WFIRST-AFTA Coronagraph Milestone 3 final report: PIAACMC mask fabrication and characterization

Brian Kern, Olivier Guyon, Rus Belikov, Dan Wilson, Rich Muller, Erkin Sidick, Bala Balasubramanian, John Krist, Ilya Poberezhskiy

March 27, 2015

1. Introduction

Phased-Induced Amplitude Apodization Complex Mask Coronagraph (PIAACMC) was selected as the backup technology for the WFIRST-AFTA coronagraph instrument. In agreement with NASA headquarters and the Technology Assessment Committee (TAC), the WFIRST-AFTA coronagraph technology development milestone 3 was defined as:

> First-generation PIAACMC focal plane phase mask with at least 12 concentric rings has been fabricated and characterized; results are consistent with model predictions of 10^{-8} raw contrast with 10% broadband light centered at 550 nm.

This document describes the fabrication, characterization, and performance modeling of a PIAACMC mask. Models indicate that the first mask fabricated, in a baseline WFIRST-AFTA PIAACMC configuration with realistic wavefront errors on all optical surfaces, would be capable of observing more than 12 known radial velocity (RV) planets, assuming 0.4 mas/axis rms residual pointing errors, 1 mas diameter star, 30× post-processing gain, and a minimum SNR of 5. This meets draft Cycle 5 requirements from the Science Definition Team (SDT) established for the coronagraph instrument. There are a number of ready improvements to the mask fabrication quality, and the expected effects of each are detailed in this report.

L1 Product	Rqmt	Performance Assessment	Comment
Exoplanet images	> 12 planets	15 RV planets imaged with “as-made” PIAACMC mask	PIAACMC meets rqmt

Table 1. L1 draft Cycle 5 science requirements for WFIRST-AFTA coronagraph. Taken from WFIRST CGI L1 Requirement Summary Document, W. Traub, 10/01/2014.

2. PIAACMC on WFIRST-AFTA

PIAACMC was selected as the backup coronagraph architecture for the WFIRST-AFTA coronagraph instrument. A pair of aspheric mirrors, which apodize the input beam, is an integral part of the coronagraphic operation. These mirrors are not compatible with the primary Occulting Mask Coronagraph (OMC) architecture, so the PIAACMC optical layout is necessarily independent of the OMC layout.

The PIAACMC design discussed in this report is referred to as "Gen 2", to differentiate it from the design originally proposed during WFIRST-AFTA coronagraph architecture down-select process in November 2013. The optical layout specific to Gen 2 PIAACMC was chosen to use only a single deformable mirror (DM), to operate without pupil relay optics, and to illuminate the occulter plane with an F/80 beam (a very slow system). In addition, the Gen 2 PIAACMC occulter is entirely reflective, as contrasted with the transmissive OMC occulters. An illustration of the PIAACMC optical layout is shown in Fig. 1. The entire layout fits comfortably inside the 1 m × 1 m nominal instrument outline. The configuration of the optics feeding the instrument are subject to design revisions, but there should be no particular complications adapting this basic layout to changes in the surrounding components.

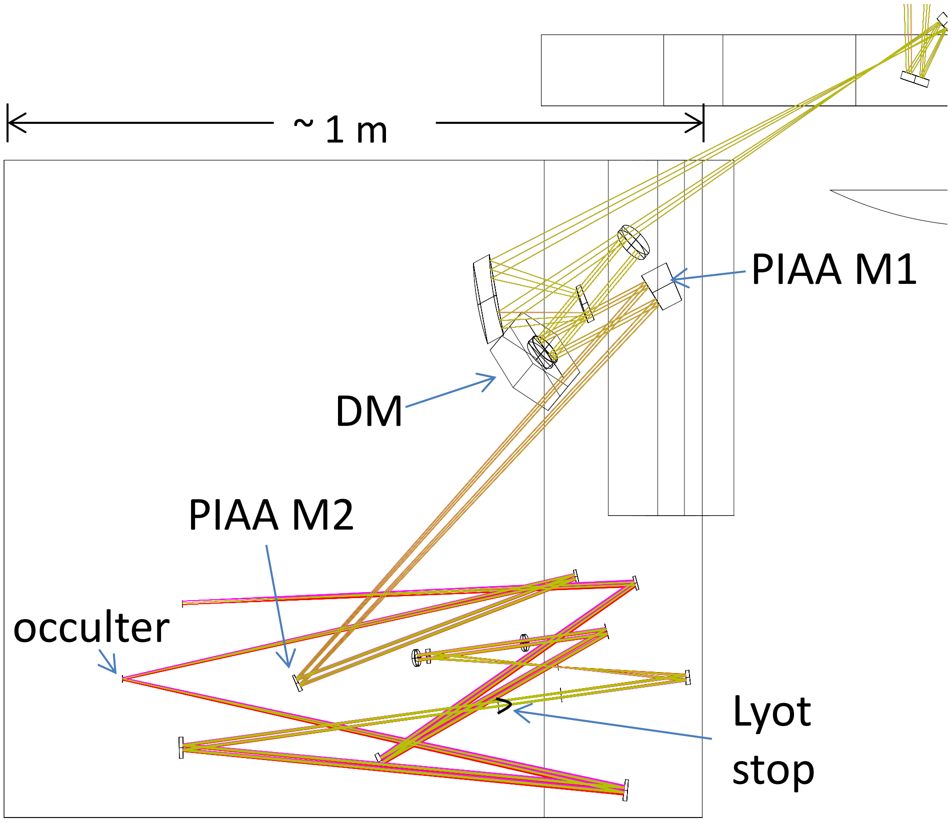


Fig. 1. Gen 2 PIAACMC optical layout. The main differences of this layout to the OMC layout is the lack of a second DM, the lack of pupil relays, beam compression before the occulter to produce a slow beam, and a reflective occulter. This layout and the drawing were provided by Hong Tang.

The choice to use only a single DM was made to simplify the instrument, as each DM and its electronics add cost and complexity, in addition to simple mass and power, to the instrument. Using a single DM typically restricts the high-contrast coronagraph field-of-view to half of the image plane. Wavefront control with high contrast over a full 360-degree field-of-view typically requires two DMs. All analysis in this report is for wavefront control over only half of the image plane.

To categorize the optical differences between PIAACMC and OMC, there are three noteworthy components: the aspheric PIAA mirrors, the reflective PIAA occulting mask, and the elongated Lyot stop. The PIAA mirrors for Gen 2 have significantly milder aspheric shapes than for previous PIAA designs. The mirror surfaces differ from conic surfaces by only 2.1 μm P-V, as compared with 50 μm on previously fabricated PIAA mirrors used for high-contrast testbed demonstrations. Fig. 2 shows the aspheric components of the surfaces of the Gen 2 mirrors.

Fig. 2. Aspheric components of Gen 2 PIAA mirror shapes. The P-V deviation from a conic surface is 2.1 μm.

The Lyot stop is a single component physically, but is not contained in a single plane. It can either be implemented as a small fixture (~ 30 mm deep) containing 3-4 planes of stops, or as a single elongated shape. A final decision on this construction must fold in considerations of the LOWFS optical arrangement for PIAACMC, which has not been completed.

The Gen 2 PIAACMC occulting mask is completely reflective, with a profiled surface giving rise to phase variations that are integral to the PIAACMC coronagraphic operation. The details of this mask are what compose the bulk of the later sections of this report. Notably, to operate broadband, the first-order notion of the PIAACMC operation requires a mask that has a uniform non-unity amplitude and π phase shift for all wavelengths, and whose diameter scales linearly with wavelength. This is accomplished by something analogous to pulse-width modulation of the phase shift over sub-λ/D spatial scales.

Taken as a coronagraphic system, the occulting mask stands out as being untested. This fact is the original motivation for defining this milestone, which directly addresses the viability of the PIAACMC coroangraphic concept by fabricating, characterizing, and modeling a PIAACMC occulting mask, with an eye to broadband performance.

3. PIAACMC Performance Overview

The main differentiator in performance between PIAACMC and OMC is the small inner working angle (IWA) of PIAACMC, more than a factor of 2× smaller than for OMC (1.3 λ/D compared to 3 λ/D). In addition to simply increasing the "search space" for planets, at smaller IWA there are more planets at less extreme contrasts (easier to observe) and for whom the observed planet flux is higher. Taken together, at smaller IWA there are more targets, that are easier to observe, in a shorter time. A scatter plot of RV planets is shown in Fig. 3.

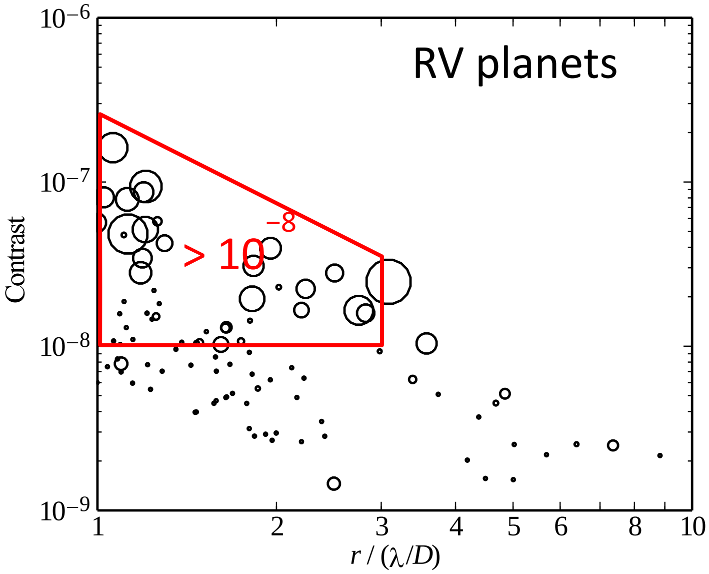


Fig. 3. Contrast of radial velocity planets. Inside of 3 λ/D, many planets have contrast $> 10^{-8}$ (easier to observe than 10^{-8}). The size of the symbols is a logarithmic representation of observed planet flux, showing that the observable planets at small *r* have higher observed fluxes. The region outlined in red simply draws attention to bright planets at small separations with contrast $> 10^{-8}$.

As can be seen in Fig. 3, there is a large population of planets at $r < 3\ \lambda/D$ whose contrast is $> 10^{-8}$ (less extreme contrast). At the time that WFIRST-AFTA Coronagraph Milestone 3 was defined, the details of the IWA and planet distributions were not folded into the relevant contrast threshold for the demonstration. While 10^{-8} contrast is a threshold for detecting planets at larger separations, at small separations ($r < 3\ \lambda/D$) many planets can be observed at larger contrasts. As such, meeting the science requirement (total number of observable planets) is considered a more relevant metric to the milestone than is an absolute threshold in contrast, irrespective of separation.

The nominal Gen 2 design, with realistic surface errors on all optics but a perfectly fabricated mask, delivers 24 planets in an observing bandpass of 10% centered at 550 nm, assuming 0.4 mas/axis residual tip-tilt and a 1 mas stellar diameter, as shown in Fig. 4. Also shown in Fig. 4 is the performance of the as-fabricated occulter, characterized as described in detail in later sections, delivering 15 planets. The performance of the as-fabricated occulter depends strongly on choices made during wavefront control (described in later section), so this single case study can be taken as a nominal example. Relative to the science requirement of 12 planets, this as-fabricated occulter succeeds in meeting the milestone criterion. The raw contrast of this choice drops below 10^{-8} at approximately 5 λ/D.

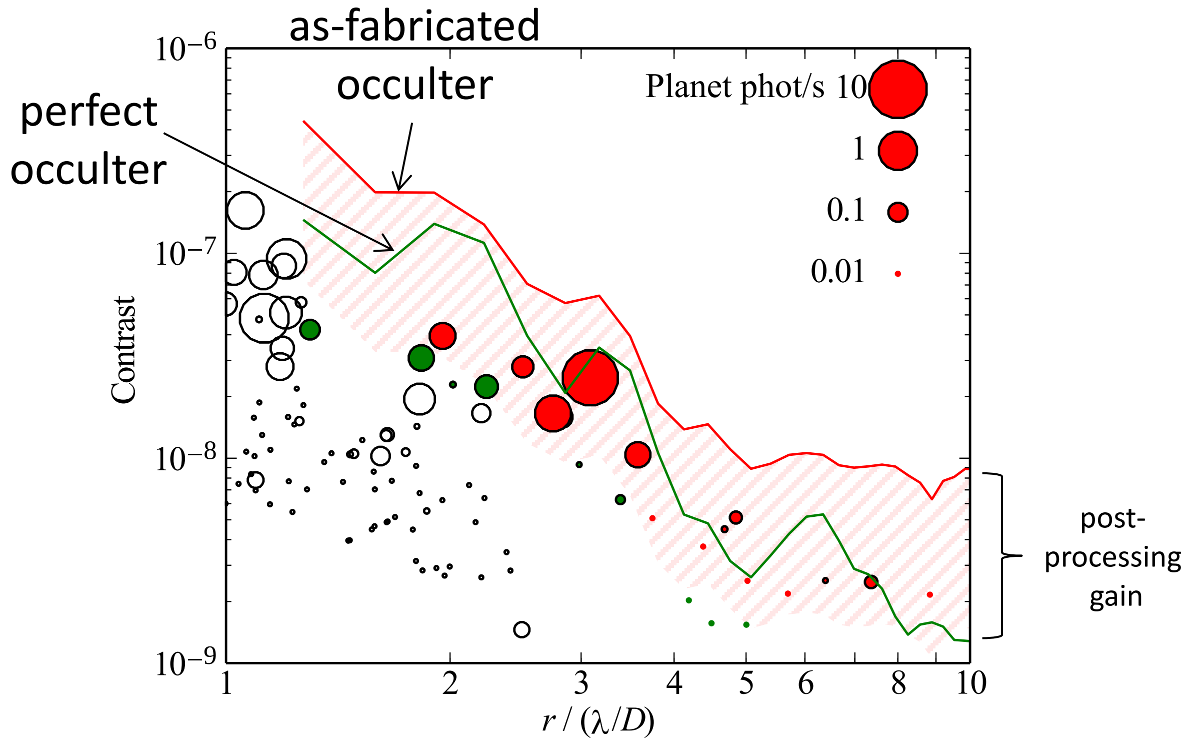


Fig. 4. Nominal contrast performance of two different occulters. The red line is the raw contrast for a nominal wavefront control choice applied to the as-fabricated occulter, the green line is raw contrast for a perfect occulter. The red symbols are planets observable with postprocessing on the as-fabricated or the perferct occulter, the green symbols are observable only with the perfect occulter. The red hatched region shows the 30× postprocessing gain.

4. Fabrication of PIAACMC Occulting Mask

The Gen 2 PIAACMC occulting mask is all-reflective, with 1259 "zones" of different heights, as shown in Fig. 5. This mask has been fabricated at JPL's Micro Devices Lab (MDL), using polymethyl methacrylate (PMMA) on a Si substrate. The height pattern was written onto PMMA using a 100 keV e-beam, with a 70 nm beam size, and 20 nm sub-pixel size. The PMMA itself can maintain 60 nm resolution in this context. Because the substrate does not need to transmit light, a Si substrate was chosen, which allows better e-beam performance relative to a glass substrate by offering better electrical and thermal conductivity during e-beam exposure. After e-beam exposure and development of the PMMA, it can be coated with any reflective coating desired – for characterization, aluminum was chosen.

Profiled PMMA optical devices with reflective coatings have been fabricated at JPL's MDL and flown in the Moon Mineralogy Mapper on Chandrayaan-1, and in CRISM on Mars Reconnaissance Orbiter. Reflective optics also have an appeal for flight missions as they avoid questions of refractive changes caused by radiation damage. Regarding contamination and yield questions, this process can readily produce arrays of occulters on 1 mm spacing, placing on the order of a hundred devices on a single substrate, in a compact space for easy mounting and selection during flight.

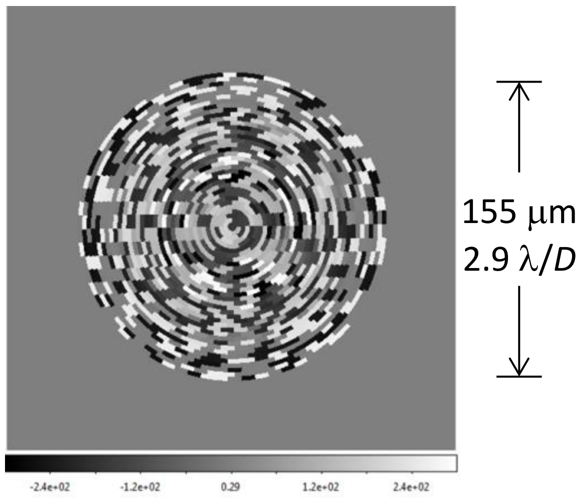


Fig. 5. Gen 2 PIAACMC occulting mask surface heights. The P-V height is 611 nm. There are 35 annular rings, each 2.2 μm wide, with a total diameter 155 μm (2.9 λ/D at λ=550 nm), comprising 1259 height zones. The reflectivity of the mask is uniform everywhere.

5. Gen2 PIAACMC Occulting Mask Characterization

Adequate characterization of the occulting mask requires measuring heights at a wide range of spatial scales. The relevant scales include the details of the transition region between zones, at the 10s of nm level, the heights of zones spanning 2 um each and extending tens of zones on a side, and heights across the nominally flat region outside the zones extending for ~ 1 mm. No single characterization technique measures heights accurately at all these spatial scales, so a combination of techniques are required for this exercise. The relevant measurement instruments are a Scanning Electron Microscope (SEM), a tapping Atomic Force Microscope (AFM), and a ZeMapper interferometric optical microscope. A table comaring these instruments is shown in Table 2.

Technique Component	Scanning Electron Microscope (SEM)	tapping Atomic Force Microscope (AFM)	ZeMapper interferometric optical microscope
Sharpness of zone transitions ("rounding")	< 20 nm spatial resolution	200 nm lateral resolution (with distortion)	1 μm lateral resolution
Individual zone heights	Hard to quantify features	< 6 nm errors at high spatial freq	Corrupted by poor lateral res
Wide-angle surrounding features	Hard to quantify features	100 μm distorted FOV, low-order errors	Good low-order sensitivity for smooth features

small

↓

large

Table 2. Relative strengths and inadequacies of measurement instruments in characterizing relevant spatial scales.

The characterization scheme used to measure on all these scales is to rely on SEM measurements to estimate the edge transitions (smallest scales, 50 nm – 10 μm), AFM for heights of individual zones (intermediate scales, 1 – 100 μm), and ZeMapper for the surrounding areas (largest scales, 10 μm – 1 mm).

5.a. SEM characterization (50 nm – 10 μm)

The SEM has ~ 20 nm spatial resolution, much finer than the 200 nm logical pixels on which the heights are specified. The SEM, viewing the mask at an angle, is good at identifying the characteristic slopes of the transition edges, but has difficulty quantifying heights of the uniform regions. Sample imaged locations are shown in Fig. 6.

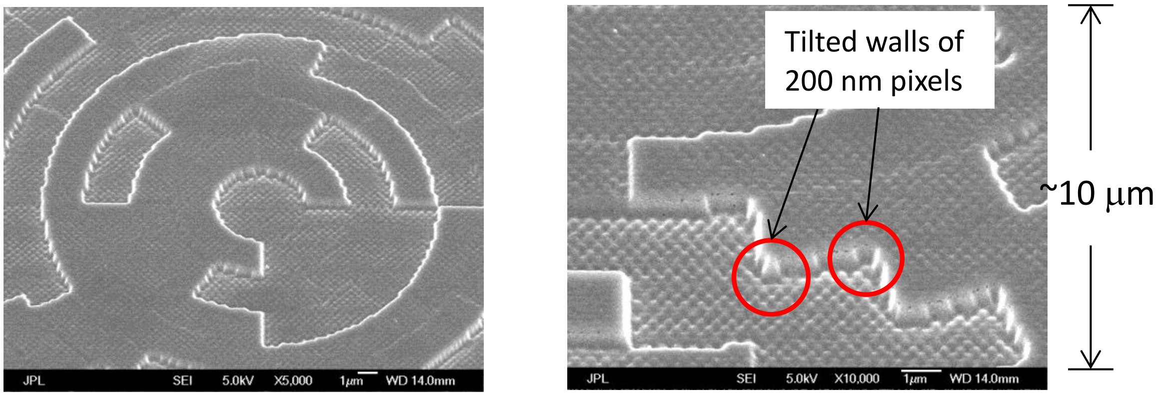


Fig. 6. SEM images of occulting mask. The individual 200 nm pixels are clearly visible, and the walls can be seen to spread by ~ 60 nm.

The main quantitative result taken from the SEM images is that the walls of the height zones spread by ~ 60 nm as they transition between zones. Also visible in the images is an “egg-crate” pattern, which is uniform and sub-wavelength ($d < \lambda/2.5$), which has no optical effect (phase or amplitude) for near-normal incidence.

5.b. AFM characterization (1 – 100 μm)

The AFM measurements, taken in tapping mode, are very good at measuring fine details of heights across a 100 μm 2-D area, but they present some artifacts that must be handled carefully to provide an accurate representation of the sample. The three factors that most affect the measurements are distortion of (x,y) positions, asymmetric errors near steep edges, and low-order height errors.

The distortion is accommodated by fitting the locations of each zone relative to the edge discontinuities. The resolution of the AFM is corrupted near the edges of the zones, so the pixels neighboring the edges are excluded from the measurement of height in each zone. The low-order terms are fit relative to the design, and subtracted. The resulting decomposition is shown in Fig. 7.

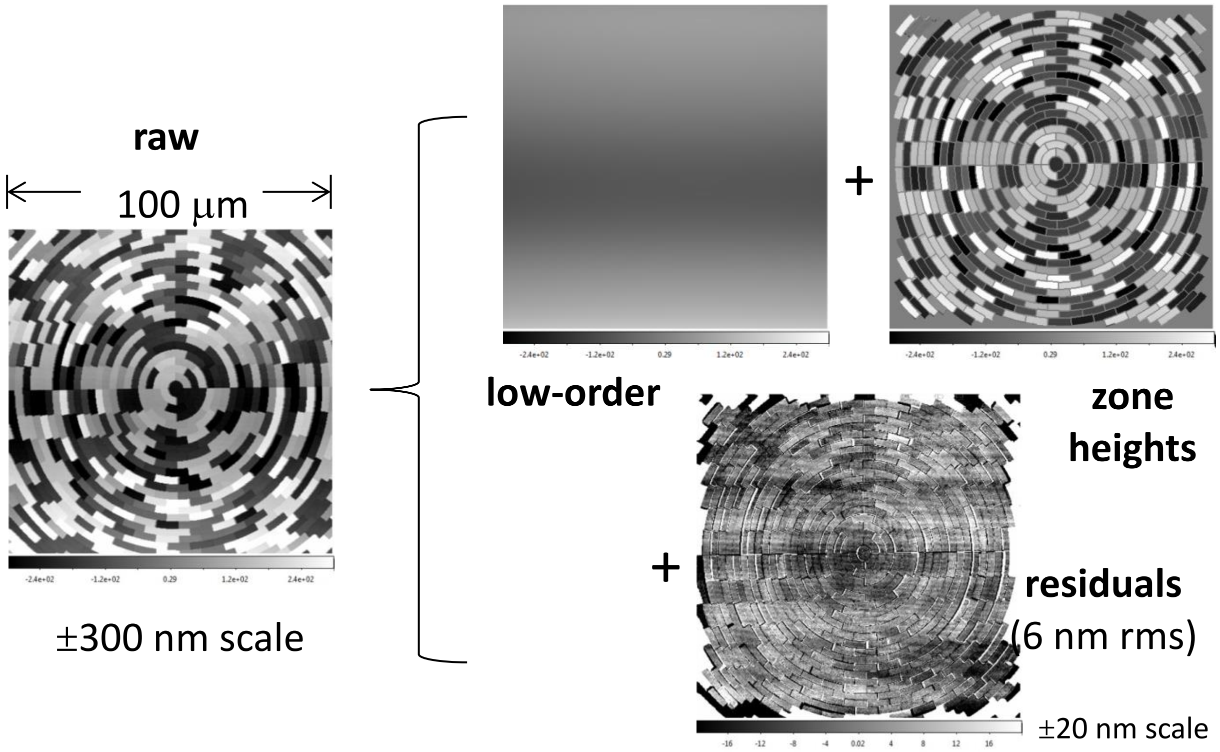


Fig. 7. Decomposition of AFM measurements into components. The left image is the raw AFM measurement, which is corrupted by lateral distortion, poor edge resolution, and low-order height errors. The top-middle image is the low-order fit to the residual heights, the top-right is the design zone heights (a single value over a zone whose location was fit to distortion and whose edges are excluded), and the lower-right image shows the residuals. The sum of the 3 right-hand images is equal to the left-hand image. The color scale for the residuals is 15x finer than the other components, representing a 6 nm rms. The other 3 images share a ± 300 nm full scale.

The residuals of the zone heights, with respect to the design values, vary by 6 nm rms, which can be compared to the P-V design heights of 611 nm. To give a graphical sense of the relative errors on a uniform scale, a 1-D cut is plotted in Fig. 8.

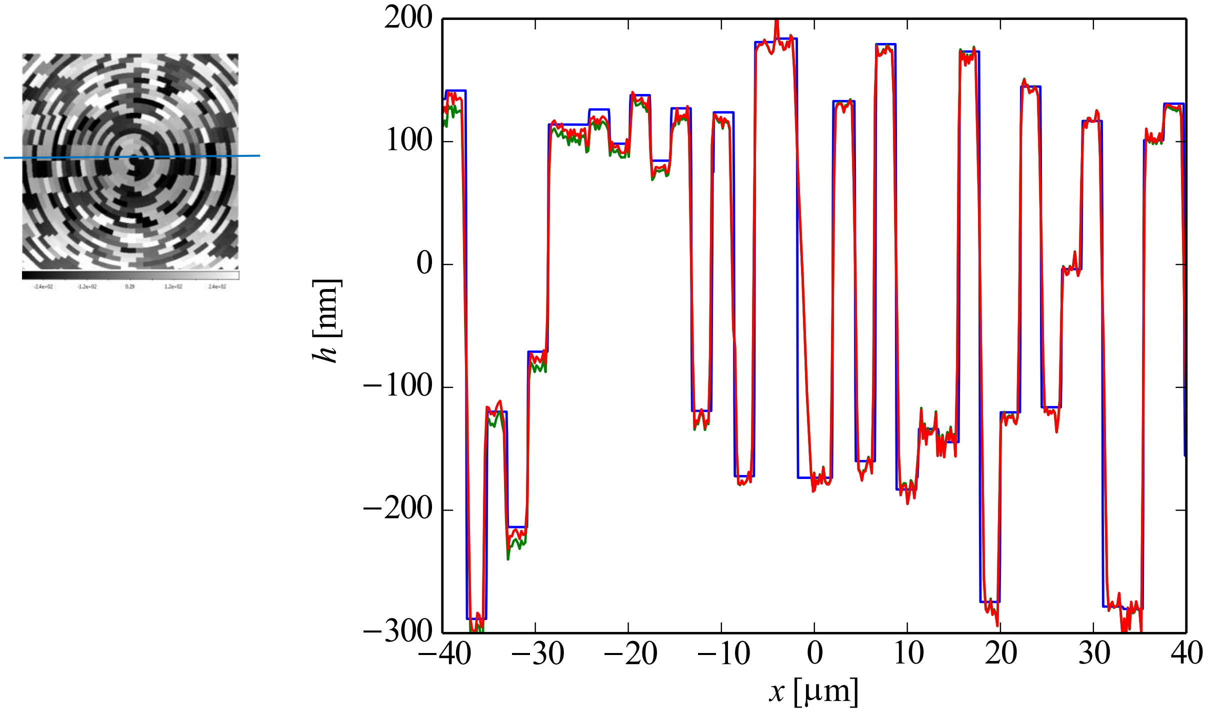


Fig. 8. 1-D cut through AFM measurements. The blue line on the left-hand image shows the *y*-location of the cut. The blue line in the plot is the design height, the green line is the raw AFM trace, and the red line is the raw AFM trace minus low-order fit. The mismatch at sharp edges, notably near *x*=0, is an artifact of the AFM measurement where the tip does not trace steep features. The SEM shows the actual edges to be far sharper than the AFM measurements.

5.c. ZeMapper characterization (10 μm – 1 mm)

The ZeMapper has the widest FOV of the three measurement techniques, which extends to approximately the limit of the relevant coronagraphic FOV. A ZeMapper image is shown in Fig. 9. The most prominent feature in the image is the grid pattern 0.5 mm on a side. This is due to the field boundaries of the 100 keV e-beam writing pattern, which can only write 0.5 mm squares before repositioning the sample for the next square. The non-telecentric beam illumination treats the edges differently from neighboring square centers. As will be described in the section on wavefront control, this effect is significant to the performance of the occulter, and should be improved by simple changes in the e-beam parameters.

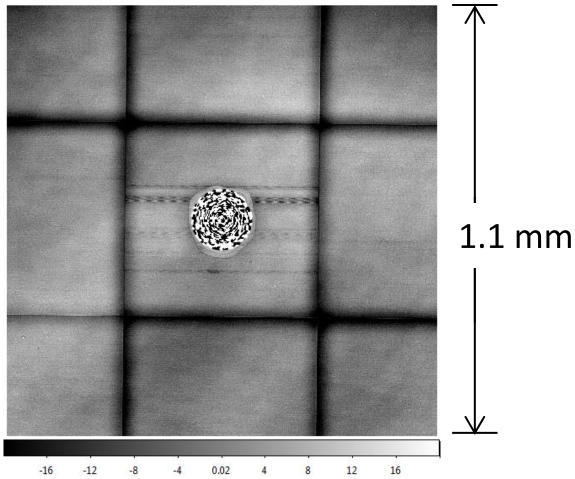


Fig. 9. ZeMapper surface height image. Full scale is ± 20 nm. The image is 1.1 mm on a side, which corresponds to about ± 7.5 λ/*D*. The grid pattern comes from the field boundaries of the e-beam writing, spaced on 0.5 mm squares.

5.d. Combination of characterization results

Measurements from these three instruments are all combined to form a single model of the occulter. The propagation models used for high-contrast modeling and wavefront control are paraxial, unfolded models, in which a reflective occulter is represented as having a complex-valued transmission map in the unfolded system.

The characterization measurements are all made at different spatial scales. No single calculation is used to combine all of these scales, but rather, each contributes to the net transmission calculation at its own, separate scale. The net transmission is a simple $t = \exp\{2\pi i\, 2\, S(x,y)/\lambda\}$, a conversion of surface height *S* to wavefront phase. The superposition of the components used to generate the full transmission is pictorially represented in Fig. 10.

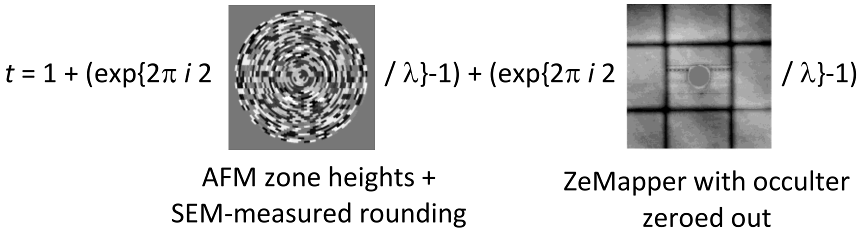


Fig. 10. Pictorial representation of equation incorporating characterization measurements together to form complex "transmission" of reflective occulter.

This decomposition is useful because the final representation of the occulter need not have information at all available spatial frequencies. The system contains a Lyot stop, which imposes a maximum transmitted spatial frequency; all component-wise occulter representations need to contain the correct information up to the maximum transmitted spatial frequency, but may be arbitrarily defined for higher spatial frequencies. This allows Fourier interpolation of the individual patterns, retaining sufficient spatial frequency information in each. This can be done with explicit DFTs with very fine spatial sampling, retaining the frequency information on a coarse grid uniform between patterns, and then use simple FFTs with relatively few elements for propagation.

At the finest scales, the AFM-measured zone heights (without the low-order artifacts) are generated on the design geometry, then convolved by a rounding kernel to have the same rounding as measured by the SEM. This produces the first term (AFM + SEM) pictured above, with zero surface height outside of the (rounded) occulter. The ZeMapper data is used exactly pixel-by-pixel as measured, with the occulting pattern multiplied by zero. Care must be taken to ensure that no two components have nonzero heights at the same locations, as the superposition is not valid by surface heights.

6. Wavefront Control Simulations

The standard optical propagation code used for evaluation of all WFIRST-AFTA coronagraphs has been PROPER (written by John Krist). The evaluation used here follows the same evaluation framework as all other "official" evaluations, where surface errors are added to each optic except the occulter (true for all coronagraphs). This propagation code (PROPER) is distinct from the wavefront control code used to solve for DM surfaces.

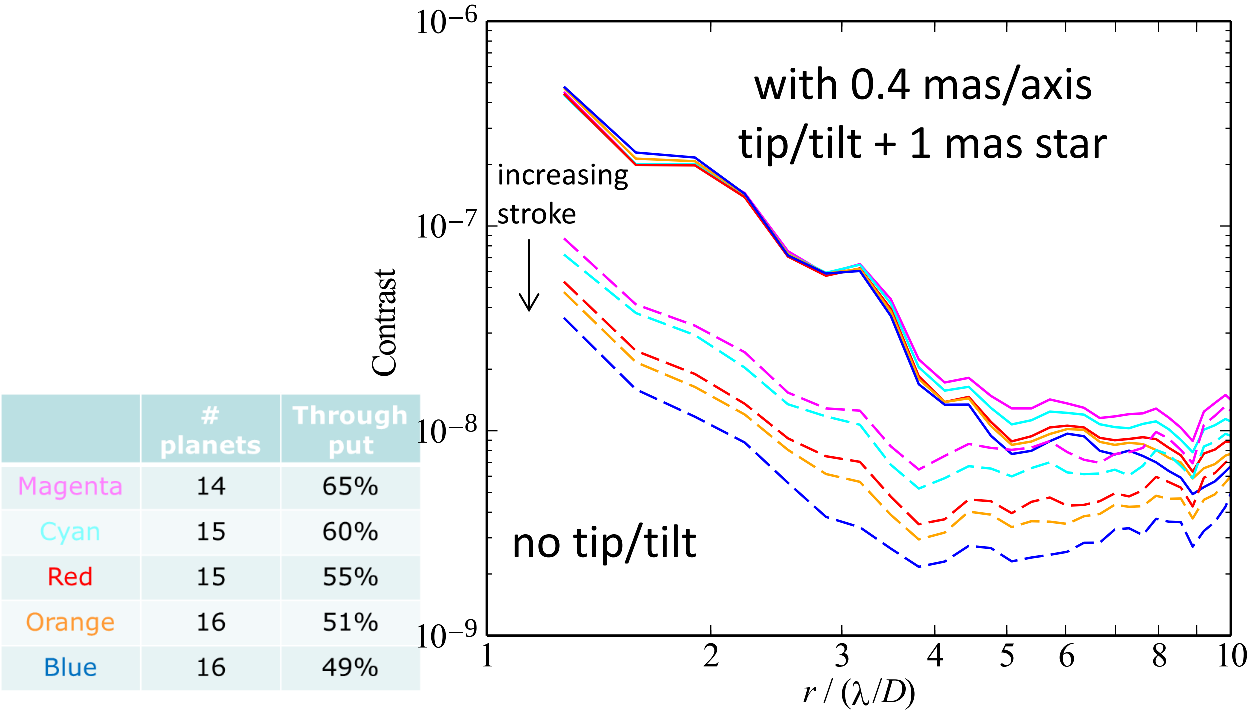
The wavefront control code is the same code that is used at JPL's High Contrast Imaging Testbed (HCIT). The combination of PROPER and the HCIT WFC code are used in a closed loop, where a DM control iteration is passed through PROPER to determine the residual E-field in the image plane, which is then fed to the HCIT WFC code to determine the next DM control iteration. The wavefront control is calculated for a point-source illumination, then after wavefront control is complete, the contrast is evaluated with different levels of residual tip-tilt errors and stellar angular diameters. The levels reported here are for 0.4 mas/axis rms residual tip-tilt, and 1 mas stellar angular diameter.

6.a. Trade between wavefront control and throughput

The improvement of contrast as wavefront control progresses involves a trade between better contrast, and planet light throughput. It is typically assumed that planet detection will be background-limited, and that only the planet light brighter than the planet half-max intensity is useful for observations. As such, a "core throughput" metric is used where the fraction of planet light brighter than half-max is compared to the total planet light. This core throughput metric is closely related to the Strehl ratio, which specifically characterizes the changes in peak brightness.

The reduction in Strehl ratio is roughly described by the Marechal / Mahajan approximation, $S \sim \exp\{-\Delta^2\}$ where Δ is the rms wavefront variation in radians. The majority of the improvement in contrast is not wavefront "correction," in which the wavefront gets flatter, but rather wavefront control, where the DM forces the PSF into a new distribution. The improvement in contrast comes from use of more DM stroke, which carries the accompanying loss of Strehl and reduced core throughput.

The final contrast is strongly dominated by residual tip-tilt errors, making the improvements from different wavefront control solutions beyond some point largely irrelevant to the final contrast. A range of wavefront control solutions, evaluated for 0.4 mas/axis residual tip-tilt and a 1 mas diameter star, is shown in Fig. 1.



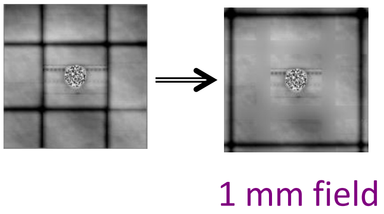
	# planets	Through put
Magenta	14	65%
Cyan	15	60%
Red	15	55%
Orange	16	51%
Blue	16	49%

Fig. 11. Contrast using as-fabricated mask, for different wavefront control solutions. The colors represent a progression of continued wavefront control, producing better contrast before tip-tilt (dashed lines) but similar small-angle contrast when tip-tilt is included (solid lines), with increasing stroke producing reduced throughput (in table on left) but more total planets. The red line is referenced as the “nominal” wavefront control case through this document.

Referencing the table included in Fig. 11, the total number of planets observable through contrast considerations increases as wavefront control continues, but the throughput (for all planets) decreases. The red line is chosen as the nominal wavefront control case, where 15 planets are observable, but the throughput has decreased to 55% of that for a perfect as-designed occulter.

6.b. Paths to improvement in occulter fabrication

Evaluating other hypothetical masks for the same considerations of throughput and planet yield, the most promising paths for fabrication improvements can be identified. Fig. 12 shows two additional hypothetical improvements and their impact on science yield.



	# planets	Through put
Nominal	15	55%
1 mm field	20	62%
Only zones	21	91%
Design	24	100%

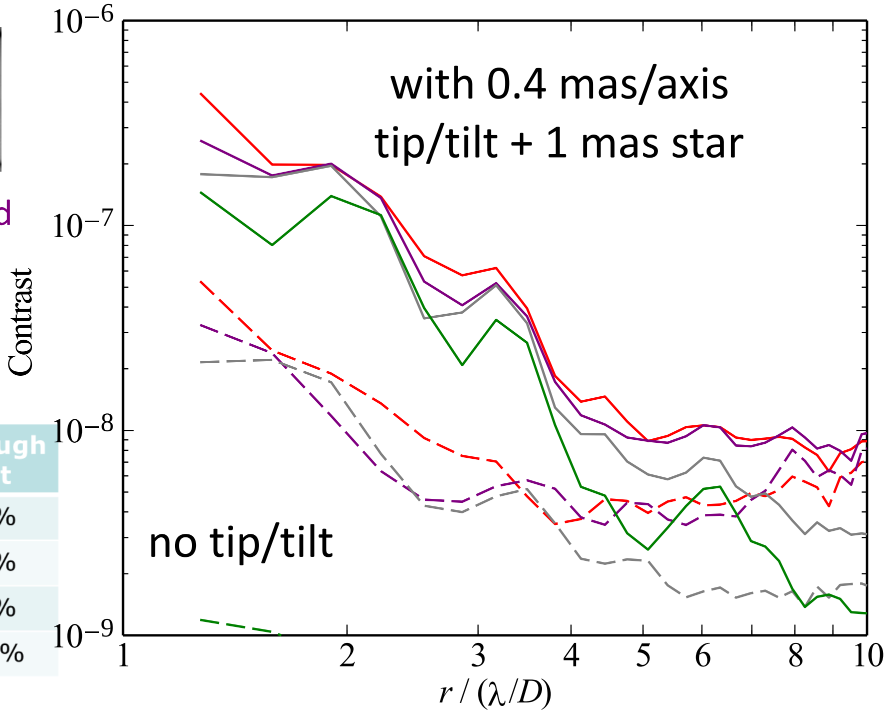


Fig. 12. Contrast of nominal as-fabricated occulter (red), perfect as-designed occulter (green), and two hypothetical fabrication improvements. The dashed lines are without tip-tilt errors, the solid lines are with residual tip-tilt errors and 1 mas stellar diameter.

The two hypothetical cases considered here are for the field boundaries to change in spacing from 0.5 mm squares to 1 mm squares (shown in the top-left insert in Fig. 12), and for there to be no errors at all in areas surrounding the zones (*i.e.*, the ZeMapper data are perfect). As can be seen from the throughput and planet yield numbers in Fig. 12, simply moving the positions of the field boundaries improves the planet yield significantly with a modest improvement in throughput. Removing all of the errors surrounding the zones delivers performance quite close to the perfect occulter, meaning that the zone errors alone are quite easy to correct with wavefront control.

The consideration of changing the field boundaries from 0.5 mm squares to 1 mm squares is motivated by the 50 keV writing capability of the e-beam at JPL. The existing mask was exposed at 100 keV, which has a 0.5 mm field boundary limit. The 50 keV writing mode has a 1 mm field boundary limit, but was unavailable at the time of the occulter fabrication because a component required for alignment at 50 keV was not functioning. The identification of the field boundary size as a limiting factor was not made until recently, so the 50 keV experiment was deferred to the next round of PIAACMC mask fabrication.

6.c. Small IWA targets and past testbed results

The contrast plots of Figs. 11-12 are shown without the radial velocity planets overlaid, because several lines were shown on each plot. Taking the nominal wavefront control solution on the as-fabricated occulter, and the perfect as-designed occulter, the radial velocity planets were overlaid in Fig. 13.

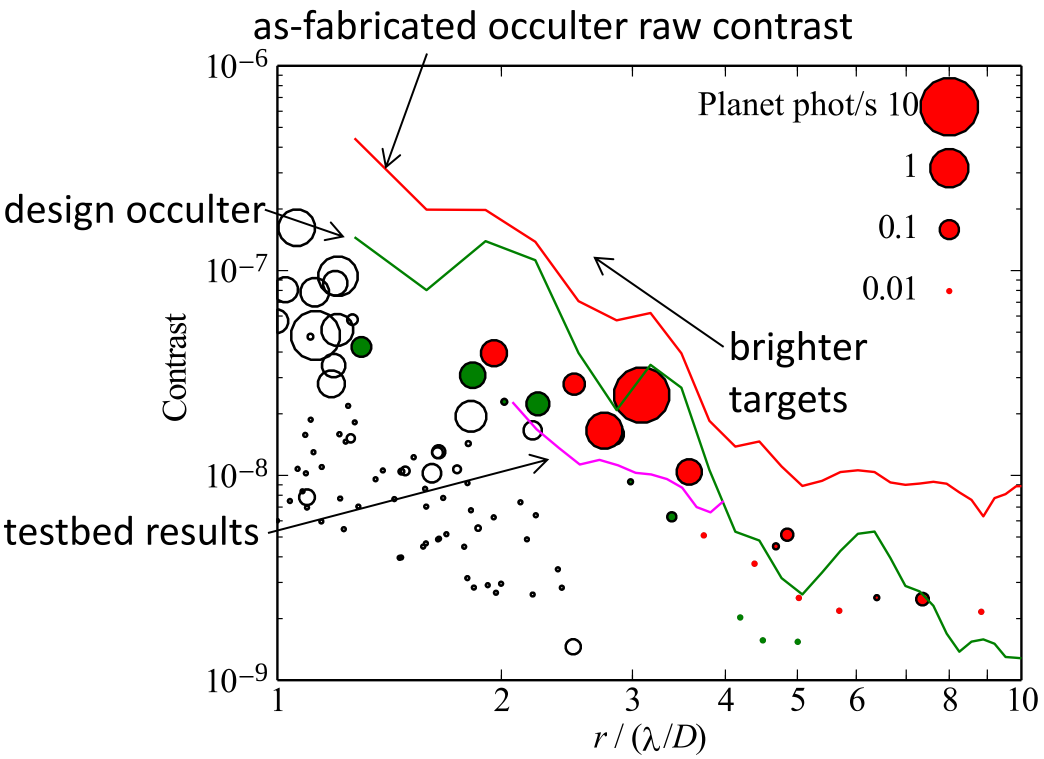


Fig. 13. Radial velocity planets plotted against nominal solution for as-fabricated occulter and as-designed occulter. Previous HCIT testbed results from an unobscured 10% broadband demonstration (at 800 nm) are shown in magenta. Planets observable with 30× postprocessing are shown in red, green planets cannot be observed by the as-fabricated occulter but can be observed by the as-desgined occulter.

What is notable about Fig. 13 is that the planets at small angles, < 3 l/D, have much brighter observed fluxes than those farther out, on the order of 10× – 100×. The advantage of PIAACMC having a small IWA is that these planets may be observed, which makes the observing scenario far more appealing because high SNR measurements can be made in a short time.

Also shown in Fig. 13 is a contrast curve from a previous testbed demonstration. This curve is not directly comparable, because it used an unobstructed pupil, and a PIAA coronagraph without the PIAACMC occulter (but with a simple hard-edged occulter). It was also measured in a 10% band centered at 800 nm, rather than the 10% at 550 nm shown for Gen2 model results.

6.d. Assumed residual tip-tilt errors

As the ultimate contrast of all these cases depends strongly on the residual tip-tilt errors, it is relevant to examine the state of modeling of the telescope to date. The current models (as of Milestone 3, December 2014), the STOP models + LOWFS predictions show residual jitter well below 0.4 mas rms per axis for most reaction wheel speeds below 42 Hz. A sample plot is shown in Fig. 14.

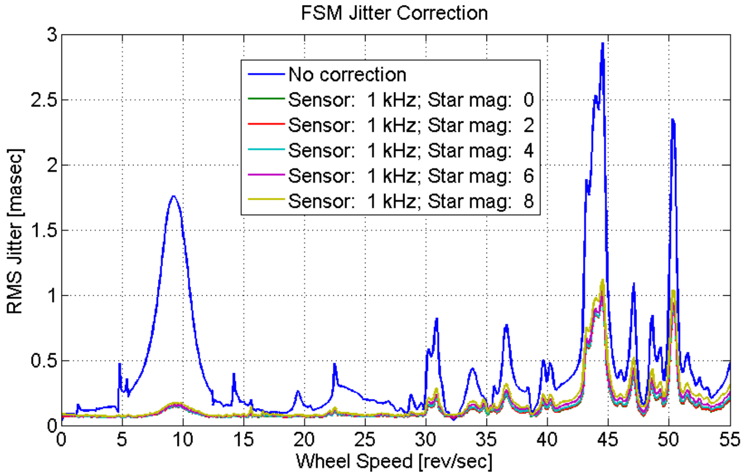


Fig. 14. STOP model + LOWFS prediction for residual tip/tilt errors from one reaction wheel, including MUF. Taken from Fang Shi, DocuShare Document-1151956, 2014-Nov-04.

As can be seen in the plot, a typical residual tip/tilt for wheel speeds < 42 Hz is about 0.1 mas rms / axis, for the worst case (*V*=8 star). Taking 4 wheels which are all at speeds < 42 Hz, residual tip-tilt errors might be 0.2 mas / axis on average, well below the 0.4 mas / axis rms used for the models in this milestone, assuming uncorrelated contributions from the 4 wheels. There may be some fraction of time spent with wheel speeds above 42 Hz, in which case some care must be taken to identify the time periods with higher tip-tilt residuals and determine, at worst, a reduced observing efficiency by eliminating those time periods.

7. Discussion

The results presented here describe a PIAACMC occulting mask fabricated at JPL's MDL. This mask, the first ever fabricated on PMMA, is modeled to deliver contrast yielding 15 radial velocity planets (nominally), exceeding the science requirement of 12 planets. This performance appears to be dominated by the e-beam field boundaries, which have a ready path for improvement, by utilizing the 50 keV writing mode of the existing e-beam facility at JPL.

What may be surprising is that the 1259 zone heights determined in the design were not particularly challenging to fabricate to sufficient tolerances by this technique – while it appears complicated, it is well within the standard operating capabilities. The first attempt to fabricate such an occulter delivered rms zone height errors of just 1% of the P-V zone heights, clearly sufficient for 10% broadband coronagraphic operation.

Acknowledgements

WFIRST-AFTA coronagraph technology development work was carried out at the Jet Propulsion Laboratory, University of Arizona, and Ames Research Center using funding from NASA SMD and STMD.