Wavefront Sensing and Control for Large Space Optics[1]

David C. Redding, Fang Shi, Scott A. Basinger, David Cohen, Joseph J. Green, Andrew E. Lowman, Catherine M. Ohara

NASA Jet Propulsion Laboratory, California Institute of Technology

4800 Oak Grove Drive, M.S. 306-438

Pasadena CA 91109

818-354-3696

david.redding@jpl.nasa.gov

Abstract — Large space optical systems will require on-orbit wavefront sensing and control systems to correct misalignments and figure errors incurred during manufacture, launch and deployment. This paper describes key WFS&C technologies developed to support the James Webb Space Telescope, Terrestrial Planet Finder, and Actuated Hybrid Mirror Telescope missions.

1. INTRODUCTION

The advantages of large aperture space telescopes are obvious — sensitivity increases (and observation time decreases) proportionately to aperture area; spatial resolution improves proportionately with aperture diameter; the orbital environment is generally stable and quiet; space telescopes can be cooled to cryogenic temperatures; and space telescopes do not need to observe through Earths semi-opaque atmosphere. These advantages will lead to revolutionary new science. The James Webb Space Telescope (JWST, see Fig. 1) will push much further back in time and distance than the Hubble Space Telescope, to image the formation and evolution of the earliest galaxies. The Terrestrial Planet Finder (TPF) will directly image earth-like planets orbiting nearby stars. Large aperture earth-observing telescopes offer potential for ever higher resolution and sensitivity for a variety of applications.
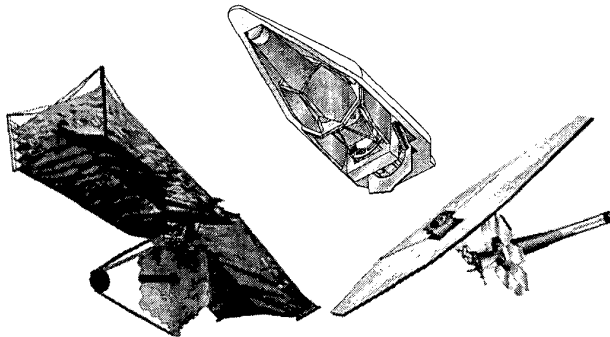

Figure 1 —JWST (aka NGST [1,2]) configurations. The left-most is the recently-selected TRW design, with 36 hexagonal segments made of beryllium or glass, each equipped with tip/tilt/piston/radius-of-curvature control.The other images represent the Government NGST Yardstick point design, as launched and as deployed.

For large aperture space telescopes to be achievable, especially those with apertures larger than the diameter of a rocket shroud, new telescope architectures are required. Making ever-bigger Hubble Space Telescopes, where optical quality is assured by utilizing massive, stiff structures and optics, quickly exceeds the throw-weight capabilities of even the largest booster. Cost and launch volume limits also keep the Hubble paradigm from being useful for larger telescopes.

Instead, missions such as JWST (Fig. 1) are moving to light-weight, deployed optics that are equipped with actuators, so that they can be adjusted after launch and cool down to compensate for optical aberrations that occur during fabrication, launch, or on orbit. This approach makes unnecessary the very massive, rigid and precise structures that traditionally have been used to assure optical quality through launch. By adopting actively controlled optics, this approach substitutes on-orbit adjustability for rigidity, volume and mass; a small computer and some precision actuators for many tons of glass; enabling much larger apertures and the exciting scientific breakthroughs that will follow.

In this paper we provide a brief overview of WFS&C techniques that we, together with colleagues from NASA Goddard Space Flight Center (GSFC), have developed for the JWST mission, and which we are further developing for the TPF and AHMT [3] missions. In contrast to ground-based adaptive optical systems, these techniques use the science cameras themselves as the WFS, rather than dedicated instrumentation. These missions provide stable environments, where residual thermal and dynamical disturbances are small, so that the WFC actuators can be run in a set-and-forget mode, turned off while the telescope performs its science observations.

We will begin by discussing the initial phasing of segmented mirrors, using the recently-selected TRW 36-segment JWST configuration as an example (see Fig. 1). The JWST primary mirror architecture features semi-rigid segments, each actuated in tilt, piston, and radius-of-curvature degrees of freedom (DOFs). Higher-order segment figure errors for this cryogenic telescope will be removed prior to launch through cryo-figuring, a procedure where the segments are measured at cryogenic temperatures but are fabricated at ambient, yielding good figure at the operating temperatures.

The paper will conclude with a look at phase retrieval WFS, where defocussed imagery is processed in a computer to measure the WF. We describe a Modified Gerchberg-Saxton algorithm that produces accurate high-resolution WFE maps using defocussed star image data. This technique is applicable to virtually any instrument. Indeed, we have developed a Phase Retrieval Camera (PRC), a small, portable instrument for laboratory testing in exactly the same mode as commercial phase-shifting interferometers. Laboratory results from the PRC and other laboratory testbeds are shown.

2. WFS&C MODES

There are 3 major tasks that a space telescope WFS&C system must perform. The first is *Initialization*, the initial alignment and figure correction control that occurs at first light. This is the subject of this paper. The other major WFS&C tasks are *Calibration*, which establishes the influence functions and WF offsets that are needed for good control for the entire observatory, and *Maintenance*, involving long-term monitoring and periodic correction of image quality and WF error.

Initialization begins immediately after telescope deployment and cooldown, with the telescope pointed at a bright, isolated guide star. For a *segmented* telescope such as JWST, this first light image will likely be broken up into multiple defocussed segment images, as sketched in Fig. 2.

The first Initialization mode is *Coarse Alignment* control, which acts to identify which blob is associated with each segment, and moves each segment image into the center of the field. Segment images that might happen to fall outside of the initial field are hunted down and also placed in the center of the field. There each segment is individually focussed. At the end of Coarse Alignment, the segments have been identified, and the initial WF error has been reduced from millimeters to 10 um or so, to within the depth-of-focus of the individual segments.

Coarse Phasing mode then takes over. Here segments are coaligned, so that their images overlap and interfere. Then the light is modulated, either spectrally using a grism, or by scanning through motion of a segment, to find the point where the interference is sharpest and the segments are co-phased. Then the segments are moved to that position. This reduces the total segment piston WF error to well under a wavelength.

Fine Phasing begins with a nearly-phased primary mirror. Fine Phasing utilizes image-based Wavefront Sensing (WFS) techniques to measure the telescope WF at the exit pupil at high resolution and to high precision. Wavefront Control (WFC) algorithms use this measured wavefront data to implement new segment actuator settings to compensate the WFE.

For a *non-segmented* telescope, the initial image is likely to be highly aberrated, but not broken into distinct blobs. The Coarse Alignment and Coarse Phasing modes are not required. Rather, initialization of a non-segmented space telescope will begin in the Fine Phasing mode.

3. COARSE ALIGNMENT — SGMENT ID

We will illustrate Initialization of a segmented space telescope using the JWST mission as an example. Performance examples draw on JWST simulations as well as data from JWST testbeds.

The process of aligning and phasing the JWST optics begins after launch, deployment and cool-down, by pointing the telescope at a bright, isolated guide star, and taking a first picture. This first-light image will likely be broken up into many large fuzzy blobs, each blob being the segment image of the guide star, formed by individually misaligned segments. Some segment images may be so misaligned as to completely miss the detector! An example of what might be expected at first light from a 36-hex JWST is shown in Fig. 2. This example assumes rather good initial alignments of well under 1 mm (1-sigma).

The first part of the Coarse Alignment task is *Segment ID*, which acts to identify which blob corresponds to which segment, and then aligns all the segment images, placing them in particular spots on the detector for further processing. Segment images that initially fall off of the detector are found by scanning the segment in an ever expanding spiral until they are detected. Segment ID occurs before the fine guidance system is turned on, so the line of sight experiences slow drift of about 2 arcsec amplitude (1-sigma). Segment ID begins by correcting secondary mirror (SM) focus, placing the SM focus to put the segment images in best average focus.

The Segment ID algorithm uses image differentiation to match segments to blobs. This process works by taking an image, then moving a group of segments in defined patterns, then taking another image. The images are registered, to remove drift effects, then normalized, differenced and thresholded. The results are processed to correlate the blobs with the segment motions. Fig. 3 shows a simulated first light image and a second image taken immediately after actuating 6 of the 36 segments in a perturbation pattern. Writing I_{i-1} and I_i for the first and second images, where I is a 2-dimensional matrix comprised of the pixel values and i is the image sequence number, the difference image (after processing and registration) is:

$$\Delta I = I_i - I_{i-1} \tag{1}$$

The perturbation pattern applied in this case was

$$r_j = \rho \angle \theta_j, \; \theta_j \in \left(\pi/3, \frac{2}{3}\pi, \pi, \frac{4}{3}\pi, \frac{5}{3}\pi, 2\pi \right) \tag{2}$$

The magnitude ρ of the perturbation was 1 arcsec. The angles θ here refer to tilt direction in the detector coordinate system, applied using the various segment actuators. For convenience in simulation, only a 512-by-512 patch of the full JWST detector area is represented.

The difference image is processed to determine the starting and ending positions of all of the segment image blobs. The ending positions are determined by thresholding the difference image at a value ε, selected to suppress the background yet permit detection of the defocussed segment images. The positive difference image is defined (here i,j signify row and column numbers for image pixels):

$$\Delta I_{pos} = \left\{ \begin{array}{l} 1, \Delta I_{i,j} \geq \varepsilon \\ 0, \Delta I_{i,j} < \varepsilon \end{array} \right\} \tag{3}$$
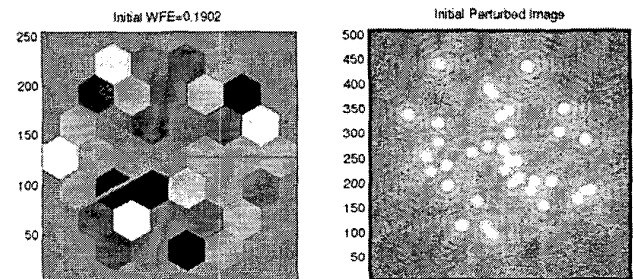
$$\tag{4}$$

Figure 2 —First-light example showing JWST pupil wavefront and image as detected.

The process is repeated for the negative blobs. The negative difference image is:

$$\Delta I_{neg} = \left\{ \begin{array}{l} 1, \Delta I_{i,j} \leq -\varepsilon \\ 0, \Delta I_{i,j} > -\varepsilon \end{array} \right\} \qquad (5)$$

The detected locations for the negative blobs example are shown as the blue boxes in the difference image in Fig. 3.

Once the start and end points are determined, it is usually a simple matter of combinatorics to match the vectors $r_{i,j}$ between the various startting and ending blobs to the commanded perturbation pattern and identify through elimination which blobs correspond to which segments. The vectors so identified are shown in Fig. 3. The identified segments are controlled to place their images at predesignated positions on the focal plane, and then Segment ID moves on to repeat the process with the next 6 segments.
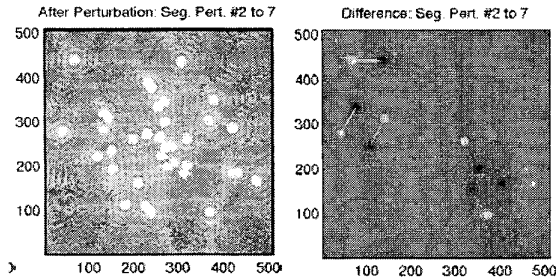

Figure 3 —Effect of Segment ID perturbation, showing image after segmetn motion and difference images for 36-hex JWST example. Commanded segment perturbations were 1 arcsec with directions 0..., 60..., 120..., 180..., 240..., 300....

Confusion of the Segment ID algorithm can occur when 2 segments images start and/or end very near other segments. These situations are identified and resolved by iterating the basic algorithm after all non-confused segments are placed under control. Once segments are identified, their images are placed at predetermined field positions in preparation for subsequent tasks.

The 36-hex example considered here was run with initial errors small enough that all of the segment images fell on the detector focal plane. This is the expected case, but is not the limit for Segment ID capture range. Segments with initial errors so large that their segment images fall off of the focal plane will not be identified in the initial image differentiation process, but the fact that they missed the detector will be determined. This triggers a second phase of Segment ID: *Segment Capture*. To capture these segments, larger tilt scan steps are taken, ordered so that the segment spot makes a spiral, incremented in units 90% of the detector field of view, until the spot falls onto the detector and is identified and controlled. Capture can find segments that are misaligned by large angles, perhaps degrees, though the process does take time. Very large angles may also introduce a confusion factor, coupling in light from other stars.

Figures 4 and 5 demonstrate Segment Capture using the JWST Wavefront Control Testbed 2 (WCT-2) [4]. WCT-2 is a small-optics hardware testbed, located at NASA GSFC, which provides a development and test platform for JWST WFS&C procedures. It is equipped with a small, 3-segment primary mirror, as well as a tertiary deformable mirror, illuminated by a filtered white-light source, and imaged on a CCD camera.

Experiments are set up by systematically aberrating the optics, tilting and pistoning the segments and deforming the DM. In this particular experiment, the segments were initially set up with random tilt errors large enough that the initial image had no light in it. Each segment was captured using the spiral scan technique, ending with segment 3, as shown in Figs. 4 and 5.
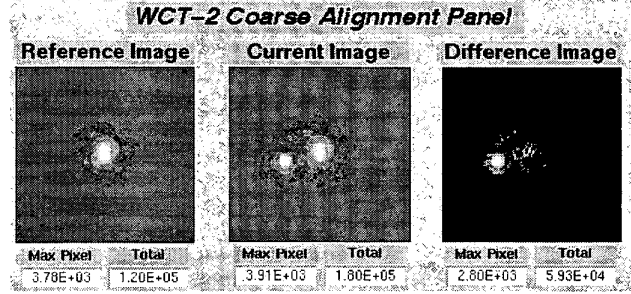

Figure 4 —WCT-2 Segmant Capture example, showing images taken at first and second spiral scan steps. Reference image shows segment images from segments 1 and 2, which were identified and coaligned earlier in the process. Current Image was taken after stepping segment 3 in tilt, from out of the field of view, into the field of view. Difference image is used to centroid segment 3 and put it under control. Figure is taken from screen-shots of the WFS&C Executive Software during experiment execution.
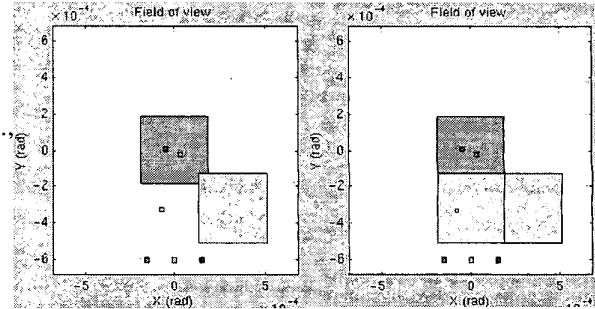

Figure 5 —WCT-2 Segment Capture example, showing approximate initial segment locations (small colored squares) and detector field of view (red square). Scanning of the segment is visualized as scanning of the field of view (green squares). Left plot corresponds to the Reference Image in Fig. 4; right plot to the Current Image. Segment is detected in second step of a spiral scan. Field-of-view overlap is 15%. Figure is taken from screen-shots of the WFS&C Executive Software during experiment execution.

Segment Capture as run in the lab using WCT-2 is performed on an interactive basis. That is, the segment is moved, then the image is analyzed, then the segment is moved again; the process repeats until the segment is captured. This approach works well for automated operation. It is not efficient for the more closely-supervised operations planned for JWST, where the time to take an image, downlink it, process it, determine new commands, review and approve new commands, uplink commands, and repeat the cycle, may be quite long. For on-orbit Segment Capture it is faster to work in a batch mode, where each off-chip segment will be scanned through a long sequence of tilts with images taken after each command. The whole sequence of images for each segment will be recorded, downlinked and processed together to locate the segments.

4. COARSE ALIGNMENT: SEGMENT FOCUS

Coarse Alignment continues, once Segment ID has identified each segment, by correcting the focus of the segments using focus scan techniques. This section describes a simple, large capture-range *Segment Focus* algorithm. It operates on those segments whose focus errors exceed the segment depth of focus, as determined by the width of the segment image. Segment Focus is not capable of phasing the segments to within a fraction of a wavelength, as Segment Focus cannot distinguish between segment piston and segment radius of curvature errors. That task is left for Coarse Phasing control.

Segment Focus works by pointing each segment image, in turn or in parallel, at particular target locations to clear it from confusion by the other segment images. The centroid x_0, y_0 of the segment is determined, and the encircled energy (E) is measured from the image I, according to 3 different encircled energy metrics defined by different circle radii. The encircled energy for the i th metric corresponding to radius r_i is:

$$E_i = \sum_{(x-x_0)^2 \le r_i^2} \sum_{(y-y_0)^2 \le r_i^2} I(x - x_0, y - y_0) \qquad (6)$$

The sign of the focus error is determined by pistoning the segment a small amount and repeating Eq. 6. If E decreases, the focus error is in the direction opposite to that of the applied perturbation.

The radii defining the 3 E metrics are chosen to provide overlapping linear gain regions in the E vs. r curve, as illustrated in Fig. 6. This allows unambiguous control computation anywhere within the total capture region.

An example illustrating Segment Focus performance is provided in Fig. 7. These examples were taken from experiments using the WCT testbed in its WCT-1 configuration, with the 3-segment mirror swapped for a single deformable mirror used as an aberrator. This configuration provides large dynamic range, using the camera defocus stage to impart errors that are many times the aperture depth of focus. As shown, all 3 cases were focussed to within their depth of focus by the 10th step.
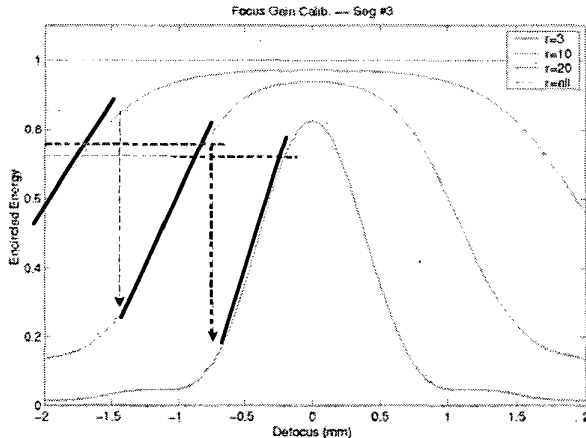

Figure 6 —Segment Focus gain calibration curves for WCT-2., showing encircled energies at different radii vs. defocus. Saturation thresholds are plotted with different colors (dashed lines). Each encircled energy metric has a linear region, capped by its saturation threshold. When the encircled energy reaches its saturation threshold, the algorithm uses the next inner radius encircled energy as the metric for focusing.

This Segment Focus algorithm is simple and robust enough to operate automatically using on-board computation. JWST will likely use a different approach, proposed by the JWST TRW/ Ball contractor team, that requires ground processing, but that permits a higher level of supervision and permits some level of global alignment to help place the SM correctly. Such an approach, processing all 36 separated segment images simultaneously using phase retrieval or prescription retrieval, is capable of determining segment focus and other aberration terms with as few as 2 or 3 images.
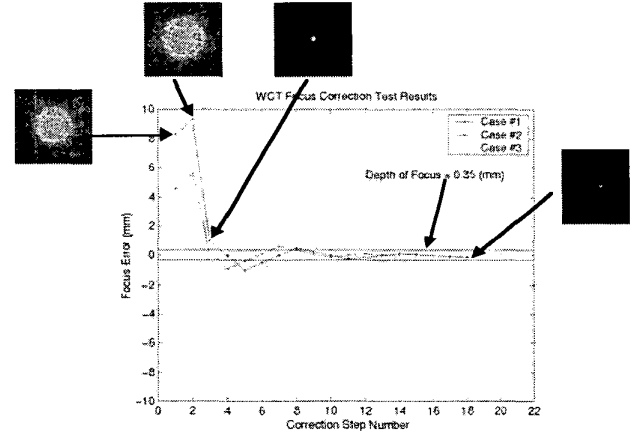

Figure 7 —Segment Focus examples from the WCT-1 testbed.

5. COARSE PHASING VIA DISPERSED FRINGE SENSING

Following Coarse Alignment, the primary mirror segments will be aligned in angle, with piston errors reduced to well within the segment depth of focus, about 128 um (wavefront) for the 36-hex example. The segments are thus co-focusssed, but not coherently phased at all wavelengths. That is the task for *Coarse Phasing*.

Coarse Phasing utilizes a spectral dispersion technique to modulate the phase differences due to segment piston errors. A telescope with 3 segments, one of which is displaced in piston by an amount δL, is sketched in Fig. 8. A *Dispersed Fringe Sensor* (DFS) is created by inserting a grism into the beam prior to the detector focus. The DFS generates a spectrum, dispersing the light from the 2 coaligned segments up and down the detector, so that the light at any point on the detector is at a single wavelength. At some points along the spectrum, the wavelength is coherent with the piston difference, and the light from the 2 segments adds constructively, creating a bright fringe. At other points, the light adds destructively, creating a null. The result is a fringe pattern of oscillating intensity, whose period is a direct measure of the piston difference.

To illustrate the DFS, consider the field at the focus. Ignoring aberrations and diffraction and writing E_0 for the illumination from the 2 aligned reference segments, and E_1 for the illumination from displaced segment, the field at each point x along the spectrum at the detector is:

$$E = E_0 \exp\left(\frac{2\pi i L}{\lambda(x)}\right) + E_1 \exp\left(\frac{2\pi i (L + \delta L)}{\lambda(x)}\right) \qquad (7)$$

The dispersing element is a grism placed in the filter wheel, with dispersion nearly linear with λ:

$$\lambda = \lambda_x x + \lambda_0 \qquad (8)$$

The intensity pattern that results is:

$$I = I_0\left[1 + \gamma\cos\left(\frac{2\pi\,\delta L}{\lambda(x)} + \phi_0(y)\right)\right] \qquad (9)$$

Here $I(x,y)$ is the DFS fringe intensity along the dispersion coordinate x, γ is the fringe visibility, and ϕ_0 is a phase constant that depends on where the DFS fringe is extracted — a position along the direction (y) that is perpendicular to the dispersion (x). The fringe modulation period is directly related to δL: the larger δL is, the shorter the period. When δL goes to zero, the fringe goes flat. The period is extracted by fitting the observed spectrum to Eq. 9 as a function of x and y. The sign of δL is evident from the angle of the dark bands.
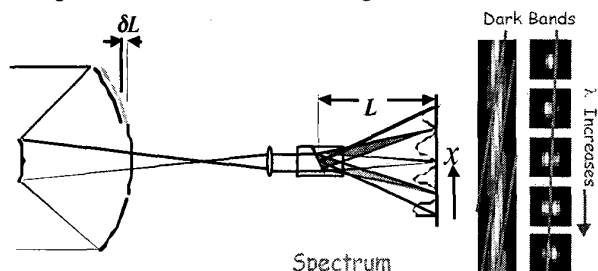

Figure 8 —Dispersed-fringe sensor is created by inserting a grism in the imaging optical path. The fringes show diagonal bands that define the magnitude and sign of the piston difference (here between 2 segments of equal area).

We have explored 2 different DFS algorithms. In the *Single-Trace* DFS algorithm, the intensity of the central row of pixels along the dispersion direction is used as the main DFS signal. We write this as $I(x,y=y_0)$, or $I(x,y_0)$. The central row is identified as having the brightest illumination. If the detected spectrum does not line up precisely to the detector axes, the spectrum is rotated in software to align with the rows and columns of a new image matrix. The 2 pixel rows adjacent to the central row also have similar modulations; however, there is a constant phase shift ($\phi_0(y)$) between adjacent rows of pixels. The sign of this phase shift determines the sign of wavefront piston δL. We write the lower and upper row signals as $I(x,y_{0-1})$ and $I(x,y_{0+1})$, respectively.

To determine the piston magnitude, the single-trace DFS algorithm solves for the four parameters in the fringe equation shown in Eq. (9), I_0, γ, ϕ_0, and absolute piston error $|\delta L|$, from $I(x,y_0)$, using the least squares method. This value of $|dL|$ is then applied to solve for I_0, γ, and ϕ_0 from $I(x,y_{0-1})$ and $I(x,y_{0+1})$. The sign of the piston (sign(δL)) is determined by solving for the row phases $\phi_0(y_{0-1})$ and $\phi_0(y_{0+1})$ between the signals $I(x,y_{0-1})$ and $I(x,y_0)$, and between $I(x,y_{0-1})$ and $I(x,y_0)$.

We also developed a more-accurate *multi-trace* DFS algorithm. This approach does not take the central row $I(x,y_0)$ as the sole signal for determining the piston. Rather, multiple rows are solved simultaneously , and a weighted average, reflecting the condition of the signals, is used to decide the piston. This approach lowers the systematic error caused by the dependence of DFS piston detection to row position (perpendicular to the dispersion direction) at which the DFS signal is extracted. Both algorithms are described in [5] and [6].

The measured piston value and sign is used directly by the executive software to control the actuators to remove the piston error and hence phase the segmented mirrors. The solved parameters I_0 and γ are used to monitor the quality of DFS detection. Erroneous solutions of wavefront piston usually will have unrealistic values of I_0, ϕ_0 and γ, and this information is fed back the executive software to set warning flags.

When the piston error is small, the DFS signal period begomes very long, longer than the measured spectrum, so that there is less than 1 modulation cycle over the spectrum. The spectrum flattens out. This is the desired post-control condition, as illustrated in Fig. 9, which shows a typical Coarse Phasing result from the WCT-2 testbed.
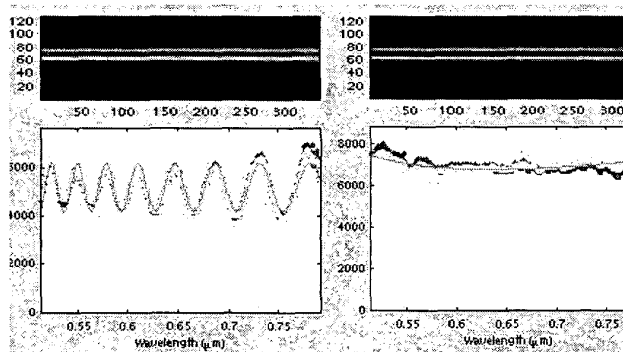

Figure 9 —DFS fringes from WCT-2, taken before (left) and after (right) correction. Top row shows the raw data; the bottom row shows the recovered fringes and the match to the fitted model (red curve). The modulation seen in the first fringe indicates a piston difference of -5.56 um between segments 2 and 3. Segment 1 is tilted away to provide a reference spectrum (top streak in the raw images). After correction, the modulation is nearly eliminated, confirming that the piston difference has been corrected. Additional examples are provided in [5] and [6].

When a DFS signal with little or no modulation is encountered before control is applied, it can mean 1 of 3 things. It most likely means that the segments under scrutiny are already phased, and no further control is required. It may mean that the segments are so far out of phase that the DFS signal is not resolved by the finite pixel spacing — that the period of the oscillations are smaller than a pixel. This condition is unlikely for segments that have been through the Coarse Alignment control. The third possibility is that the grism dispersion axis is oriented in such a way, relative to the baseline between the segments under test, that the modulation is poorly visible. This condition can be avoided or reduced by design.

The first condition, that the segments are already phased, can be simply tested, if desired. To obtain a high-accuracy measurement of a small piston error δL, a predetermined piston phase ΔL is added to the segment being tested, so that total wavefront phase error from the piston $\Pi = \Delta L + \delta L$ is large enough to form a high-visibility DFS signal. For example, if the WCT-2 has a small piston error of 0.2 μm, the DFS fringe intensity shows only a slope trend — a small part of the sine wave. If we directly fit for the fringe parameters we will get wrong numbers. However, if we add a piston of 3.0 μm to the segment (moving the segment by this known amount of piston) then the total piston is now 3.2 μm. At this piston, the fringes will now have a few complete modulations and DFS can see the 3.2 μm wavefront piston to an accuracy of better

5

than 0.1 μm. By detecting and correcting the 3.2 μm piston the system will reduce the original 0.2 μm wavefront piston to under 0.1 μm.'

The third condition in which little or no modulation may be seen occurs when the grism dispersion axis is not oriented nearly perpendicularly to the baseline connecting the centroids of the segments (or groups of segments) being tested. As shown in Fig. 10, fringe visibility depends strongly on the angle between the grism dispersion axis and the baseline between 2 segments. For the case of WCT-2, with a sparse aperture made up of small, widely separated segments, increasing grism angle shows a sharp drop in visibility within −5 deg of perpendicular, dropping to zero at −30 deg. For the case of our 36-hex example, testing adjacent segments, fringe visibility is strong for positive piston error for grism angles from -120$_i$ to about 30$_i$, and then drop to zero. For negative piston error, visibility is strong from about -30$_i$ to 120$_i$. This asymmetry is due to the sidebands of the segment PSFs filling the fringes, depending on the relative angles of the baseline and dispersion. See [6] for further details.
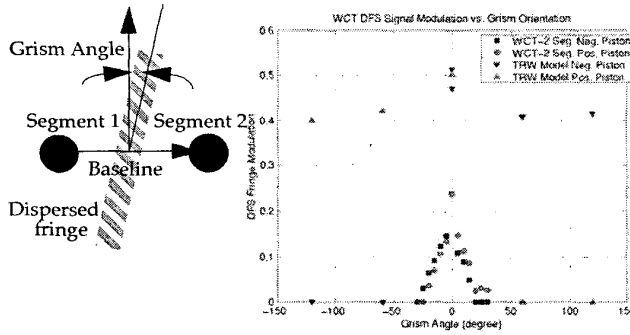

Figure 10 —DFS fringe visibility against the rotation angle of the grism for the sparse-aperture WCT-2 (measured data) and for adjacent segments in the TRW 36-hex configuration (simulation). 0$_i$ is the position at which the dispersion is perpendicular to the segments baseline. The test was done for both positive piston (red circles) and negative piston (blue asterisks). Beyond the grism angle range from -30$_i$ to +30$_i$, the DFS visibility remains at zero. Performance vs. grism angle improves if segment separation is decreased.

Clearly it is desirable to maximize the visibility of fringes between segments we wish to phase, and just as clearly, it is desirable to minimize the visibility of fringes between segments we are not currently phasing. Careful sequencing of operations in a complex, multi-segment aperture like our 36-hex example permits DFS to be effective in phasing the full aperture using either 2 grisms (the expected case) or 1 grism. Provision of 2 grisms, perhaps in 2 different channels of the NIRcam, provides required redundancy. The dispersion axes of the provided grisms can be set to be perpendicular to each other, in such a way that either grism can be used to phase the PM by itself, while using both grisms improves efficiency and performance. In the next few paragraphs we outline the process of phasing the 36-hex example using 2 crossed grisms.

The preferred location of the segment spots for 2-grism Coarse Phasing for the 36-hex example is shown in Fig. 11. Here the 36 segments are grouped in rows and columns to permit optimal fringe visibility and efficient parallel operations. The sequence is to phase all segments individually into rows, and then to phase the rows to each other as groups. Initially all 36 segment spots are clustered at parking lot positions on

the FOV. Then 7 segments, those along a vertical column through the center of the aperture (Fig. 11) are spread out evenly across a part of the FOV. These segments will be the references for phasing the rows of segments across the aperture. These rows are indicated by red outlines on Fig. 11. The rest of the segments are kept out of the way, in positions that are clear of the spectra.
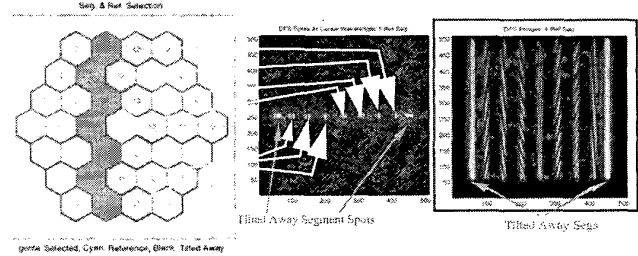

Figure 11 —Aperture showing segment organization for DFS phasing, 36-hex example. Cyan segments serve as references for phasing their respective rows of segments, and magenta segments are those selected for phasing in this step. Center image shows norrow-band spots formed by spreading the reference spots across the FOV, and then coaligning the selected spots for phasing. The selected spots are those corresponding to the magenta segments in the same row. The right image shows the dispersed fringes after the grism is inserted into the beam.

DFS Coarse Phasing begins by moving 7 more segment spots to overlie the reference spots, as shown in Fig. 11. Then (after being sure the segment spots are well aligned) the first grism, with a dispersion axis along the detector y-axis, is moved into the beam. The resulting spectra are shown in Fig. 11. This image is processed to produce analyzed fringes (Fig. 12), and the piston errors that are so determined are corrected by moving the segments. This first step requires 1 DFS image, 14 segment actuations, and phases 2 segments per row to their reference segments.
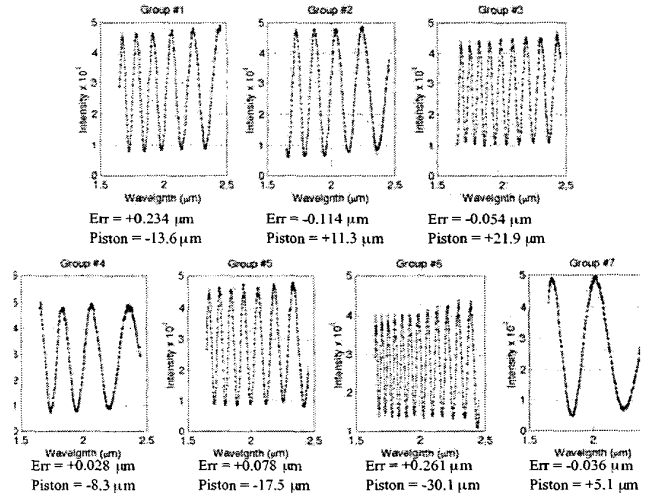

Figure 12 —Processed fringes from first DFS image.

This phasing into rows process is repeated, bringing in the next adjacent segments in each segment row, taking a DFS image, processing to determine the piston errors, actuating the segments to correct the piston errors, and repeating until all of the segments in each row are cophased. The fifth and final round of row phasing for the 36-hex example has only 3 seg-

ments remaining unphased in their rows, is shown in Fig. 13. By now, all segments are coaligned with the others in their respective rows. The spectra from the upper and lower rows, which have by this point been phased, are flat. Residual errors are obscured in the noise in the processed fringes (Fig. 14).
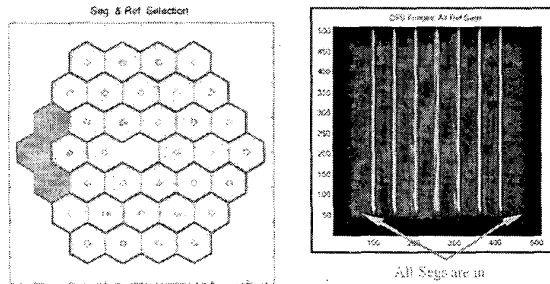

Figure 13 —Aperture and DFS image for fifth and final round of phasing the segments into rows, for the 36-hex example. After this step they will be cophased into rows.
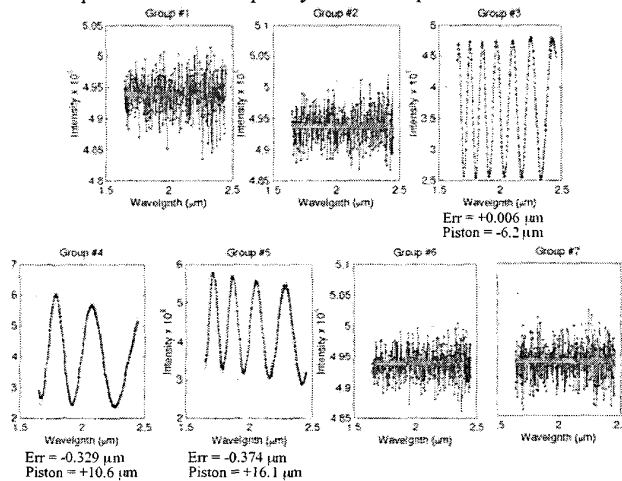

Err = +0.006 μm
Piston = -6.2 μm

Err = -0.329 μm
Piston = +10.6 μm

Err = -0.374 μm
Piston = +16.1 μm

Figure 14 —Processed fringes from fifth DFS image.

Once the process of phasing the segments into rows is complete, the first grism is swapped out and the second grism, with dispersion axis perpendicular to that of the first, is brought in to support phasing the rows to each other. Whereas the first grism had a vertical dispersion axis, which is ideal for phasing horizontally-separated segments, the second grism has a horizontal axis. This makes it ideal for phasing the rows of segments to each other, as they are separated vertically.

For phasing the rows to each other, the initial reference is the center row of segments. The process begins by tilting the next row of segments to coalign their spots with the reference row. The resulting spectra and processed fringe are shown on Fig. 15. Note the much stronger signal seen in the spectra. Moving the segments collectively as rows increases the collecting area quite substantially. The piston error is corrected by moving all of the segments in the lower row to align with the reference segments. The next row is then phased to the first and second. Again the signal strength grows as more segments are coherently aligned. The final step phases the final row to all of the others, as indicated in Fig. 16.

This example of 2-grism phasing for a 36-hex aperture started with a total WFE of 190 μm, and ended with all segments phased, with a total WFE of 484 nm. This result is well within

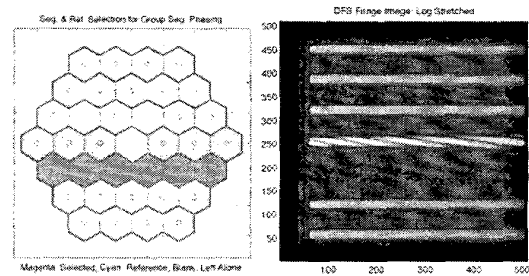the capture range of the Fine Phasing control to follow. The minimum number of images required was 11.

Figure 15 —Aperture and spectra for phasing second row to the first. This is the 6th DFS image...
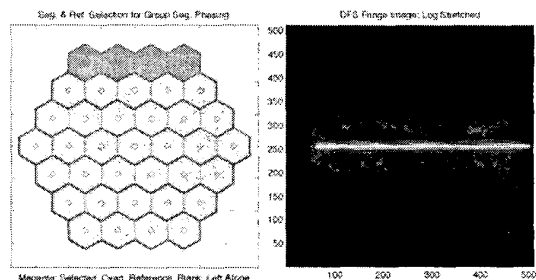

Figure 16 —Aperture and spectrum for phasing last row. This is the 11th DFS image.

Err = -0.147 μm
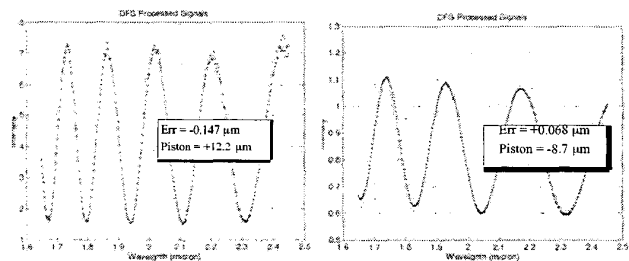Piston = +12.2 μm

Err = +0.068 μm
Piston = -8.7 μm

Figure 17 —Fringes from the 6th and 11th DFS images.

The number of images and actuations required to complete Coarse Phasing will grow in certain circumstances, including:

• No modulation seen after bringing in a segment or row of segments. As discussed earlier, if the selected segment or group of segments starts out phased to the reference, no modulation will be seen in the fringe. This may necessitate pistoning the segment or group of segments to dephase it and generate a strong signal that can be used to confirm that it is indeed phased.

• Actuator errors. If the segment actuators do not perform as specified (resolution of 10 nm and accuracy of ??? nm over long commands) then multiple steps may be required to ensure best phasing. In particular, if the segment tilt alignment drifts, then Coarse Alignment control may be required to recover, adding several more actuations per segment. Actuator error effects are minimized by operations that keep the number and size of actuations at small values. These conditions are not expected under normal operations.

• Thermal drift and other destabilizing effects. If the temperature varies much more than is expected, causing deformation of the telescope backup structure, then the phasing process might be interrupted by drift of the segment align-

ments. This may in turn necessitate using Coarse Alignment control to recover, adding more actuations. These conditions are not expected, as drifts at this level would make scientific operations very difficult!

- Single-grism phasing. If one grism is used rather than 2 crossed grisms, a slightly more complex operation is required. Here the visibility of some observations will be dependent on the sign of the piston, as shown in Fig. 10. For some of the steps in 1-grism phasing deliberate piston offsets will be introduced, to eliminate the ambiguity the sign effect introduces, adding a few more images and actuations.

6. FINE PHASING: WAVEFRONT SENSING

Fine Phasing picks up where the Coarse Phasing leaves off, with the PM phased to within a wave or two. Fine Phasing corrects the final alignments and simultaneously commands deformation states interior to the segments to achieve diffraction-limited imaging. For the 36-hex example, segment deformation control is limited to segment radius-of-curvature actuation; for the government team Yardstick configuration and the WCT testbed, a 349-actuator deformable mirror provides a higher degree of figure control.

Fine Phasing accomplishes its task using 2 main sub-functions. Image-based *WF Sensing* (WFS) is used to measure high-resolution maps of the wavefront, using imagery from the science cameras. WFS is the topic of this section. *WF Control* (WFC) takes the measured phase and processes it to determine actuator commands, commanding segment tilt, piston, focus (and higher-order deformations if so equipped) to minimize a *WF Error* (WFE) metric. WFC is discussed in the next section. Secondary mirror alignments will also be decided during Fine Phasing, based on WFS results from multiple field points.

Several techniques were initially considered for WFS. These included phase-shifting interferometric WF sensors, shearing interferometers, Shack-Hartmann sensors, and others. Image-based WFS using focus-diverse phase retrieval was found to have several significant advantages. Because it uses the science cameras directly, it does not require a dedicated, complex instrument, a major cost savings compared to some other techniques. In fact, it can be run in any of the imaging cameras planned for JWST, using the telescope secondary mirror piston actuation to provide defocussed images. This provides versatility, reduces the risk that an instrument failure would lead to a mission failure, and allows for the separate calibration of each instrument. It can be used in multiple field points, which provides leverage for separating telescope alignment errors from figure errors. It is also simpler and less expensive.

Image-based WFS using phase retrieval has excellent accuracy, as will be described later in this paper. Experiments conducted in the visible using WCT and other testbeds have shown repeatibility in air of better than 1nm, and accuracy (compared to phase shifting interferometers) of better than 4 nm. Good performance was observed in the IR (3.5 μm wavelength) during the recent SIRTF environmental testing as well. Demonstrated capture range exceeds a few waves — and at the longer JWST wavelengths, this can approach 50 μm peak-to-valley capture range. It has excellent spatial resolution and is quite robust to jitter and other blurring effects. Unlike the usual Shack-Hartmann sensor, it accurately measures segment piston across segment gaps, and can resolve true piston errors

exceeding λ using 2 or 3 observations at different wavelengths.

The baseline WFS approach is focus-diverse phase retrieval operating on multiple defocussed images simultaneously. Defocussed images are excellent encoders of optical aberrations. To see this, consider the 1-mirror system sketched in Fig. 18. Here a small bump on the mirror shifts the focus of a patch of the beam away from the nominal focus. The effect, after defocussing by moving the detector towards the mirror to take an intra-focal image, is to decrease the illumination in the corresponding part of the beam. The effect in an extra-focal image is to concentrate the illumination, now in the flipped location. These images are accurately predicted by a computational model relating the pupil (the mirror, with its bump) to the images through a phase shift (representing the defocus) and a Fourier transform, which propagates the beam to the detector plane. This model can be iterated, varying the parameters defining the mirror surface, to fit the defocussed images, matching the data and determining the precise shape of the bump. This process generalizes to permit measurement of any number of bumps distributed across the mirror surface, so that it can measure the full pupil wavefront.
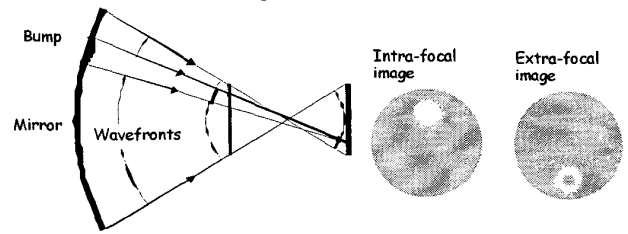

Figure 18 —Defocussed images provide clear signatures of optical aberrations.

The model that is used, and the parameters that are exercised in the iteration, can vary. The JWST baseline approach for Fine Phasing is to use Modified Gerchberg-Saxton (MGS) iterative-transform phase retrieval, for which the parameters are the thousands of individual WF samples that comprise a WF map, for its high resolution and excellent accuracy. Other algorithms are also very useful.

The classic Gerchberg-Saxton phase retrieval iteration [7] forms the basis of the MGS algorithm. It assumes an imaging system whose ~ number is known, with known detector pixel size and a known aperture mask, illuminated with monochromatic light at a known wavelength. It operates on a single, nominally in-focus image, iterating a pupil phase map between image and pupil planes, enforcing consistency with a measured image at the image plane and enforcing aperture constraints at the pupil plane. It starts with a random guess at the phase in the exit pupil, represented as the matrix θ_0. The initial pupil complex amplitude matrix f combines the phase and a pupil amplitude mask M (the ¥ denoting element-by-element multiplication):

$$f_0 = M \bullet e^{j\theta_0} \qquad (10)$$

The mask M has unit amplitude within the area of the aperture and is 0 elsewhere. The pupil field is propagated forward to the image plane by a Fourier (or Fresnel) transform, denoted F(¥). The resulting image-plane complex amplitude matrix is:

$$F_0 = F(f_0) \qquad (11)$$

8

Now the amplitude of the image-plane field is replaced by the square root of the image data I_0, and this new field is back-propagated to the pupil plane, where it is masked by the aperture function M:

$$f_1 = M \cdot F^{-1}\{\sqrt{I_0}\exp[j\ \text{angle}(F_0)]\} \qquad (12)$$

Here f_1 provides an estimate of the pupil complex amplitude; the angle of f_1 is the estimated pupil phase; and the amplitude of f_1 is the estimated amplitude distribution. This estimate is improved by iterating Eqs. 10-12 as indicated on Fig. 19, until subsequent iterates do not change by more than some small value. At iteration i, with i large, the predicted image $I_i = F_i' \cdot F_i$ is likely to compare very closely to the data I_0, and the pupil phase estimate $\theta_i = \text{angle}(f_{i-1})$ is at least consistent with the image.
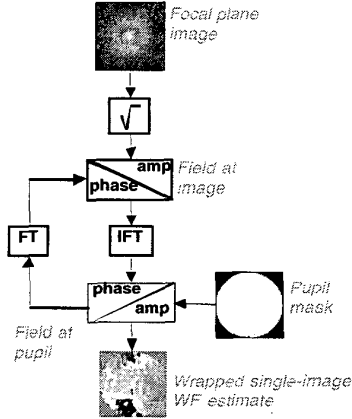

Figure 19 —Gerchberg-Saxton phase retrieval algorithm.

This algorithm is effective in generating WF estimates that reproduce the image data. These single-image phase estimates are not unique, however. Even very simple aberrations will have a sign ambiguity if retrieved from only one image. Pure defocus of a point-symmetric aperture, for instance, produces the same image on either side of the true focal point. Other aberrations, such as segment piston errors, may be less ambiguous. The algorithm is also very susceptible to detection noise, jitter, blurring and other defects, especially when used with in-focus images. It suffers from phase wrapping problems, if the objective is to estimate surface errors exceeding half a wave. Several simple modifications improve these and other problems.

The ambiguity problem is resolved by using 2 or more defocussed images, as shown in Fig. 20. In this we follow Misel and others. Misel proposed an iterative-transform method that iterates phase between 2 images taken with different focus settings [8]. The Roddiers implemented a modified Misel algorithm that explicitly enforces pupil constraints [9]. Others use the same principle in a variety of algorithms [11—21].

The effect of simultaneously processing 2 or more images with known phase shifts between the images is to provide additional data without introducing new unknowns. It is important that the diversity phase $\Theta_{DIV\ k}$ for the kth image is accurately known. The deliberately induced diversity phase $\Theta_{DIV\ k}$ can be predicted from the beam parameters, or estimated from a prescription-based optical model, or determined through calibration. The phase imparted may not be pure defocus, depending on how it is implemented. The cleanest approach is simply to piston the focal plane, but other approaches are often more practical. Lenses mounted in the camera filter wheel are the preferred method for JWST NIR-cam, and these lenses may have their own aberrations. For other JWST cameras, as for telescopes such as SIRTF and Hubble, defocus can be induced using telescope secondary mirror motion. JWST systematic aberrations will also vary with spacecraft attitude, even if the apparent field, which is corrected by a fast-steering mirror, is fixed. Field angle and defocus-induced aberrations can and should be removed from the iteration by using correct values for $\Theta_{DIV\ k}$. The MGS control software predicts $\Theta_{DIV\ k}$ using an embedded computer model of the optical system, operating on a detailed as-built prescription of the optical system to predict $\Theta_{DIV\ k}$ as a function of whatever diversity is introduced.

In the same vein, MGS incorporates an explicit priori estimate of the systematic phase into the phase iteration. This prior phase, signified Θ_{PRIOR}, is added to each image iterate prior to computing the image and subtracted before computing the pupil phase. It is common among all images in the MGS outer loop. It represents what is known of the phase aberrations of the system prior to the retrieval. This knowledge may come from system modeling, from interferometric or Shack-Hartman WF measurements, from prescription retrieval or parametric phase retrieval, from previous MGS estimates based on other data, or from previous MGS estimates based on the same data. The last instance defines an outer outer loop to the MGS, which allows for successive retrievals to refine the estimate to the point where the image iterates become very small, even if the aberrations are large. The total phase estimate from an MGS retrieval is the sum of the averaged iterates from each image inner loop and the prior phase:

$$\Theta_i = \Theta_{PRIOR} + \frac{1}{N}\sum_{i=1}^{N}\theta_i \qquad (13)$$

Another modification is to use pupil image data to constrain the field estimate f_i in the standard G-S iteration, rather than the aperture mask M more typically used in astronomical phase retrieval. Use of pupil (or Fourier plane) intensity has been described in the context of electron microscopy in [11]. The starting guess pupil amplitude combines the square root of the pupil intensity with the random phase matrix θ_0:

$$f_0' = \sqrt{I_{PUP\ k}} \cdot \exp(j\theta_0) \qquad (14)$$

Pupil images can be taken using a small flip-in lens located in front of the science detector. They show the pupil amplitude variation and precise aperture shape, as incurred at that particular field point. By substituting the square root of the pupil image data for the estimated pupil amplitude, useful information is added and many unknowns are removed from the iteration. The result is improved accuracy and convergence. Pupil images do not need to be taken with every image set, as they tend not to be affected by small aberrations.

The MGS uses a single-image Gerchberg-Saxton iteration for each (focus and field diverse) image, and then mixes the results to produce a joint estimate. All inner loops are referred to a common (in-focus) phase setting, with differences captured by. the diversity phase. For the kth inner loop, $\Theta_{DIV\ k}$ is added to the iterating phase θ_i and the prior phase Θ_{PRIOR} (discussed later) before propagating to the image plane, where it is constrained by the focal plane image I_k:

$$F_i = F\{f_i \cdot \exp[j(\Theta_{DIV\ k} + \Theta_{PRIOR})]\} \qquad (15)$$

$$F_i' = \sqrt{I_k}\exp[j\ \text{angle}(F_i)] \qquad (16)$$
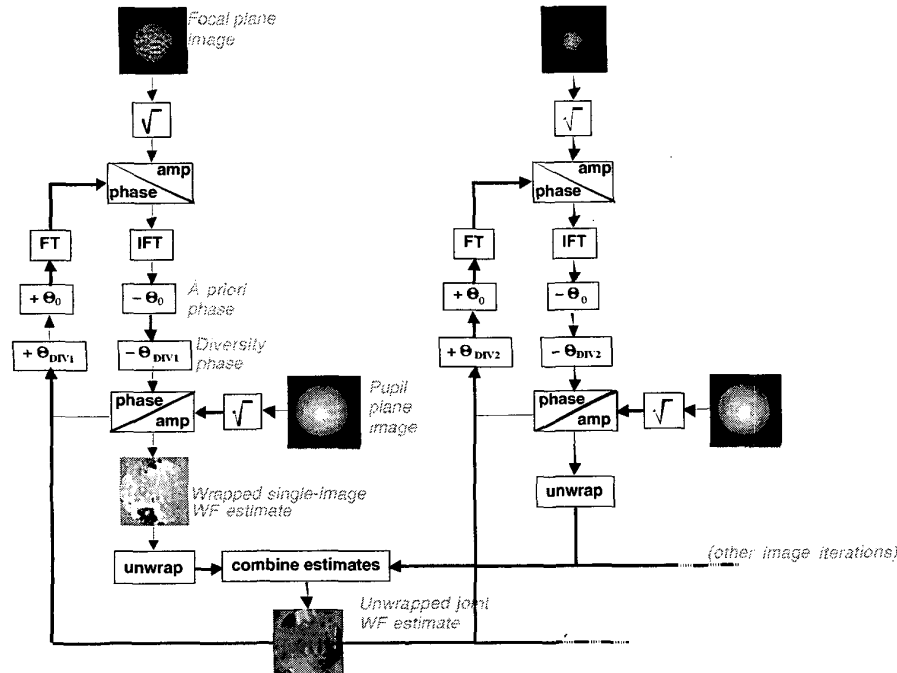
9

Figure 20 —MGS outer-loop iteration.

The diversity and prior phases are subtracted after back-propagating to generate the next phase iterate:

$$f_i = \mathrm{F}^{-1}(F_i') \cdot \exp[j(-\Theta_{DIV\ k} - \Theta_{PRIOR})] \qquad (17)$$

$$f_i' = \sqrt{I_{PUP\ k}} \cdot \exp[j\ \mathrm{angle}(f_i)] \qquad (18)$$

An example from WCT-1 showing typical images, pupil and estimate from WCT is shown in Fig. 21.As mentioned above (Fig. 18) defocussed images encode optical aberrations in a straight-forward way. They have other good properties as well. One is improved sensitivity to higher spatial-frequency aberrations. An in-focus PSF puts almost all of its signal into a few tightly clustered pixels that correspond to the lowest spatial frequencies of the WF error. Higher spatial-frequency aberrations scatter light away from the core of the in-focus PSF, creating faint signals that can be orders of magnitude fainter than the core, perhaps vanishing into the background. Defocussing pulls the higher-frequency induced structures back into a wider core, where the illumination is much more constant over all spatial frequencies. This lowers the contrast between low- and high-spatial frequency signals, improving the high-frequency resolution and sensitivity of the iteration.

This point is illustrated in Fig. 22, which shows the WF estimation error from a single-image iteration vs. the amount of defocus, for the WCT-1. Note that the best defocus value varies as a function of spatial frequency, with small defocus being better for low spatial frequencies, and large defocus being best for high spatial frequencies. An optimal mixing of the estimates from different image iterations will weight the spatially distributed estimates according to the spatially distributed estimate error, however it is not necessary to do this to meet the JWST WFS accuracy requirements. The optimal mixing filter will also likely be a function of the particular aberrations in the system when the image was taken, as the aberrations

govern the distribution of light (and hence signal) in the image. This is an area of continuing study.
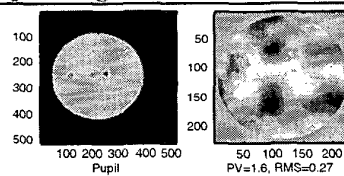
Images
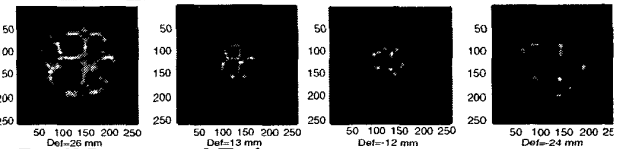
Pupil Image and Estimate

Figure 21 —Images, pupil image, and resulting WF estimate for a typical WCT-1 experiment. Random aberrations were induced using a 69-actuator aberrator deformable mirror (DM), for subsequent correction by a 349-actuator corrector DM.

Another good property of defocussed images is improved robustness to blurring effects, such as jitter, undersampling, detector charge leakage and spectral bandpass. Some of these effects can be represented as the convolution of a blur kernel with the ideal image. If the blur is larger than the in-focus PSF, it will smooth the aberration signals excessively, leading to noisy or unsuccessful retrievals. The effect of increased defocus is to enlarge the aberration signatures — the wisps, tendrils, blobs, streaks and distortions that signal hills and valleys in the optics — while the blur kernel remains the same size. Put another way, defocussing shrinks the size of the blur kernel compared to the image, reducing its contribution to WF sensing error. In general, more blurring will favor larger defocus settings.
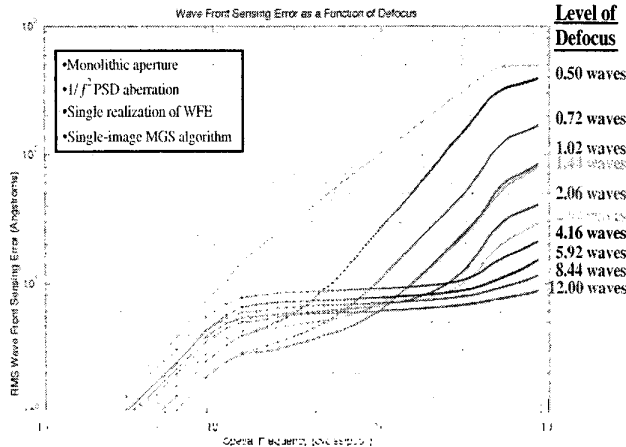
Figure 22 —WF sensing error vs. defocus and spatial frequency, for single images with low noise and no jitter.

The blurring can be measured and incorporated into the retrieval process, either by deconvolving the blur from the data images, or by inserting a blur/deblur step into the algorithm. The latter modifies the transformation from the pupil to the image plane to include a convolution at the image plane; a deconvolution occurs on the inverse transformation.

If feasible, defocussed images should be taken with exposure times that increase with defocus, so that the brightest pixel in each frame is just below saturation. This approach greatly increases the SNR of the aberration signatures in each image. If this is not done, the full advantage of defocussing will not be realized, leading to a preference for smaller defocus values and reduced high spatial-frequency sensitivity and resolution.

Image selection and image preprocessing turn out to be very important determinants of WFS accuracy. There are several considerations for obtaining the very best WFS performance:

- Noise cleanup. Calibration and removal of dark current and QE variations using standard methods.
- Image centering. The wander of the images on the CCD, due to non-systematic tilts or effects such as defocus mechanism misalignment couples strongly to perceived coma, and must be estimated and removed prior to phase retrieval processing, either through recentering the image or through adjustment of Θ_{DIVk}.
- Defocus determination. A mismatch between the actual and assumed defocus in the iteration leads to ringing artifacts in the estimate. These can be eliminated by estimating and compensating through Θ_{DIVk} or other parameters.
- Image deblurring. Image blurring from jitter, CCD effects, etc. can lead to ringing artifacts. Deconvolution or other methods can eliminate these errors.
- Image selection. Using symmetric defocus image pairs is important to reducing effects of model mismatch and noise [23]. Including an in-focus image is important for sparse segmented apertures such as WCT-2. The best amount of defocus will vary with the aberrations, noise, blurring, and desired spatial resolution, subject to photometric and operational constraints.
- Narrow band-pass filters are preferred.

We have tested the MGS algorithm in a wide array of facilities, including large telescopes such as SIRTF, and small-optics testbeds such as the WCT. Figures 23 and 24 show results from a head-to-head comparison of phase retrieval with standard interferometry in the task of measuring the figure of a small mirror. We used the NGST Phase Retrieval Camera (PRC) [22], a small, portable light source/camera combination set up much like a commercial phase-shifting interferometer, and compared results with a Zygo interferometer. The test was a differential measurement of the deformation of a 2˝ optic with a set screw. Basic repeatibility of the PRC and Zygo measurements was about 2 nm each, and the final results are about 3.5 nm (or about λ/200), as shown in the figures. Further detail is provided in [23].
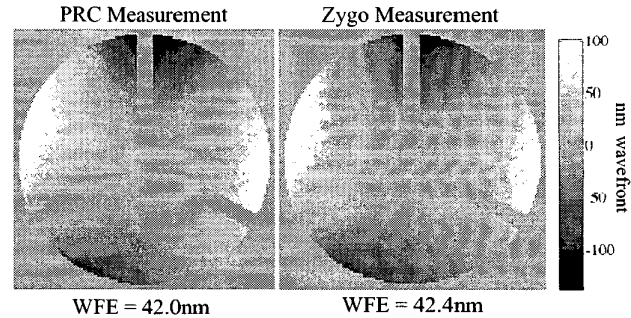

WFE = 42.0nm   WFE = 42.4nm

Figure 23 —Phase measurements derived using Phase Retrieval Camera (PRC) [22] image-based WFS and interferometrically measured WF for a small optics test.

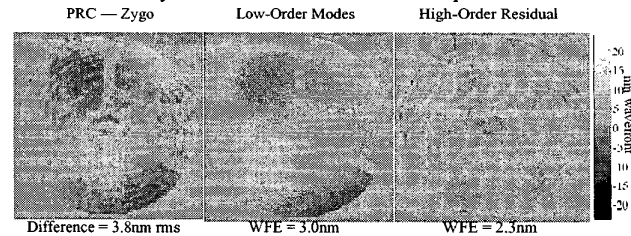Difference = 3.8nm rms   WFE = 3.0nm   WFE = 2.3nm

Figure 24 —Detail of the difference between the PRC and interferometric measurements of a small optics test piece. Most of the difference is in low spatial frequency modes,and includes the effects of mirror figure drift, alignment repeatability error, drift in instrument aberrations and lab seeing.

Further work in high-accuracy image-based WFS is continuing under the sponsorship of NASAs Terrestial Planet Finder (TPF) project. Figure 25 shows data taken during buildup of the TPF High Contrast Imaging Testbed, showing the difference between MGS phase retrieval and interferometry is again about λ/200, this time in an absolute measurement that was single-pass for phase retrieval, and double-pass for the interferometer. Subsequent experiments on this facility show consistent repeatibility for MGS of better than λ/1000 in spatial frequencies from 0 to 20 cycles/aperture. This facility will be placed in a vacuum tank for testing in a more space-like environment sometime in 2003.

A fourth MGS example is provided in Figs. 26 and 27, which show WFS results from the WCT-2 testbed, including estimates taken before and after a segment 3 piston command. The estimates resolve figure errors interior to each segment as well as the piston and tilt errors due to misalignment. The segment influence is pure piston, with very little residue from the higer spatial frequency figure errors, confirming the high

repeatability of the process. MGS is very effective with segmented systems, as shown, modulo λ. Piston differences that exceed λ can be resolved by multi-color techniques, by enforcing simple consistency rules between estimates at different wavelengths or by combining estimates to create long synthetic wavelength data.
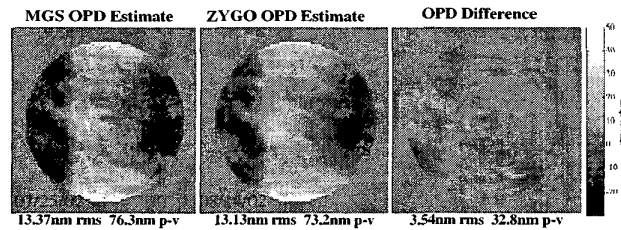
MGS OPD Estimate   ZYGO OPD Estimate   OPD Difference

13.37nm rms 76.3nm p-v   13.13nm rms 73.2nm p-v   3.54nm rms 32.8nm p-v

Figure 25 — Another test comparison between phase retrieval and interferometry. The phase maps shown here were measured during buildup of the Terrestrial Planet Finder projects s High Contrast Imaging Testbed (HCIT) [24]. The (very small) differences are due to lab seeing, non-common optics (the interferometer reference shpere), drift occuring in the 3 weeks between measurements, and some residual MGS ringing through imperfect removal of blurring effects.
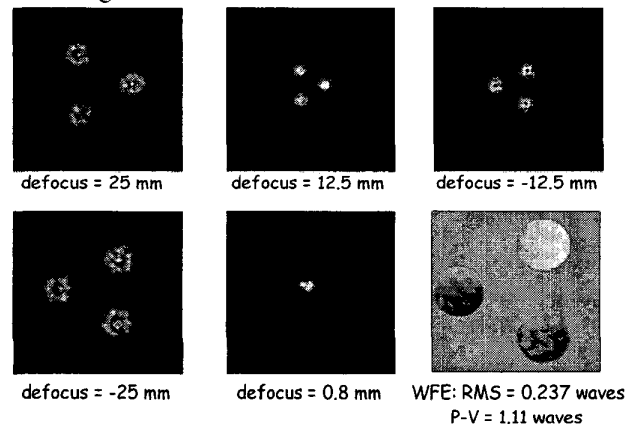

defocus = 25 mm   defocus = 12.5 mm   defocus = -12.5 mm

defocus = -25 mm   defocus = 0.8 mm   WFE: RMS = 0.237 waves
P-V = 1.11 waves

Figure 26 — Image-based WFS example from WCT-2 testbed. Five images were processed to produce the estimate in the lower left frame, including one very near focus, where interference among the segment spots is stronger.

The final major MGS modification is to incorporate phase unwrapping into the iteration. Many cases of interest have WF errors that exceed 1 wave peak-to-valley at the phase retrieval wavelength, causing them to wrap, in that the retrieved phase approaches the actual phase modulo 2π. Unwrapping is a means of reversing the modulo operation. The MGS algorithm incorporates 2 algorithms. The first is a simple raster algorithm that starts in the center of each segment and unwraps each line of pixels from center to edge. It uses 3 numerically phase-shifted angle realizations of each line to identify large phase jumps, which are resolved contextually. A voting scheme, comparing the unwrapped estimates from all images in a retrieval, helps further to identify phase jumps due to wrapping.

The second MGS unrapping algorithm is a more sophisticated (and hence slower) LP norm method [25, 26], which utilizes the observation that phase wrapping preserves the phase gradient almost everywhere. To unwrap, it develops a pixelwise fit such that the gradient of the fit matches the gradient of the

wrapped phase, except at wrap boundaries. It uses a quality map, which is the magnitude of the gradient of the wrapped phase, developed from gradients of the unwrapped phase. Highest-quality pixels are adaptively selected based on a histogram of all quality-mapped values. The fit to the remaining pixels is determined as the solution to a least-squares problem using a conjugate-gradient technique. Details of both approaches are provided in [25].
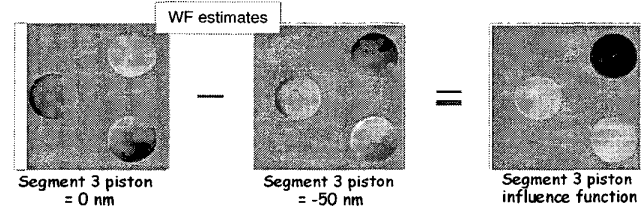
WF estimates

Segment 3 piston   Segment 3 piston   Segment 3 piston
= 0 nm   = -50 nm   influence function

Figure 27 — Difference of segmented WFs shows influence of segment piston command (WCT-2 data).

Figure 28 shows 2 large-aberration examples taken from WCT experiments. Here the injected aberration was imparted using the 69-actuator aberrator DM and corrected using the 349-actuator corrector DM. Magnitude of the injected error was limited by the aberrator DM stroke. Confirmation of the correctness of the retrieved phase is provided by the fact the the control computed from it effectively flattens the WF.
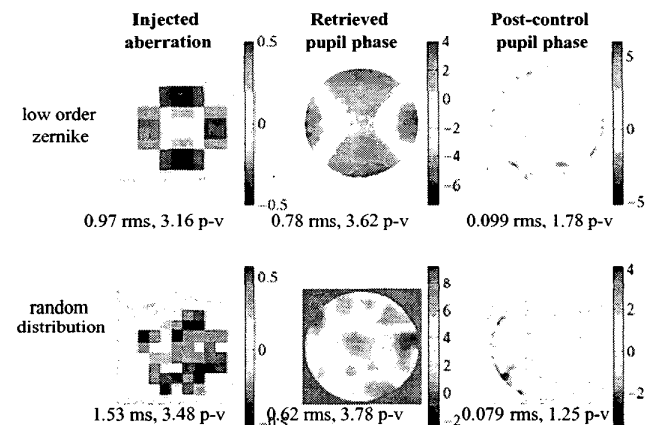
Injected   Retrieved   Post-control
aberration   pupil phase   pupil phase

low order
zernike

0.97 rms, 3.16 p-v   0.78 rms, 3.62 p-v   0.099 rms, 1.78 p-v

random
distribution

1.53 ms, 3.48 p-v   0.62 rms, 3.78 p-v   0.079 rms, 1.25 p-v

Figure 28 — High dynamic WFS/WFC examples from WCT-1. Left column shows the aberrator DM commands for the 2 cases. Retrieved phase in each case is shown in the middle column. Post WFC phase,

Other algorithms than MGS are also used. *Prescription retrieval* [20,21,27], which drives a parametric model of the optical train to match out-of-focus images, provides a useful complement to the MGS approach. Where MGS is monochromatic and requires phase unwrapping, prescription retrieval uses continuous-surface representations of the optics and can work broad-band to match highly aberrated data without phase unwrapping. Prescription retrieval uses optical-prescription parameters — surface radius, conic, and deformations; surface position and orientation — in a ray-trace and physical-optics model capable of accurately capturing all systematic error effects for an optical system. Prescription retrieval is more effective than MGS for working with very high-aberrration data, or data that is noisy, undersampled or broad-band. It is also more effective for WFS from in-focus images, and so is the basis of the PSF Monitoring function for Maintenance control [28]. Being based on optical prescription

parameters, it is capable of separating the contributions of multiple optics to the single WF, making it useful for SM Alignment.

In testing, phase retrieval WF sensing has proven to be robust and reliable, even in conditions of relatively large turbulence and jitter errors. Figure 29 illustrates this by showing the WF sensing error vs. jitter for a Monte Carlo analysis of a 7-hex segmented configuration of the WCT. Each point on this plot represents 100 trials with different noise and aberrations, through the entire image generation and processing sequence. Above 0.2 pixels jitter, the WFS error drops smoothly, reaching $\lambda/50$ at 1 pixel jitter and $\lambda/30$ at 2 pixels. These results compare with WCT performance as indicated. Note that 4 images are better than 2, though this may be more because larger defocus values are used. The flattening of the curve at 0.2 pixels is likely due to the undersampling of the PSF in WCT. Simulations with critically-sampled PSFs show significantly better WFS performance.
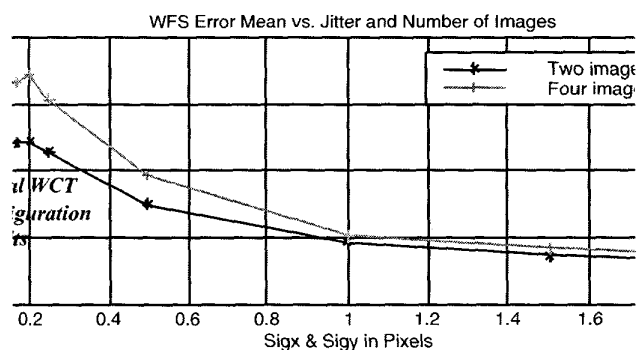
WFS Error Mean vs. Jitter and Number of Images

Figure 29 —WF sensing performance vs. jitter blur diameter in pixels (simulation).

Now let us return to the 36-hex example of the previous sections. Figure 30 shows the WFE due to segment aberrations left over after the Coarse Phasing process. The PM segment phasing errors left over from Coarse Phasing are 360 nm in this example, about $\lambda/4$ at the sensing wavelength of 2 μm. Total WFE is 5.25 μm when referenced to the true telescope axis. Much of this WFE is tilt, due to residual SM alignment errors of about 8 arcsec in tilt and 110 μm in decenter.

There are several ways in which the SM and tertiary mirror (TM) misalignments can be estimated and compensated during initialization. Gross SM misalignment can be estimated during the Coarse alignment stage, through prescription retrieval processing of separated segment imagery [27]. This will reduce SM misalignments to a level acceptable for initialization, and it is assumed in setting the SM misalignment numbers used in this example. A more thorough process for setting SM, TM and PM alignments will be pursued during post-Fine Phasing Calibration. Using field and focus diverse imagery, preferably including data from all of the instruments, a globally optimal setting for all of the telescope optics can be achieved.

This example continues by taking simulated data images, as shown in Fig. 31. The defocus settings were chosen for symmetry and good sensitivity to the full range of aberration spatial frequencies. Wavelength was 2 μm, with image detection noise as indicated in the figure caption. Since a full prescription of the NIRcam was not available at the time this example

was run, the images were taken at the telescope TMA focus, assuming sampling of 2.4 pixels across the PSF core.
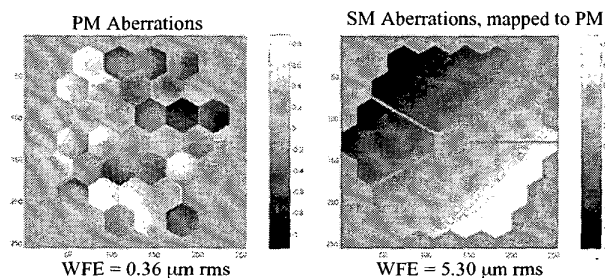
Figure 30 —Starting-point aberrations for 36-hex Fine Phasing example. Total WFE is 5.25 μm, shared between the SM and PM aberrations. The SM aberrations are dominated by tilt, which is not observed in the tilt-removed MGS images.
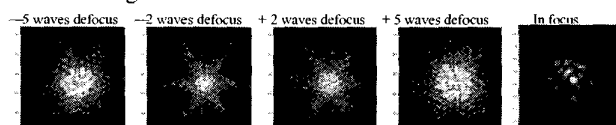
Figure 31 —Images for Fine Phasing, 36-hex example (simulated data). Defocussed images are displayed on a log stretch; the in-focus image is shown on a linear stretch. Initial WFE is 5.25 μm including tilt. Imagery was taken at 2 μm wavelength, using a single frame per defocus position (no coadding). Source flux is 9.1e7 photons, read noise is 10 e—/pixel, and dark current is 0.1 e—/sec/pix. Images are run to near full well.

The images were preprocessed using the image processing tools built into the WFC Executive software we use to run our various testbeds. For these frames, the background was subtracted, and the images were centered. Standard MGS phase retrieval processing was performed, again using the Executive. Results with and without a final phase unwrapping step are shown in Fig. 32. Comparison with the known total WFE is provided in Fig. 33, which shows agreement to the tilt-removed WFE good to 33 nm rms, or $\lambda/60$. Global tilt is not explicitly estimated at this stage.

7. FINE PHASING: WAVEFRONT CONTROL

Having taken good WF data, it remains to decide how to compensate the measured aberrations. This is the task for WFC: processing the estimated wavefront and using it to determine new actuator commands for the segments, the SM and any other controllable optic.

The wavefront estimate is retrieved in the form of a large 2-dimensional matrix, mapped spatially across the telescope exit pupil. Each entry of the matrix is the average WF phase in the corresponding small areal patch. The wavefront estimate matrix can be *vectorized* by stacking columns of the matrix on top of each other, producing the wavefront vector w. Consider as an example a wavefront matrix W sampled 5-across. It covers a circular pupil, so some entries of the matrix are always zero. The others are ordered as follows:

$$\text{order}(W) = \begin{bmatrix} 0 & 0 & 5 & 0 & 0 \\ 0 & 2 & 6 & 10 & 0 \\ 1 & 3 & 7 & 11 & 13 \\ 0 & 4 & 8 & 12 & 0 \\ 0 & 0 & 9 & 0 & 0 \end{bmatrix} \quad (19)$$

The w vector stacks the entries of W, deleting the zero entries:

$$\text{order}(w^T) = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 \end{bmatrix} \quad (20)$$

Thus $W(2,3) = w(6)$, and $w(11) = W(3,4)$. An indexing array is created to map the matrix to the vector indices, allowing easy conversion of W to w and vice versa.
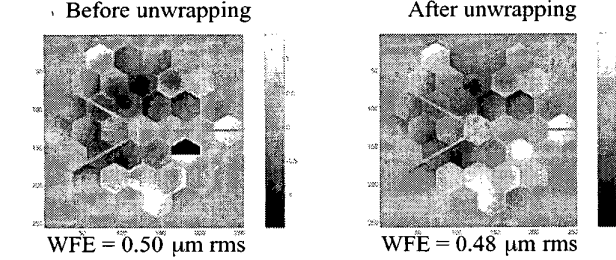

Before unwrapping   After unwrapping

WFE = 0.50 µm rms   WFE = 0.48 µm rms

Figure 32 —Estimate for 36-hex example. Estimated total WFE is 484 nm (rms).

Input Aberrations (SM+PM)   MGS Estimate (unwrapped)   Difference (tilt subtracted)

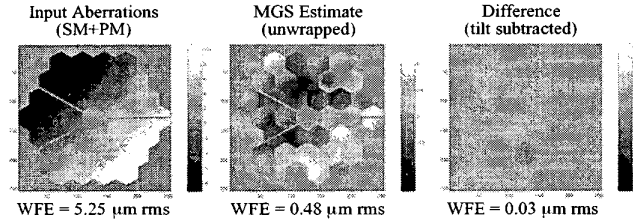WFE = 5.25 µm rms   WFE = 0.48 µm rms   WFE = 0.03 µm rms

Figure 33 —Comparison of retrieved WF to actual WF for 36-hex example. Actual WF includes PM and SM aberrations. MGS is insensitive to tilt; when tilt is subtracted from the actual WF, the agreement between the actual and estimate is within 33 nm.

The wavefront w is affected by changes to the optical state of the telescope. We define a state vector x, whose entries are the most-likely contributors to abberations of JWST, things like:

- Segment motion, in 6 degrees of freedom (DOF) per segment
- SM motion in 6 DOF
- TM motion in 6DOF
- Segment surface figure errors represented by 15 Zernike polynomials per segment
- Segment figure errors represented by a high-resolution surface map (hundreds of samples/segment)
- Similar terms for instrument optics
- Telescope pointing angles
- FSM pointing angles

The wavefront w is also strongly affected by control actuations, such as segment tilts, segment pistons, segment radius-of-curvature changes, or controlled mirror deformations. We define a vector u to capture these inputs.

Detailed nonlinear ray-trace and diffraction-based models are used in general to represent the effect of optical state variations and control commands on the wavefront, beam direction, and image quality. Well-established methods and computer tools make the determination of w from x and u an accurate and efficient process [29]. We have already shown many results generated by such simulations. For analysis purposes, though, some simplification is in order.

The wavefront w is — for small motions —*linear* function of the optical states x and the control commands u. Small here is really the range of perturbations over which the small-angle approximation is valid for a ray-trace through the optical system. This range is entirely consistent with the sort of perturbations that remain following Coarse Phasing. The linearized model of w takes the form:

$$w = \frac{dw}{dx}x + \frac{dw}{du}u + w_{nom} \quad (21)$$

Here the sensitivity matrices dw/dx and dw/du capture the influence of the individual entries in the x and u vectors on w. They can be computed through numerical differentiation, by recording the nominal wavefront w, and then poking individual states or actuators and recording the new w. This can be done using the model or by using the hardware — see the next section on calibration. The normalized influence is the difference of the pre- and post-poke w vectors, divided by the magnitude of the poke. Fig. 34 shows influence matrices computed using the 36-hex example. A complete set of matrices accumulates the influence of each optical state and each control. The w_{nom} vector is the nominal w that is present by design when the telescope is unaberrated. It is not usually zero, but rather has some small astigmatism, left in as a result of balancing aberrations between the various telescope optics to meet the telescope field performance objectives.
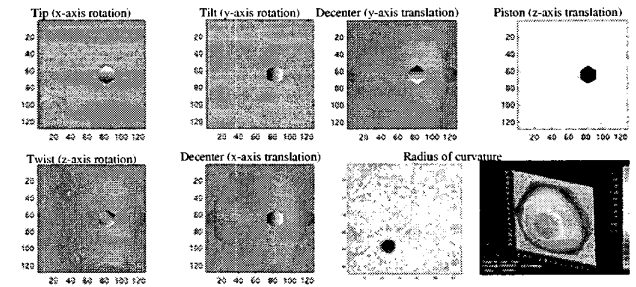

Tip (x-axis rotation)   Tilt (y-axis rotation)   Decenter (y-axis translation)   Piston (z-axis translation)

Twist (z-axis rotation)   Decenter (x-axis translation)   Radius of curvature

Figure 34 —Example influence matrices for segment rigid-body states and radius-of-curvature control, for the 36-hex example. The tip, tilt and piston DOFs appear in both x and u, as states and as controls. The decenter and twist DOFs are uncontrolled states, and appera in x only. The radius of curvature is a control only, appearing in u. Segment deformation states are included in x but are not shown here.

An unconstrained least-squares control that seeks to minimize the wavefront error squared is:

$$u = -G(w_{est} - w_{tgt}) = -\left[\left(\frac{dw}{du}\right)^T\frac{dw}{du}\right]^{-1}\left(\frac{dw}{du}\right)^T F(w - w_{tgt}) \quad (22)$$

Here F is an arbitrary spatial filter matrix, and w_{tgt} is the target WF for the control, which may be simply w_{nom}, but more generally will include calibrated offsets designed to assure balanced WFE across the multiple instruments of the observatory, as discussed in the next section. The estimate is equal to the true WF plus some WFS error, denoted as the vector dw :

$$w_{est} = w + dw \quad (23)$$

The residual WF, after application of this control to an arbitrary starting WF w_0, is w_f:

$$w_f = \left(I - \frac{dw}{du}G\right)w_0 + \frac{dw}{du}Gw_{tgt} - \frac{dw}{du}Gdw + \frac{dw}{du}du \quad (24)$$

Here du is the actuation error and I is the identity matrix.

The residual WF goes to w_{tgt} when $(dw/du)G = I$ and $dw = du = 0$. The first condition occurs when the actuation completely controls the sampled WF w. This is not the usual case. Optical actuators have finite sampling and so are spatial frequency bandpass-limited. Errors with frequency content beyond the bandpass of the WFC actuators are in the null space of $(dw/du)G$ and so are not compensated. In terms of the optical state x, the WF residuals are:

$$w_f = \left(I - \frac{dw}{du}G\right)\frac{dw}{dx}x + \frac{dw}{du}Gw_{tgt} - \frac{dw}{du}Gdw + \frac{dw}{du}du \qquad (25)$$

The previous control term, multiplying u_0, goes to zero when Eq. 22 is used. The residual here goes to zero in the same circumstances as for the arbitrary initial wavefront (Eq. 24); additionally the residuals show tolerance for optical states x that are orthogonal to the wavefront or that affect the wavefront only in ways that the actuators fully control.These equations provide the basis for WF error analysis using covariance analysis.
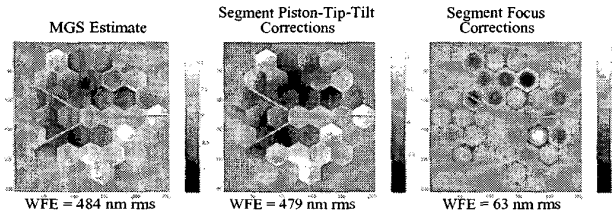

MGS Estimate   Segment Piston-Tip-Tilt Corrections   Segment Focus Corrections

WFE = 484 nm rms   WFE = 479 nm rms   WFE = 63 nm rms

Figure 35 —Controls computed for the 36-hex example of Figs. 31 and 32. The left frame shows the retrieved phase, the center and right frames show the negative of the control commands (Gw in Eq. 22) for the segment tilt and piston DOFs (center) and for the segment radius-of-curvature actuators (right).The simple linear control law of Eqn. 22 is noniterative and exact, subject to assumptions of linearity, and ignoring noise. Nonlinearities and constraints, such as relative and absolute actuator stroke limits, can be explicitly encompassed using different control formulations, most of which require iterative solution [30].

Another 36-hex example is shown in Fig. 37. This example was computed using a different telescope based on the early NGST Govt. team Yardstick point-design configuration, and includes a small quartenary deformable mirror (DM) [32] with 349 actuators. The DM is quite effective in correcting aberrations, even though it is a continuous surface device and the PM is segmented. The experiment begins by deforming and misaligning the segments, so the initial WF error is 7.5 µm RMS, with a completely broken-up in-focus image, as illustrated in the left column of the figure. The image plots here are taken at 2 µm wavelegth, sampled at ¯/28 with 27 µm pixels. The initial Strehl ratio (SR) is 0.2%. The control is computed in 2 parts, with the segments moved first, and the DM following up. The effect of the segment control is shown in the middle column — the WF error is reduced somewhat, but more significantly, the segment edges are matched to create a smooth WF that is more effectively corrected by the continuous facesheet DM. Strehl following segment-only control in this example is 10%. Finally, the DM is used to peak the Strehl to 95% (rightmost column).

A third WF control example (Fig. 38) shows the result of controlling the example of Fig. 21. The WCT AO DM was used to compensate the random aberrations induced by the random actuation of the Aberrator DM. The images show much

smoother structure, and the post-control estimate is much improved ($\lambda/15$ vs. $\lambda/4$ RMS initially). Further iteration of the control process would improve this result: depending on the magnitude of the initial Aberrator DM input, the corrector DM is able to flatten the system to about $\lambda/30$ at 633 nm wavelength. The final WF error is dominated by fitting error, or the residual WFE left when approximating the aberration surface with a finite-resolution, structured DM surface, and high spatial-frequency figure errors beyond the spatial-frequency cut-off of the DM.
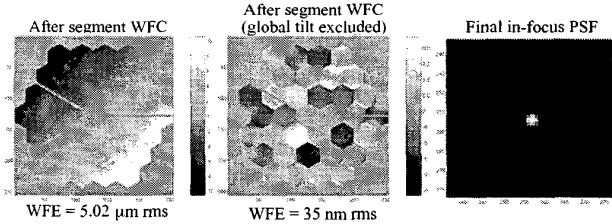

After segment WFC   After segment WFC (global tilt excluded)   Final in-focus PSF

WFE = 5.02 µm rms   WFE = 35 nm rms

Figure 36 —Wavefront and PSF after WFC for the 36-hex example. Following actuation, the tilt-removed WF shows very small residual WFE, at 35 nm (center). The excellent final PSF, with a Strehl of 95%, is shown on the right. The WF referenced to the original optical axis shows global tilt, as shown on the left. This can be identified and corrected in the Calibration control step.
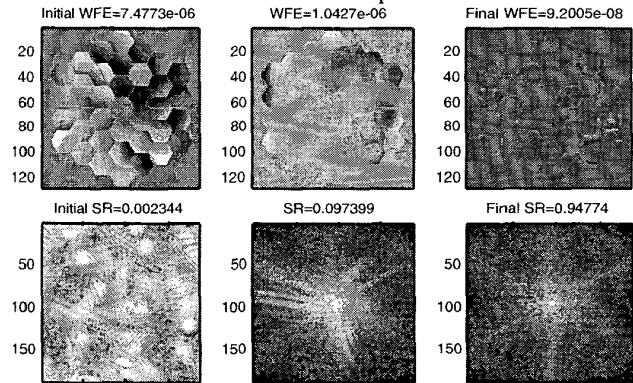

Initial WFE=7.4773e-06   WFE=1.0427e-06   Final WFE=9.2005e-08

Initial SR=0.002344   SR=0.097399   Final SR=0.94774

Figure 37 —WFC example for a 36-segment telescope with DM as well as segment rigid body control..

Defocussed images

Def=26 mm   Def=13 mm   Def=-12 mm   Def=-24 mm

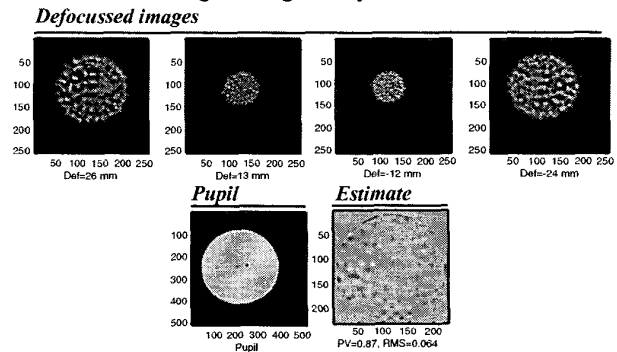*Pupil*   *Estimate*

Pupil   PV=0.87, RMS=0.064

Figure 38 —Images, pupil and WF estimate following control for example of Fig. 21 (WCT data).

8. CONCLUSION

Working with colleagues at NASA Goddard Space Flight Center, we have developed JWST WFS&C technologies to a

15

TRL (Technology Readiness Level) [31] of at least 4, signifying demonstration of the system in a laboratory environment using sub-scale components. Work on JWST WFS&C is continuing, now in partnership with TRW and Ball, the newly-selected JWST contractor team. Technologies presented here, especially Fine Phasing, are also being brought to bear on new missions such as TPF and AHMT.

9. ACKNOWLEDGEMENT

10. REFERENCES

[1] B. D. Seery, "Next generation space telescope (NGST)," Proc. SPIE 3356-01, Kona HI (1998).

[2] J. Mather, "NGST Science and Engineering Requirements," Proc. SPIE 4013-87, Munich (2000).

[3] G. Hickey, JPL, private communication.

[4] P. Petrone, S. Basinger, C. Bowers, D. Cohen, L. Burns, A. Chu, P. Davila, P. Dogota, B. Dean, J. Green, K. Ha, W. Hayden, D. Lindler, A. Lowman, C. Ohara, D. Redding, F. Shi, M. Wilson, and B. Zukowski, "Optical design and performance of the NGST wavefront control testbed," Proc. SPIE 4850-55 (2002).

[5] F. Shi, D. Redding, A. Lowman, C. Bowers, L. Burns, C. Ohara, P. Petrone, S. Basinger, W. Hayden, R. Burg, M. Wilson, and P. Davila, "NGST Segmented Mirror Coarse Phasing with a Dispersed Fringe Sensor," JPL D-24196 (2002).

[6] F. Shi, D. Redding, A. Lowman, C. Bowers, L. Burns, C. Ohara, P. Petrone, and S. Basinger, "Segmented Mirror Coarse Phasing with a Dispersed Fringe Sensor: Experiment on NGST's Wavefront Control Testbed," Proc. SPIE 4850 (2002).

[7] R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," Optik 35 (1972).

[8] D. Misel, "An Examination of an Iterative Method for the Solution of the Phase Problem in Optics and Electron Optics I, Test Calculations," J. Phys. D 6 (1973).

[9] C. Roddier and F. Roddier, "Combined approach to the Hubble Space Telescope wave-front distortion analysis," Applied Optics, Vol. 32, No. 16 (1993).

[10] R.A. Gonsalves and R. Chidlaw, "Wavefront sensing by phase retrieval," Proc. SPIE 207 (1979).

[11] J. R. Fienup, "Phase retrieval algorithms: a comparison," Applied Optics, Vol. 21, No. 15 (1982).

[12] J. Cederquist, C. Wackerman, "Phase-retrieval error: a lower bound," JOSA, (4) 9, 1987.

[13] J. Fienup, J. Marron, T. Schulz and J. Seldin, "Hubble Space Telescope Characterized by Using Phase Retrieval Algorithms," Applied Optics 32 (1993).

[14] M. Shao, M. Colavita, R. Dekany, P. Dumont, B. Hines, M. Levine, D. Redding, S. Brewer and A.

Decou, "Image Inversion Analysis of Hubble Space Telescope Optics Using a Non-Linear Least-Squares Algorithm," JPL Report (1991).

[15] J. Krist and C. Burrows, "Phase Retrieval Analysis of Pre- and Post-Repair Hubble Space Telescope Images," ST ScI Preprint No. 901 (1995).

[16] R. Lyon, J. Dorband and M. Hollis, "Hubble Space Telescope Faint Object Camera Calculated Point Spread Functions," Applied Optics 36 (1997).

[17] R.G. Paxman, T.J. Schulz and J.R. Fienup, "Joint Estimation of Object and Aberrations Using Phase Diversity," J. Opt. Soc. Am. A 9, 1072-85 (1992).

[18] M. Löfdahl and G. Scharmer, "Wavefront Sensing and Image Restoration from Focused and Defocused Solar Images," Astron. Astrophys. Suppl. Ser. 107 (1994).

[19] B.J. Thelen, R.G. Paxman, D.A. Carrara and J. H. Seldin, "Maximum *a posteriori* estimation of fixed aberrations, dynamic aberrations, and the object from phase-diverse speckle data," JOSA A , Vol. 16, No. 5 (1999).

[20] D. Redding, P. Dumont and J. Yu, "Hubble Space Telescope Prescription Retrieval," Applied Optics, V. 32, No. 10 (1993).

[21] D. Redding, A. Boden, S. Sirlin, J. Mo, R. Hanisch and L. Furey, "Optical Prescription of the HST," in *Calibrating Hubble Space Telescope: Post Servicing Mission*, A. Koratkar and C. Leitherer, eds., Space Telescope Science Institute (1995).

[22] A. E. Lowman, D. Redding, S. Basinger, D. Cohen, J. Faust, J. Green, C. Ohara, F. Shi, "Phase Retrieval Camera for Testing NGST Optics," Proc. SPIE 4850 (2002).

[23] J. Green, D. Redding, Y. Beregovski, A. Lowman and C. Ohara, "Interferometric validation of image based wavefront sensing for NGST," Proc. SPIE 4850 (2002).

[24] J. Green, D. Redding, S. Shaklan, and S. Basinger, "Extreme wave front sensing accuracy for the ECLIPSE Coronagraphic Space Telescope," Proc. SPIE 4860-48 (2002).

[25] D. Cohen and D. Redding, "NGST High Dynamic Range Unwrapped Phase Estimation," Proc. SPIE 4850 (2002).

[26] D. Ghiglia and M. Pritt, *Two-Dimensional Phase Unwrapping*, Wiley, New York (1998).

[27] S. Sirlin and D. Redding, "VSIM," NASA Tech Briefs (1997).

[28] C. Ohara, D. Redding, F. Shi, and J. Green, "PSF monitoring and in-focus wavefront control for NGST," Proc. SPIE 4850-64 (2002).

[29] *MACOS Manual*, JPL Document, 1999.

[30] A.E. Bryson and Y.C. Ho, *Applied Optimal Control*, Halsted Press, 1969.

[31] J. Mankins, http://see.msfc.nasa.gov/WorkShop/TRL Descriptions.doc

[32] M. Ealey, "Cryogenic Deformable Mirror Technology," NGST Science and Technology Exposition, Hyannis MA (1999).