Focal plane wavefront sensing and control for ground-based imaging


July 9, 2012

SPIE Astronomical Telescopes and Instrumentation
Amsterdam, Netherlands
July 1, 2012 through July 6, 2012
Focal plane wavefront sensing and control for ground-based imaging

Dmitry Savransky\textsuperscript{a}, Bruce A. Macintosh\textsuperscript{a}, Sandrine J. Thomas\textsuperscript{b}, Lisa A. Poyneer\textsuperscript{a}, David W. Palmer\textsuperscript{a}, Robert J. De Rosa\textsuperscript{c}, Markus Hartung\textsuperscript{b}

\textsuperscript{a}Lawrence Livermore National Lab, Livermore, CA USA
\textsuperscript{b}Gemini Observatory, La Serena, Chile
\textsuperscript{c}University of Exeter, Exeter, UK

ABSTRACT

We evaluate the performance of existing wavefront sensing and control techniques, including speckle nulling and electric field conjugation, and discuss their applicability to high-contrast imaging spectrographs such as the Gemini Planet Imager (GPI). These techniques can be highly useful in correcting system phase errors, and can potentially improve instrument operating efficiency by working in conjunction with the dedicated adaptive optics (AO) wavefront sensor. We discuss the specifics of our implementation of speckle suppression for GPI and present lab demonstrations with average contrast improvements from $5.7 \times 10^{-6}$ to $1.03 \times 10^{-6}$.

Keywords: AO, GPI, high-contrast imaging, focal plane wavefront control

1. INTRODUCTION

The discovery of planets orbiting HR8799 has demonstrated that direct imaging of exoplanets from the ground is not only possible, but can produce science-rich data.\textsuperscript{1} Direct detection of planets can vastly increase our knowledge of the exoplanet population of our galaxy by probing a section of the mass-period space that is difficult to capture via other detection methods such as transit photometry or doppler spectroscopy. Multiple dedicated instruments are currently under development with this specific goal in mind, including the Gemini Planet Imager (GPI). GPI, which will operate at the Gemini South observatory, will achieve deeper contrast levels than any instruments currently in operation, and will thus be able to directly detect young, Jovian and super-Jovian objects.\textsuperscript{2}

In order to achieve the required contrast levels (down to $10^{-7}$), GPI will have to correct not only for aberrations caused by the moving atmosphere, but also for imperfections in the system optics and the effects of sensor noise. To do this, GPI was designed with an adaptive optics (AO) system employing a Shack-Hartmann wavefront sensor, and a calibration interferometer to provide updated system corrections on a slower cycle than the primary AO. While the AO system is sufficient for correcting for atmospheric turbulence, it cannot correct for any optical errors that are not in its light path. The calibration interferometer is designed to measure the post-coronagraph wavefront, but generates relative phases and so needs an input reference to correct the full non-common path aberration. We therefore need another method of measuring residual noise at the focal plane and generating a set of references for the CAL and AO system. We do this by using the science instrument and the AO control optics as a focal plane wavefront sensor and apply an iterative noise suppression scheme. In this paper we describe the theory behind this approach, the specifics of its implementation for GPI, and present the results of initial lab experiments.

Send correspondence to Dmitry Savransky savransky1@llnl.gov
2. GPI OVERVIEW AND OPERATION

The Gemini Planet Imager consists of three main subsystems (illustrated in Figure 1) and a diffraction control system, whose components are spread throughout the instrument. Light from the telescope pupil, or alternatively an Artificial Source Unit (ASU), first passes through the AO system which is responsible for correcting phase errors introduced into the wavefront by varying atmospheric conditions and internal surface errors. The control portion of the AO system consists of a MEMS deformable mirror (DM), a piezo-electric DM and a tip-tilt stage. Wavefront sensing is done with a visible light CCD behind a 43x43 lenslet array. A variable-size spatial filter is included in this leg of the beam path to remove aliasing errors.\(^3\)

The remaining IR light passes through a pupil plane apodizer, forming the first stage of the apodized-pupil lyot coronagraph (APLC) that serves as the primary diffraction control mechanism\(^4\) and continues to the calibration interferometer (CAL). The CAL system interferes on-axis light passing through the focal plane occulting mask (the second component of the APLC) with a portion of the reflected off-axis light in order to reconstruct the post-coronagraph wavefront and generate reference centroid offsets that are then incorporated in the main AO loop.\(^5\)

Finally the remaining IR light enters the science instrument, an integral field spectrograph (IFS), where it is passed through a pupil plane Lyot stop (the final APLC component) and then through a focal plane lenslet array and disperser, after which it is imaged with a HAWAII 2RG detector.\(^6\) The IFS assembly also includes a deployable camera that can be used for imaging the IFS entrance pupil. Note that GPI is configurable to operate in multiple different modes, with several different diffraction control mechanisms available, and multiple configurations possible for each one. The configuration described here is the default for acquiring spectral data, and was used for all of the experiments described in this paper.

![GPI schematic](image-url)

Each of the principal subsystems is controlled via a dedicated computer, with a top level computer (TLC) used for commanding actions. All moveable components internal to GPI are mounted on motorized stages or filter wheels, and so all configuration changes and alignments can be performed without physically accessing the system. These four machines implement a physical shared memory block that stores information about the state of every component, and can be directly queried by external agents with sufficient access. Data collected by the IFS (and other internal imaging systems) is stored on a networked volume hosted by a separate, fifth, computer.
This machine also runs the GPI data reduction pipeline, which performs standard astronomical reduction on the raw IFS images, and reconstructs dispersed images into 3-dimensional data cubes (where wavelength is the \(z\) axis).\(^7\) The standard operating procedure for acquiring the data used for the experiments described below is as follows:

1. An observation is configured and commanded via an API to the TLC instrument sequencer.

2. The TLC passes the command to the IFS computer, which carries out the observation and writes the data, in FITS format, to the network mounted volume.

3. The data reduction pipeline detects a new data file and, using the FITS header keywords to establish the configuration the data was generated under, runs through a standard reduction.

4. The reduced data is written out as a new FITS file on the same network mounted volume.

5. The new reduced data file is detected and read in by the machine commanding the original observation.

The experiments described here were all performed with the AO system running in closed loop, but without any updated reference centroids from the CAL system. The phase offsets calculated from the focal plane wavefront estimation were sent to the AO controller in the form of additional reference centroids, and were thus applied as part of the closed loop, removing any errors that could potentially be introduced by an open loop model of the AO control surfaces.

### 3. SPECKLE SUPPRESSION

The idea of using a DM to cancel out semi-static noise realizations in science images, and thereby create a lower noise ‘dark hole’ region, was first introduced in Ref. 8, and further developed in Refs. 9 and 10. Treating the imaging system as a linear operator \(F\) such that,

\[
F : E_o(u, v) \rightarrow E_i(x, y),
\]

where \(E_o(u, v)\) is the electric field transmitted by the object and \(E_i(x, y)\) is the image electric field (with image coordinates \(x, y\)), it can be shown that \(F\) is a scaled Fourier transform of the input field for the APLC (and a variety of other Lyot-type coronagraphs).\(^11\) The electric field at the DM (which is assumed conjugate to the apodized pupil) is given by

\[
E_o(u, v) = A(u, v)e^{i\phi_\lambda(u, v)}e^{2\pi i\psi(u, v)/\lambda},
\]

where \(A\) is the apodization function of the pupil, \(\phi_\lambda\) is the (wavelength \(\lambda\) dependent) complex pupil aberration function and \(\psi\) is the DM phase function. Scattering from each of the sinusoidal components of \(\phi_\lambda\) generates an intensity contribution in the final image plane, and these are commonly referred to as ‘speckles’.\(^12\) Note that unlike \(\phi_\lambda\), \(\psi\) is purely real, as the DM operates only in phase, and a single DM cannot be used to correct amplitude aberrations over the entire image plane. It is possible, however, to correct for both amplitude and phase aberrations in half of the image plane with a single DM, creating a ‘half dark hole’, or to use two DMs in series to correct all aberrations in the full image.\(^13\) As the DM has a finite number of actuators and finite physical extent, there is a maximum spatial frequency that can be corrected, and therefore a maximum size to the speckle-cleared region that can be produced.

The mapping between spatial frequency \((k, l)\) and image plane pixel location \((x, y)\) is given by a single linear scale factor, \(d\), that is found empirically by placing a ‘waffle’ pattern on the DM—driving neighboring actuators in opposite directions to produce speckle at the highest controllable spatial frequency—and then finding the pixel locations of the four generated intensity regions (see Figure 2). These same spots are also used to center and de-rotate the image. The pixel distance of these spots from the image center, scaled by the maximum spatial frequency \(N\) (equal to half the number of actuators along the DM), yields the value of \(d\). Note that this mapping is symmetric about the center of the image plane, so that every spatial frequency corresponds to two locations in the image. Thus, a pure sinusoid applied to the DM generates a speckle at the location determined by the sinusoid’s spatial frequency (period and rotation), with intensity determined by the sinusoid’s amplitude, as
Figure 2. (Left) GPI focal plane coronagraphic image with flat DM. The bright boxed regions are astrometric calibration spots generated by a grid pattern on the apodizer. The dashed box indicates the high contrast region created by the diffraction control system and the dashed circle shows the extent of the focal plane mask. (Right) Image with waffle pattern applied to DM. The resulting speckles are circled.

Figure 3. (Left) A pure sinusoid ($k = l = 9$) applied to the DM producing (Right) speckles at the corresponding location in the image (circled). The summed intensities in these regions are referred to as $I_1(k, l)$ and $I_2(k, l)$.

Illustrated in Figure 3. We label the total intensity associated with each spatial frequency as $I_1(k, l)$ and $I_2(k, l)$ on each side of the image plane, respectively.

The task of speckle suppression is to find a DM surface shape (or equivalently a value for $\psi$) that will cancel as much of the phase aberration across some subsection of the pupil as possible. In the most basic approach, speckle nulling, the phase of specific speckles is found by iteratively applying different phases via the DM and tracking speckle intensity variations. After finding the phase that minimizes the intensity of a given speckle, we save the corresponding DM phase setting and proceed to the next highest-intensity speckle.

The problem as stated in Equation (2) is infinite-dimensional with no exact solution. To map to the controllable space and to make the problem tractable, $\psi$ is typically approximated as a finite sum of some basis functions, such as the device-specific influence functions (the shapes assumed by the DM when voltage is applied.
In this case, the weighting coefficients of the solution map directly to the desired DM voltage settings, so DM commands are generated automatically. Unfortunately, the influence functions are often difficult to derive, and this approach assumes that superposition holds everywhere, which is sometimes false, depending on the specific DM technology employed. We avoid this problem by applying the desired phases in closed loop, and using a forward model of the wavefront sensor, rather than the DM, to create the desired set-point. This means that our calculated DM shapes are parameterized by spatial frequency and have physical units of microns.

A refinement of speckle nulling, known as energy minimization, reformulates the problem to one of minimizing the total energy in the dark hole rather than finding conjugate surfaces for individual speckles. The total energy in a region $S$ is given by:

$$E_S = \int_S F\{E_0\} (F\{E_0\})^* \, du \, dv$$

(3)

where $(\cdot)^*$ represents the complex conjugate operator. This expression can be minimized in a variety of ways, including gradient hill-climbers, or Gauss-Newton iteration. This approach addresses the problem of energy beyond the maximum controllable spatial frequency, which is aliased in the dark hole and cannot be corrected for with speckle nulling. Unfortunately, it also represents a very difficult search problem, and is not guaranteed to achieve the desired contrast in finite time. Two other schemes have been developed to address this.

The first, electric field conjugation (EFC), discretizes the field into a finite number of points in the image plane and solves for the desired DM surface shape by linearizing the DM phase contribution, such that:

$$e^{2\pi i \psi / \lambda} \approx e^{2\pi i \psi_{k-1} / \lambda} \left(1 + \frac{2\pi i}{\lambda} \delta \psi_k\right),$$

(4)

where $\psi_{k-1}$ is the previous phase estimate in an iterative scheme, and $\delta \psi_k$ is a (small) correction to this estimate. This formalism transforms the search problem into a simpler root-finding problem, although it is also not always numerically tractable. The second approach, stroke minimization, instead seeks to minimize the coefficients of the DM surface expansion rather than the electric field, which is equivalent to minimizing the DM commands (or stroke). The minimization is performed under the constraint of a maximum contrast value to ensure that the dark hole produced meets contrast requirements. In this method, Equation (3) is also expanded about the DM surface coefficients, so that both the minimization and constraint functions are linear. While these methods will most likely be adapted for GPI in the future, for our initial experiments with speckle suppression we decided to implement speckle nulling since previous work with these algorithms had shown that nulling could produce good results in GPI’s target contrast regimes with a much lower computational burden.

4. WAVEFRONT ESTIMATION

All of these approaches require an estimate of the phase of the electric field. Because the aberrations we are trying to fix are exactly those caused by non-common path errors and sensor noise, the estimation must be performed at the focal plane. One way to do so is to place random phase terms on the DM and measure the resulting intensity variation in the focal plane images. The mapping between DM input phase and image plane intensity is found in a similar fashion, by putting sine waves of known amplitude on the DM and measuring the resulting intensity. Because it would take too long to perform this test at every one of the $N \times N$ controllable spatial frequencies, instead we sample only the $N$ spatial frequencies such that $k = l$. These are fit to the Gaussian function

$$a + b + e^{(r/c)^2},$$

(5)

where $r$ is the ‘radial spatial frequency’, $r = \sqrt{k^2 + l^2}$. The fit is then used to interpolate intensity values for the remaining modes, and these are used for mapping DM phase to intensity. Figure 4 shows the measured intensities resulting from applying sinusoids to the DM, and the resulting fit. As the amplitude of the input sine wave was kept constant, the variations in resulting intensity are a product of the forward modeling of the wavefront sensor used in generating the actual DM commands. The fit essentially flattens out the model, ensuring that amplitude maps to intensity correctly at all spatial frequencies.
Figure 4. Points marked with an ‘x’ are intensity measurements at spatial frequencies $k = l$. The solid line with triangle markers is the fit to these points. Note that values for $k = l < 4$ are not used in the fit as they represent spatial frequencies that fall behind the focal plane mask.

At each iteration of the nulling algorithm 8 phases probes are applied to the DM and the phase of specific speckles is found by fitting to the resultant intensities at the speckle locations. The probe shapes are generated by summing sine waves corresponding to the spatial frequency of speckles we wish to remove, with phase offsets uniformly distributed in $[0, 2\pi]$ (and an initial random offset for every speckle), as shown in Figure 5. As we are primarily interested in the phase that minimizes a given speckle, we can simply fit a sinusoid to the normalized summed intensities ($\langle I_1 + I_2 \rangle$) corresponding to the spatial frequency of the speckle we wish to cancel, as in Figure 6. Alternatively, we can exploit the symmetry of the focal plane image to try to fit both phase and amplitude. We can fit sinusoids independently to $I_1 + I_2$ and $I_1 - I_2$. As any intensity variations will be due to amplitude aberrations, the difference in the initial phase fits can be used to establish a grid of amplitude and phases that is then searched for the best-fit amplitude and phase. This procedure is demonstrated in Figure 7.
Figure 6. The normalized measured intensities are represented by the solid line with diamond markers, while the best fit sinusoid is given by the dashed line. The plus marker indicates the best fit phase.

Figure 7. The solid black line with diamond markers represents the summed measured intensities \( (I_1 + I_2) \) and the black dashed lines are \( I_1 \) and \( I_2 \), respectively. The gray dashed lines represent best fits to \( I_1 + I_2 \) and \( I_1 - I_2 \) and the black dash-dotted line represents the total amplitude and phase fit to the data. The plus sign denotes the best found amplitude (mapped to intensity) and phase.

This is a variation of the standard DM diversity algorithm where a series of DM shapes \( \{\psi_k\} \) are applied to produce a set of intensities \( \{I_k\} \). Approximating the electric field of the \( k \)th image as

\[
E_k \approx \mathcal{F} \{A e^{\phi_{\lambda}}\} + \mathcal{F} \{e^{2\pi i \psi_k/\lambda} - 1\},
\]

and assuming that the first intensity, \( I_0 \), corresponds to a flat DM setting, we can write

\[
I_k - I_0 = |\mathcal{F} \{e^{2\pi i \psi_k/\lambda} - 1\}|^2 + (\mathcal{F} \{A e^{\phi_{\lambda}}\})^* \mathcal{F} \{e^{2\pi i \psi_k/\lambda} - 1\} + \mathcal{F} \{A e^{\phi_{\lambda}}\} \left(\mathcal{F} \{e^{2\pi i \psi_k/\lambda} - 1\}\right)^*.
\]
Using the whole image set to generate an overdetermined system, we then solve for the real and imaginary parts of the focal plane electric field as

\[
\begin{bmatrix}
\Re(\mathcal{F}\{Ae^{\phi_x}\}) \\
\Im(\mathcal{F}\{Ae^{\phi_x}\})
\end{bmatrix} = \frac{1}{2} \begin{bmatrix}
\Re(\mathcal{F}\{e^{2\pi i \psi_1/\lambda} - 1\}) & \Im(\mathcal{F}\{e^{2\pi i \psi_1/\lambda} - 1\}) \\
\vdots & \vdots \\
\Re(\mathcal{F}\{e^{2\pi i \psi_k/\lambda} - 1\}) & \Im(\mathcal{F}\{e^{2\pi i \psi_k/\lambda} - 1\})
\end{bmatrix}^\dagger \begin{bmatrix}
I_1 - I_0 - |\mathcal{F}\{e^{2\pi i \psi_1/\lambda} - 1\}|^2 \\
I_k - I_0 - |\mathcal{F}\{e^{2\pi i \psi_k/\lambda} - 1\}|^2
\end{bmatrix},
\] (8)

where the \(\dagger\) represents the left-pseudoinverse operator.\(^{15}\)

The estimation and correction is run in closed loop using proportional-integral control. Note that this control loop is entirely separate from the AO control, running at a much slower frequency, and only interacting with the AO loop by providing reference offsets. The basic nulling algorithm is thus:

1. Take a reference image with a flat (or previous best) DM setting.
2. Identify the highest intensity speckles in the desired dark hole region and find the corresponding spatial frequency.
3. Select the highest intensity controllable speckles up to the maximum number of modes set per iteration, discarding neighboring spatial frequencies to avoid cross-talk.
4. Take 8 more images with varying phase settings supplied to the AO loop via reference centroids.
5. For each speckle, fit the phase. The sum of these fits is the new reference offset.
6. Apply the control law to the new command and add to the previous best setting.

5. RESULTS

![Figure 8. (Left) Initial focal plane image. (Right) Focal plane image after four iterations of speckle nulling.](image)

As an initial test of focal plane wavefront control for GPI, we applied the speckle-nulling algorithm, estimating phases only, to correct for speckles in \(r \in [10, 17]\) (approximately 0.35 to 0.71 as), correcting a maximum of 100 modes at each iteration. The control loop was run with a gain of 0.5 and an integrator leak of 0.001. Images were taken in H band, using a 1.5 s integration time, and the phase estimates were all performed using the 1.5 to 1.512 \(\mu\)m channel of the reconstructed image cube. Figure 8 shows the difference in focal plane intensities after three iterations of the algorithm, and Figure 9 shows the corresponding improvement in image contrast. Contrast here...
is defined as the $5\sigma$ contrast, i.e., five times the standard deviation of the intensity in an annulus of radius equal to the angular separation. The average contrast in the dark hole region improved from $1.1\times10^{-5}$ to $5.5\times10^{-6}$.

The majority of the contrast improvement in this experiment was due to the removal of the vertical structure seen in Figure 8 that was introduced by the DM itself, operating with an old reference centroid. It should also be noted that these experiments were performed without any external aberrations. However, while atmospheric turbulence will add additional noise, high-contrast imaging is primarily limited by internal quasi-static errors and so on-sky performance is bounded by the system performance in the absence of aberrations.

The experiment was repeated after new reference centroids had been generated for the AO loop (creating a much better starting contrast), with the same parameters, except that the dark hole was redefined to $r \in [10, 17]$ (approximately 0.35 to 0.6 as), and the integration time was increased to 15 s to compensate for the decreased intensity in the dark hole. Figure 10 shows the difference in focal plane intensities after four iterations of the algorithm, and Figure 11 shows the corresponding improvement in image contrast. The average contrast in the dark hole region improved from $5.7\times10^{-6}$ to $1.03\times10^{-6}$, with a minimum contrast of $7\times10^{-7}$.

We found that after four iterations, the average contrast in the dark hole did not significantly improve, although contrast values at particular radial locations continued to drop. This could indicate that most of the remaining speckle was due to amplitude aberration, or that we were simply approaching the photon noise limit.
in our measurements and improved performance could still be achieved with increased integration times. The phase settings derived in this experiment were reapplied to the DM after several hours had passed, and produced the same final average contrast, indicating very good stability throughout the system.

6. CONCLUSIONS

We have demonstrated iterative speckle suppression for GPI using the speckle nulling algorithm, and have achieved average contrast improvements of up to a factor of 5.5 in only 3 to 4 iterations of the algorithm. We have also demonstrated the application of DM phase offsets while running the adaptive optics system in closed loop, and the utility of employing a forward wavefront sensor model over traditional DM influence functions. Furthermore, the resulting PSFs were relatively stable and achromatic, indicating that the noise can be further attenuated by ADI and SDI techniques.16

Future steps will include testing other speckle suppression algorithms such as EFC, and improving the efficiency of our wavefront estimation. As we have demonstrated the ability to fit both the amplitude and phase of the electric field from the image plane intensity, one possible refinement is to only attempt to correct phase aberrations and not waste control effort on amplitude speckles. At the same time, in order to deepen the contrast level in the dark hole and extend the inner working angle inwards, we will need to correct aberrations at low-frequency modes, which may require coupling the iterative speckle suppression algorithm with other methods of image sharpening.
ACKNOWLEDGMENTS

The authors would like to thank the entire GPI collaboration, and especially the GPI integration and testing team for their support and dedication. Portions of this work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. This document number is LLNL-CONF-563503.

REFERENCES