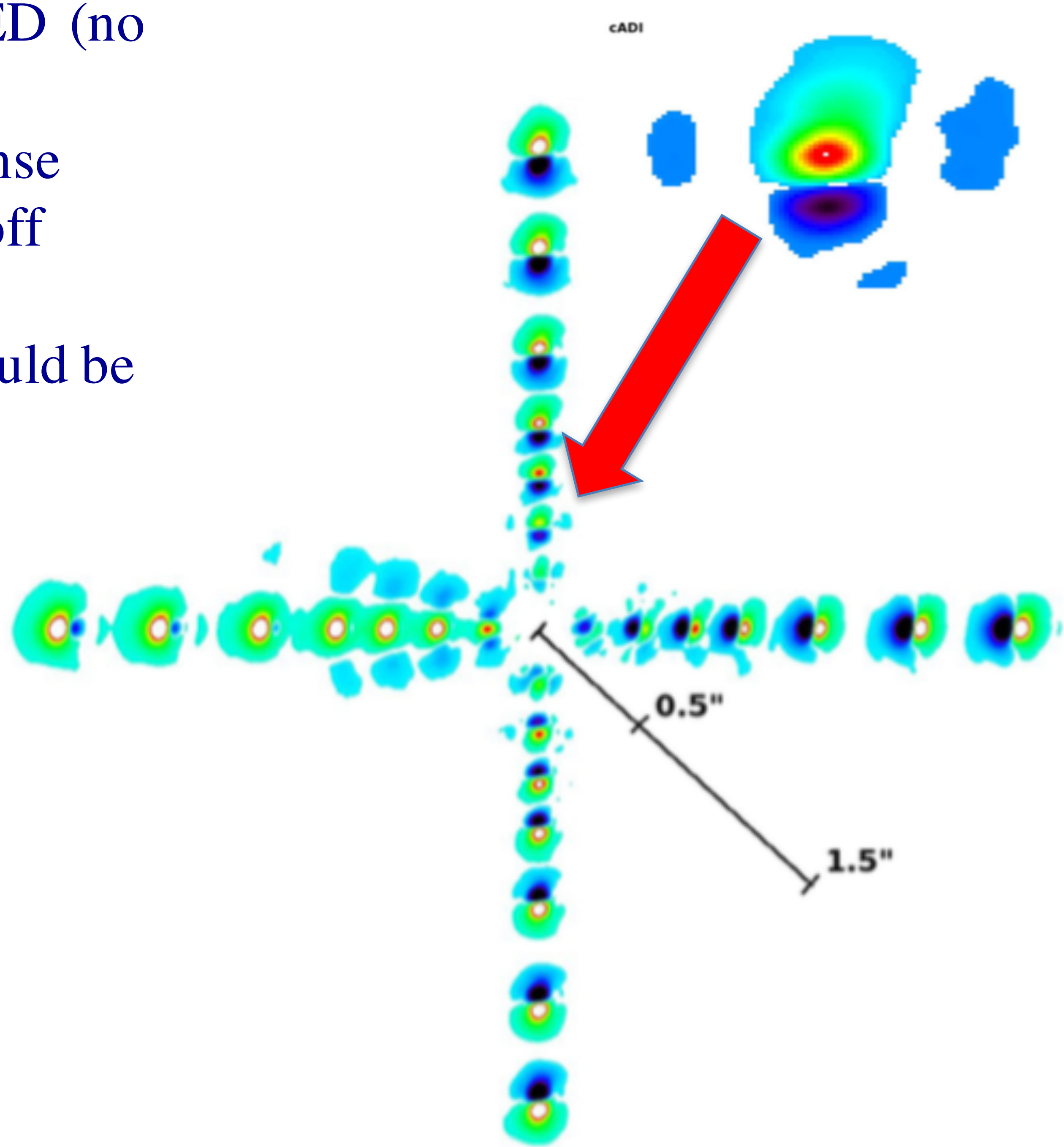


Fig. 11. Same as Fig. 3 with cADI processing in addition to SDI.

The Astrophysical Journal, 792:97 (11pp), 2014 September 10 doi:10.1088/0004-637X/792/2/97

FUNDAMENTAL LIMITATIONS OF HIGH CONTRAST IMAGING SET BY SMALL SAMPLE STATISTICS

D. Mawet[1], J. Milli[1], Z. Wahhaj[1], D. Pelat[2], O. Absil[3,10], C. Delacroix[3], A. Boccaletti[4], M. Kasper[5], M. Kenworthy[6], C. Marois[7], B. Mennesson[8], and L. Pueyo[9]
[1] European Southern Observatory, Alonso de Cordóva 3107, Vitacura, Santiago, Chile
[2] LUTH, Observatoire de Paris, CNRS, UPMC and University Paris Diderot, 5 place Jules Janssen, F-92195 Meudon, France
[3] Département d'Astrophysique, Géophysique et Océanographie, Université de Liège, 17 Allée du Six Août, B-4000 Liège, Belgium
[4] LESIA, Observatoire de Paris, CNRS, UPMC and University Paris Diderot, 5 place Jules Janssen, F-92195 Meudon, France
[5] European Southern Observatory Headquarters, Karl-Schwarzschild-Strasse 2, D-85748 Garching bei München, Germany
[6] Leiden Observatory, Leiden University, P.O. Box 9513, 2300-RA Leiden, The Netherlands
[7] NRC, Herzberg Institute of Astrophysics, Victoria, BC V9E 2E7, Canada
[8] Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA
[9] Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA
Received 2014 January 16; accepted 2014 July 7; published 2014 August 21

ABSTRACT

In this paper, we review the impact of small sample statistics on detection thresholds and corresponding confidence levels (CLs) in high-contrast imaging at small angles. When looking close to the star, the number of resolution elements decreases rapidly toward small angles. This reduction of the number of degrees of freedom dramatically affects CLs and false alarm probabilities. Naively using the same ideal hypothesis and methods as for larger separations, which are well understood and commonly assume Gaussian noise, can yield up to one order of magnitude error in contrast estimations at fixed CL. The statistical penalty exponentially increases toward very small inner working angles. Even at 5–10 resolution elements from the star, false alarm probabilities can be significantly higher than expected. Here we present a rigorous statistical analysis that ensures robustness of the CL, but also imposes a substantial limitation on corresponding achievable detection limits (thus contrast) at small angles. This unavoidable fundamental statistical effect has a significant impact on current coronagraphic and future high-contrast imagers. Finally, the paper concludes with practical recommendations to account for small number statistics when computing the sensitivity to companions at small angles and when exploiting the results of direct imaging planet surveys.

Key words: methods: statistical – techniques: high angular resolution

Online-only material: color figures

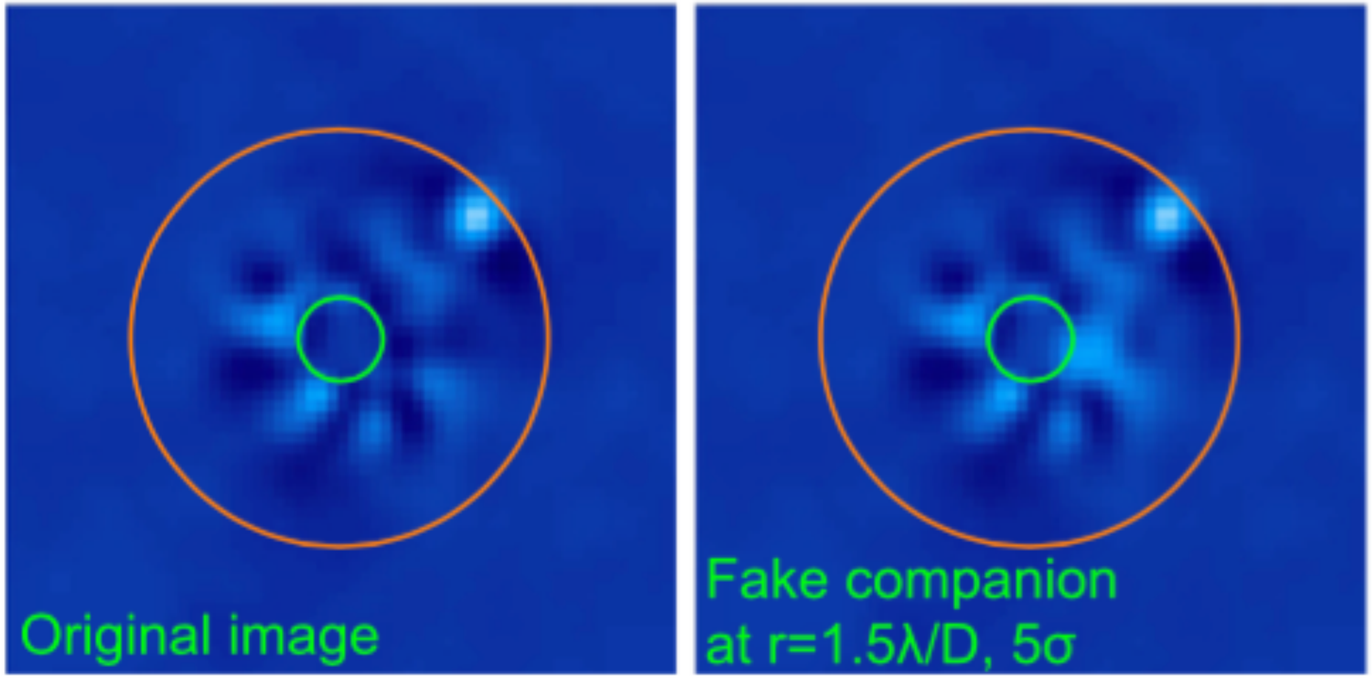
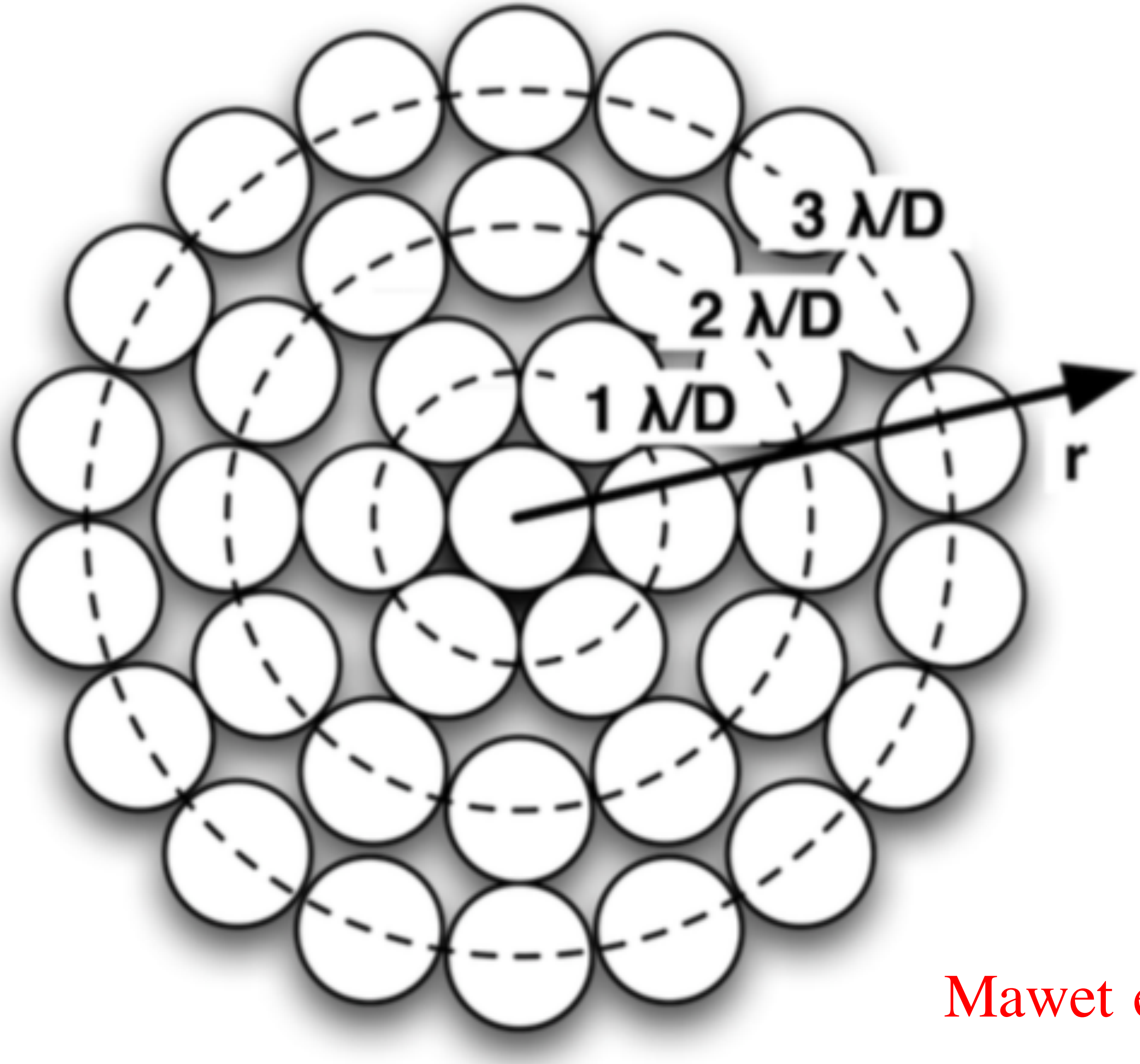


Figure 3. β Pictoris contrast curve (top, continuous curve) and image (bottom left, north is not up) taken with NACO in the L band (Absil et al. 2013), both corrected for the ADI-PCA data reduction throughput. The small green circle is of radius $r = 1\lambda/D$, while the big orange circle is of radius $r = 5\lambda/D$. A fake planet was injected at $r = 1.5\lambda/D$ (to the right of the green circle) at the 5σ throughput-corrected contrast level as presented in Absil et al. (2013). This 5σ fake companion is supposedly yielding a solid detection, rejecting the null hypothesis at the 1–3 $\times 10^{-7}$ CL, assuming normally distributed noise. This is clearly not the case here because of the effect of small sample statistics at small angles. The FPF curve (dashed line) traces the increase of false alarm probability (or equivalently, the decrease of CL) toward small angles. Note that the scale of the y axis is unique, the contrast and FPF curves being dimensionless. Both quantities are related but have different meanings (see the text for details).

Mawet et al. (2014)



Mawet et al. (2014)

Figure 4. Number of resolution elements at a given radius r, is $2\pi r$ (here shown for r ranging from 1 to 3 λ/D). At close separation, the speckle PDF nature is likely varying drastically as a function of r because of the well-known sensitivity of the PSF to low-order aberrations, especially after a coronagraph.

Then, what can be done? The planets look like speckles and the speckles look like planets.

Or do they?

- In 2012 Szymon Gladysz told me about his work (JOSAA 27, A64, 2010) involving AO images in which the exposures were not too long to completely average over the turbulence statistics (which evolves on 10^{-3} s time-scales).
- His work established that, at short exposure stellar speckles and planets have different statistical distributions of the intensity in the image.
- I tried to understand why…

The Astrophysical Journal, 767:21 (11pp), 2013 April 10 doi:10.1088/0004-637X/767/1/21

UTILIZATION OF THE WAVEFRONT SENSOR AND SHORT-EXPOSURE IMAGES FOR SIMULTANEOUS ESTIMATION OF QUASI-STATIC ABERRATION AND EXOPLANET INTENSITY

Richard A. Frazin
Department of Atmospheric, Oceanic and Space Sciences, University of Michigan, Ann Arbor, MI 48109, USA; rfrazin@umich.edu
Received 2012 November 20; accepted 2013 February 13; published 2013 March 21

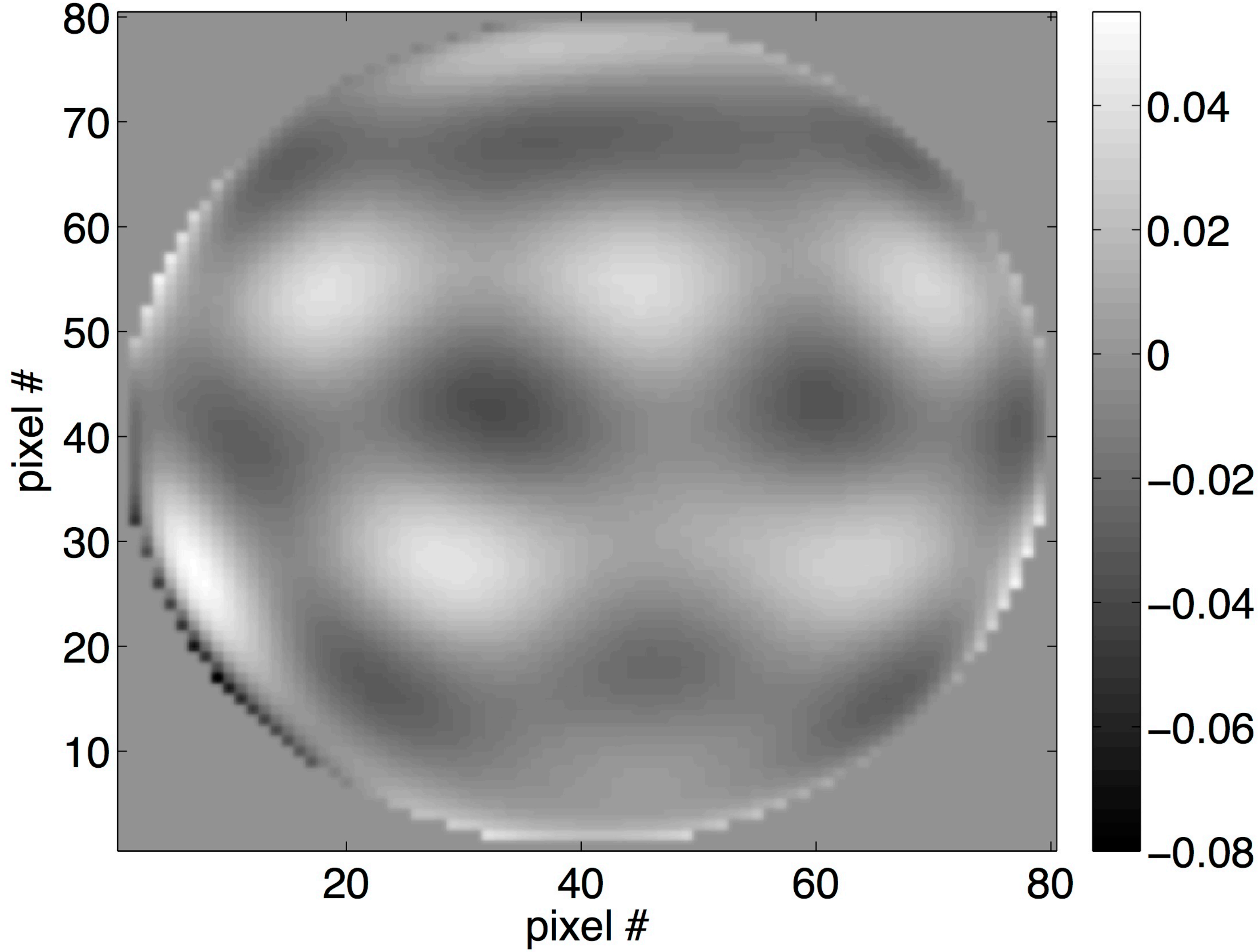
ABSTRACT

Heretofore, the literature on exoplanet detection with coronagraphic telescope systems has paid little attention to the information content of short exposures and methods of utilizing the measurements of adaptive optics wavefront sensors. This paper provides a framework for the incorporation of the wavefront sensor measurements in the context of observing modes in which the science camera takes millisecond exposures. In this formulation, the wavefront sensor measurements provide a means to jointly estimate the static speckle and the planetary signal. The ability to estimate planetary intensities in as little as a few seconds has the potential to greatly improve the efficiency of exoplanet search surveys. For simplicity, the mathematical development assumes a simple optical system with an idealized Lyot coronagraph. Unlike currently used methods, in which increasing the observation time beyond a certain threshold is useless, this method produces estimates whose error covariances decrease more quickly than inversely proportional to the observation time. This is due to the fact that the estimates of the quasi-static aberrations are informed by a new random (but approximately known) wavefront every millisecond. The method can be extended to include angular (due to diurnal field rotation) and spectral diversity. Numerical experiments are performed with wavefront data from the AEOS Adaptive Optics System sensing at 850 nm. These experiments assume a science camera wavelength λ of 1.1 μ, that the measured wavefronts are exact, and a Gaussian approximation of shot-noise. The effects of detector read-out noise and other issues are left to future investigations. A number of static aberrations are introduced, including one with a spatial frequency exactly corresponding the planet location, which was at a distance of $\approx 3\lambda/D$ from the star. Using only 4 s of simulated observation time, a planetary intensity, of $\approx$1 photon ms^{-1}, and a stellar intensity of $\approx 10^5$ photons ms^{-1} (contrast ratio 10^5), the short-exposure estimation method recovers the amplitudes' static aberrations with 1% accuracy, and the planet brightness with 20% accuracy.

My Coronagraph Simulations

- I started with a series of 4000 measured wavefronts from the AEOS AO system (thanks to Lewis Roberts at JPL)
- Then I simulated how a simple stellar coronagraph would respond to these wavefronts
- I included “unknown” aberration in the optical system, including a sinusoidal term with a spatial frequency that placed a speckle exactly over the simulated planet.

Aberration (pupil plane)



Aberration used in simulation (sinusoid + Zernicke polynomials).

Image of Star w/ Aberration

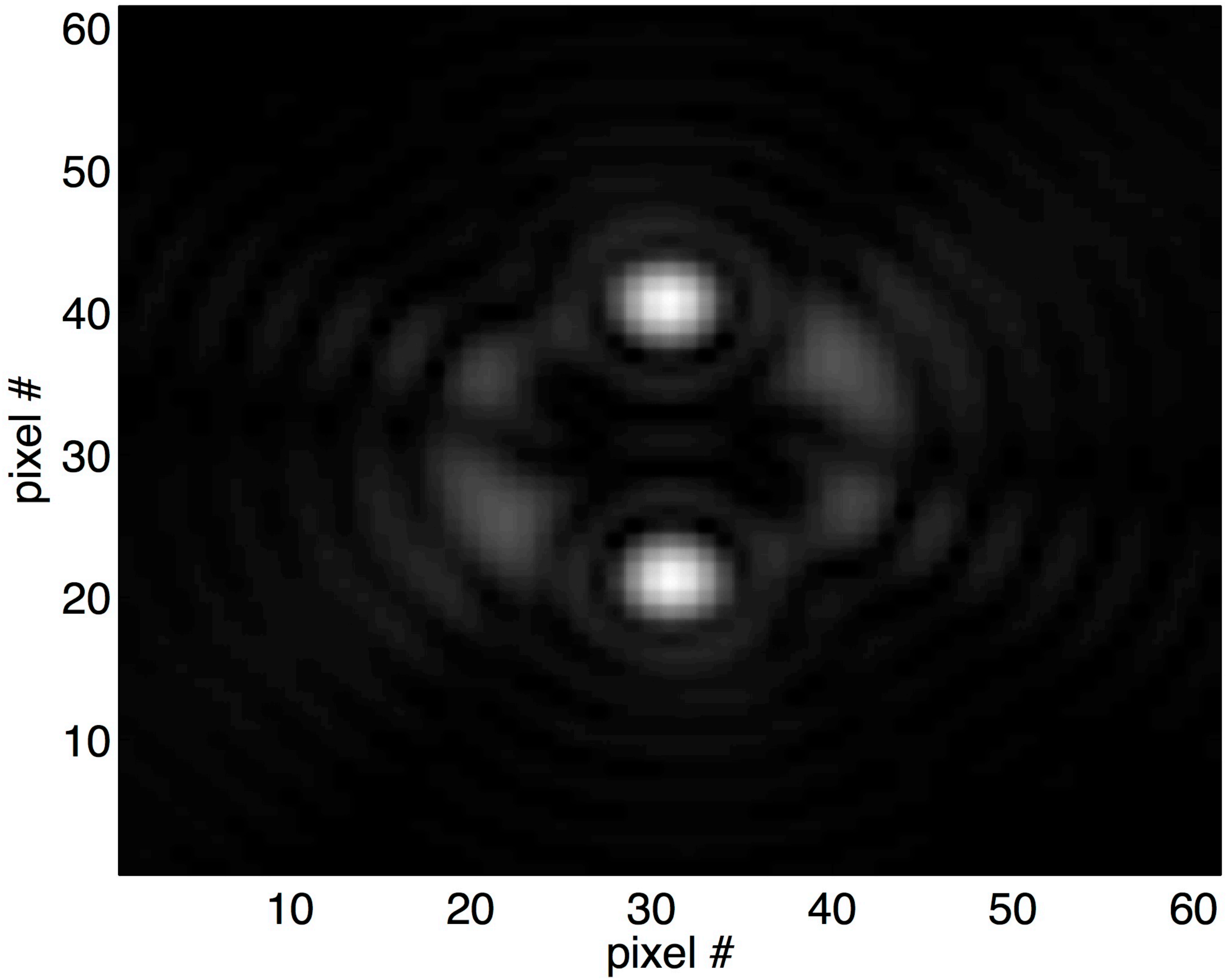
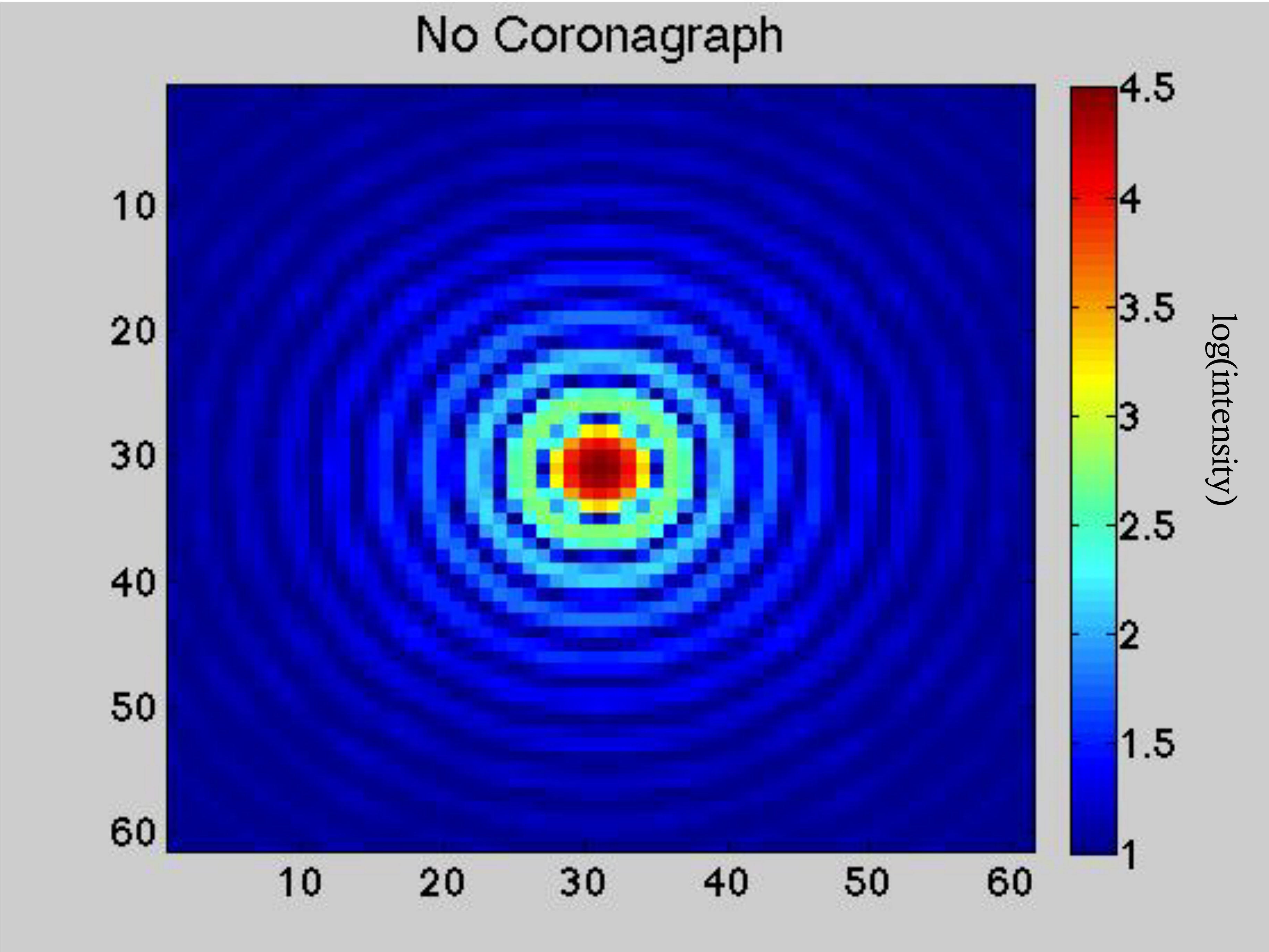
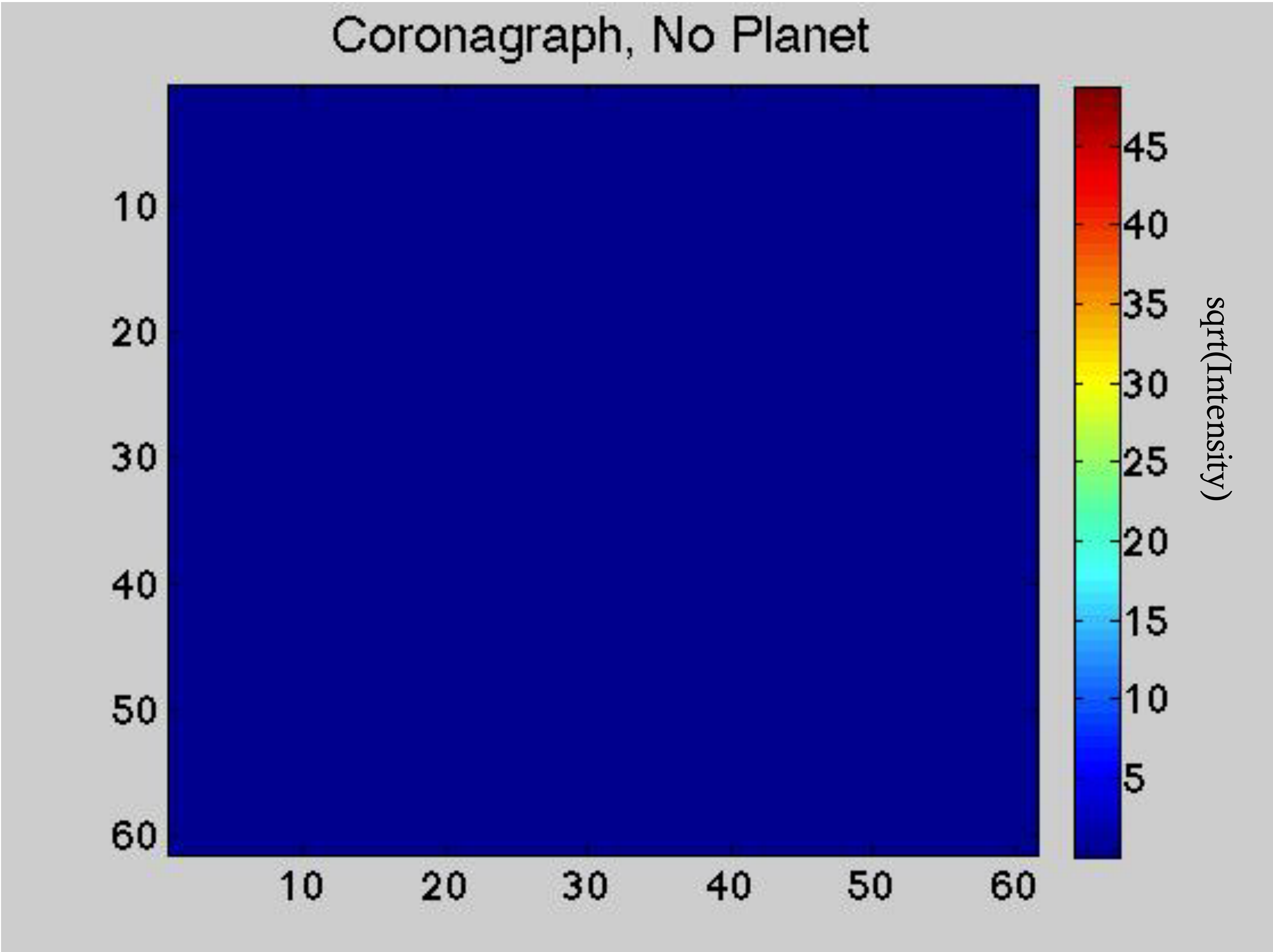


Image plane manifestation of aberration used in simulation (flat wavefront). One of the dots is exactly coincident with a planet.

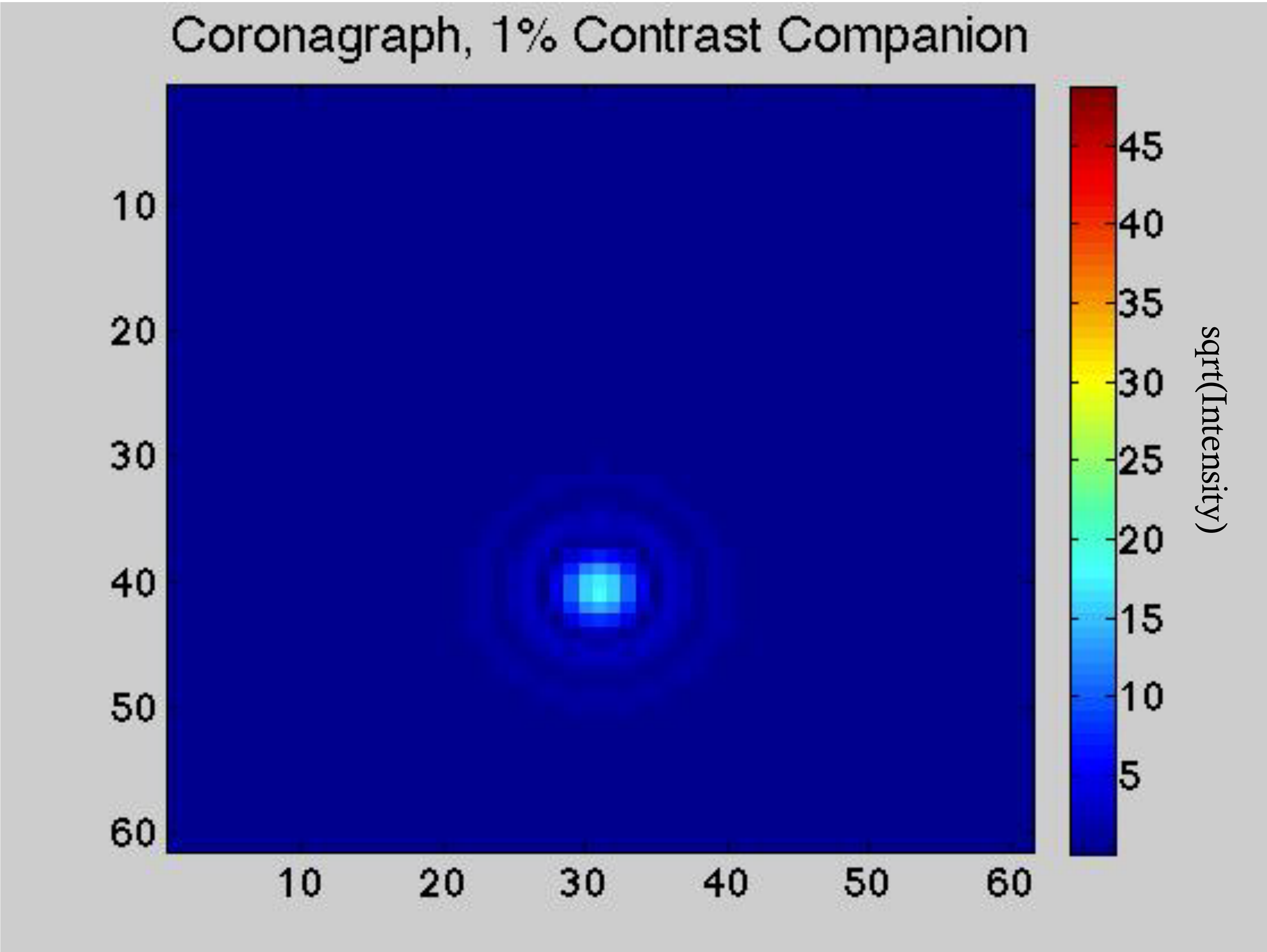
Movie: No Coronagraph



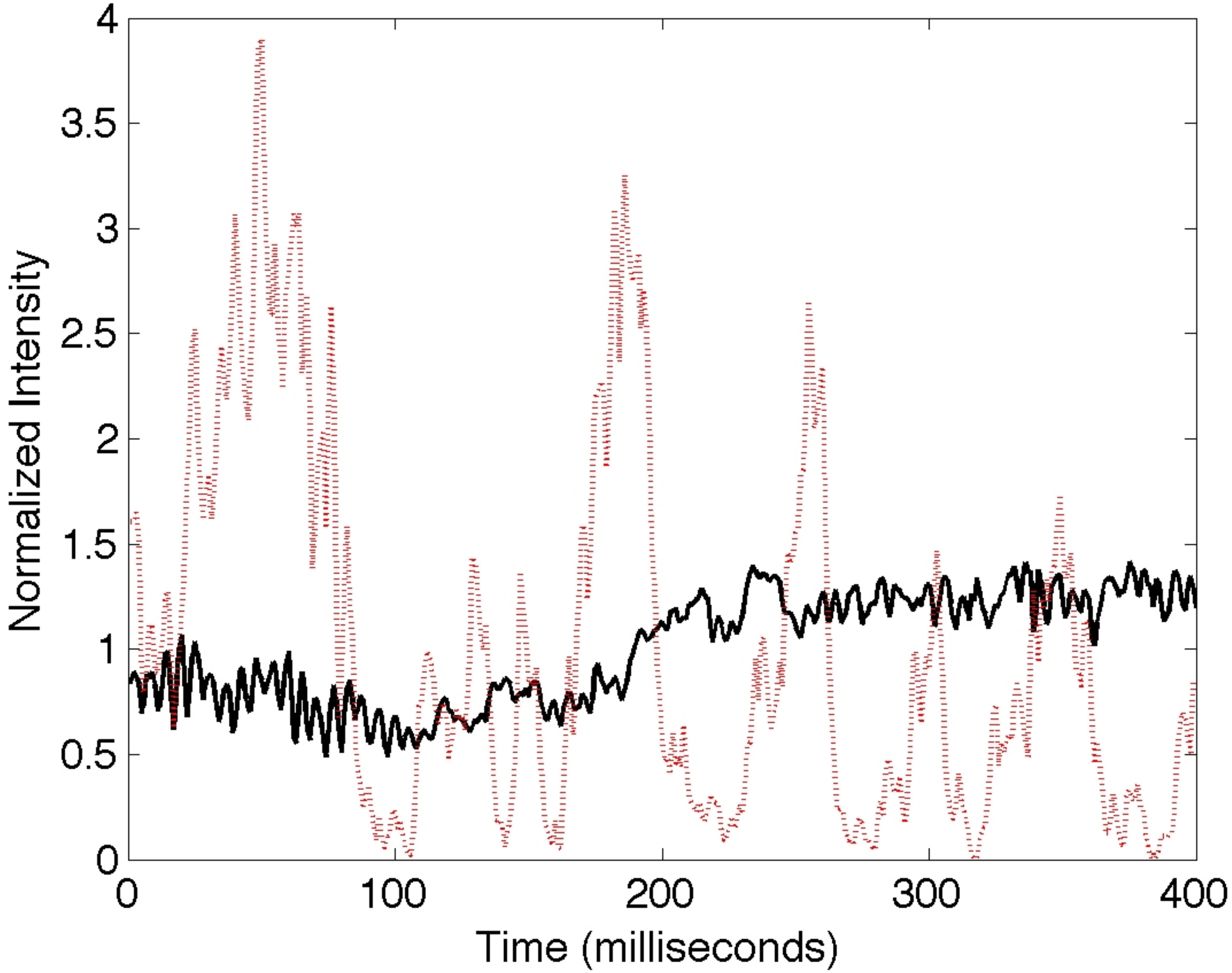
Coronagraph (no companion)



Coronagraph (Companion 1%)



Coronagraph Simulations



Red: stellar speckle intensity (normalized at planet position. Black: Planet intensity (normalized) at same position.

I demonstrated this effect analytically using physical optics arguments in my 2013 ApJ.

Comprehensive Solution: Statistical Inference

- I showed mathematically that the wavefront sensor data stream and millisecond exposures can be used to simultaneously determine the aberrations and the planetary image self-consistently.
- Later, I demonstrated that this approach can take into account subtle effects of polarizing elements (e.g., mirrors) in the telescope

Statistical Inference

$$\underbrace{I(\rho_{ij}, t_k)}_{\text{pixel value}} = \underbrace{n(\rho_{ij}, t_k)}_{\text{noise}} + \underbrace{\mathcal{A}(\rho_{ij}, t_k) + \mathcal{F}(\rho_{ij}, t_k) \cdot \boldsymbol{p} + \mathcal{B}(\rho_{ij}, t_k) \cdot \boldsymbol{a} + \mathcal{G}(\rho_{ij}, t_k) \cdot \boldsymbol{g}}_{\text{system model containing wavefront information}}$$

variable	status	description
$I(\rho_{ij}, t_k)$	measured	science camera intensity at pixel (i, j) in frame k
n	modeled	random process describing noise in measurements
$\boldsymbol{p}$	inferred	coefficient vector specifying planetary image
$\boldsymbol{a}$	inferred	coefficient vector of NCPA functions (inc. WFS bias)
$\boldsymbol{g}$	inferred	uncalibrated WFS gain coefficient vector
$\mathcal{A}$	modeled	atmospheric speckle image
$\mathcal{F}$	modeled	atmospheric speckle convolution kernel
$\mathcal{B}$	modeled	function to describe NCPA speckles
$\mathcal{G}$	modeled	function for uncalibrated WFS gains

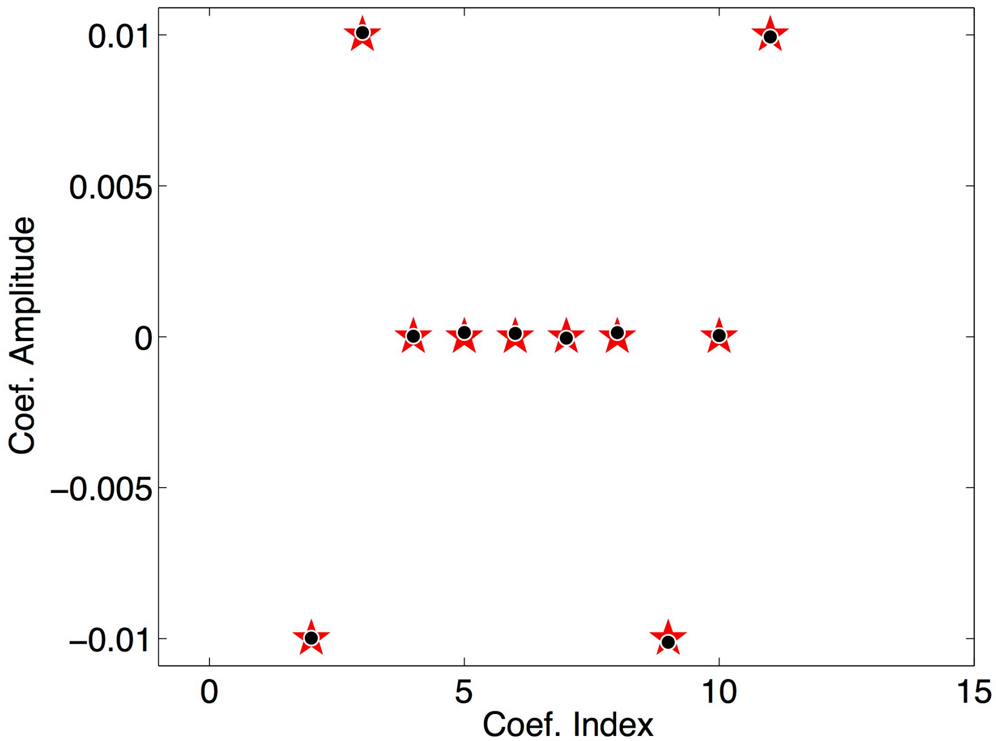
Simulation Parameters

- Incl. shot noise
- Star/planet brightness 10^5 (total)
- Planet brightness 1 photon/millisecond
- Star/planet brightness at planet position

  > 500 (averaged over time)
- Sinusoidal aberration creates speckle over planet
- Planet located at $\approx 3\lambda/D$
- 4 s of AOES data (4000 wavefronts)
- Since there is no rotational (spectral, or other) information, no existing other method is capable of separating the planet from the speckle.

Simulation Results

- 20% accuracy of planetary brightness
- Near perfect estimation of aberration coefficients
- Other simulations show graceful response to detector readout noise

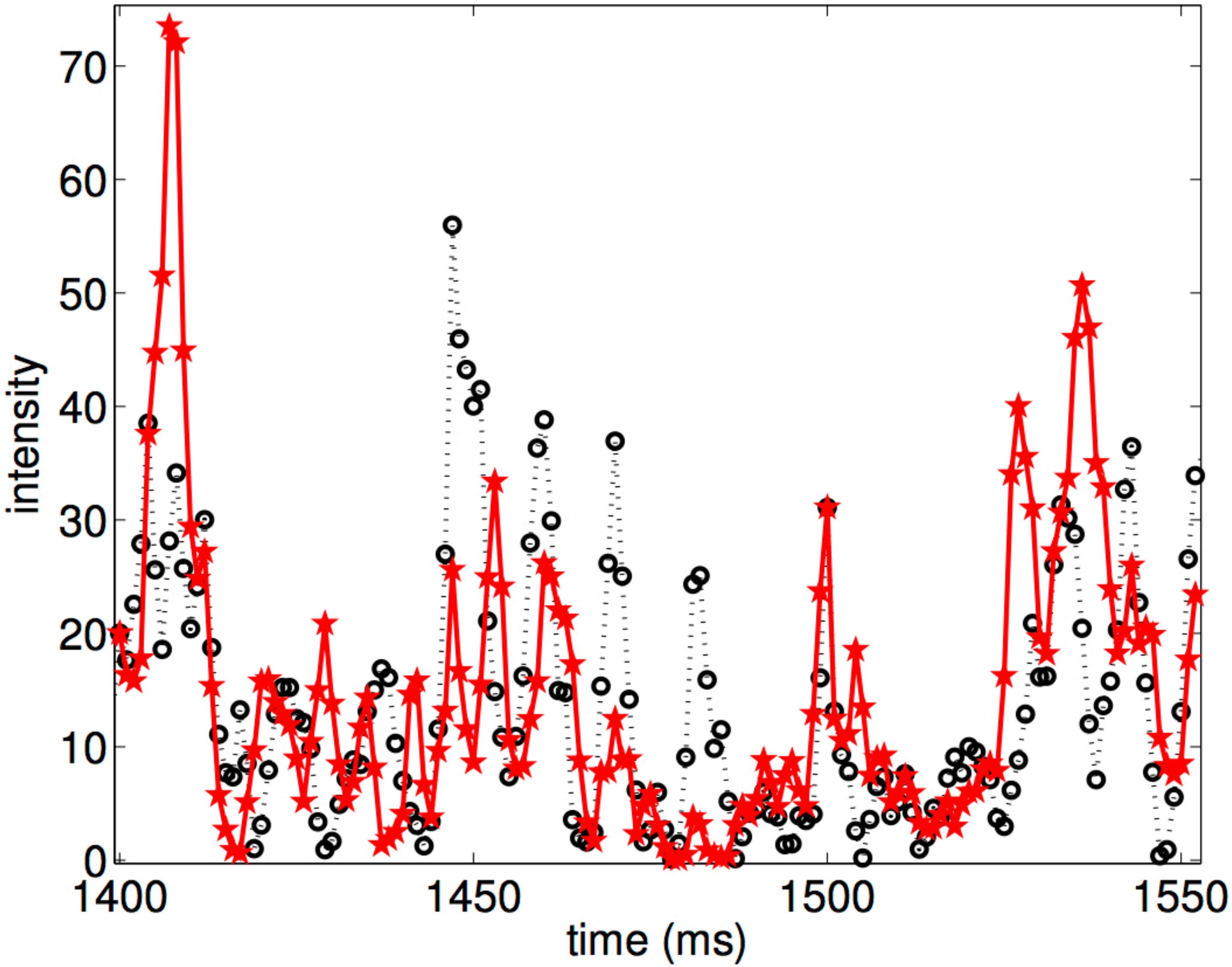


Stars: actual aberration coef.
Points: determined coef. Value via statistical inference

Can Also Include:

- Diurnal rotation constraints (used by ADI)
- Multi-wavelength constraints (used by SDI)
- Polarization constraints (used by PDI)
- Dark hole strategies, but improved to handle multi-planar aberration and multiple DMs
- High frequency vibration detection!

Incidental Benefit: Vibration Detection with FP Sensing



Frazin, SPIE 2014

- Turbulent modulation of speckle caused high frequency vibrations. Red: 10 Hz Black: 100 Hz.

Potential Hurdles

- Detector readout noise – New generation of NIR detectors is capable of kHz readout and about 1 e noise per pixel.
- Need precise calibration of WFS – Solve for bias and gain errors (as shown in equation)
- 1 kHz rate → 1 M images in 17 m. Huge data processing demand – Sequential estimation based on Kalman filtering
- Complicated but interesting statistical issues arising from WFS measurement error…

Journal of the Optical Society of America A
OPTICS, IMAGE SCIENCE, AND VISION

Statistical framework for the utilization of simultaneous pupil plane and focal plane telemetry for exoplanet imaging. I. Accounting for aberrations in multiple planes

Richard A. Frazin

Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, Michigan 48109, USA (rfrazin@umich.edu)

Received 30 November 2015; revised 15 February 2016; accepted 29 February 2016; posted 29 February 2016 (Doc. ID 254879); published 22 March 2016

A new generation of telescopes with mirror diameters of 20 m or more, called extremely large telescopes (ELTs), has the potential to provide unprecedented imaging and spectroscopy of exoplanetary systems, if the difficulties in achieving the extremely high dynamic range required to differentiate the planetary signal from the star can be overcome to a sufficient degree. Fully utilizing the potential of ELTs for exoplanet imaging will likely require simultaneous and self-consistent determination of both the planetary image and the unknown aberrations in multiple planes of the optical system, using statistical inference based on the wavefront sensor and science camera data streams. This approach promises to overcome the most important systematic errors inherent in the various schemes based on differential imaging, such as angular differential imaging and spectral differential imaging. This paper is the first in a series on this subject, in which a formalism is established for the exoplanet imaging problem, setting the stage for the statistical inference methods to follow in the future. Every effort has been made to be rigorous and complete, so that validity of approximations to be made later can be assessed. Here, the polarimetric image is expressed in terms of aberrations in the various planes of a polarizing telescope with an adaptive optics system. Further, it is shown that current methods that utilize focal plane sensing to correct the speckle field, e.g., electric field conjugation, rely on the tacit assumption that aberrations on multiple optical surfaces can be represented as aberration on a single optical surface, ultimately limiting their potential effectiveness for ground-based astronomy.

Statistical Framework for Utilization of Simultaneous Pupil Plane and Focal Plane Telemetry for Exoplanet Imaging, Part II: Implications of Wavefront Measurement Error for Regression Variables

RICHARD A. FRAZIN[‡*]

‡ Dept. of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI 48109
* E-mail: rfrazin _at_ umich.edu

under review (JOSA-A)

Compiled June 10, 2016

This series of articles is part of effort to transcend the limitations of differential imaging of exoplanets by combining simultaneous millisecond imaging in both the science camera (SC) and the wavefront sensor (WFS). The impetus for this is the construction of extremely large telescopes (ELTs) and the arrival of a new generation of high-cadence, ultra-low noise near infrared detectors. This article, the second in the series, builds on the structure *mise en place* in the first in order to express the SC image in terms of the field impinging on the WFS, thereby providing a direct connection between the measurements made in both subsystems. The expression for the SC image in terms of the field impinging on the WFS is used to construct a system of linear regression equations from which the unknown optical aberrations (in multiple planes) and the planetary image can be estimated simultaneously. The so-constructed regression equations require an estimate of the residual phase, produced by analysis of the WFS data stream, which will be subject to significant uncertainties. It is shown that uncertainties in the residual phase estimate not only add "noise" to the coefficients in the regression equations, but also create important biases in the estimates of these quantities, all of which must be taken into account in the regression analysis. Simulations are shown in order to illustrate the bias effects.

Regression Challenges
(simple example):

measured intensity at detector

unaberrated stellar speckle intensity

propagation operator (integration implied)

residual phase (AO residual)

$$I_{\mathrm{m}}(\boldsymbol{\rho}_l, t_i) - I_0(\boldsymbol{\rho}_l, t_i) = jI_\star \mathrm{Y}(\boldsymbol{\rho}_l, \boldsymbol{r})\mathrm{Y}^*(\boldsymbol{\rho}_l, \boldsymbol{r}') \times \left[\sum_{l=0}^{N-1} (a_l \psi_l(\boldsymbol{r}) - a_l \psi_l(\boldsymbol{r}'))\right] \exp j[\phi_{\mathrm{r}}(\boldsymbol{r}, t_i) - \phi_{\mathrm{r}}^*(\boldsymbol{r}', t_i)] \,, \quad \textbf{(36)}$$

(unknown) aberration coefficient

(pupil plane) aberration expansion function, e.g., Zernike polynomial

Eq. (36) can be put in the canonical form of a linear system of equations, $\mathbf{y} = \mathbf{H}\mathbf{x}$, in which $\mathbf{y}$ and $\mathbf{H}$ and $\mathbf{x}$ are given by:

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_0 \\ \vdots \\ \mathbf{y}_i \\ \vdots \\ \mathbf{y}_{T-1} \end{bmatrix}, \mathbf{H} = \begin{bmatrix} \mathbf{H}_0 \\ \vdots \\ \mathbf{H}_i \\ \vdots \\ \mathbf{H}_{T-1} \end{bmatrix} \text{ and } \mathbf{x} = \begin{bmatrix} a_0 \\ \vdots \\ a_{N-1} \end{bmatrix}, \quad \textbf{(37)}$$

where the index i corresponds to the time-stamp t_i, and T is the number of SC exposures. In ms imaging the size of this system quickly becomes unwieldy, necessitating sequential solution methods, such as Kalman filtering.

where

$$\mathbf{y}_i = \begin{bmatrix} I_{\mathrm{m}}(\boldsymbol{\rho}_0, t_i) - I_0(\boldsymbol{\rho}_0, t_i) \\ \vdots \\ I_{\mathrm{m}}(\boldsymbol{\rho}_l, t_i) - I_0(\boldsymbol{\rho}_l, t_i) \\ \vdots \\ I_{\mathrm{m}}(\boldsymbol{\rho}_{L-1}, t_i) - I_0(\boldsymbol{\rho}_{L-1}, t_i) \end{bmatrix} . \quad \mathbf{(38)}$$

The (l, k) element of the matrix $\mathbf{H}_i$ is given by:

$$h_{i,l,k} = -2I_\star \Im\{\Upsilon(\boldsymbol{\rho}_l, \boldsymbol{r})\Upsilon^*(\boldsymbol{\rho}_l, \boldsymbol{r}')\psi_k(\boldsymbol{r}) \times \exp j\left[\phi_{\mathrm{r}}(\boldsymbol{r}, t_i) - \phi_{\mathrm{r}}^*(\boldsymbol{r}', t_i)\right]\} , \quad \mathbf{(39)}$$

remember I_0 ?

$$I_0(\boldsymbol{\rho}_l, t_i) \equiv I_\star \Upsilon(\boldsymbol{\rho}_l, \boldsymbol{r}) \Upsilon^*(\boldsymbol{\rho}_l, \boldsymbol{r}') \exp j[\phi_\mathrm{r}(\boldsymbol{r}, t_i) - \phi_\mathrm{r}^*(\boldsymbol{r}', t_i)], \quad \textbf{(35)}$$

How do you estimate the AO residual, $\phi_\mathrm{r}(\boldsymbol{r}, t_i)$, for the regression equations ?

You need to get it from the WFS data stream.

Assume the WFS gives us some estimate of $\phi_{\mathrm{r}}(\boldsymbol{r}, t_i)$. Let us call it $\hat{\phi}_{\mathrm{r}}(\boldsymbol{r}, t_i)$.

What happens when we replace $\phi_{\mathrm{r}}(\boldsymbol{r}, t_i)$ with $\hat{\phi}_{\mathrm{r}}(\boldsymbol{r}, t_i)$ in the expression for I_0?

In other words, what happens if:

$$I_0(\boldsymbol{\rho}_l, t_i) \equiv I_\star \Upsilon(\boldsymbol{\rho}_l, \boldsymbol{r}) \Upsilon^*(\boldsymbol{\rho}_l, \boldsymbol{r}') \exp j \left[\phi_\mathrm{r}(\boldsymbol{r}, t_i) - \phi_\mathrm{r}^*(\boldsymbol{r}', t_i)\right]$$

$$\Downarrow$$

$$\hat{I}_0(\boldsymbol{\rho}_l, t_i) \equiv I_\star \Upsilon(\boldsymbol{\rho}_l, \boldsymbol{r}) \Upsilon^*(\boldsymbol{\rho}_l, \boldsymbol{r}') \exp j \left[\hat{\phi}_\mathrm{r}(\boldsymbol{r}, t_i) - \hat{\phi}_\mathrm{r}^*(\boldsymbol{r}', t_i)\right]$$

?

- In post-analysis, i.e., after the WFS makes the measurement of $\hat{\phi}_{\mathrm{r}}(\boldsymbol{r}, t_i)$, $\hat{\phi}_{\mathrm{r}}(\boldsymbol{r}, t_i)$ is no longer considered a random variable.

- Instead, the random variable is the unknown, true value: $\phi_{\mathrm{r}}(\boldsymbol{r}, t_i)$.

- The governing probability distribution is: $\mathcal{P}\left[\phi_{\mathrm{r}}(\boldsymbol{r}, t_i) | \hat{\phi}_{\mathrm{r}}(\boldsymbol{r}, t_i)\right]$.

- Our analysis requires the 1<u>st</u> and 2<u>nd</u> order moments of $\mathcal{P}\left[\phi_{\mathrm{r}}(\boldsymbol{r}, t_i) | \hat{\phi}_{\mathrm{r}}(\boldsymbol{r}, t_i)\right]$.

Considering $\mathcal{P}\left[\phi_{\mathrm{r}}(\boldsymbol{r}, t_i) | \hat{\phi}_{\mathrm{r}}(\boldsymbol{r}, t_i)\right]$:

- Let $\mathcal{E}$ be the expectation operator for this distribution.
- Define $\phi_{\mathrm{b}}(\boldsymbol{r}, t_i)$ as the 1st moment or BIAS. In practice, it might not depend (much) on $\boldsymbol{r}$ or t.
- Define the estimator error as: $\phi_{\mathrm{r}}(\boldsymbol{r}, t_i) - \hat{\phi}_{\mathrm{r}}(\boldsymbol{r}, t_i) = \phi_{\mathrm{b}}(\boldsymbol{r}, t_i) + \delta_{\mathrm{r}}(\boldsymbol{r}, t_i)$.
- Then, $\mathcal{E}[\delta_{\mathrm{r}}(\boldsymbol{r}, t_i)] = 0$, and the 2nd order moments are given by: $\mathcal{E}[\delta_{\mathrm{r}}(\boldsymbol{r}, t_i)\delta_{\mathrm{r}}^*(\boldsymbol{r}', t_k)]$, $\mathcal{E}[\delta_{\mathrm{r}}(\boldsymbol{r}, t_i)\delta_{\mathrm{r}}(\boldsymbol{r}', t_k)]$, and $\mathcal{E}[\delta_{\mathrm{r}}^*(\boldsymbol{r}, t_i)\delta_{\mathrm{r}}^*(\boldsymbol{r}', t_k)]$.

For simplicity, assume $\phi_{\mathrm{b}}(\boldsymbol{r}, t_i) = 0$, and consider $\mathcal{E}\{\exp[j\phi_{\mathrm{r}}(\boldsymbol{r}, t_i)]\}$:

$$
\begin{aligned}
&\mathcal{E}\{\exp[j\phi_{\mathrm{r}}(\boldsymbol{r}, t_i)]\} = \\
&\qquad \mathcal{E}\{\exp j[\hat{\phi}_{\mathrm{r}}(\boldsymbol{r}, t_i) + \delta_{\mathrm{r}}(\boldsymbol{r}, t_i)]\} = \\
&\qquad \exp j[\hat{\phi}_{\mathrm{r}}(\boldsymbol{r}, t_i)]\,\mathcal{E}\{\exp[j\delta_{\mathrm{r}}(\boldsymbol{r}, t_i)]\} \\
&\approx \exp j[\hat{\phi}_{\mathrm{r}}(\boldsymbol{r}, t_i)]\,\mathcal{E}\{1 + \overbrace{j\delta_{\mathrm{r}}(\boldsymbol{r}, t_i)}^{0} - \delta_{\mathrm{r}}^2(\boldsymbol{r}, t_i)/2\} \\
&\qquad = \exp j[\hat{\phi}_{\mathrm{r}}(\boldsymbol{r}, t_i)]\,\{1 - \mathcal{E}[\delta_{\mathrm{r}}^2(\boldsymbol{r}, t_i)]/2\}
\end{aligned}
$$

Thus, due to its nonlinearity, this function has an attenuation (assuming that δ is mostly real and small enough for the Taylor expansion to be valid).

Similarly, consider $\mathcal{E}[I_0(\boldsymbol{\rho}_l, t_i)]$:

$$\mathcal{E}[I_0(\boldsymbol{\rho}_l, t_i)] =$$

$$I_\star \Upsilon(\boldsymbol{\rho}_l, \boldsymbol{r}) \Upsilon^*(\boldsymbol{\rho}_l, \boldsymbol{r}') \mathcal{E}\{\exp j[\phi_\mathrm{r}(\boldsymbol{r}, t_i) - \phi_\mathrm{r}^*(\boldsymbol{r}', t_i)]\}$$

$$= I_\star \Upsilon(\boldsymbol{\rho}_l, \boldsymbol{r}) \Upsilon^*(\boldsymbol{\rho}_l, \boldsymbol{r}') \exp j[\hat{\phi}_\mathrm{r}(\boldsymbol{r}, t_i) - \hat{\phi}_\mathrm{r}^*(\boldsymbol{r}', t_i)] \times$$

$$\left\{1 - \frac{\mathcal{E}[\delta_\mathrm{r}^2(\boldsymbol{r}, t_i)]}{2} - \frac{\mathcal{E}[\delta_\mathrm{r}^{*2}(\boldsymbol{r}', t_i)]}{2} + \mathcal{E}[\delta_\mathrm{r}(\boldsymbol{r}, t_i) \delta_\mathrm{r}^*(\boldsymbol{r}', t_i)]\right\}.$$

Thus, the integrand of I_0 has an attenuation, assuming δ is real and small enough for the Taylor expansion to be valid.

Similar considerations apply to the other terms of the **H** matrix in Eqs. (37) and (39), because they all contain the same exponential.

Simulation of Bias Effect on I_0

- Assume telescope with D = 8 m
- 32 x 32 WFS measuring AO residual with variance of 0.3 rad^2, corresponding to an "expected" Strehl of exp(-0.3) = 0.74.
- The phase error δ was taken to be a set of 574 x 574 random numbers, variance 0.17 rad^2 (24 deg RMS)
- Sim1: correlation length l_c = 8.5 cm (1/3 WFS pixel)
- Sim2: correlation length l_c = 25 cm (1 WFS pixel)
- Each simulation had 1000 independent realizations of δ.

Simulated WFS measurement and pupil mask

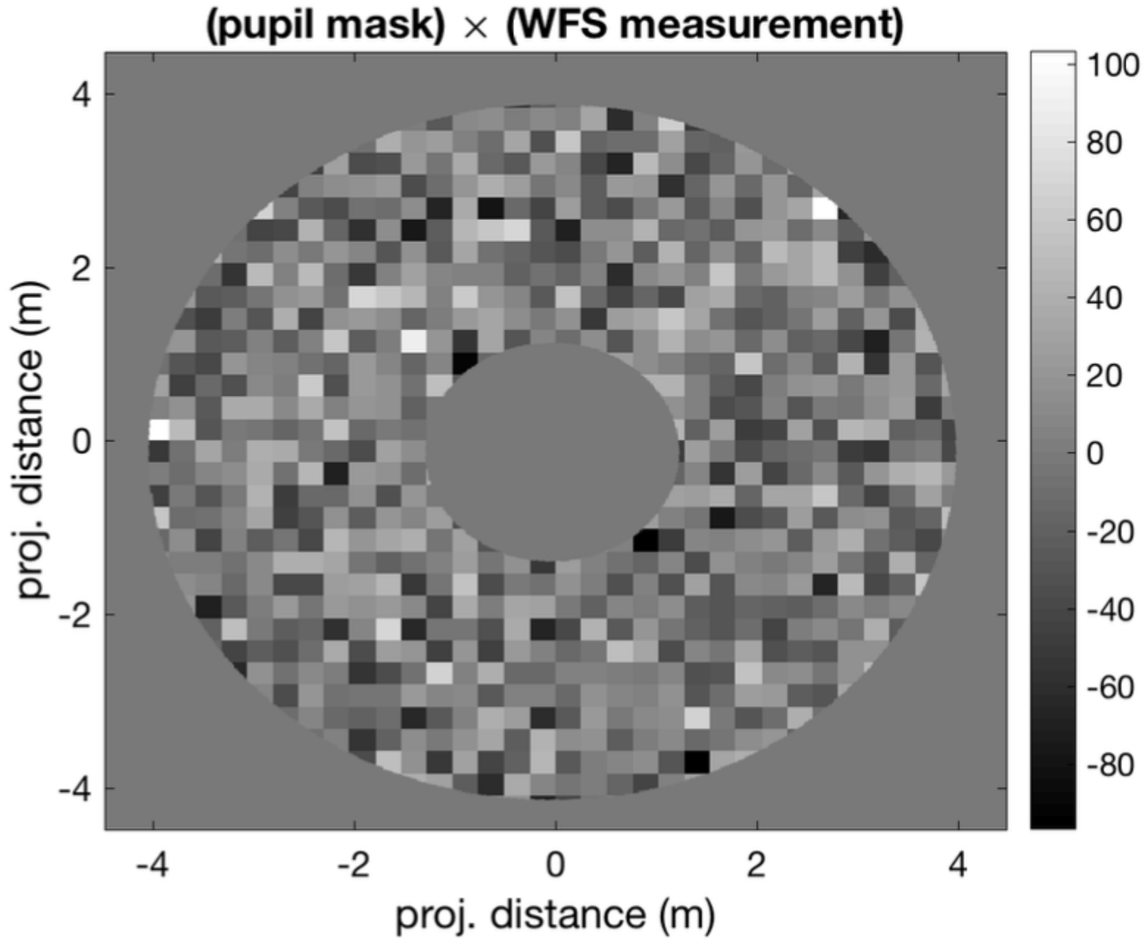


Fig. 2. Pupil plane mask and simulated 32×32 residual phase estimate, $\hat{\phi}_{\mathrm{r}}(\boldsymbol{r})$, (color scale units are degrees). The variance of $\hat{\phi}_{\mathrm{r}}(\boldsymbol{r})$ is 0.3 rad^2, which would correspond to a Strehl ratio of 0.74 in the absence of error in $\hat{\phi}_{\mathrm{r}}(\boldsymbol{r})$. The pupil mask has a projected diameter of 8 m.

Resulting I_0 image

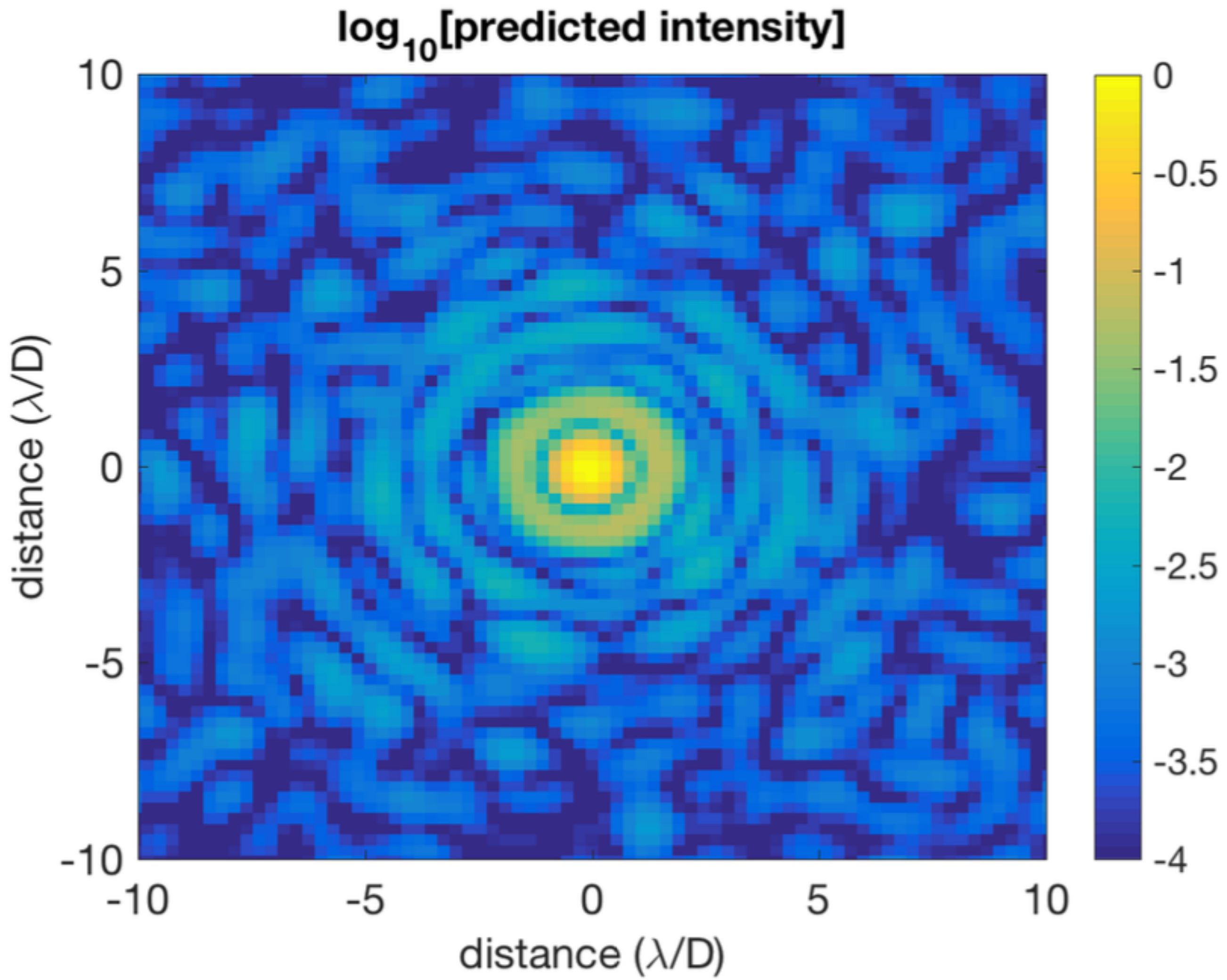


Fig. 3. Calculated value of the stellar speckle intensity $\hat{I}_0$ in the image plane, based on a (simulated) WFS estimate of the residual phase $\hat{\phi}_\mathrm{r}$, which is shown in Fig. 2. Color scale corresponds to the $\log_{10}$ of the normalized intensity.

Images of $\hat{I}_0(\boldsymbol{\rho})/\mathcal{E}[I_0(\boldsymbol{\rho})]$

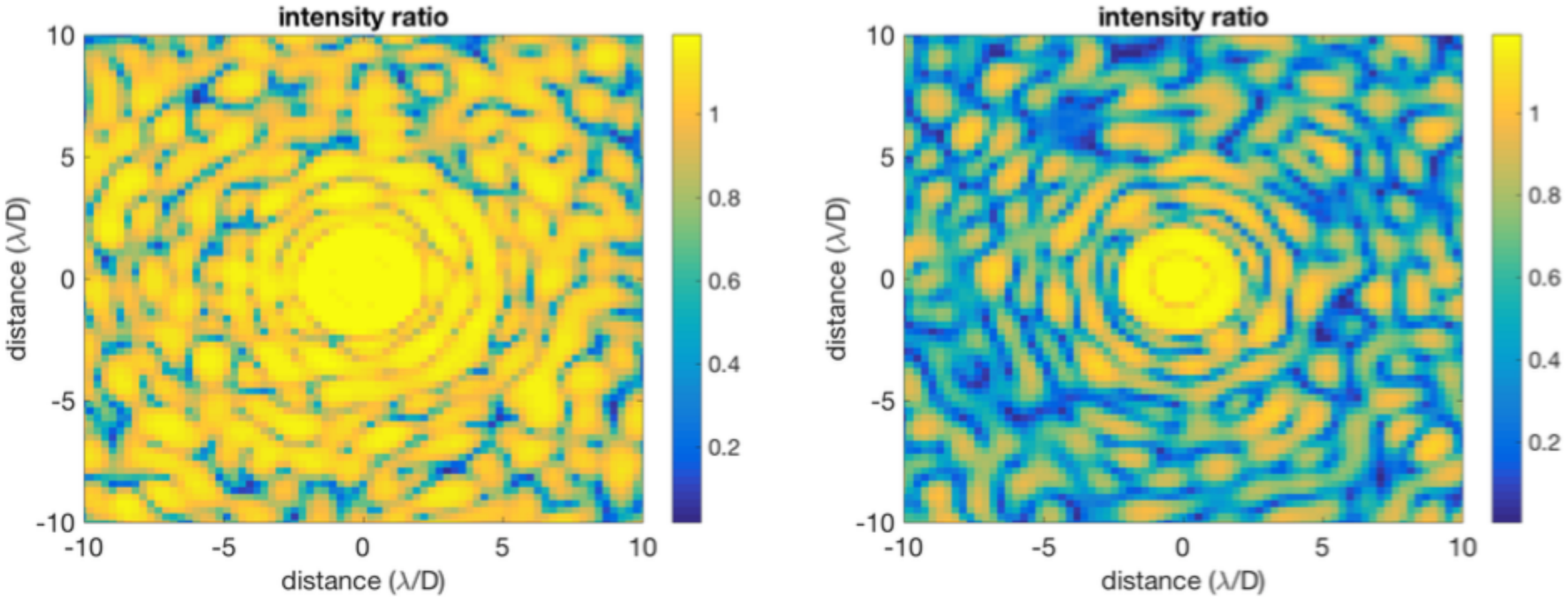


Figure 4. Ratio of the estimated value of the stellar speckle intensity $\hat{I}_0$ (shown in Fig. 3) to the expectation of the true value $\mathcal{E}(I_0)$. Based on 1000 Monte Carlo realizations of $\delta_{\mathrm{r}}(\boldsymbol{r})$, which is the error in the estimate of the residual phase $\hat{\phi}_{\mathrm{r}}(\boldsymbol{r})$. *Left pannel:* correlation length $l_c = 8.5$ cm, *Right pannel:* $l_c = 25$ cm. The difference in the two panels shows how the bias effect depends on the correlation structure of the error in the estimate of the residual phase.

Line plots of $\hat{I}_0(\rho)/\mathcal{E}[I_0(\rho)]$

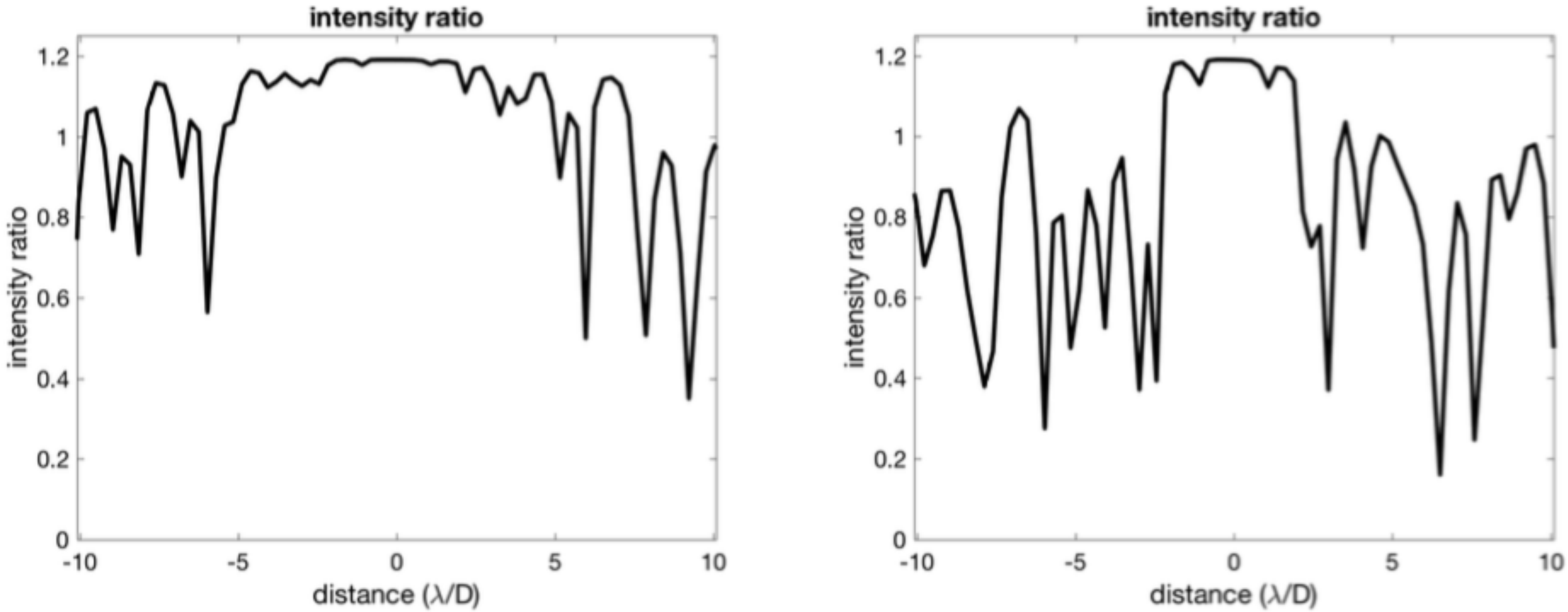


Figure 5. Line plots corresponding to a horizontal cut through the center of the images in Fig. 4. *Left pannel:* correlation length $l_c = 8.5$ cm, *Right pannel:* $l_c = 25$ cm.

What About the Variance?

- We have seen that y and H in the regression eqs. will be biased due their nonlinear dependence on $\phi_{\mathrm{r}}(\boldsymbol{r}, t_i)$.
- But, what about the random part of the error in in y and H, which is a correlated form of noise in these quantities?
- My latest JOSA-A paper provides expressions for the covariance of I_0 and H, in terms 2nd order stats of the δ functions …

Consider the following single-parameter linear regression problem:

$$y_k = h_k x + \epsilon_k \, , \; k = 0, \ldots , K - 1 \, ,$$

where $\{y_k\}$ are observed data points, x is the unknown parameter we want to determine and the $\{h_k\}$ are the true, unknown values of the independent variable (regressor). Now, although we don't know the values of $\{h_k\}$ we do have estimates of these quantities $\{\hat{h}_k\}$ at our disposal. For simplicity assume

$$h_k = \hat{h}_k + \eta_k \, ,$$

where the $\{\eta_k\}$ are the errors in the estimates.

Our problem is

$$\begin{aligned} y_k &= h_k x + \epsilon_k\,, \; k = 0, \ldots, K-1\,, \\ h_k &= \hat{h_k} + \eta_k\,. \end{aligned}$$

Further assume that the errors $\{\eta_k\}$ are independent of $\{h_k\}$. What happens if we simply ignore the errors $\{\eta_k\}$ and treat the estimates $\{\hat{h_k}\}$ like the true values $\{h_k\}$ in least-squares regression? Our estimate of x, $\hat{x}$, is then:

$$\hat{x} = \frac{\frac{1}{K}\sum_k \left(\hat{h_k} - \overline{\hat{h}}\right)\left(y_k - \overline{y}\right)}{\frac{1}{K}\sum_k \left(\hat{h_k} - \overline{\hat{h}}\right)^2}$$

where $\overline{\hat{h}}$ and $\overline{y}$ are the sample means. Does $\hat{x} \to x$ in the large sample limit?

No!

As $M \to \infty$ (large sample limit)

$$\hat{x} \to \frac{\text{cov}(\hat{h_k}, y_k)}{\text{var}(\hat{h_k})} = \frac{x\sigma_h^2}{\sigma_h^2 + \sigma_\eta^2} = x\frac{1}{1 + \sigma_\eta^2/\sigma_h^2},$$

where $\sigma_h^2 = \mathcal{E}[(h_k - \overline{h})^2]$ and $\sigma_\eta^2 = \mathcal{E}[\eta^2]$ are variances. The factor in red is the dilution factor. Thus, in this simple context we can expect $\hat{x}$ to tend to lead to an estimate of x that is too small.

Errors in Variables Modeling

- We cannot know the true value of the residual phase $\phi_{\mathrm{r}}(\boldsymbol{r}, t_i)$, which on which the independent variables (IVs) of the regression problem depend in a nonlinear fashion.
- This nonlinearity creates biases in the estimated values of the IVs.
- Even without this bias, the random component of the error in the IVs can cause biases in the results of the statistical inference, which can be unacceptable.
- This brings us into an active area of statistics research called EiV modeling.

Conclusions

- Differential Imaging (ADI, SDI, etc.) methods have fundamental problems, notably time-variable aberrations.
- They cannot make use of simultaneous ms data streams in the SC and WFS.
- I have presented a promising statistical inference method that simultaneously determines the time-dependent aberrations and the planetary image by leveraging the ms data streams in the WFS.

Conclusions, con’t

- The statistical inference method can incorporate all information sources used by differential imaging methods (e.g., diurnal rotation, multi-wavelengths)
- In principle, it overcomes the fundamental limitations in differential imaging
- However, the statistical inference itself is still in very early development and a number of challenges need to be overcome:
  - Need good knowledge of WFS estimation error
  - Difficult regression issues
  - Size/complexity of computation (Kalman filters)
  - Optimal hardware design

C’est Tout