

Data Post-Processing and Algorithm Development for the WFIRST Coronagraph: FY2015 Report

Marie Ygouf, Neil T. Zimmerman,
Laurent Pueyo, Marshall D. Perrin, Rémi Soummer^(PI)
Space Telescope Science Institute
3700 San Martin Dr, Baltimore, MD 21218



Abstract

The WFIRST Coronagraph Instrument (CGI) will require advanced data processing techniques to reach its scientific goals. No astronomical instruments have operated yet in the planet-to-star contrast regime of the proposed CGI (between 10^{-9} and 10^{-8}). Therefore, the evaluation of post-processing algorithms must rely on trials applied to a mixture of simulations and laboratory data. The Jet Propulsion Lab awarded a subcontract to the Space Telescope Science Institute to carry out such investigations, on images produced by end-to-end flight models and testbed measurements. In this report, we summarize the activities completed during FY2015 in fulfillment of that contract, comprised of three main tasks:

1. Post-processing algorithm development and testing for simulated broadband and multi-wavelength images of the WFIRST-AFTA coronagraph.
2. Post-processing algorithm development and testing for broadband and multi-wavelength coronagraphic data obtained at JPL's High-Contrast Imaging Testbed (HCIT), under static aberration conditions.
3. Algorithm developments for the analysis of integral field spectrograph data.

On both simulated and laboratory images, we find the KLIP algorithm is significantly more robust than classical PSF subtraction against temporal variations present in the wavefront aberrations. On the simulated images acquired in the broadband/HLC mode of the CGI, we achieve contrast gains of 10 to 30, enabling the detection of injected planets with contrasts as low as 0.8×10^{-9} . For simulated IFS data acquired in the characterization/SPC mode of the CGI, the KLIP algorithm in combination with forward modeling enables an unbiased retrieval of a Jupiter-like spectrum at a contrast of 4.2×10^{-9} . Our analysis of multi-wavelength laboratory testbed data suggests that KLIP will mitigate the challenges of broadband wavefront control for the IFS: KLIP naturally compensates for the higher speckle noise floor at the short- and long-wavelength ends of the filter bandpass, flattening the wavelength-dependence of the final, post-processed sensitivity.

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 4 |
| 2 | Simulated coronagraph data | 5 |
| 2.1 | Data processing | 6 |
| 2.1.1 | Reduction techniques | 6 |
| 2.1.2 | Forward modeling | 7 |
| 2.2 | Results and analysis | 8 |
| 2.2.1 | Detection in the imaging channel with an HLC | 8 |
| 2.2.2 | Characterization in the IFS channel with an SPC | 11 |
| 3 | Laboratory coronagraph data | 14 |
| 3.1 | Broadband coronagraphic data | 14 |
| 3.1.1 | Data set description, selection and fake planet simulation | 14 |
| 3.1.2 | Reductions and analysis | 16 |
| 3.2 | Multi-wavelength coronagraphic data | 18 |
| 3.2.1 | Data set description | 18 |
| 3.2.2 | Detector background noise | 20 |
| 3.2.3 | Frame arrangement for reference differential imaging | 21 |
| 3.2.4 | Contiguous sub-sequence RDI | 24 |
| 3.2.5 | Interleaved sub-sequence RDI | 28 |
| 3.2.6 | RDI using probe and regularization images together | 30 |
| 3.3 | Summary of multi-wavelength RDI results | 30 |
| 4 | Algorithm developments for IFS data analysis | 32 |
| 5 | Conclusions and Perspectives | 35 |
| 6 | Acknowledgments | 38 |
| 7 | Acronym list | 39 |

1 Introduction

The development of algorithms to process high-contrast images has attracted strong interest from observers in recent years. The *principal component analysis* (PCA) approach to PSF subtraction is now widely used by the exoplanet imaging community to improve the scientific yield of data acquired both on the ground and in space. The algorithm known as Karhunen-Loève Image Projection (KLIP) is a specific implementation of PCA [1]. Our research group developed KLIP in the context of the ALICE project [2, 3], conducting a comprehensive reprocessing of the entire HST NICMOS coronagraphic archive. KLIP enables significant contrast improvements over classical PSF subtraction when applied to HST data, with results often approaching the fundamental photon noise limits. KLIP has also been applied with great success to NICMOS and WFC3 data of the HR 8799 planetary system [1, 4]. Based on its performance on HST-NICMOS data, the architecture of ALICE pipeline is the baseline for JWST coronagraph data analysis. Since PCA-based methods such as KLIP are today’s standard for high-contrast imaging pipelines, there is a strong motivation to study their applicability to the WFIRST CGI.

In addition to KLIP and similar PCA-based methods such as PynPoint [5] and S4 [6], a number of related post-processing algorithms for high-contrast imaging have been published in the last decade, including LOCI [7], SOSIE [8], and methods based on the Hotelling Observer [9]. All of these approaches share a common foundation in that they are based on calculations of covariances between large imaging data sets [10]. As a consequence, the ultimate contrast gains achieved using these various algorithms are comparable, provided that the choice of free parameters such as optimization zones and selection of input images, specific to each implementations, are chosen adequately. Therefore, the conclusions we draw in this report, based on the performance of the KLIP algorithm, apply to a broad family of post-processing algorithms.

The general application of PCA to the PSF subtraction problem can be summarized as follows: first, a library of reference images is decomposed into a set of principal components that capture the temporal and/or spatial variations of the PSF. Then, the PSF of the target star in a scientific image is estimated with a linear combination of the principal components, optimized in the least-squares sense. By subtracting this PSF estimate, the majority of the star’s light, including its associated speckle noise pattern, is removed from the scientific image, enhancing the visibility of exoplanets and circumstellar debris structures.

The specific PSF subtraction problem we study here is *reference differential imaging* (RDI). RDI assumes that for a given coronagraph observation of a science target, a library of reference coronagraphic images from one or more different stars can be used to estimate the science target coronagraphic PSF. This relies on the assumption that a hypothetical astrophysical signal in the target image is unlikely to be present or mimicked in the reference images. Classical PSF subtraction, sometimes called coronagraphic PSF template subtraction (CPSFTS) [11], uses a single reference

star and can be seen as a simplest case for RDI. See e.g. [12] for a general review of techniques in this area.

Crucial issues are 1) to understand to what extent the stability and the amount of wavefront aberrations will impact post-processing for WFIRST and 2) to develop spectral extraction techniques, including compensation for spectral biases induced by the PSF subtraction. After a brief description of currently available simulated data (section 2) and a summary of the reduction techniques used (section 2.1), we address the following questions (section 2.2):

- How does the temporal stability of aberrations impact the achievable contrast after post-processing?
- What will be the impact of the total amount of aberrations?
- What could be the impact on post-processing of instrumental control strategies such as the use of a Low Order Wavefront Sensor (LOWFS)?
- What error terms can impair the quality of extracted planetary spectra?

The results described in this report have been presented at several conferences: the 225th meeting of the AAS in Seattle, WA [13], the 2015 “In the Spirit of Bernard Lyot” conference in Montréal, Canada [14], the 2015 SPIE conference in San Diego, CA [15], and the 227th meeting of the AAS in Kissimmee, FL [16, 17].

2 Simulated coronagraph data

We considered noise-free data sets simulated by the Jet Propulsion Laboratory (JPL) using a diffractive model for the WFIRST coronagraph [18], with temporal variability derived from thermal models generated by the Goddard Space Flight Center (GSFC). We considered first and third iterations of thermal models with associated Observation Scenarios (OS), labelled OS1 and OS3 by the WFIRST project. OS1 corresponds to a fall 2014 model for WFIRST in a geosynchronous orbit; for this data set, we artificially doubled and quadrupled the coma to study the robustness of post-processing. OS3 is a spring 2015 design for a quasi-L2 orbit (i.e. without Earth influence) for which we studied the performance with and without a LOWFSC subsystem. These datasets each simulate successive observations of two stars: (1) first a bright star targeted for dark hole generation (β UMa; A1IV, $V = 2.37$); (2) this is followed by observations of the science target (47 UMa; G1V, $V = 5.04$).

Simulated data were provided for each of the baseline WFIRST coronagraphs: the HLC in the broadband imaging channel for both observing scenarios OS1 and OS3, and the SPC in the integral field spectrograph channel (IFS), for the OS3 scenario only. There are 88 images for β UMa and 321

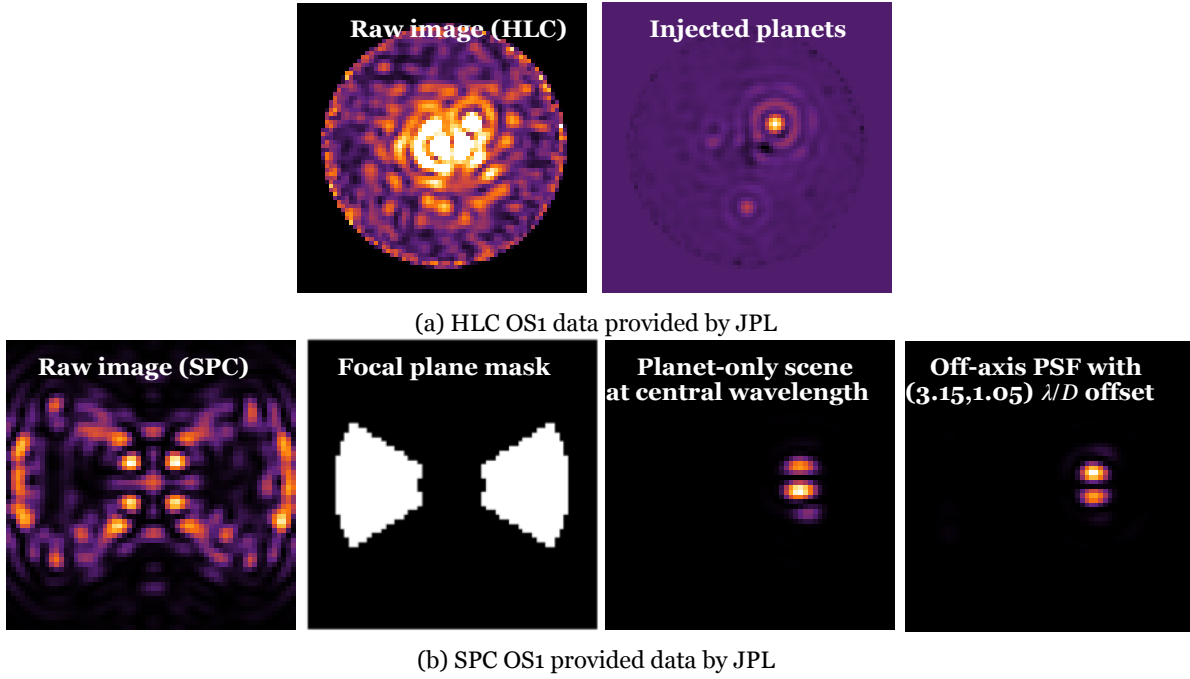


Figure 1: **HLC and SPC OS1 simulated data sets provided by JPL.** The simulated observation scenario 1 (OS1) correspond to an observatory thermal model assuming a geosynchronous orbit. (a) HLC OS1 dataset from left to right: raw long exposure image of the science target (47 UMa) and injected planets with contrasts of 5.7×10^{-9} , 10^{-9} and 0.8×10^{-9} with respective separations of $3.1 \lambda/D$, $2.9 \lambda/D$ and $5.3 \lambda/D$. (b) SPC OS1 dataset, from left to right: raw long exposure image of the science target (47 UMa), bow-tie mask (focal plane mask sampled at the resolution of the final image), planet-only scene at central wavelength (contrast of 4.2×10^{-10}), and an unocculted PSF with offsets of $3.15 \lambda/D$ and $1.05 \lambda/D$ and X and Y respectively at $0.8 \mu\text{m}$ with only two visible lobes, the third being hidden by the bow-tie mask.

for 47 UMa for the OS1 scenario, and 8 images for β UMa and 17 for 47 UMa for the OS3 scenario. The number of images in the reference datasets is significant because it sets the maximal number of eigenmodes available in a PCA analysis. Synthetic planets were added to the 47 UMa images in all data sets. Figure 1 shows example images from two simulated data sets: HLC (Figure 1a) and SPC (Figure 1b). Note the particular shape of the off-axis PSF in the image simulated with the SPC. Depending on the position of the planet in the field, its shape is more or less truncated by the bow-tie mask in the focal plane, which makes the spectrum extraction more challenging. For more details on these simulations, refer to the WFIRST Coronagraph special issue of the Journal of Astronomical Telescopes, Instruments, and Systems (JATIS) [19].

2.1 Data processing

2.1.1 Reduction techniques

We applied classical PSF subtraction and KLIP to these simulated data sets. In this case, the classical PSF subtraction consists of a simple subtraction of the collapsed and normalized reference data

from the normalized long exposure raw image sequence. Before reducing the data with KLIP, we subtract the average value of the science and reference data sequences so that they have zero spatial mean. We collapse the science data sequence (47 UMa) to obtain a long exposure raw image. We then apply the KLIP algorithm, which 1) decomposes the reference data into principal components or KL modes; 2) projects the long exposure raw image into these modes to create a synthetic reference; and 3) subtracts the synthetic reference from the target. In the case of multi-wavelength data cubes from the IFS channel, we perform the same process for each wavelength independently in such a way that no diversity from spectral differential imaging (SDI) is used.

The number of modes K that we use to create the synthetic reference affects the reduction performance. Past experience has shown that observing the changing appearance of a given point source candidate with varying number of KL modes is a useful (though imperfect) diagnostic of whether that candidate is a likely exoplanet or a persistently “stubborn” residual speckle. Speckle subtraction increases with the number of modes but an increasing part of the signal is removed as well. This arises because the principle component vectors in general will have some nonzero dot product with an off-axis PSF planet signal. The loss of planet flux due to this effect is termed “over-subtraction bias” or “algorithm throughput”.¹ Therefore, there is a trade-off to optimize the detection. This trade-off depends on several parameters, including the kind of astrophysical signal, its flux and separation, and the stability of instrumental aberrations during the observation. As a result, it is not possible to derive a general optimal value for the number of KL modes. Rather, the algorithm throughput must be calibrated through a rigorous quantitative analysis of the PSF subtraction process in order to yield unbiased spectrophotometry of planets.

2.1.2 Forward modeling

We therefore use forward modeling [1, 20, 21] to estimate the photometry at each wavelength and extract the exoplanet spectrum from the SPC OS1 distribution. In ref. [1], Soummer et al. described an optimal way to estimate an exoplanet’s photometry and astrometry using KLIP. This “forward modeling” process consists in propagating a template of the off-axis PSF through the reduction algorithm using the same set of KL eigenmodes in order to calibrate the over-subtraction effect. The result is a calculated flux ratio as a function of field position and number of KL modes, which can then be applied as a correction factor to detected planet fluxes. For IFS data, this calculation must be repeated for each wavelength channel.

¹Over-subtraction bias is closely related to but distinct from self-subtraction bias which arises in ADI-style PSF “roll” subtractions. Self-subtraction is caused by some fractional contribution of the science signal itself in the reference library, albeit at a different position angle. However even in cases of pure RDI where there is no planetary signal in the reference library, over-subtraction bias will remain present. While the two biases have slightly different root causes, the methods for calibrating their effects on the planet flux are similar.

2.2 Results and analysis

2.2.1 Detection in the imaging channel with an HLC

We compared the results of our reductions for the two simulated observing scenarios OS1 (Figure 2) and OS3 (Figure 3) by plotting on the same dynamic range: the long exposure raw image; the classical PSF subtraction reduction; the KLIP reduction with 4 modes; and the KLIP reduction with the maximal number of modes for OS1 (88 modes) and OS3 (8 modes). All the reductions show an improvement in contrast compared to the raw image by at least a factor of 2 and up to a factor of 6 locally, even with increasing amounts of coma or without LOWFSC.

The level of detection for OS1 is quite similar in the three different reductions, with two of the three synthetic planets detected (Figure 2.a). Recall that the amount of coma was artificially increased in copies of the OS1 data set. Based on this, Figure 2 gives insight into the robustness of post processing to increasing low-order aberrations, and demonstrates that increasing numbers of KL modes can be required in situations with higher amplitude WFE variations. Even though the thermal model used for the simulations is the same in the three OS1 data sets, KLIP with $K = 88$ modes does a better job than the classical PSF subtraction at cleaning the residual speckles when the amount of coma increases. The case of coma multiplied by 4 (Figure 2.c) shows this most clearly, as the KLIP reduction with 88 modes is the only one that enables the detection of two of the three synthetic planets. KLIP is more robust than the classical PSF subtraction to increasing amounts of aberrations, enabling very similar detections to the case with lower aberrations represented in Figure 2.a.

In the case of OS3, even though the number of reference images used to build the reference library is much smaller (8) than in the OS1 case (88), KLIP still enables detecting all three synthetic planets. In contrast only two planets are seen with the classical PSF subtraction (Figure 3.a). As has often been seen in other contexts such as HST NICMOS data, classical subtraction leaves residual speckles at small separation angles, which KLIP is more effective in removing. These residual speckles can be traced to temporal wavefront variations, mostly on relatively low spatial frequencies, which the classical subtraction does not account for. These are the spatial frequencies that the LOWFS aims to sense and control.

Figure 3 then demonstrates the impact of the LOWFSC on post-processing. The stabilization of aberrations from the LOWFS improves the performance of all PSF subtractions, and decreases the differences between the two reduction techniques. Without LOWFSC, the detection level is significantly better with KLIP (3 planets detected with 8 modes) than with classical PSF subtraction (2 planets detected). With the LOWFS, the classical subtraction performs nearly as well as KLIP, with both methods providing enough improvement compared to the raw data to detect all three simulated planets. Indeed, if we look at the contrast gain (that is, the ratio of the raw to final contrasts) as a function of separation for the cases with and without LOWFSC (Figure 4), we can

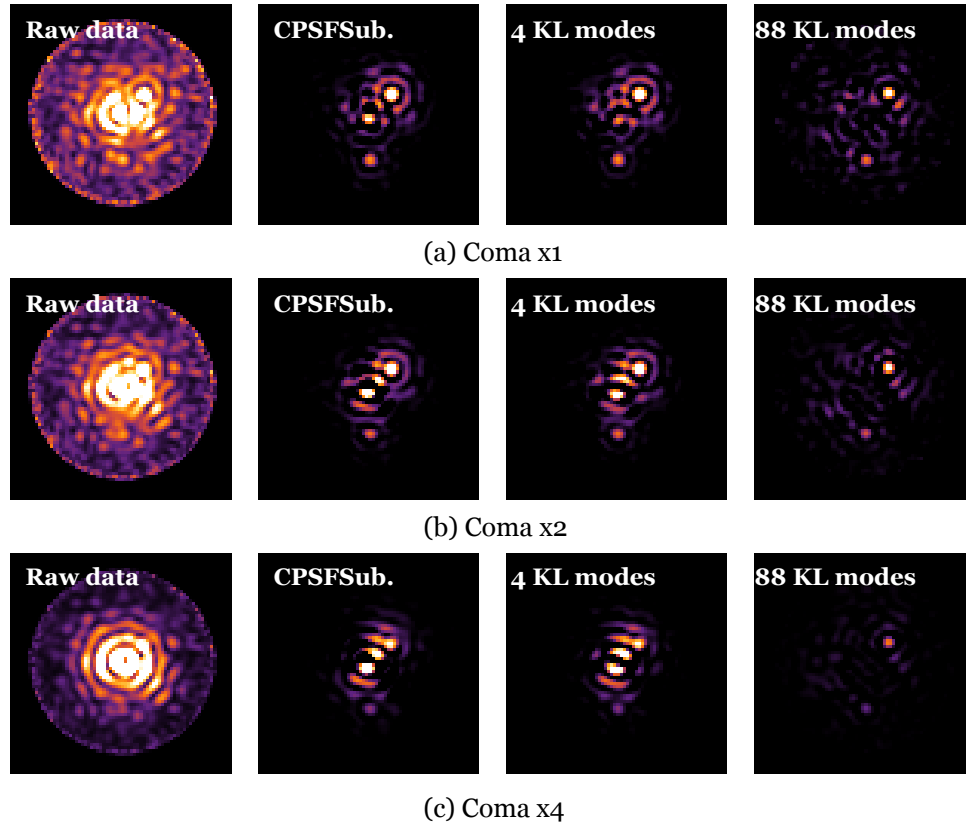


Figure 2: **HLC OS1 simulated data sets - Impact of the amount of aberrations on post-processing.** Comparison of the noiseless raw data and three different reductions for increasing amounts of coma. Top: (a) coma $\times 1$ (reference amount). Middle: (b) coma $\times 2$. Bottom: (c) coma $\times 4$. From left to right with same dynamic range: raw image (long exposure); image processed with classical PSF subtraction; image processed with KLIP 4 modes; and image processed with KLIP 88 modes, corresponding to maximum number of images. The two techniques show a relatively similar PSF subtraction compared to the raw image for the reference amount of coma (coma $\times 1$), the contrast improvement with 88 KL modes is better than with the classical PSF subtraction when the amount of coma increases. A single planet is detected at low SNR in the coma $\times 1$ raw data, so post processing is a significant improvement, enabling 2 planets detected. KLIP renders the detection more robust to the level of low-order aberrations in this case.

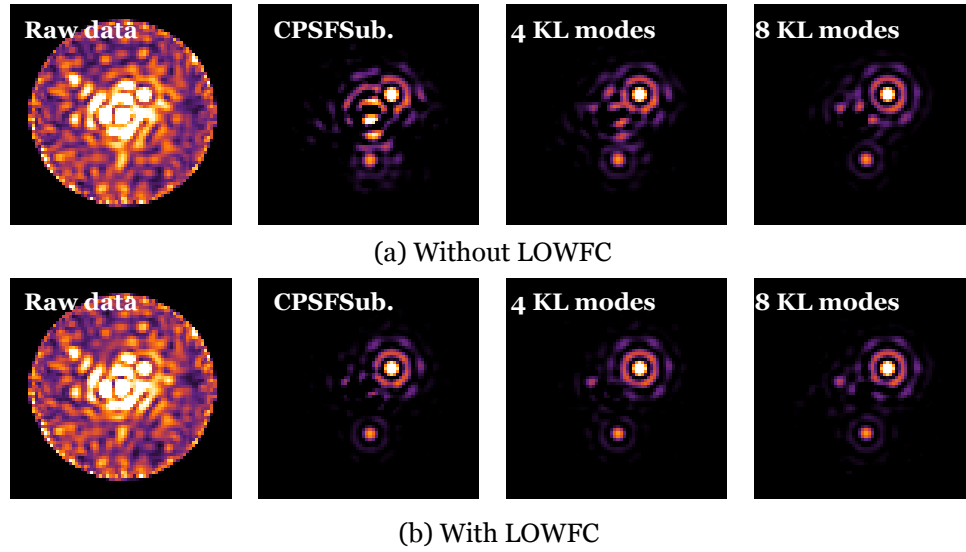


Figure 3: **HLC OS3 simulated data sets - Impact of the LOWFSC on post-processing.** Comparison of the noiseless raw data and three different reductions. Top: (a) data without LOWFSC. Bottom: (b) data with LOWFSC. From left to right with same dynamic range: raw image (long exposure); image processed with classical PSF subtraction image processed with KLIP 4 modes and image processed with KLIP 8 modes. The two techniques show a relatively good PSF subtraction compared to the raw images for both the cases without and with LOWFSC. The reduction quality is improved by the KLIP reduction compared to the classical PSF subtraction; the more KL modes used, the better the result.

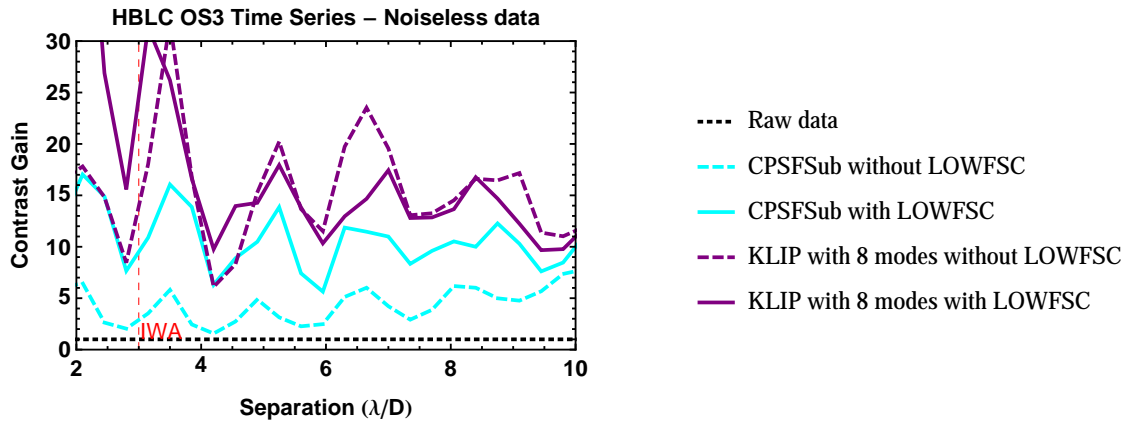


Figure 4: **HLC OS3 simulated data sets - Contrast gain curves.** Comparison of contrast gain curves as a function of separation for data without and with the LOWFSC. Gains are comprised between 5 and 30, and are typically more important at small separations close to the IWA. In all cases KLIP outperforms the classical PSF subtraction. The use of a LOWFS to control low-order aberrations leads to substantially improved contrast gains for the classical subtraction, with a smaller effect on the KLIP subtraction gains.

quantify the performance of the reductions. For the reduction with the classical PSF subtraction in the case without LOWFSC, the improvement in contrast is 3–6. This increases to 6–15 in the case of classical subtraction with LOWFSC. For KLIP we measure a contrast improvement of 10× or better at almost all separations, with or without LOWFSC. For both the cases with and without LOWFSC, the best performing KLIP reduction uses 8 modes (maximal for the reference data set), which enables a contrast gain up to 30 at small separations. This is particularly important because small angular separations are most critical for science. This result is consistent with theoretical understanding of speckle noise properties in coronagraphic images [22, 23], where the higher noise variance at short angular separation can be better compensated if a larger reference library with PCA is used.

2.2.2 Characterization in the IFS channel with an SPC

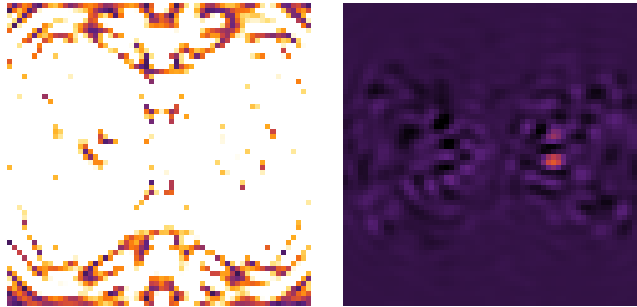


Figure 5: **SPC+IFS OS1 simulated data set.** Comparison of the raw data and an 88-KL mode reduction for the central wavelength with the same dynamic range. [Left] raw image (long exposure) and [right] image processed with 88 KL modes. The injected planet is detected with a high SNR (≈ 10) signal-to-noise ratio on the right of the central occultation.

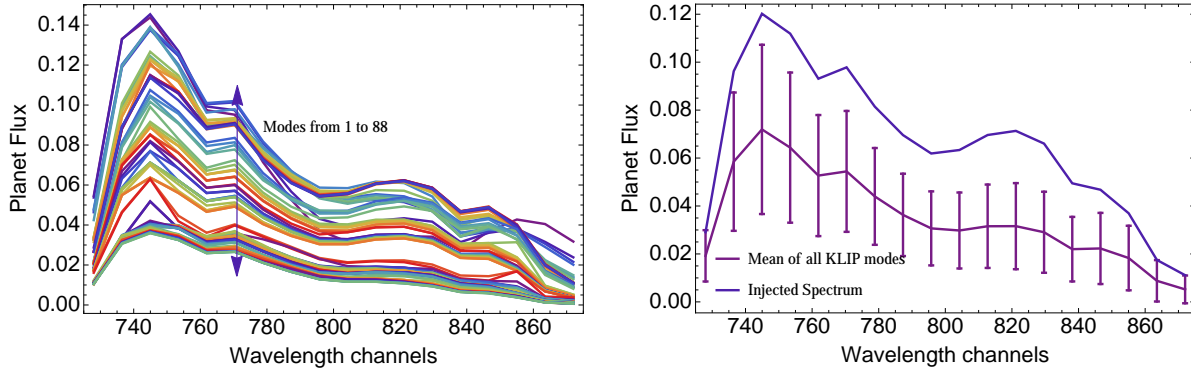
For characterization in the IFS channel with an SPC, we focused on the spectrum extraction of a Jupiter-like exoplanet with a contrast of 4.2×10^{-10} , assuming that the position of the planet is already well-known from prior observations (e.g. a detection in the imaging channel). We carried out a quantitative study to understand the efficiency of the algorithm at subtracting the speckles, and the effects on the algorithm throughput, with the goal of eventually developing a rigorous error budget for spectral extraction with the WFIRST CGI IFS.

Figure 5 shows an example of a KLIP reduction with $K = 88$ modes for the central wavelength. The injected planet was detected with a high SNR (≈ 10). After the reduction, we estimated the planet flux at each wavelength using a matched filter with the wavelength-dependent planet-only scene as a template. We thus assumed here that we have a perfect calibration of the PSF shape and of its wavelength evolution.

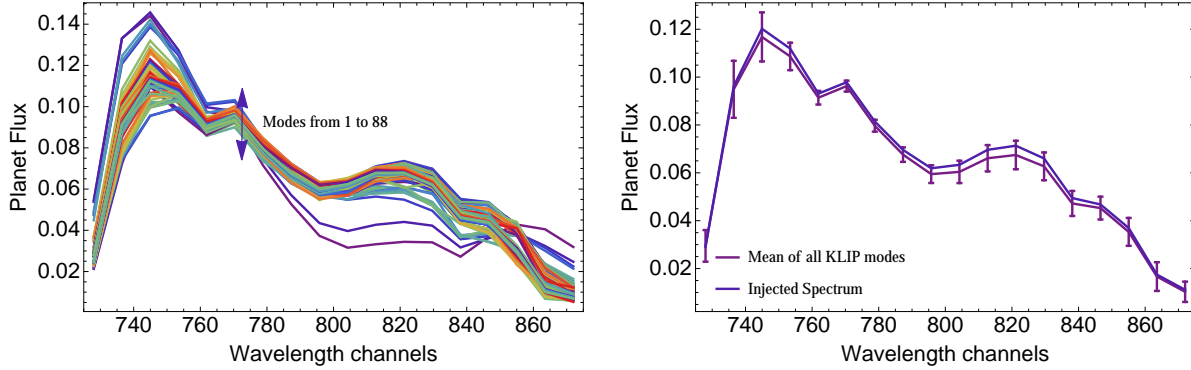
We performed spectrum extractions with and without forward modeling to evaluate the impact of the algorithm throughput on the flux estimation. Figure 6 shows on the same plot a set

of 88 spectra extracted with the different possible reductions with KLIP (i.e. using K from 1 to 88 principal components). The right panel presents the average of those 88 spectra along with the standard deviation, which provides a conservative proxy for uncertainty due to systematic biases from the subtraction. Without forward modeling (cf. Figure 6a), the spectrum is significantly biased downward (factors of 3–5 reduction in planet flux), and the dispersion is large (up to $\sim 50\%$ of the apparent planet flux). This is due to the increased throughput loss from over-subtraction bias with increasing number of modes. The photometry bias disappears with forward modeling (cf. Figure 6b). Spectra with the lowest number of modes are somewhat discrepant from the injected planet spectrum, due to the relatively limited PSF subtraction when few modes are used, but as the number of modes is increased the extracted spectrum converges to the injected spectrum. The remaining difference between the two spectra is mostly unbiased. The dispersion of the extracted spectra is reduced substantially as well. This calculation demonstrates that the forward modeling process can indeed enable both accurate and precise extraction of planetary spectra.

Figure 7 illustrates how a less-challenging contrast (i.e., reduced amount of residual speckles relatively to the planet flux) improves the spectrum extraction. For the sake of simplicity, we chose to increase the planet flux by a factor of 10 instead of decreasing the level of residual speckles, all else being equal. This was equivalent to extracting the spectrum of a planet with a contrast of 4.2×10^{-9} . In this case the injected spectrum is recovered very precisely, with little scatter in the extracted spectra with varying number of modes, and negligible error in the average over the modes. This spectral extraction demonstrates that after forward modeling our spectral extraction accuracy is only limited in cases of residual un-subtracted speckles; it is not biased intrinsically. Thus, in the more-challenging contrast case, the algorithm’s efficiency at subtracting the speckles determines the overall accuracy of the spectrum extraction process. At higher SNR, calibration of algorithm throughput becomes less important. Of course in practice IFS datacube reconstruction noise, photon noise, and detector noise will contribute additional limiting terms to the spectrophotometric error budget.



(a) Without forward modeling



(b) With forward modeling

Figure 6: **SPC OS1 simulated data set - Impact of forward modeling.** [Left] 88 spectrum extractions represented on the same plot (as many as different possible reductions with KLIP). [Right] The blue curve represents the mean and standard deviation of these estimations and the purple curve represents the injected spectrum. Forward modeling (b) can be used to calibrate the algorithm throughput using a template PSF at the location of the planet, and therefore compensate for the dispersion and bias compared to (a).

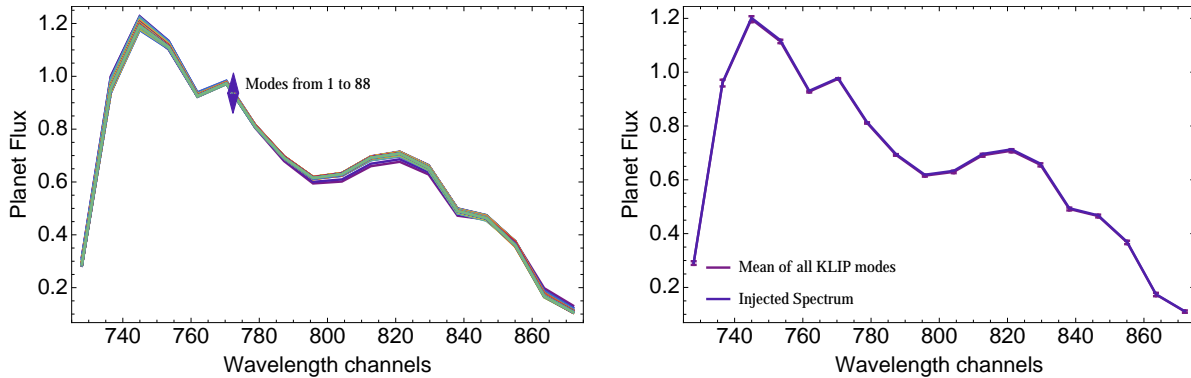


Figure 7: **SPC OS1 simulated data set - Impact of the level of residual speckles.** [Left] 88 spectrum extractions represented on the same plot (as many as different possible reductions with KLIP). [Right] The blue curve represents the mean and standard deviation of these estimations and the purple curve represents the injected spectrum. When the level of residual speckles does not affect the forward modeling, the calibration of the KLIP throughput is optimal and there is therefore no bias in the spectrum estimation. In practice other sources of noises (detector, photon, etc.) will contribute additional limiting factors.

3 Laboratory coronagraph data

3.1 Broadband coronagraphic data

3.1.1 Data set description, selection and fake planet simulation

In February 2015 we received one broadband dataset recorded on the static SPC testbed at HCIT (Fig. 8), as well as an unocculted PSF to simulate a fake planet (Fig. 9). The coronagraph design installed on the testbed at this time was the first-generation characterization shaped pupil, with a bow-tie-shaped dark hole extending over an opening angle of 60 deg over angular separations 4–12 λ/D [24, 25].

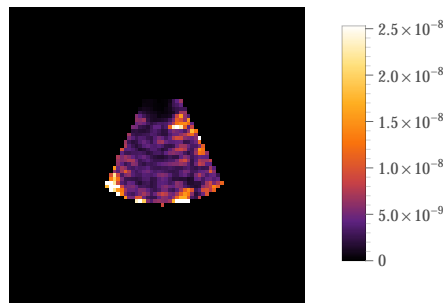


Figure 8: Example of a raw image in the mock RDI data sequence.

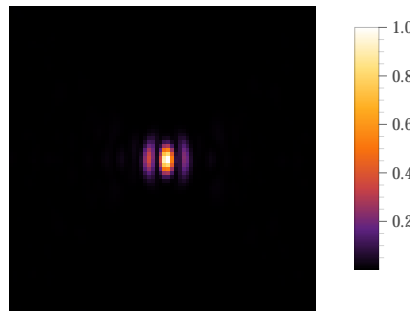


Figure 9: Unocculted PSF used to inject a fake planet in the science target cube.

The data set, consisting of 1011 FITS files, was acquired during the execution of a wavefront control loop. Therefore, the deformable mirror actuator settings change between each exposure, and the overall speckle intensity decreases gradually with time. Following the electric field conjugation algorithm, for each control iteration the testbed measurements alternate between two modes of operation: deformable mirror probes, and the sweep of the regularization parameter imposed on the actuator solution. Each FITS file contains either 12 or 14 image arrays depending on the iteration type, with half the arrays in each iteration being focal plane images and the other half being a representation of the actuator positions when each image was taken. The regularization category

offers a larger volume of stable, high-contrast images. Therefore, before applying any PSF subtraction algorithm, we assemble the images from the regularization sweeps (each with 6 ascending parameter values) into a separate data cube, excluding the DM probe images. Then, from each regularization iteration, we chose the focal plane image with the best contrast, i.e. minimum averaged flux on the analyzed sub-region. After this selection and cropping procedure, a three-dimensional array (505 exposures, 500×500 pixels) served as our “raw” data set.

To examine the statistical properties of the data set, we computed the cross-correlations between the first exposure and each other exposure:

$$C^i = \frac{1}{N} \sum_{x,y} \frac{[I^i(x,y) - \bar{I}^i] [I^1(x,y) - \bar{I}^1]}{\sigma_{I^i} \sigma_{I^1}}. \quad (1)$$

I^i and \bar{I}^i are respectively the i -th exposure of the data set and its average, N is the total number of pixels in the images, and σ_{I^i} is the standard deviation of the exposure I^i . We plotted the evolution of the correlation with respect to the number of the exposure (Fig. 10). Because each exposure of

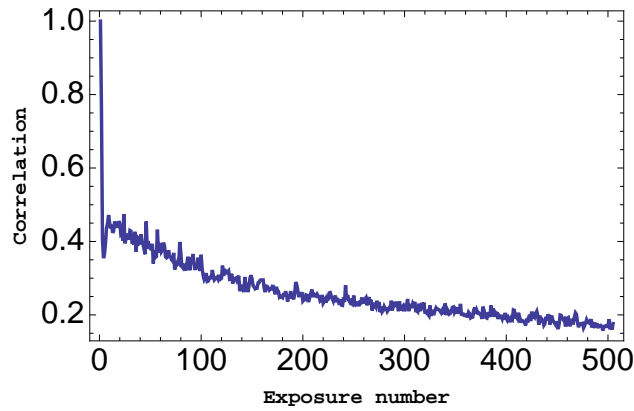


Figure 10: **Correlation analysis of the “raw” image sequence.** Evolution of the correlation with time, during reference acquisition. The correlation shown here is that of *the first exposure of the sequence* with each subsequent exposure in our extracted dataset. After a rapid initial evolution the PSF stabilized; while the correlation of later images is only ~ 0.2 with the first image, the correlation of the later images with each other is much higher.

the data set corresponds to the successive times at which the images were taken, the evolution of the correlation with the number i of the exposure can also be seen as an evolution of the correlation with time. The plot shows that the correlation decreases with time and tends to stabilize after at least 180–200 iterations.

After this statistical analysis, we eliminated the nine first images of the data set corresponding to the most unstable wavefronts. The data set was then divided into two parts: a first part of the cube is used for reference (243 exposures); the second part of the cube was used as science target (252 exposures). The science target cube was then duplicated to create two different data sets: without and with a fake planet. To simulate the fake planet, we used the image of unocculted PSF

of the figure 9 that we injected a contrast of 2×10^{-9} .

3.1.2 Reductions and analysis

Before reducing the data with KLIP, we subtract the average values of the target and references over the algorithm search zone to obtain the zero-mean input images required by the PCA decomposition. We collapse the target cube to obtain a long exposure raw image, and then apply the KLIP algorithm. Figure 11 shows the results of several reductions as a function of modes used, for the dataset without the injected planet. We show on the same dynamic range the long exposure raw image (all exposures collapsed), the classical PSF subtraction reduction, and KLIP reductions with 2, 20 and 243 modes. Both techniques show a significant PSF subtraction compared to the raw image with averaged speckle suppression gains up to 13.5 using the maximum number of KL modes (i.e. 243). However, because the algorithm throughput also decreases with increasing number of modes, this does not correspond to an actual contrast gain.

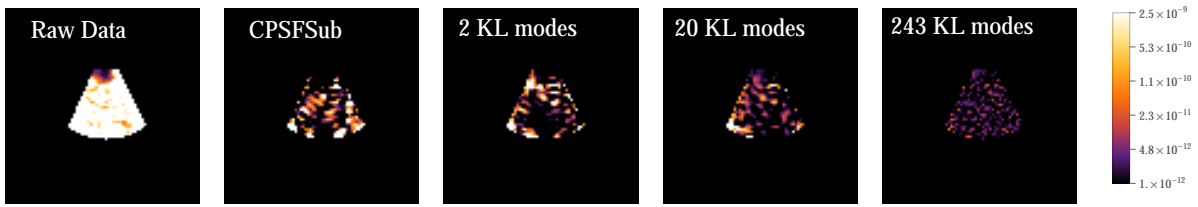


Figure 11: **Reductions without injected planet.** From left to right: raw co-add of all science images, classical PSF subtraction residual image, and KLIP subtraction residuals with the number of KL modes K equal to 2, 20, and 243 respectively. The gains in speckle suppression, averaged on the entire analyzed sub-regions are respectively: 1.5, 2.8, 4.8 and 13.5. Note that this does not correspond to an actual contrast gain because the algorithm throughput is not included here.

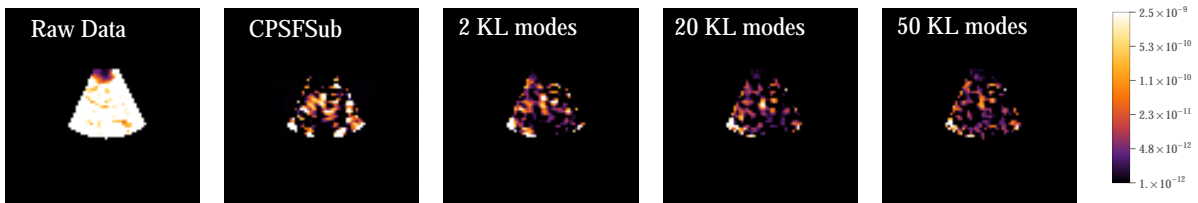


Figure 12: **Reductions with planet of contrast 2×10^{-9} .** From left to right: raw coadd of all science images, classical PSF subtraction residual image, KLIP subtraction residual with 2 KL modes, KLIP subtraction residual with 20 KL modes and KLIP subtraction residual with 50 KL modes. The gains in contrast, averaged over the entire analyzed sub-region, are respectively 1.5, 3.8, 5.4 and 6.2. The best detection for the planet corresponds to the reduction with 20 KL modes.

We then repeated this analysis for the data with a synthetic planet at a contrast of 2×10^{-9} , and we estimated its flux using the forward modeling technique as described above [1, 20, 21]. In order to obtain an accurate estimation of the flux of the planet, we had to refine the selection of references as well as the number of KL modes to use. Using the 50 most correlated references to

the collapsed target cube, we obtain a detection using $K = 20$ KL modes. While a higher number of modes improves speckle subtraction, it also leads to increased flux over-subtraction for the planet. Forward modeling is therefore essential to characterize the planet. Figure 12 shows the results of the reductions. The classical PSF subtraction does not enable the detection of the planet in this particular case. We present the 1σ contrast curves as a function of separation for the different reductions in Figure 13. Typical contrast gains are in the range 2–5, as shown in Figure 14. The gains are not spatially uniform but rather are periodic or “spiky”, presumably related to periodic diffractive structure and the characteristic speckle size of λ/D . At a few certain separations, little to no contrast gains are seen with any of the algorithms (e.g. $\sim 7.2 \lambda/D$ in Fig. 14), but KLIP with 20 modes achieves gains $\gtrsim 2$ over a large majority of the field.

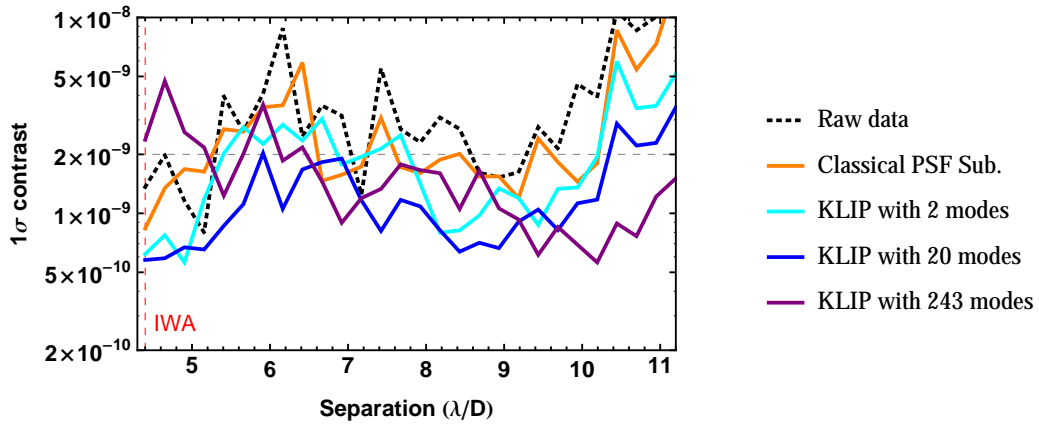


Figure 13: **Contrast curves.** 1σ contrast curves as a function of separation. These curves take into account the KLIP throughput. KLIP enables better contrasts than classical PSF subtraction. Typical gains from post processing here are in the 2–5 range.

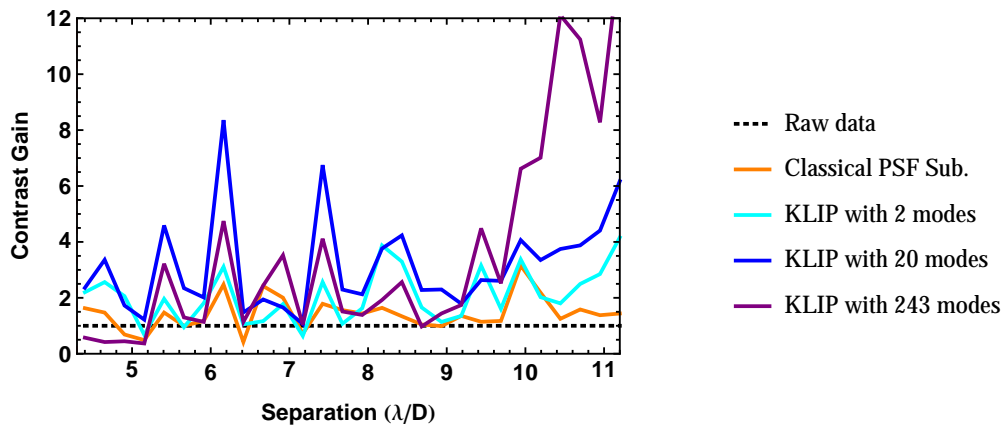


Figure 14: **Contrast gains.** Contrast gains as a function of separation. These curves take into account the KLIP throughput. KLIP enables better contrasts than classical PSF subtraction.

To summarize for this data set, both classical PSF subtraction and KLIP yield contrast gains, but KLIP performs better than the classical PSF subtraction, enabling contrast gains around 3 averaged

for all separations and up to 8 locally. This result confirms the advantage of the KLIP algorithm over the classical PSF subtraction when significant wavefront changes occur between the target and reference acquisition, which is the case here. However the relative performance improvement of KLIP over classical subtraction is less than the gains seen at lower contrast levels (e.g. HST or AO coronagraphy). We showed that a planet injected at a contrast of 2×10^{-9} and with a separation of $8 \lambda/D$ is detectable using 20 KL modes. This is the first time that the KLIP algorithm has been tested on WFIRST-like experimental data obtained with deformable mirrors and a SPC coronagraph in broadband light.

3.2 Multi-wavelength coronagraphic data

3.2.1 Data set description

In June 2015 we received two multi-wavelength datasets recorded on the static SPC testbed at HCIT, again with the first-generation characterization shaped pupil. The images were acquired in five 2% sub-bands spanning a 10% bandpass centered at 550 nm. The corresponding center wavelengths of the five channels (which we hereafter refer to as channel 1–5) are therefore 528, 539, 550, 561, and 572 nm. Two independent wavefront control runs were executed: one with both sides of the bow-tie region controlled, and the other operating on only one side of the dark bow-tie region. To date we have concentrated our efforts on the one-sided data set, which achieved a significantly deeper absolute (raw) contrast than the two-sided control run. Therefore, in this report we present our results only for the one-sided multi-wavelength data set.

The images were acquired during a wavefront control loop similar to the one described above in Section 3.1; the only essential difference being the repetition of exposures for each wavelength channels on a given DM actuator setting. Therefore, as in the case of the broadband SPC data set, before applying any PSF subtraction algorithm we assembled the images from the regularization sweeps into a separate data cube, excluding the DM probe images. There are many fewer control iterations present in the multi-wavelength data set: 12 instead of the 505 carried out for the broadband data set. Therefore, in order to keep a reasonably high volume of data to analyze, we reject only the worst image in each six-value regularization sweep (generally, the first and smallest parameter value), as opposed to rejecting all but the best regularization parameter.

We restrict our analysis to a dark sub-region within the bow-tie for which we found that the speckle field appears statistically homogeneous at each wavelength. After this selection and cropping procedure, a four-dimensional array (60 exposures, 5 wavelength channels, 62×57 spatial pixels) serves as our “raw” data set. In Figure 15, we plot the standard deviation (std. dev.) within our bow-tie sub-region, as it evolves over time for all of the regularization frames.

In each frame, we also measure the background detector noise by taking the standard deviation of a 60×60 box of pixels adjacent to the coronagraph field of view, in a region where the field

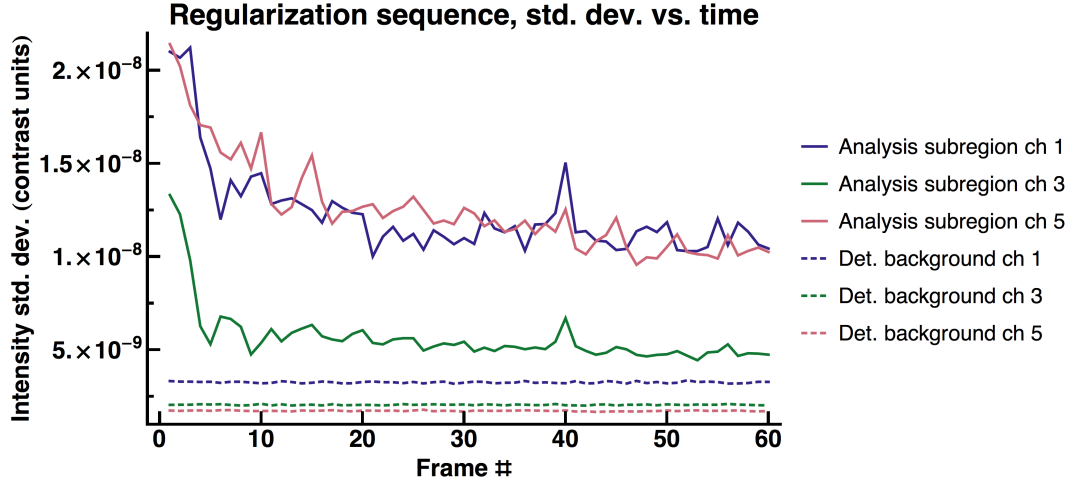


Figure 15: Intensity standard deviation versus time in the raw bow-tie sub-region of the data set for the one-sided bow-tie, at three wavelengths. We only include exposures from the regularization parameter sweep. For comparison, we plot the standard deviation of an adjacent, non-illuminated region of the focal plane to estimate the background detector noise in each frame.

stop mask prevents illumination. Here we expect the CCD pixel values to be random and lacking significant correlated structure, and determined by some combination of read noise, residual dark current and bias, and bit quantization. In Figure 15, alongside the intensity in the analysis sub-region, we plot the background level versus regularization frame number. The background standard deviation is nearly constant over time, but different for each channel. That is because the HCIT team has already normalized each image to contrast units, a calibration that is specific to each wavelength.

In Table 1 we list more properties of the raw data cube, at each wavelength listing the intensity standard deviation in the final frame of the regularization sequence, as well as the time-averaged standard deviation measurement in the aforementioned background box, and the ratio of these quantities, which we label the speckle-to-background ratio. This is lowest for channels 2 and 3, and increases for the shortest and longest channels. In the last column we list the least significant bit (LSB), or the discretization interval of the CCD readout, again in units of contrast. We inferred the LSB from the clustering interval of the CCD pixel value distribution. The downward trend in LSB with wavelength reveals that when the images were rescaled to contrast units, the longer wavelength images were scaled down the most, presumably to compensate for higher light source flux. That compensation is also consistent with the downward trend we see in the background standard deviation versus wavelength, also apparent in Figure 15. Therefore, there are two coinciding factors that boost the speckle-to-background ratio of the longest wavelength channel: higher light source flux, and higher wavefront error.

| chan | λ (nm) | Intensity std. dev., analysis (speckle) sub-region, final regular. frame | Intensity std. dev., background sub-region | Ratio (speckle -to-bkg.) | LSB |
|------|----------------|--|--|--------------------------------|------------------------|
| 1 | 528 | 1.04×10^{-8} | 3.26×10^{-9} | 3.21 | 2.42×10^{-10} |
| 2 | 539 | 6.19×10^{-9} | 2.70×10^{-9} | 2.29 | 2.01×10^{-10} |
| 3 | 550 | 4.74×10^{-9} | 2.05×10^{-9} | 2.31 | 1.52×10^{-10} |
| 4 | 561 | 6.24×10^{-9} | 1.90×10^{-9} | 3.28 | 1.35×10^{-10} |
| 5 | 572 | 1.02×10^{-8} | 1.73×10^{-9} | 5.94 | 1.27×10^{-10} |

Table 1: Intensity of the analysis sub-region at the end of the regularization sequence, and the background (bkg.) detector noise in an adjacent non-illuminated region of the focal plane. The speckle-to-bkg. ratio is highest at the longest wavelength, consistent with a larger wavefront error. The last column is the least significant bit in units of contrast, inferred from the periodic clustering of readout values.

3.2.2 Detector background noise

We carried out additional analysis of the detector background to anticipate what limits it might impose on the final contrast. At first, we tried to model the non-illuminated background box samples as independent draws from a normal distribution, discretized according to the inferred detector quantization (LSB in Table 1). Although this model can reproduce an individual background box, it fails further scrutiny. In Figure 16 we plot the average of the background regions in all 60 regularization frames (panel b), and then the result of splitting the background stack into two 30-frame sequences, averaging them, and taking their difference (panel c). Note that in the 60-frame average, the std. dev. remains much higher than would be expected if there were an approximate $1/\sqrt{N} \simeq 1/7.7$ scaling; instead the std. dev. in the box is reduced by less than a factor of 2. Meanwhile, in the difference of two averaged sequence halves, the std. dev. is a factor of ~ 50 below that of an individual frame. The simplest explanation for this behavior is a significant temporal correlation in the noise pattern. If we maintain the assumption of zero spatial correlation, we can try modeling the background box with two components: one component a pattern fixed in time determined by drawing the value of each pixel from a normal distribution, and add to that a pattern that is random in time as well as spatially, also determined by a normal distribution. At the flux scaling of the central wavelength channel, we infer that the std. dev. of the time-constant pattern is 1.24×10^{-9} , while the std. dev. of the distribution with no temporal correlation is 1.64×10^{-9} . Note that in quadrature these account for the overall std. dev. of all samples, 2.05×10^{-9} .

We generated a numerical realization of this two-component noise model, again rounding each pixel sample in accordance with the testbed CCD quantization. The lower half of Figure 16 shows the model closely resembles an individual frame (panel d versus a), the average of 60 frame realizations (e versus b), as well as the difference of averaged halves of the 60-frame sequence (f versus c). For an additional check, in Figure 17 we plot the histogram of the two-component model against the measured background samples from all 60 of the 60×60 -pixel boxes. The match is satisfactory, if we neglect the slight positive skew of the data value distribution. In later analysis, we assume the

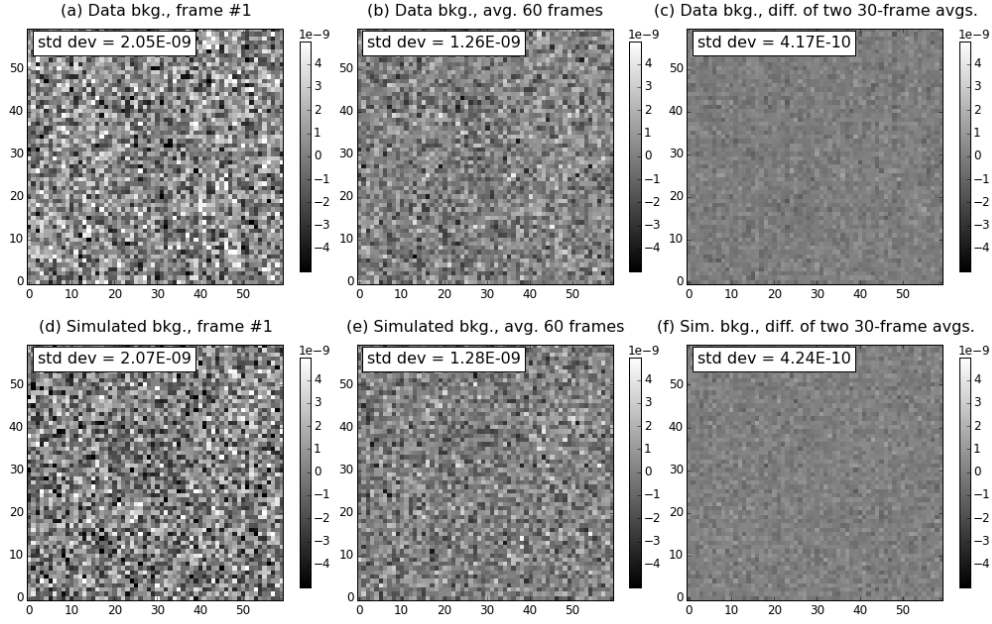


Figure 16: Detector background noise properties: (a) the background sample box from one regularization frame; (b) the average of all 60 frames; (c) the result of a 30/30 split at the halfway point, averaging the two halves, and then differencing. (d)–(f) are same as (a)–(c) above but simulated using the two-component (uncorrelated plus temporally correlated) noise model we describe in the text of Section 3.2.2. The two-component model we propose, combining a constant random background with one that is uncorrelated in time, is the simplest one that offers reasonable agreement with the standard deviations measured at each step.

same uncorrelated noise component (1.64×10^{-9}) applies in the field of view of the coronagraph. For each wavelength channel, this std. dev. level is scaled according to the relative value of the LSB in Table 1.

3.2.3 Frame arrangement for reference differential imaging

To perform RDI with the testbed data set, we must arbitrarily split our data into reference and target sub-sequences. We divided our image sequence in three ways, to test different limiting cases: (i) *Contiguous* regularization sub-sequences: one contiguous block is treated as the reference sub-sequence, and the succeeding block is treated as the target sub-sequence:

$\{A_1 A_2 A_3 \dots A_{m-1} A_m B_1 B_2 \dots B_{n-1} B_n\}$ maps to $\{A_1 A_2 \dots A_m\}, \{B_1 B_2 \dots B_n\}$ (see left-hand side of Figure 18).

(ii) *Interleaved* regularization sub-sequences: consecutive images are alternately assigned to the reference or target sub-sequence: $\{A_1 B_1 A_2 B_2 \dots A_n B_n\}$ maps to $\{A_1 A_2 \dots A_n\}, \{B_1 B_2 \dots B_n\}$ (see right-hand side of Figure 18).

(iii) Dark probe–regularization combination: use the EFC-corrected images from the probe iteration steps as targets, and all 60 regularization frames as references. Chronologically, one dark probe image is acquired at each of the 12 control iterations; here we group them as if they were an

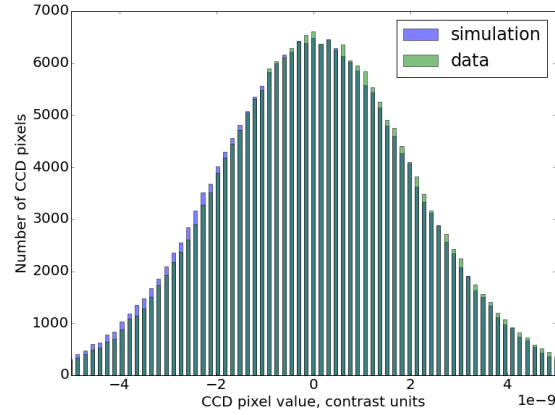


Figure 17: CCD readout value histogram of the background noise sample region at the center wavelength, including 60×60 pixels over all 60 regularization frames (green). We compare this distribution with the numerical noise model described in the text (blue), to check its ability to reproduce the bulk distribution. The gaps between the histogram bars reflect the actual clustering of values due to the CCD quantization (for this wavelength channel’s flux calibration, roughly 13 levels across a contrast increment of 2×10^{-9}).

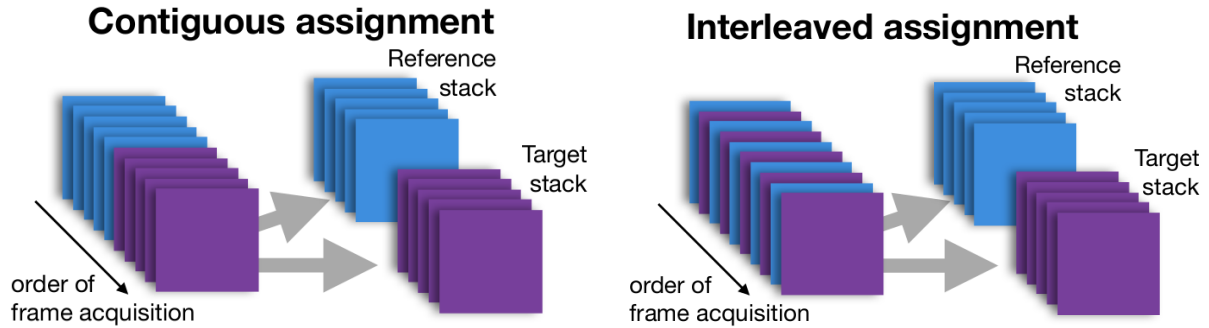


Figure 18: Diagram of two of the frame arrangement schemes we applied to the multi-wavelength testbed data: contiguous and interleaved assignment of images to reference and target stacks.

independent observation.

The contiguous regularization scenario is relatively pessimistic in terms of speckle subtraction, since in this case the typical number of control iterations separating the reference and target images is higher, resulting in dissimilar DM settings, and a correspondingly different speckle pattern. This is evident in Figure 19, where we plot the correlations between target images and reference images (this correlation formula is defined in Equation 1). For the contiguous regularization arrangement, shown at the top row, the first target image is well correlated with the last reference image, but is poorly correlated with the first reference images, chronologically the most distant. The last image in the target stack is distant from all of the reference images, and the correlation curve does not reach as high of a plateau. Conversely, for the interleaved regularization sub-sequence, the last target image has a high correlation across all wavelength channels with the last third of the reference stack.

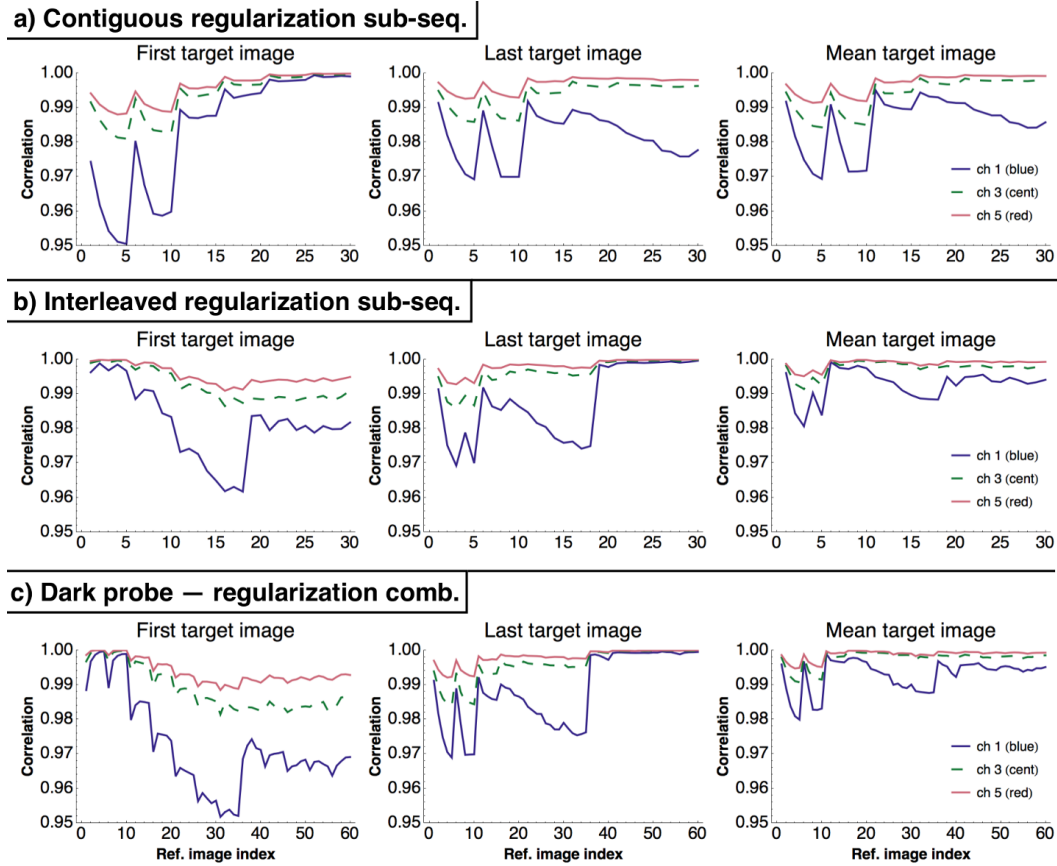


Figure 19: Normalized cross correlation values between target images and each image in the reference image stack, for the three RDI configurations. The target images used are the first (left column of figures), last (middle column) and the mean (right column). The horizontal axis in each sub-figure is the index across the reference image dataset. The colors indicate wavelengths in the same manner as in Fig.15.

Regardless of the image arrangement, Figure 19 reveals a consistent trend in wavelength: for any given pair of images, the correlation is lowest at the short wavelength (blue curve), and highest at the long wavelength (red curve). There are two effects that enhance the correlation of the long-wavelength channel. First, for a given change in the wavefront at the coronagraph entrance, the effect on the PSF is proportionally smaller at longer wavelengths, because in the linear phase aberration regime, the change in the focal plane field depends inversely on wavelength. The second effect is the fact that the ratio of the speckle intensity to background noise is highest in the long-wavelength channel, as illustrated before in Figure 15 and Table 1. As noted above, that is due to a higher light source intensity in the testbed in the long-wavelength channel. This effect combines with a weaker dark hole null at the edge channels (evident in Figure 15), originating from a poorer wavefront correction. Therefore, the signal in the long-wavelength channel tends to be dominated by bright, stable speckles. The speckle-to-background ratio is also high in the short-wavelength channel, but there the speckle pattern will be more sensitive to the wavefront changes caused by evolving DM actuator settings.

There is a quasi-periodic structure in the correlation curves of the contiguous regularization arrangement, apparent in all of the top row plots of Figure 19. This is due to the cycling of the regularization parameter over five values at each control iteration. The overall correlation pattern of the interleaved and dark probe–regularization arrangements are similar since in both cases the target images are acquired contemporaneously to the references, in the latter case more sparsely.

3.2.4 Contiguous sub-sequence RDI

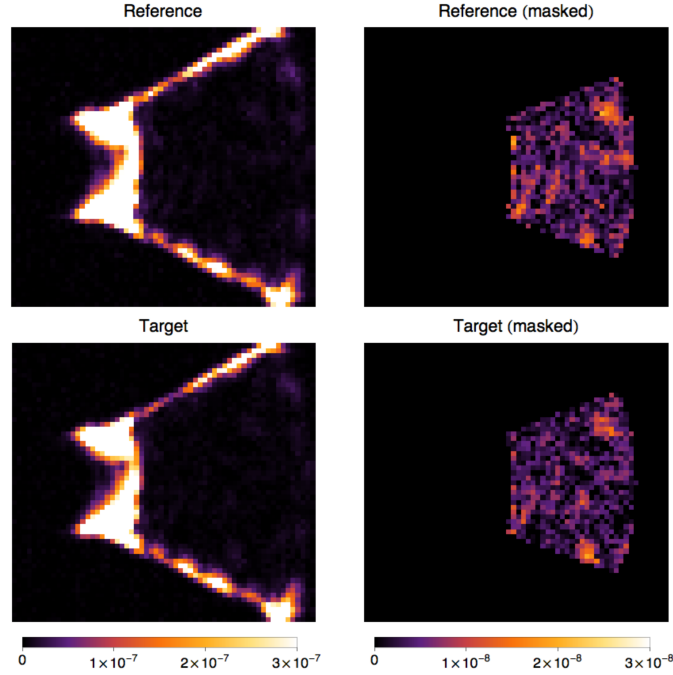


Figure 20: Examples of individual raw images in the mock RDI data sequence, at the center wavelength. Top row: a typical reference image, before and after masking off the analyzed sub-region. Bottom row: a typical target (science) image, before and after masking off the analyzed sub-region. The intensity in the part of the field of view excluded from the analysis, along the interior and radial edges of the bow-tie field stop, is over 30 \times brighter than the analyzed sub-region.

Figure 20 shows examples of individual raw images used in the contiguous RDI arrangement, one from the reference sub-sequence and one from the target sub-sequence. Each image is shown first with a relatively wide linear intensity stretch, and then a narrow stretch restricted to the masked sub-region that we analyze. The analysis sub-region extends from 5–9 λ_0/D in angular separation from the star (where λ_0/D is a center wavelength channel resolution element), with an opening angle of 40 degrees.

In Figure 21 we show how the std. dev. intensity evolves over time within the raw reference and target sub-sequences. We use the last 30 exposures of the original (60 exposure) regularization data set for the target sub-sequence, and the first 30 images for the reference sub-sequence. Figure 22 illustrates the result of performing KLIP subtraction on the co-added target image, at the center

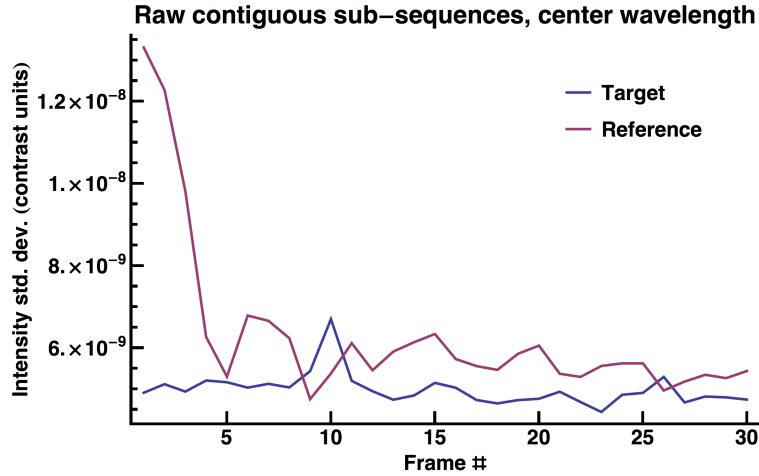


Figure 21: Bow-tie sub-region standard deviation in the images assigned to the reference and target subsequences.

and long wavelength channels, respectively. As a sanity check we include the result of a classical PSF subtraction, which is identical as expected to subtracting only the first KL mode. With 10 KL modes (annotated $K = 10$), the residual intensity is noticeably suppressed in both channels. However, the change in the residual intensity between $K = 1$ to $K = 10$ is more pronounced at the long-wavelength channel. In other plots we see that the contrast gain is consistently higher at the long-wavelength end. The long-wavelength channel data reduction is illustrated again in Figure 23, this time comparing an individual target frame to the time-average of the target stack, and showing the first 6 KL modes used to estimate the PSF.

To optimize the KLIP subtraction, we compute the final contrast as a function of the number of KL modes. We inject a multi-wavelength image of the off-axis coronagraph PSF (also provided by the JPL HCIT team) to take into account over-subtraction, calibrating the contrast at 63 locations in the field of view (7 position angles at 9 radial separations). Figure 24 shows that the contrast floor in each channel hits a plateau after ~ 5 KL modes, with little or no improvement visible beyond that. The best final contrast is reached in the center channel, close to 1×10^{-9} ; the edge channels, 1 and 5, are poorest, reaching between 1.6 to 2.2×10^{-9} .

Another way of examining the efficacy of KLIP subtraction is to plot the contrast gain, that is the ratio of the raw to the final residual contrast, as a function of number of KL modes. Here, in Figure 25, the optimal number of reference KL modes for each wavelength channel is most easily discerned. The optimal K number is different for each channel: for channels 1–5 they are 4, 2, 8, 11, and 6 modes, respectively.

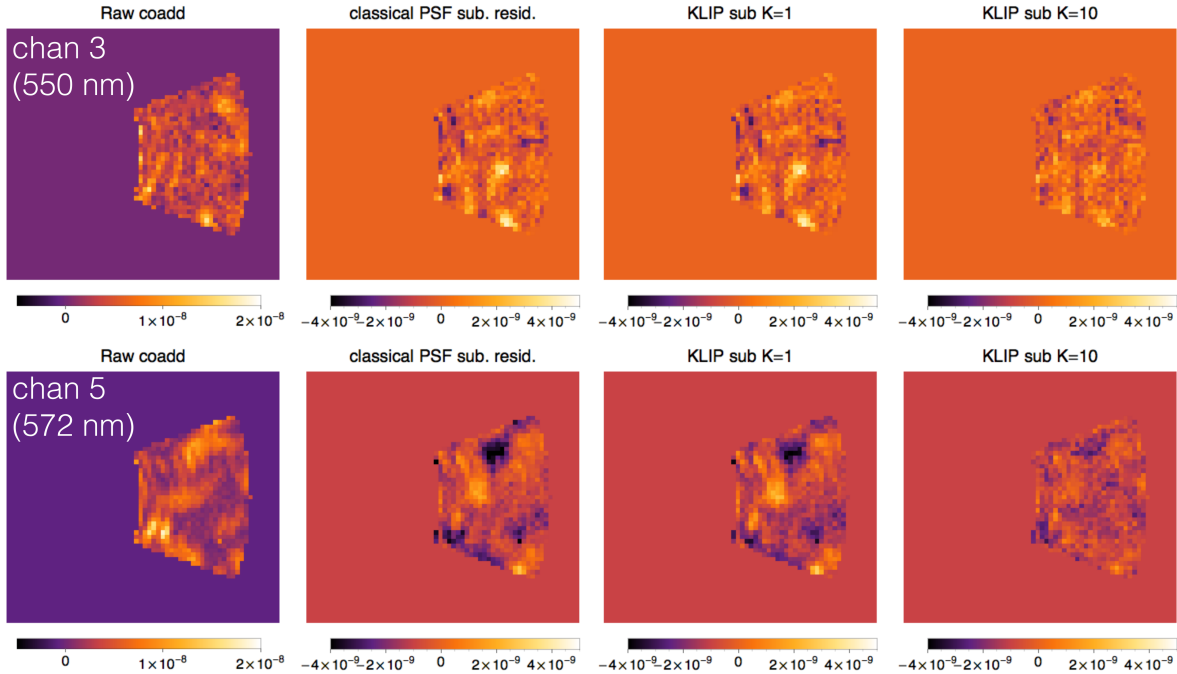


Figure 22: Images of the KLIP subtraction results for the contiguous RDI case, at the central wavelength channel and the longest wavelength channels. From left to right: raw co-add of all science images, classical PSF subtraction residual image, KLIP subtraction residual with the first KL mode, and KLIP subtraction residual with 10 KL modes.

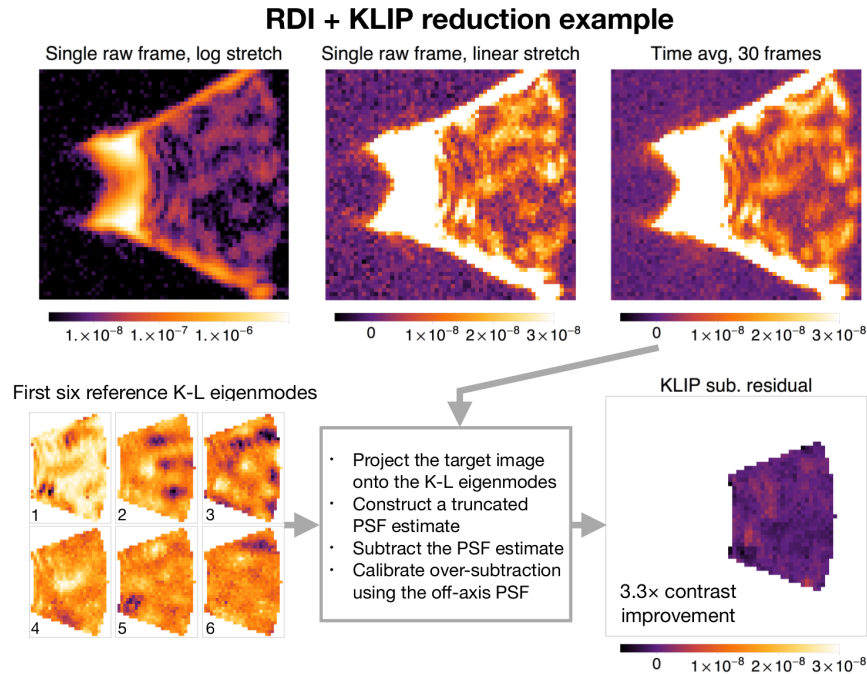


Figure 23: Illustration of the full, optimized RDI reduction procedure for channel #5 ($\lambda = 572$ nm), using reference and image stacks formed from contiguous assignment of regularization frames. First, the target stack is averaged over time, producing the image at the top right. The images at the lower left show the result of decomposing the reference stack into six KL modes, or principal components. The PSF estimate is determined by the projection of the mean-subtracted, time-averaged target image onto these KL modes. After subtracting, and calibrating for over-subtraction, we measure a factor of 3.3 gain in contrast with respect to the raw time-average.

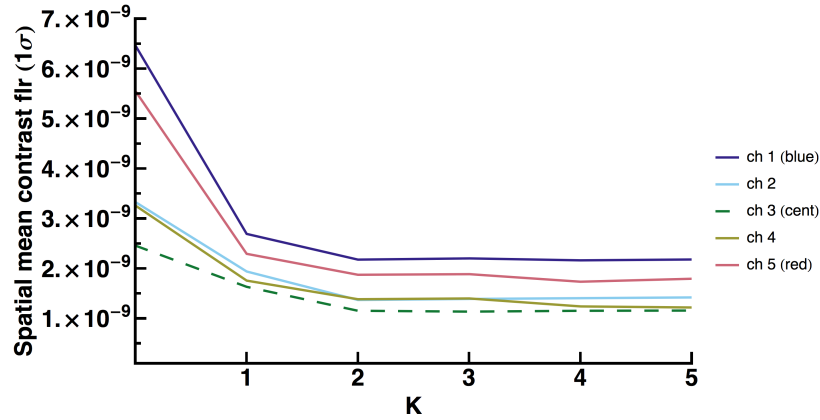


Figure 24: Throughput-corrected, spatially-averaged contrast versus number of retained KL modes for the contiguous RDI case. The achieved contrast is plotted separately for each wavelength channel. The value at the origin ($K = 0$) of each curve is the contrast in the raw co-add. The final contrast hits a plateau after only a few reference KL modes, with very little improvement, if any, beyond 5 modes. The central wavelength shows the best contrast of all five channels both before and after PSF subtraction.

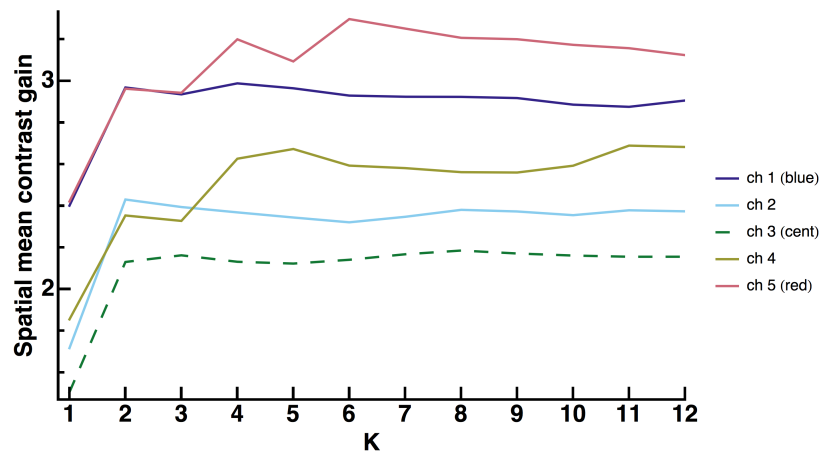


Figure 25: The ratio of the raw contrast to the contrast after KLIP subtraction, plotted against the number of KL modes, for the contiguous RDI case. This plot shows that the edge channels, where the raw speckle intensity is highest, benefit the most from KLIP. There is considerable variation in the optimal number of KL modes, ranging from 2 to 11 depending on the wavelength. But consistent with the prior figure, the contrast gains are relatively flat overall for $K > 5$.

3.2.5 Interleaved sub-sequence RDI

When the reference and target subsets are arranged by alternating between consecutive regularization exposures, the two raw sub-sequences show a more uniform trend in std. dev. In Figure 26 we plot the evolution of the std. dev. intensity in the analysis subregion at the center wavelength.

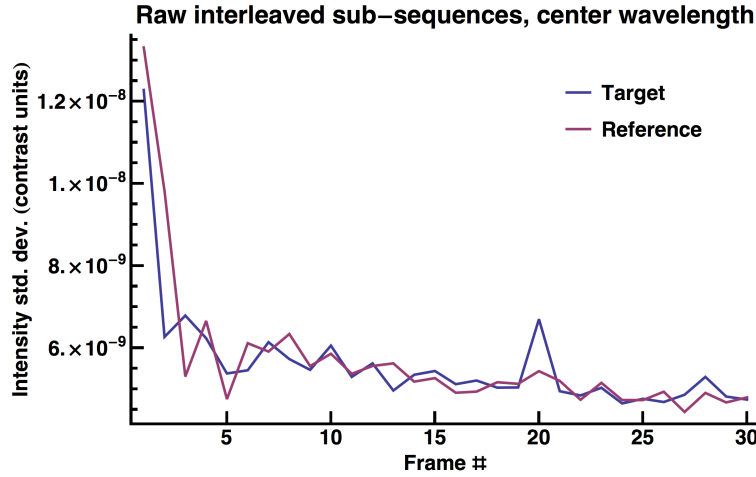


Figure 26: Bow-tie sub-region std. dev. in the images assigned to the reference and target sub-sequences, for the interleaved RDI arrangement. Unlike the contiguous RDI arrangement (Figure 21), the two sub-sequences trace a similar trend in intensity, since the pair of regularization images at a given index were acquired in direct succession.

Because the correlation between temporally neighboring pairs of target and reference images is relatively high, we are effectively testing speckle subtraction with a more stable PSF. As expected, the final contrast floor improves over contiguous RDI at all channels, as we show in Figure 27. The final contrast floor (1σ) ranges between about 6×10^{-10} and 1×10^{-9} .

In Figure 28, we see the contrast gain peaks at a smaller number of KL modes than for the contiguous RDI arrangement; the optimum K number is 2 for all wavelength channels. By assigning every other image to the target and reference stacks, we have ensured that the time averages of the reference and target data sets are very close. This makes the use of a single KL mode (or equivalently classical subtraction) very effective. In this case contrast is maximized with $K = 2$, that is, classical subtraction plus a single additional component to capture time variability. This is particularly true in the long-wavelength channel, where the speckle stability is highest. There the contrast gain factor is 8–10, much better than for all other channels which have contrast gains of 4–6. The higher relative efficacy of the subtraction in this case is just enough to compensate for the higher initial raw contrast, so the final contrast in the longest channel equals that achieved in the central wavelength channel. This is not the case in the shortest-wavelength channel, which has the second-best contrast gain but even after subtraction has worse contrast by about a factor of 1.5 than the central channel.

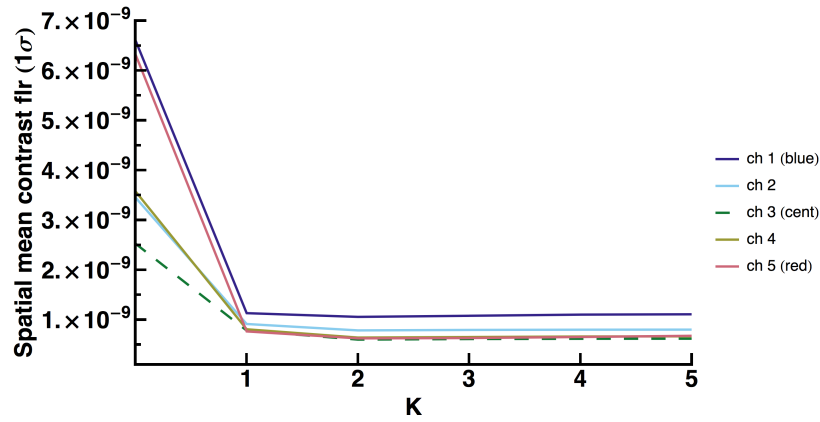


Figure 27: Throughput-corrected, spatially-averaged contrast versus number of retained KL modes for the interleaved RDI case. The achieved contrast is plotted separately for each spectral channel. The value at the origin ($K = 0$) of each curve is the contrast in the raw co-add. Despite factor of ~ 3 disparities in raw contrast between the channels, after subtraction the final contrasts all converge to within a factor of 2, reaching between 6.0×10^{-10} and 1.1×10^{-10} . In the interleaved RDI arrangement, the reference and target image stacks trace a similar wavefront evolution, so their time-averages are similar. As a result PSF subtraction (both classical and KLIP) is in general more effective than in the contiguous RDI case.

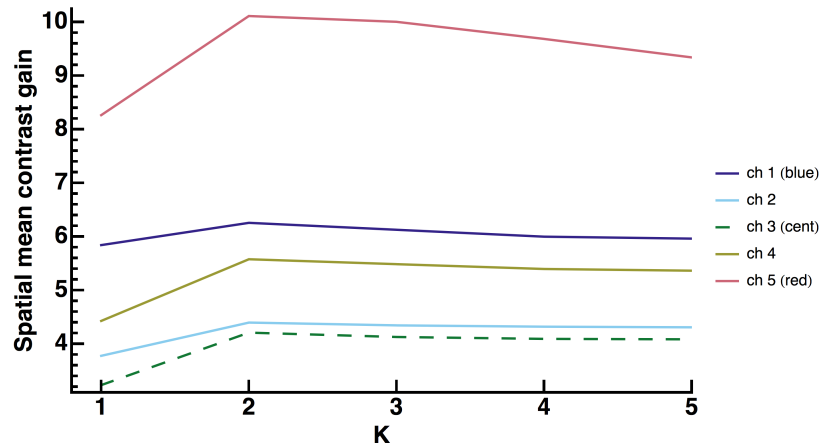


Figure 28: The ratio of the raw contrast to the contrast after KLIP subtraction, plotted against the number of KL modes, for the interleaved RDI case. We find an optimum subtraction with 2 KL modes at all wavelengths. The contrast gain factor is highest in the long-wavelength channel, where the speckle pattern is brightest with respect to the background noise, and most stable over time.

3.2.6 RDI using probe and regularization images together

The third RDI arrangement we tested is the case of using the small set of EFC-corrected probe images to form a target stack, and assigning the full regularization sequence for the reference stack. This scenario produced results with properties intermediate between the contiguous and interleaved RDI arrangements. Instead of repeating the same plots, we list the essential results of the probe–regularization combination in Table 2, along with those of the the previously described RDI arrangements.

| λ (nm) | I. Contiguous Regularization Seq. | | | | II. Interleaved Regularization Seq. | | | | III. Probe–Regularization Seq. | | | |
|----------------|-----------------------------------|----------------------|------|-----|-------------------------------------|-----------------------|------|-----|--------------------------------|-----------------------|------|-----|
| | raw (1σ) | fin. (1σ) | gain | K | raw (1σ) | fin. (1σ) | gain | K | raw (1σ) | fin. (1σ) | gain | K |
| 528 | 6.5×10^{-9} | 2.2×10^{-9} | 3.0 | 4 | 6.6×10^{-9} | 1.1×10^{-9} | 6.3 | 2 | 6.8×10^{-9} | 1.4×10^{-9} | 4.8 | 1 |
| 539 | 3.3×10^{-9} | 1.4×10^{-9} | 2.4 | 2 | 3.5×10^{-9} | 7.9×10^{-10} | 4.4 | 2 | 3.4×10^{-9} | 1.0×10^{-9} | 3.4 | 2 |
| 550 | 2.5×10^{-9} | 1.1×10^{-9} | 2.2 | 8 | 2.5×10^{-9} | 6.0×10^{-10} | 4.2 | 2 | 2.5×10^{-9} | 7.8×10^{-10} | 3.2 | 2 |
| 561 | 3.3×10^{-9} | 1.2×10^{-9} | 2.7 | 11 | 3.6×10^{-9} | 6.4×10^{-10} | 5.6 | 2 | 3.6×10^{-9} | 9.6×10^{-10} | 3.7 | 7 |
| 572 | 5.6×10^{-9} | 1.7×10^{-9} | 3.3 | 6 | 6.3×10^{-9} | 6.3×10^{-10} | 10.1 | 2 | 6.4×10^{-9} | 1.1×10^{-9} | 5.6 | 3 |

Table 2: Result summary of multi-wavelength data post-processing. For each RDI arrangement, we list the std. dev. of the raw target co-add, the final contrast after optimized KLIP subtraction, the contrast gain factor, and the optimal number of KL modes.

3.3 Summary of multi-wavelength RDI results

As we previously noted in Section 3.2.2, there is a significant level of spatially uncorrelated background noise in the CCD images. Across the wavelength channels, the ratio of speckle to background std. dev. in an individual raw image ranges from 2.3 to 5.9. One component of this noise is temporally correlated, while the other is present in each observation as a random pattern, consistent with read noise. We take this uncorrelated noise level, as inferred from the background sample boxes, and compare it to the contrast floor after KLIP subtraction. If we are combining N_t target images and subtracting a weighted average of N_r reference images, we can assume the residual uncorrelated noise is $\sigma_u \sqrt{\frac{1}{N_t} + \frac{1}{N_r}}$. Taking $N_r = N_t = 30$ and $\sigma_u = 1.64 \times 10^{-9}$ (appropriate for the center wavelength flux calibration), the uncorrelated noise floor of the final image from either the contiguous or interleaved RDI result should be about 4.2×10^{-10} . The interleaved RDI scenario approaches this contrast level (6.0×10^{-10} in center channel), so background noise may be a contributing factor, although not necessarily the dominant limitation.

Another question along these lines is whether we would benefit from a higher volume of data. In Figures 30 and Figure 29, we alter the contiguous and interleaved RDI trials slightly. We keep the same reference stack, but we vary the number of target images, to see how quickly the contrast is depressed, and what the trend is at the final image. In the contiguous RDI arrangement, there is little improvement beyond 20 images. Perhaps, by the 20th target frame, the speckle patterns represented in the reference stack are dissimilar enough that additional target images gain little benefit. For the interleaved RDI arrangement, on the other hand, the contrast improves with each

step before we run out of data at image number 30. Therefore, it appears likely that we would reach a deeper final contrast with a higher volume of images acquired under the same conditions. As before, however, the result would be highly dependent on how we assign images to the reference and target stacks.

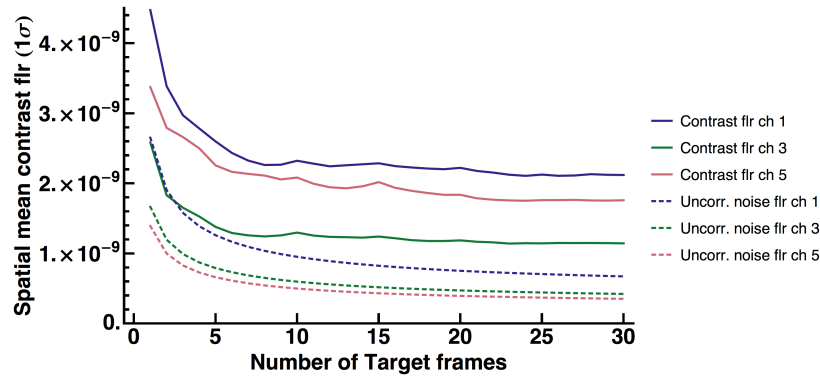


Figure 29: Contrast floor versus number of target images, contiguous RDI case. The noise floor attributed to uncorrelated background noise is plotted alongside the final contrast, taking into account the number of reference and target frames, and the varying background noise between channels.

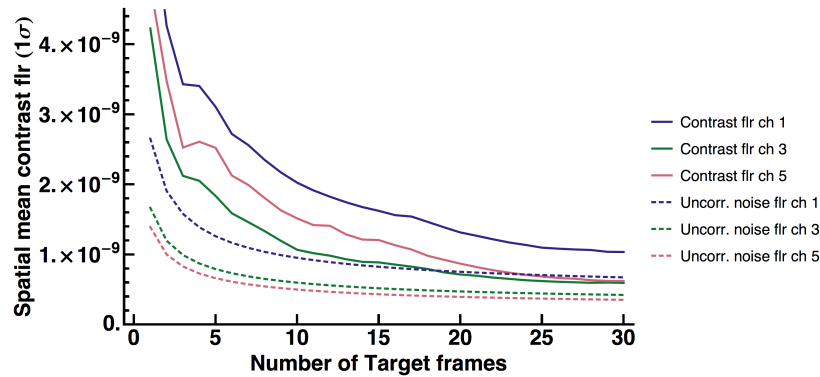


Figure 30: Contrast floor versus number of target frames, interleaved RDI case. The noise floor attributed to uncorrelated background noise is plotted alongside the final contrast, taking into account the number of reference and target frames, and the varying background noise between channels.

4 Algorithm developments for IFS data analysis

Reduction artifacts and/or systematic calibration biases in spectral cubes may set limits to how well PSFs can be subtracted, and to how well the spectra of planets can be measured. For this reason the careful assembly of the spectral datacubes will be an essential step of the process for the CGI IFS on WFIRST.

This calibration process will be investigated through the PISCES prototype IFS being developed by GSFC for use with the HCIT. The second task of this contract supports STScI involvement in partnership with GSFC and the WFIRST project in preparation for the use of PISCES at HCIT, building on lessons learned from current IFS instruments such as the Gemini Planet Imager (GPI). Due to resource constraints and competing priorities, in this fiscal year this task was smaller in scope compared with the other two tasks (but could potentially be expanded in the future once PISCES is operational, if necessary).

As this task is primarily advisory/collaborative in nature, technical content from this task is primarily incorporated into GSFC PISCES software and documents, including two recent conference presentations [26, 27]. This present report briefly summarizes contributions and progress.

The most complex tasks in an IFS data pipeline are wavelength calibration and data cube assembly, which involve mapping the precise location of key wavelengths for each lenslet across the field of view, and then using that knowledge perform an accurate extraction of each spectral element from the 2D detector image. To efficiently achieve these key steps, the PISCES team has adapted software from the GPI data pipeline [28, 29] as described in Ref [26] section 4. Perrin provided advice and guidance to the GSFC team regarding the use and adaptation of the GPI data pipeline. The GPI data pipeline’s algorithms for wavelength calibration from multiple narrow band images, and for data cube assembly based on that calibration, were successfully adapted to work on simulated PISCES data. To accommodate the nonlinear dispersion of PISCES while still achieving the required accuracy, the wavelength calibration code was extended to use a higher order polynomial model. This code is now working and available to support commissioning of PISCES at HCIT. Perrin also supported the PISCES team in discussions of the measurement and use of microlens PSFs based on a library of calibration monochromatic images obtained with a tunable calibration source. This is a promising method for delivering very high quality reconstructed data cubes which is especially well suited to a high stability space instrument. See [26] for more details.

On 2015 May 18, the project held the so-called “5th Stanford meeting” as a teleconference focused on the IFS for CGI. Perrin organized a discussion on IFS calibrations, specifically the processing steps and instrument calibration knowledge needed to convert data from raw detector arrays to spectrophotometrically accurate datacubes. Calibration methods and lessons learned from all of the ground-based high contrast IFSes were summarized, based on inputs from D. Mawet (for the SPHERE IFS), L. Pueyo (Project 1640), T. Groff (CHARIS), as well as Perrin’s first-hand knowledge

of the GPI IFS. These discussions continued intermittently in planning telecons for PISCES.

A key issue is that all existing high contrast IFSes make extensive use of multiple types of internal calibration lamps. Calibrations obtained in this manner include:

1. lenslet geometry and wavelength calibrations, via Xe and Ar arc lamps for GPI, and fixed or tunable laser sources for SPHERE and CHARIS.
2. Flat fielding, including both low-spatial-frequency components from the optical system before the lenslet array (via broadband flat lamps for all instruments) and high-spatial-frequency components from either the lenslet itself or detector pixel sensitivity variations (which are generally difficult to measure after the lenslet array).
3. Measurements of individual lenslet PSFs either monochromatically or in broadband light, for use in accurate spectral extraction and/or calibration of cross-talk between adjacent spectra.

For WFIRST CGI, we *recommend that careful consideration be given to whether one or more calibration lamps are needed in the IFS*. Specifically at least one near-monochromatic source would enable on-orbit mapping of the dispersed spectral geometries (spectral locations and wavelength solutions). This could relax fabrication and launch vibe tolerances, which might otherwise be required to maintain the lenslet-to-detector positional mapping to better than a pixel across the launch. Alternatively, it might be possible to use astrophysical emission nebulae as a wavelength calibration source on orbit, but this would require telescope on-sky time rather than being an internal calibration. HST spectrographs have generally chosen to include internal calibration lamps. This should be studied further during Phase A. We also *recommend that the detector used in the CGI IFS should be calibrated as thoroughly as possible on the ground prior to being installed behind the lenslet array, at a minimum obtaining high SNR flat fields at all wavelengths of interest*. Such measurements will be essentially impossible to obtain on orbit, but since pixel quantum efficiencies tend to be relatively stable over time, it is likely that ground calibrations will suffice.

Also essential are astrometric and photometric calibrations of the IFS, which the ground-based IFSes obtain using reference “satellite spots” to enable measurements of a planet’s brightness and astrometry relative to the occulted central star, as well as periodic observations of calibration star fields and/or lab measurements of calibration pinhole grids. The post-processing techniques we study in this report depend on being able to accurately align the individual coronagraphic observations. The greater stability in space and the ability of the CGI LOWFS to correct for slow drifts means that per-image registration should be less necessary for WFIRST. Further simulations work for the CGI should assess quantitatively how well astrometry relative to the star can be assessed, and what type(s) of on-sky calibration observations will be needed to support astrometric calibrations. In particular, while we expect the star to be centered relative to the coronagraph to $\ll \lambda/D$ accuracy, for accurate astrometry we must also know the precise location of the coronagraph center

relative to the IFS lenslet array (and to the imager detector pixels) at a similar level. We note that in addition to impacting the efficacy of PSF subtractions, achieving as high as possible astrometric precision directly contributes to mission science by enabling measurements of planetary orbital motions.

As noted in section 2 of this report, the forward modeling process requires knowledge of the off-axis planetary PSF. While this could be obtained from numerical modeling, for the highest accuracy empirical measurements are desired. The planned ND4 filter in the CGI will enable such measurements on orbit in any of the CGI modes, including with the IFS. While the purpose of this is enabling target acquisitions with the imager, provisions should be made for periodically obtaining IFS observations of unocculted stars with the ND4 for use in forward modeling. At the precision needed, the off-axis PSF should be very stable and will likely need calibration no more often than annually.

5 Conclusions and Perspectives

We have applied PSF subtraction algorithms, including classical subtraction and the PCA-based KLIP algorithm, to reduce and analyze both simulated and experimental images and both broadband and multi-wavelength with a view to detecting and characterizing exoplanets with the future space-based WFIRST coronagraph instrument.

- *For simulated imaging data with HLC coronagraph*, we report contrast gains from 10 to 30 on broadband images, enabling detection of 3 injected planets with a range of contrasts from 5.7×10^{-9} to 0.8×10^{-9} . The stability of aberrations has a large impact on the relative efficacy of the two methods. The performance of classical PSF subtraction is limited by thermal variations induced by the observing scenario. If these variations are mitigated by a LOWFSC system, classical subtraction achieves contrast gains of 6–15, but without a LOWFS classical subtraction is limited to contrast gains < 5 . On the other hand, KLIP is more robust to these variations and achieves contrast gains > 10 on the simulated data either with or without the LOWFS. KLIP is also more effective at compensating for the higher levels of wavefront instability in datasets where we scaled up the aberrations. While the use of a LOWFS for WFIRST CGI is planned and the observatory is being designed to be very stable, given the inevitable uncertainties about on-orbit stability before flight, the KLIP subtraction appears to provide useful robustness against higher levels of temporal instability.
- *For simulated IFS data with the SPC coronagraph*, we demonstrated that the KLIP algorithm in combination with the forward modeling technique enables an unbiased retrieval of a Jupiter-like spectrum at a contrast of 4.2×10^{-9} . The quality of spectrum extraction depends on several factors, including the reduction technique that we use, the reduction parameters such as number of KL modes, and prior knowledge of input parameters such as the planet's expected astrometric position and the PSF of an off-axis source. Forward modeling is essential to correct for the algorithm throughput loss. Furthermore, PSF subtraction and spectral extraction will be limited by the quality of the input datacubes, so we assessed necessary calibration inputs for the IFS and have contributed to development of the IFS pipeline for use with the PISCES prototype. These results set the stage to develop an error budget estimation methodology for spectrum extraction. Future work will further quantify the dominant contributors, the requirements for calibrations, and the impact of the reduction parameters; such an error budget will be valuable when confronting PSF subtraction algorithms with experimental IFS data from PISCES in the near future.
- *For broadband coronagraph lab data with SPC coronagraph*, the reduction of broadband coronagraphic data obtained in JPL's HCIT under static conditions demonstrated contrast gains up to 3 averaged on one side of the bow-tie region and up to 8 locally. The speckle

noise statistics in the chosen data set are such that KLIP enables at least $2\times$ better contrast gains than classical PSF subtraction. With a carefully chosen strategy, we were able to detect an injected planet with a contrast of 2×10^{-9} at a separation of $8 \lambda/D$. The optimization involved selecting the 50 most correlated images of the reference cube as the reference library and choosing 20 modes for use in PSF subtraction. This strategy will be refined to take into account the spectral extraction when processing future HCIT data. Further evaluation should continue as the temporal characteristics likely to apply to WFIRST CGI data become clearer, both through integrated modeling and the planned dynamical environment tests at HCIT.

- *For multi-wavelength coronagraphic lab data with SPC coronagraph* obtained in JPL’s HCIT under static conditions confirms that the way the reference and target subsets are arranged has an important impact on the reductions. The current study is constrained by the fact that we are assembling surrogate “target” and “reference” datasets out of HCIT experimental data originally taken for another purpose, which imperfectly at best capture the temporal evolution appropriate for data in flight. Nonetheless by assessing several different ways of assembling the “target” and “reference” datasets we were able to partially explore the relevant parameter space. Future work could refine this arrangement to better take into account realistic WFIRST observing scenarios, particularly as additional test dataset become available for the dynamic test environment.

This study has also led us to better understand the chromatic behavior of the speckles and its impact on the reductions. It has long been recognized that a significant challenge for coronagraphic space missions is generating and maintaining a dark hole in *broadband* light; contrast is typically best at a specific monochromatic wavelength and degrades toward either longer and shorter wavelengths. This behavior is rooted in fundamental physical scaling laws and in practice appears likely to limit observations to $\leq 10\%$ wide spectral bandwidths, with decreasing contrast performance near the edges of that bandwidth. We found that PSF subtraction may help to overcome this: the higher intensity speckles in the extreme wavelength channels are *temporally relatively static*, and therefore subtract well. Regardless of which way we assembled the “target” and “reference” sequences from the HCIT data, we found the greatest contrast gains in the extreme short and long wavelength channels, and after the PSF subtraction process the dispersion in contrasts between the 5 wavelengths was reduced compared to the dispersion in contrasts of the multi-wavelength raw data. This behavior is reasonable: the degradation of contrast away from the central wavelength is a property of the static optical solution to the EFC process, and would pertain even in a perfectly temporally stable system. Therefore it is well characterized in the reference data set and removed in PSF subtraction. Higher order effects mean that this works more effectively for the reddest channel than for the bluest one, but nonetheless even for that blue channel the post-subtraction contrast becomes relatively close to that of the central wavelength. We caution that we

have as yet only assessed this behavior for a single experimental dataset, but it is plausible that this will turn out to be more generally applicable. If so, this is likely to be highly relevant for WFIRST and beyond, and may help mitigate some of the challenges of broadband dark hole control.

We are now confident that using existing PSF subtraction algorithms such as KLIP, we can achieve contrast gains of 10–30 for broadband simulated images using an HLC coronagraph and 2–4 for both broadband and multi-spectral experimental HCIT data obtained using a SPC coronagraph. These are the best results ever obtained on post-processed WFIRST-like images. Using the probe regularization sequence on HCIT data we reach a final 1σ average contrast of 1×10^{-9} or better depending on the wavelength channel. These sequences represent somewhat artificially an actual observing sequence, but is something worth exploring to take advantage of the all the data available during dark hole optimization, therefore potentially improving observatory efficiency. With the acquisition of longer and more detailed sequences with HCIT, in particular during “dynamic” testing, we will be able to investigate the optimal sets of targets and references along the optimization iterations, and to develop the observing scenario together with post processing.

Under the assumption that selecting contiguous sub-sequences of the HCIT data yields target and reference cubes whose correlation properties are perhaps most representative of conditions in flight, we can extrapolate to the flight mission performance. In this case, based on the results presented in this report we conclude that contrast gains from post-processing between 2 and 4 are likely for WFIRST CGI, for an observing scenario involving one reference star to calibrate the speckle field of a science target observed subsequently. This value may be taken as a lower limit. If the DM actuator settings are frozen between reference and target stars, as was the case in the OS1 and OS3 simulations, then the results of the interleaved RDI arrangement of multi-wavelength lab data might offer a more relevant prediction, especially in the presence of LOWFSC that will stabilize the wavefront. Indeed, the contrast gains achieved for the interleaved testbed images, ranging from 4 to 10 depending on wavelength, approach the factors of 10 or more we found for the simulated HLC observation.

Further gains in post-processing are potentially achievable in other ways. For instance, observations with the telescope rolled to two isothermal positions (e.g. the $\pm 13^\circ$ ADI case described in recent observing sequence ops concepts) may increase the available pool of references while maintaining high correlation in PSF properties. At the other temporal extreme, it might prove to be the case that PSFs taken across longer time spans would still be sufficiently correlated to contribute useful diversity to the reference pool and enable improved subtractions. While this has been demonstrated for Hubble operating at 10^{-5} contrast, it is more speculative here and perhaps unlikely in the case of WFIRST at 10^{-9} . These topics should be revisited as observatory optical modeling and operations planning continue to mature.

6 Acknowledgments

This WFIRST study was carried out at Space Telescope Science Institute (STScI) under subcontract No. 1506553, funded by the Jet Propulsion Laboratory (JPL), which is managed for NASA by the California Institute of Technology. The authors are grateful to John Krist for providing simulations used in this study, and to Bertrand Mennesson, John Krist, Bijan Nemati and Gautam Vasisht for helpful discussions.

7 Acronym list

AAS - American Astronomical Society
 AFTA - Astrophysics Focused Telescope Assets
 ALICE - Archival Legacy Investigations of Circumstellar Environments
 CPSFTS - Coronagraphic PSF template subtraction
 DM - Deformable Mirror
 GSFC - Goddard Space Flight Center
 HCIT - High-Contrast Imaging Testbed
 HLC - Hybrid Lyot Coronagraph
 HST- Hubble Space Telescope
 IFS - Integral Field Spectrograph
 JATIS - Journal of Astronomical Telescopes, Instruments, and Systems
 JPL - Jet Propulsion Laboratory
 JWST - James Webb Space Telescope
 KLIP - Karhunen-Loève Image Projection Algorithm
 LOWFC - Low-Order Wavefront Control
 LOWFSC - Low-Order Wavefront Sensing and Control
 NICMOS - Near Infrared Camera and Multi-Object Spectrometer
 OS - Observing Sequence
 PCA - Principal Component Analysis
 PSF - Point Spread Function
 RDI - Reference Star Differential Imaging
 SDI - Spectral Differential Imaging
 SPC - Shaped Pupil Coronagraph
 SPIE - Society of Photographic Instrumentation Engineers
 STScI - Space Telescope Science Institute
 WFC3 - Wide Field Camera 3
 WFIRST - Wide Field InfraRed Survey Telescope

References

- [1] R. Soummer, L. Pueyo, and J. Larkin. Detection and Characterization of Exoplanets and Disks Using Projections on Karhunen-Loève Eigenimages. *Astrophys. J. Lett.*, 755:L28, August 2012.
- [2] R. Soummer, M. D. Perrin, L. Pueyo, É. Choquet, C. Chen, D. A. Golimowski, J. Brendan Hagan, T. Mittal, M. Mörchen, M. N'Diaye, A. Rajan, S. Wolff, J. Debes, D. C. Hines, and G. Schneider. Five Debris Disks Newly Revealed in Scattered Light from the Hubble Space Telescope NICMOS Archive. *Astrophys. J. Lett.*, 786:L23, May 2014.

- [3] E. Choquet, J. B. Hagan, L. Pueyo, M. D. Perrin, D. C. Hines, C. Chen, G. Schneider, J. Debes, D. Golimowski, N. Reid, T. Mittal, M. Moerchen, M. N'Diaye, A. Rajan, S. Lonsdale, and R. Soummer. Archival Legacy Investigation of Circumstellar Environments using KLIP algorithm on HST NICMOS coronagraphic data. In M. Booth, B. C. Matthews, and J. R. Graham, editors, *IAU Symposium*, volume 299 of *IAU Symposium*, pages 30–31, January 2014.
- [4] A. Rajan, T. Barman, R. Soummer, L. Pueyo, J. Patience, J. Brendan Hagan, B. Macintosh, C. Marois, and Q. M. Konopacky. Detection and characterization of the atmospheres of the HR 8799 b and c planets with high contrast HST/WFC3 imaging. In *American Astronomical Society Meeting Abstracts*, volume 225 of *American Astronomical Society Meeting Abstracts*, page #323.07, January 2015.
- [5] A. Amara and S. P. Quanz. PYNPOINT: an image processing package for finding exoplanets. *Mon. Not. R. Astron. Soc.*, 427:948–955, December 2012.
- [6] R. Fergus, D. W. Hogg, R. Oppenheimer, D. Brenner, and L. Pueyo. S4: A Spatial-spectral model for Speckle Suppression. *ApJ*, 794:161, October 2014.
- [7] D. Lafrenière, C. Marois, R. Doyon, D. Nadeau, and É. Artigau. A New Algorithm for Point-Spread Function Subtraction in High-Contrast Imaging: A Demonstration with Angular Differential Imaging. *ApJ*, 660:770–780, May 2007.
- [8] C. Marois, B. Macintosh, and J.-P. Véran. Exoplanet imaging with LOCI processing: photometry and astrometry with the new SOSIE pipeline. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 7736 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 1, July 2010.
- [9] Luca Caucci, Harrison H. Barrett, Nicholas Devaney, and Jeffrey J. Rodríguez. Application of the hotelling and ideal observers to detection and localization of exoplanets. *J. Opt. Soc. Am. A*, 24(12):B13–B24, Dec 2007.
- [10] D. Savransky. Sequential Covariance Calculation for Exoplanet Image Processing. *ApJ*, 800:100, February 2015.
- [11] G. Schneider, M. D. Silverstone, E. Stobie, J. H. Rhee, and D. C. Hines. NICMOS Coronagraphy: Recalibration and the NICMOS Legacy Archive PSF Library. In *Hubble after SM4. Preparing JWST*, page 15, July 2010.
- [12] D. Mawet, L. Pueyo, P. Lawson, L. Mugnier, W. Traub, A. Boccaletti, J. T. Trauger, S. Gladysz, E. Serabyn, J. Milli, R. Belikov, M. Kasper, P. Baudoz, B. Macintosh, C. Marois, B. Oppenheimer, H. Barrett, J.-L. Beuzit, N. Devaney, J. Girard, O. Guyon, J. Krist, B. Mennesson, D. Mouillet, N. Murakami, L. Poyneer, D. Savransky, C. Véraud, and J. K. Wallace. Review of small-angle coronagraphic techniques in the wake of ground-based second-generation adaptive optics systems. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 8442 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 4, September 2012.
- [13] M. Ygouf, R. Soummer, M. D. Perrin, L. Pueyo, M. N'Diaye, and B. Macintosh. Post-processing methods for high-contrast imaging in the context of the WFIRST-AFTA telescope. In *American Astronomical Society Meeting Abstracts*, volume 225 of *American Astronomical Society Meeting Abstracts*, page #338.30, January 2015.
- [14] M. Ygouf, L. Pueyo, R. and. Soummer, M. Perrin, B. Mennesson, J. Krist, E. Cady, G. Vasisht, B. Nemati, and B. Macintosh. PSF-subtraction for the WFIRST-AFTA telescope. In *In the Spirit of Lyot 2015*, June 2015.
- [15] M. Ygouf, L. Pueyo, R. and. Soummer, M. Perrin, R. van der Marel, and B. Macintosh. Data Processing and Algorithm Development for the WFIRST-AFTA Coronagraph: Reduction of Noise Free Simulated Images, Analysis and Spectrum Extraction with Reference Star Differential Imaging. In *Society of Photo-Optical Instrumentation*

- Engineers (SPIE) Conference Series*, volume 9605 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, August 2015.
- [16] M. Ygouf, L. Pueyo, N. T. Zimmerman, R. Soummer, M. D. Perrin, B. Mennesson, J. E. Krist, G. Vasisht, B. Nemati, and B. Macintosh. PSF subtraction for the WFIRST-AFTA coronagraph. In *American Astronomical Society Meeting Abstracts*, volume 227 of *American Astronomical Society Meeting Abstracts*, page 147.18, January 2016.
 - [17] N. T. Zimmerman, M. Ygouf, L. Pueyo, R. Soummer, M. D. Perrin, B. Mennesson, E. Cady, and C. Mejia Prada. Post-processing images from the WFIRST-AFTA coronagraph testbed. In *American Astronomical Society Meeting Abstracts*, volume 227 of *American Astronomical Society Meeting Abstracts*, page 147.17, January 2016.
 - [18] J. E. Krist. End-to-end numerical modeling of AFTA coronagraphs. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 9143 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 0, August 2014.
 - [19] J. Krist, B. Nemati, and B. Mennesson. Numerical modelling of the proposed WFIRST-AFTA coronagraphs and their predicted performances. *Journal of Astronomical Telescopes, Instruments, and Systems*, 2(1), January 2016.
 - [20] L. Pueyo, R. Soummer, J. Hoffmann, R. Oppenheimer, J. R. Graham, N. Zimmerman, C. Zhai, J. K. Wallace, F. Vescelus, A. Veicht, G. Vasisht, T. Truong, A. Sivaramakrishnan, M. Shao, L. C. Roberts, Jr., J. E. Roberts, E. Rice, I. R. Parry, R. Nilsson, S. Luszcz-Cook, T. Lockhart, E. R. Ligon, D. King, S. Hinkley, L. Hillenbrand, D. Hale, R. Dekany, J. R. Crepp, E. Cady, R. Burruss, D. Brenner, C. Beichman, and C. Baranec. Reconnaissance of the HR 8799 Exosolar System II: Astrometry and Orbital Motion. *ArXiv e-prints*, September 2014.
 - [21] L. Pueyo. Detection and characterization of exoplanets using projections on Karhunen-Loève eigenimages: forward modeling. *submitted to Astrophys. J.*
 - [22] C. Aime and R. Soummer. The Usefulness and Limits of Coronagraphy in the Presence of Pinned Speckles. *ApJL*, 612:L85–L88, September 2004.
 - [23] R. Soummer, A. Ferrari, C. Aime, and L. Jolissaint. Speckle Noise and Dynamic Range in Coronagraphic Images. *ApJ*, 669:642–656, November 2007.
 - [24] A. J. E. Riggs, N. Zimmerman, A. Carlotti, N. J. Kasdin, and R. Vanderbei. Shaped pupil design for future space telescopes. volume 9143 of *Proc. SPIE*, August 2014.
 - [25] E. Cady, Camilo Mejia Prada, Xin An, K. Balasubramanian, Rosemary Diaz, N. Jeremy Kasdin, Brian Kern, Andreas Kuhnert, Bijan Nemati, Ilya Poberezhskiy, A. J. Eldorado Riggs, Roger Zimmer, and Neil T. Zimmerman. Demonstration of high contrast with an obscured aperture with the WFIRST-AFTA shaped pupil coronagraph. *SPIE JATIS*, 2(1), January 2016.
 - [26] Jorge Llop Sayson, Nargess Memarsadeghi, Michael W. McElwain, Qian Gong, Marshall Perrin, Timothy Brandt, Bryan Grammer, Bradford Greeley, George Hilton, and Catherine Marx. Pisces: high contrast integral field spectrograph simulations and data reduction pipeline, 2015.
 - [27] Qian Gong, Michael McElwain, Bradford Greeley, Bryan Grammer, Catherine Marx, Nargess Memarsadeghi, Karl Stapelfeldt, George Hilton, Jorge Llop Sayson, Marshall Perrin, Richard Demers, Hong Tang, Brian Kern, and Janan Ferdosi. Prototype imaging spectrograph for coronagraphic exoplanet studies (pisces) for wfirst/afta, 2015.

- [28] M. D. Perrin, J. Maire, P. Ingraham, D. Savransky, M. Millar-Blanchaer, S. G. Wolff, J.-B. Ruffio, J. J. Wang, Z. H. Draper, N. Sadakuni, C. Marois, A. Rajan, M. P. Fitzgerald, B. Macintosh, J. R. Graham, R. Doyon, J. E. Larkin, J. K. Chilcote, S. J. Goodsell, D. W. Palmer, K. Labrie, M. Beaulieu, R. J. De Rosa, A. Z. Greenbaum, M. Hartung, P. Hibon, Q. Konopacky, D. Lafreniere, J.-F. Lavigne, F. Marchis, J. Patience, L. Pueyo, F. T. Rantaky  , R. Soummer, A. Sivaramakrishnan, S. Thomas, K. Ward-Duong, and S. Wiktorowicz. Gemini Planet Imager observational calibrations I: Overview of the GPI data reduction pipeline. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 9147 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 3, July 2014.
- [29] S. G. Wolff, M. D. Perrin, J. Maire, P. J. Ingraham, F. T. Rantaky  , and P. Hibon. Gemini planet imager observational calibrations IV: wavelength calibration and flexure correction for the integral field spectograph. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 9147 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 7, August 2014.