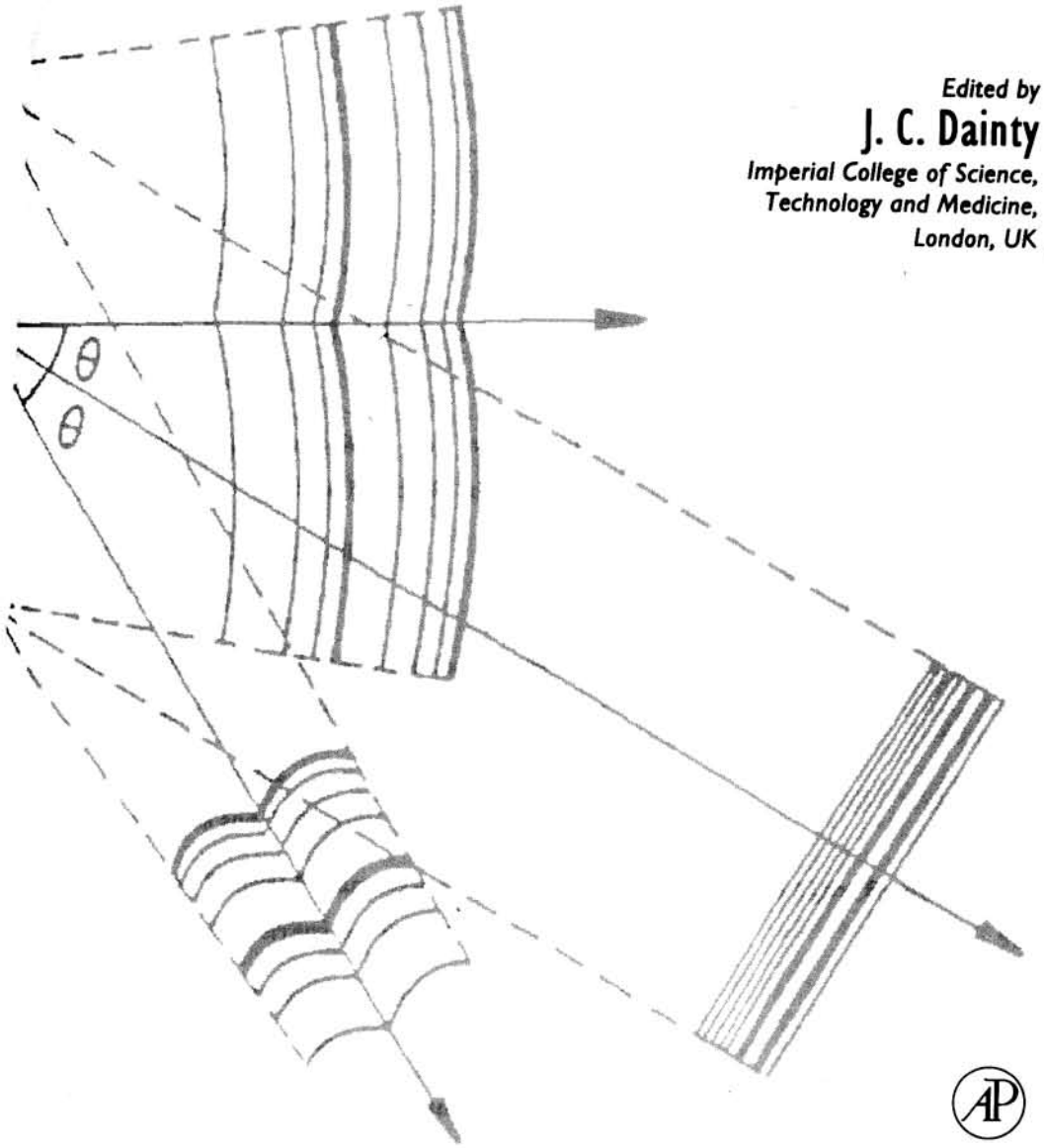
Current Trends in Optics



Edited by
J. C. Dainty
Imperial College of Science, Technology and Medicine, London, UK

ACADEMIC PRESS
Harcourt Brace & Company, Publishers
London San Diego New York
Boston Sydney Tokyo Toronto

1994

CHAPTER TWENTY

Diagnosing the Aberrations of the Hubble Space Telescope

J.R. Fienup

20.1. INTRODUCTION

Soon after the U.S. National Aeronautics and Space Administration (NASA) launched the Hubble Space Telescope (HST) into orbit in the spring of 1990, its operators discovered that it was producing severely blurred images. They recognized the point-spread function (the image of a point-like star taken through a narrowband spectral filter) as having the characteristics of spherical aberration. NASA quickly called together panels of experts to determine what had gone wrong. The commission chaired by John Allen from the Jet Propulsion Laboratory (JPL) analysed the records saved from the manufacturer of the telescope and the still-existing optical equipment used to test the telescope's primary mirror during polishing. They called these records and equipment, most of them a decade old, the fossil evidence, since they were the tangible remains of the manufacture of the telescope still on earth. The commission determined that the reflective null assembly, used with an interferometer to measure the primary mirror during its figuring and polishing, contained a mis-spaced element that could account for most of the spherical aberration.

An accurate characterization of the telescope is of great importance. To correct the telescope and restore it to its designed performance, we must know exactly what is wrong with it. However, the fossil evidence was not entirely satisfactory. Could there be other sources of error besides the mis-spaced reflective null? Furthermore, it is no longer practical to put the telescope in an interferometer and test it, since it is orbiting 500 km above the earth's surface.

An alternative to interferometry for determining aberrations (or, more generally, measuring wave fronts) is phase retrieval. Instead of putting the optical system in an interferometer, we can shine coherent light through the system and measure the intensity of the pattern of light that it produces in some plane. Using an iterative computer algorithm, we can determine what aberrations, when imposed on an optical wave front and propagated through the system, could result in the measured intensity.

Starting in June 1990, several groups of researchers rushed to apply a variety of phase retrieval techniques to characterize the aberrations of the HST. Their efforts played an important role in determining the optics required to correct the aberrations. NASA scheduled the installation of the replacement optics in the HST for late 1993.

In the remainder of this chapter we briefly review phase retrieval for wave front sensing in Section 20.2, describe how phase retrieval was adapted to the special circumstances of the HST in Section 20.3, show some example results in Section 20.4, and draw conclusions and give opinions in Section 20.5. Further details and descriptions of other approaches to characterizing the HST can be found in references [1] and [2].

20.2. WAVE-FRONT SENSING BY PHASE RETRIEVAL

We can accomplish wave-front sensing in several ways, including by interferometry, by a Hartmann (wave-front slope) sensor, or by phase retrieval [3]. In the case of the HST, we have most interest in phase retrieval because it uses the data collected by the cameras available in the telescope. Phase retrieval for wave-front sensing is mathematically similar to the problem of reconstructing an image from the modulus of its Fourier transform, which we described in the first volume of this book series [4]. For wave-front sensing by phase retrieval we typically assume that we have information about the wave front in more than one plane as it propagates. An optical wave front (coherent optical field) $f(x, y)$ in one plane propagates to become a wave front $F(u, v) = \mathcal{F}[f(x, y)]$ in a second plane, where $\mathcal{F}[\bullet]$ may be a Fresnel or Fourier transform, for example. Both wave fronts are generally complex-valued, for example, $f(x, y) = |f(x, y)| \exp[i\theta(x, y)]$, where $|f(x, y)|$ is the modulus (the square root of the intensity), and $\theta(x, y)$ is the phase, of the wave front. In neither plane do we know the phase of the wave front. In the (u, v) plane we measure the intensity $|F(u, v)|^2$. In the (x, y) plane we either measure the intensity $|f(x, y)|^2$ or know the shape of an aperture in that plane through which the wave front has passed. The latter is equivalent to knowing the intensity if the wave front originated from a point source, in which case the intensity would be a constant within the aperture and zero outside it. In other cases we may know only that the intensity is zero outside the aperture, and not know the intensity distribution within the aperture – that is, we have a support constraint.

Researchers have investigated several approaches to determining the phases of the wave fronts $F(u, v)$ and $f(x, y)$ from their intensities. Gerchberg and Saxton [5,6] pioneered an iterative transform approach for a similar problem in electron microscopy. Their algorithm iteratively transforms back and forth between the two planes, repeatedly replacing the computed modulus with the measured modulus (or setting the modulus to zero outside the support) in each plane before transforming the wave front back to the other plane. We have made extensions of, and improvements to, that algorithm [7]. Several groups have applied the iterative transform algorithm to optical wave-front sensing as well [8–10].

A second approach to wave-front sensing by phase retrieval is to use a gradient-search algorithm [7,9]. We digitally propagate a model of the wave front from the aperture plane to the plane where the intensity measurement was made. We define an error metric that measures the difference between the detected intensity distribution and the

distribution we predict from the model of the wave front. We can specify the model of the wave front either by the coefficients of a polynomial expansion of the phase function [9] or by a point-by-point description of the phase function [7]. The number of unknown parameters might be a few to a few dozen in the case of the polynomial coefficients, and a few hundred to several thousand for the point-by-point description of the phase. We compute the partial derivative of the error metric with respect to each of the parameters of the model to form the gradient. Then we can use any standard gradient search routine to minimize the error metric. Examples include steepest descent, which is usually the slowest of the algorithms, and conjugate gradient, which is much faster. We assume that the model that minimizes the error metric has a phase that is close to the true phase of the wave front. We make the gradient search approach computationally efficient by using an analytic expression for the gradient [7].

20.3. PHASE RETRIEVAL FOR THE HST

For phase retrieval to work we must have constraints on the wave front in two planes and must be able to digitally propagate the wave front between those two planes. A smeared image of an isolated point-like star detected by a CCD array in the focal plane of the HST gives us the intensity distribution $|F(u, v)|^2$, which is one constraint. Light from a single distant star striking the entrance pupil would be in the form of a plane wave with constant intensity. Therefore knowing the shape of the pupil function gives us the aperture-plane intensity distribution $|f(x, y)|^2$, which is the second constraint.

Although the light emitted by the star is spatially and temporally incoherent, the detected light is effectively coherent. If the star is unresolved by the 2.4 m aperture of the telescope, then the wave front from the star is spatially coherent over the aperture, and if the star is imaged through one of the narrowband spectral filters in the camera's filter wheel, then the wave front is quasi-monochromatic as well. Consequently we can digitally propagate a coherent wave front $f(x, y)$ from the entrance pupil to the image plane, where the wave front would become $F(u, v)$. The major aberrations of the HST are in the primary mirror, so we can include the aberrations in the phase of the wave front $f(x, y)$ in the entrance pupil.

We can most conveniently determine the parameters of the propagation integrals by the ABCD matrix method [11]. We calculate the coefficients of the ABCD matrix from the spacings, thicknesses and curvatures of the optical elements in the telescope or from paraxial ray-traces through a model of the optical system.

For the Wide-Field/Planetary Camera (WF/PC) mode of the HST, the digital propagation of a wave front through the system is complicated by the nature of the relay telescope. The main telescope, the Optical Telescope Assembly (OTA), consisting of a 2.4 m primary mirror and a secondary mirror, forms an image too large for a single CCD array to detect. Therefore several different cameras look at different parts of the field of view, each using a small relay telescope to form an image on a different CCD array. For each of the WF/PC relay telescopes, the central obscuration and the spiders (struts) holding it in place absorb light and we must include them in the wave-front propagation calculations. Unfortunately the WF/PC obscurations are in a plane that is not conjugate to the obscurations associated with the OTA. This causes two difficulties when we try to

compute accurately the wave front $F(u, v)$ in the image plane from the aberrated wave front $f(x, y)$ in the entrance pupil.

The first difficulty is that the apparent lateral position of the WF/PC obscurations (with respect to the OTA obscurations) depends on the location of the star within the field of view. Furthermore, early on we discovered that the translation of the obscurations differed from what was designed for the telescope. For this reason we had not only to retrieve the phase errors but also to infer the position of the WF/PC obscurations from the measured images of stars.

The second difficulty is that simply computing the shape of the exit pupil as the product of the obscurations in the OTA and the WF/PC is not quite accurate enough, because the obscurations are not in planes conjugate to one another. Instead we must first propagate the wave front $f(x, y)$ from the entrance pupil to the plane of the WF/PC obscurations, multiply by the transmittance of the obscurations, and then propagate that wave front to the image plane. To compute the first propagation, we need a large (about 2048×2048) fast Fourier transform (FFT) to satisfy the sampling requirements. We greatly reduce this size by performing that propagation in two steps: first, propagate $f(x, y)$ to the focal plane of the OTA; second, propagate from there to the WF/PC obscurations. Then the sizes of the two FFTs can be as small as 256×256 or 512×512. Therefore we must use three propagations (FFTs) to compute $F(u, v)$ from $f(x, y)$ accurately and efficiently.

When propagation through a complicated system having additional obscurations is necessary, we must modify both the gradient search and iterative transform algorithms. Let $g(x, y) = |f(x, y)| \exp[i\theta(x, y)]$ be our estimate (model) of $f(x, y)$, where $\theta(x, y)$ is our estimate of the phase error, and let $G(u, v)$ be the result of propagating $g(x, y)$ to the image plane. We wish to minimize the error metric

$$E = \sum_{u,v} W(u,v)\Big[\big|G(u,v)\big| - \big|F(u,v)\big|\Big]^2 \tag{20.1}$$

where $|F(u, v)|$ is the measured modulus and $W(u, v)$ is a weighting function that is zero at the locations of bad detector pixels. For the partial derivative of E with respect to a sample value of $\theta(x, y)$, we have [12]

$$\frac{\partial E}{\partial \theta(x,y)} = 2\,\mathrm{Im}\Big\{g(x,y)g^{w*}(x,y)\Big\} \tag{20.2}$$

where $g^w(x, y)$ is the wave front obtained by propagating the wave front

$$G^w(u,v) = W(u,v)\left[\big|F(u,v)\big|\frac{G(u,v)}{\big|G(u,v)\big|} - G(u,v)\right] \tag{20.3}$$

backwards through the optical system from the detector plane to the entrance pupil. Alternatively, since the aberrations are largely a smooth function, with spherical aberration dominating, we can advantageously express the aberrations in terms of a polynomial expansion:

$$\theta(x,y) = \sum_{j=1}^{J} a_j Z_j(x,y)\_ \tag{20.4}$$

For the HST, the panel agreed to employ the modified Zernike polynomials orthonormal over an annular aperture with a 0.330 obscuration ratio, which are given in reference [13], Appendix A. The partial derivative of E with respect to a polynomial

coefficient is [12]

$$\frac{\partial E}{\partial a_j} = 2\,\mathrm{Im}\left\{\sum_{x,y} g(x,y)Z_j(x,y)g^{w*}(x,y)\right\} \tag{20.5}$$

With these analytical expressions for the partial derivatives, we can compute the entire gradient of the error metric with just two propagations through the system, no matter how many parameters we use to describe the phase.

For complicated optical systems we must replace the familiar four steps of the iterative transform algorithm [5,7] by the following [12], at the kth iteration:

1. Propagate an input wave-front estimate, $g_k(x, y)$, to the detector plane, giving $G_k(u, v)$.
2. Compute $G_k^w(u, v)$ from $G_k(u, v)$ by equation (20.3) above.
3. Inverse propagate $G_k^w(u, v)$ back to the input plane, giving $g_k^w(x, y)$. Then compute

$$g_k^p(x,y) = g_k^w(x,y) + g_k(x,y) \tag{20.6}$$

4. Form the new input wave front, $g_{k+1}(x, y)$, from $g_k^p(x, y)$, $g_k(x, y)$, and the support constraint, using any version of the iterative transform algorithm [7].

For example, consider the case where we know only a support constraint. If $m(x, y)$ is unity within the support and zero outside it, then for the error-reduction (extended Gerchberg–Saxton) algorithm, step 4 is

$$g_{k+1}(x,y) = m(x,y)\,g_k^p(x,y) \tag{20.7}$$

and for the hybrid input–output algorithm, step 4 is

$$g_{k+1}(x,y) = m(x,y)g_k^p(x,y) + [1 - m(x,y)][g_k(x,y) - \beta g_k^p(x,y)] \tag{20.8}$$

where β is a feedback parameter [7,12].

In practice, we found that the iterative transform algorithm by itself worked poorly for this application, in contrast to its success for image reconstruction [7]. The gradient search algorithms optimizing the coefficients of a Zernike-polynomial expansion of the phase-error function worked much better. This is true probably because we can fairly well approximate the phase-error by just a few polynomial coefficients, and constraining the solution to be smooth by the use of a few polynomials proved to be an effective *a priori* constraint. After finding an initial phase-error estimate using polynomial coefficients, we could then use that phase-error function as the initial estimate to the iterative transform algorithm to find the fine-scale structure of the phase error. This worked because we started the iterative transform algorithm with an estimate close to the true solution.

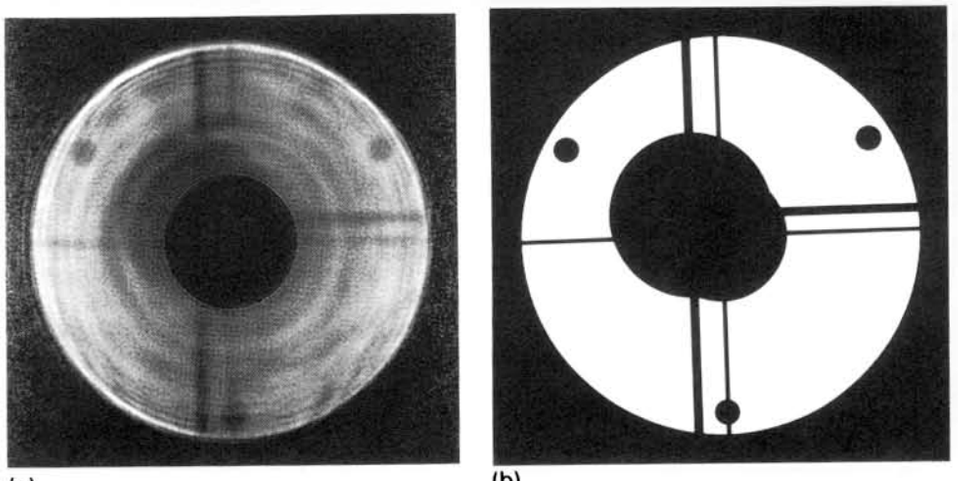
The actual retrieval procedure was complicated by the fact that the effective aperture function was not known, owing to the unexpected shift of the WF/PC pupil function relative to the OTA pupil function. A misalignment of the optical axis of the WF/PC relay telescope relative to the optical axis of the OTA is the probable cause of this problem. Consequently we used a boot-strapping approach to characterize the system. We first crudely determined the aberrations using an approximate value for the shift of the WF/PC pupil function. Next we used the iterative transform algorithm together with the estimated aberrations to estimate the translation of the WF/PC obscurations (as will be shown later). Then with the new pupil function we estimated the aberrations again, and so on. At the end we used the iterative transform algorithm to estimate a point-by-point phase map that includes the fine-scale phase errors.

For several months after the discovery of the error in the HST, the phase retrieval groups continued to add more features to their algorithms to model more fully the physics of the HST and to allow a better fit between the computed point-spread functions (PSFs) and the measured PSFs. The features that could be included in the system model include (1) multiple planes of diffraction, (2) reconstructed (as opposed to designed) pupil functions, (3) telescope jitter during the integration time of a specific PSF, (4) optimization over imperfectly known system parameters, such as plate scale (arc-seconds per pixel), or at least the use of the latest (field-position-dependent) estimates of these parameters, (5) finite optical bandwidth, (6) geometrical effects not modelled by Fresnel propagation (but that could be modelled by ray tracing), (7) the effects of undersampling by the CCD pixels and integration by CCD pixels over finite areas, (8) statistical model of the noise and bias properties of the detected PSFs, (9) higher-order phase terms beyond 22 Zernike polynomials, (10) knowledge of locations of glitches (e.g., bad pixels) in the measured data, (11) measured flat fields, (12) accounting for the possibility of aberrations in both the OTA primary and the OTA secondary, (13) knowledge of the ray-trace design aberrations as a function of field position, and (14) accounting for the possibility of non-point-like stars. Also being included to improve the solution are (15) different PSFs simultaneously and (16) the phase maps reduced from the fossil interferograms of the finished OTA primary and secondary mirrors. Also necessary is the subtraction of the spherical aberration estimated to be present in the camera in order to arrive at the spherical aberration of the OTA. With increasing sophistication, the algorithms can require prodigious amounts of computation (many giga-floating-point operations per reconstruction, which are manageable with modern computers).

Various phase-retrieval groups made several other phase-retrieval innovations. One having far-reaching potential for optical testing is "prescription retrieval" [14]. Like the phase retrieval algorithms described above, prescription retrieval minimizes the difference between a measured optical intensity distribution and one predicted by a computer model to arrive at an accurate model. But it computes the intensity distribution using a ray-trace computer program, and the parameters over which it optimizes are the optical design parameters common to ray-tracing software with which the optical system was designed. It more directly retrieves the parameters of interest to the optical designer. Thus to test an optical system or component, one could shine light through it, measure the intensity distribution in one or more arbitrary planes, and use prescription retrieval to determine the actual parameters of the system or component that would be consistent with the measurements. This approach could replace interferometry and eliminate the expense of null lenses necessary for testing aspheric elements by interferometry.

20.4. EXAMPLE RESULTS

Figure 20.1 shows an example of the reconstruction of the pupil function. We accomplished this by one iteration of the simplified iterative transform algorithm, as follows. We started with an estimate of just the spherical aberration across an annular



(a) (b)

Figure 20.1 *Pupil reconstruction. (a) Pupil function reconstructed by one iteration of the iterative transform algorithm; (b) model of pupil function inferred from the reconstructed pupil.*

aperture to form an aperture-plane wave front. We transformed the wave front to the detector plane; replaced the magnitude of that wave front with the square-root of the measured intensity; and transformed the resulting wave front back to the aperture plane, giving the result shown in Figure 20.1(a). This result gives a clear indication of the additional obscurations in the optical system that we did not include in the original wave front: (1) the four spiders (struts holding the secondary mirror) of the OTA, (2) the three pads (bolts holding the primary mirror in place) near the edges of the aperture, (3) the offset secondary obscuration in the relay telescope, and (4) the three spiders holding the secondary mirror of the relay telescope (yes, it has only three). By measuring the positions of the spiders we inferred the appropriate shift of the WF/PC obscurations, a model of which is shown in Figure 20.1(b) superimposed on the OTA obscurations.

Most of the early phase retrieval results were hampered by either poor signal-to-noise ratio, too-wide an optical bandwidth, too much telescope jitter, or having the image taken too close to the paraxial focus. Nominally in-focus images have a bright core, or main lobe, in the PSF. In order to avoid saturating the detector in the main lobe, the side lobes receive relatively few photons. However, the phase retrieval algorithms are most sensitive to the fine structure in the sidelobes of the PSF. Consequently, images taken far from focus, having lower dynamic range – having the energy spread more uniformly in the area of the PSF – were much more suitable for phase retrieval [13].

One of the most suitable images from the "HARP1A" collection was taken through a narrowband filter centred at 889 nm through planetary camera number 6 at a position well out of focus, and was designated PC6-F889N_P2. Using the polynomial-fitting algorithm we obtained the coefficients of spherical aberration (the eleventh Zernike polynomial) given in Table 20.1. We found that the value of plate scale that minimized the error metric was larger than the one originally assumed; S. Brewer, who used ray tracing of the designed system, later found that larger value to be close to the true value [13]. As can be seen from the results in Table 20.1, as

Table 20.1 Retrieved values of spherical abberation coefficient (μm RMS wave front deviation) for PC6-F889N_P2

a_{11}	Reconstruction algorithm version
–0.28	Single-plane diffraction
–0.295	Multiple-plane diffraction
–0.299	Multiple-plane diffraction, optimizing over plate scale

the level of sophistication and accuracy in modelling the system increased, our estimate of the magnitude of the spherical aberration increased as well. For this reason all the earliest results with phase retrieval, which used single-plane diffraction and too small a value of plate scale, may be biased toward underestimation of spherical aberration.

The relationship between a_{11} and the conic constant, κ, of the primary mirror of the OTA that would produce that aberration is $\kappa = -1.0023 + 0.043841a_{11}$. For example, $a_{11} = -0.299$ μm corresponds to a conic constant of –1.0154.

20.5. CONCLUSION AND REMARKS

Phase-retrieval groups estimated values of the spherical aberration that were substantially larger in magnitude than that predicted by the initial fossil evidence, the mis-spaced element in the null corrector. This discrepancy motivated NASA to perform further tests that they would otherwise have skipped due to their cost. These tests revealed that indeed there were additional small errors: the curvatures of the mirrors in the null corrector were slightly different from their design [1]. This additional fossil evidence brought their predictions closer to those of the phase retrieval results. The Jet Propulsion Laboratory (JPL) also searched through their records of the WF/PC relay optics, and they found that the relay optics also suffered from a small amount of spherical aberration which was equivalent to –0.0010 in conic constant on the OTA. Since the phase retrieval results would predict the combined spherical aberration of the OTA and WF/PC, we subtract this spherical aberration from that predicted by phase retrieval to arrive at an estimate of the spherical aberration of the OTA equivalent to a conic constant of –1.0144, which was closer still to the fossil evidence. When the effort to characterize the telescope was ended due to lack of further funds, the spherical aberration of the OTA estimated by phase retrieval was still slightly larger in magnitude than that predicted by the fossil evidence, which gave a conic constant of –1.0139 [15]. The committees deciding on the prescription to be used to correct the spherical aberration of the HST chose to use the aberration implied by the fossil data rather than that estimated by phase retrieval. The difference is small, so by using either estimate the telescope would be largely corrected. Nevertheless, after the astronauts install the correction optics (currently scheduled for December 1993), it will be interesting to see whether the HST is slightly under-corrected, as predicted by the phase retrieval results.

Note added in proof

The December 1993 Hubble repair mission was a success. The image quality has vastly improved. Researchers have analysed the new data in detail, and have concluded that the actual conic constant of the Optical Telescope Assembly is –1.0144 [C. Burrows and J. Krist, "Phase Retrieval Analysis of Pre- and Post-Repair Hubble Space Telescope Images," submitted to *Applied Optics*, 1994]. This new result agrees exactly with the value we predicted from our earlier phase retrieval results. Although the telescope appears to be not fully corrected for spherical aberration, the residual error is slight and not of great concern.

REFERENCES

1. *Space Optics* feature issue, *Appl. Optics* **32**, 1677–1804 (1993).
2. C. Roddier and F. Roddier, *Appl. Optics* **32**, 2992–3008 (1993).
3. "Wave-front sensing by phase retrieval" section, in *Wavefront Sensing, Proc. SPIE* **351** (1982).
4. J.R. Fienup, in *International Trends in Optics* (ed. J.W. Goodman), Academic Press, Boston (1991), pp. 407–422.
5. R.W. Gerchberg and W.O Saxton, *Optik* **35**, 237–246 (1972).
6. W.O. Saxton, *Computer Techniques for Image Processing in Electron Microscopy*, Academic Press, New York (1978).
7. J.R. Fienup, *Appl. Optics* **21**, 2758–2769 (1982).
8. R.H. Boucher, in *1980 International Optical Computing Conference* (ed. W.T. Rhodes), Proc. SPIE **231**, 130–141 (1980)
9. R.A. Gonsalves, *J. Opt. Soc. Am.* **66**, 961–964 (1976).
10. J.N. Cederquist, J.R. Fienup, C.C. Wackerman, S.R. Robinson and D. Kryskowski, *J. Opt. Soc. Am. A* **6**, 1020–1026 (1989).
11. A.E. Siegman, *Lasers*, University Science Books, Mill Valley, CA (1986), Chapters 15 and 20.
12. J.R. Fienup, *Appl. Optics* **32**, 1737–1746 (1993).
13. J.R. Fienup, J.C. Marron, T.J. Schulz and J.H. Seldin, *Appl. Optics* **32**, 1747–1767 (1993).
14. D. Redding, in *Technical Digest on Space Optics for Astrophysics and Earth and Planetary Remote Sensing, 1991*, Optical Society of America, Washington, D.C. (1991), Vol. 19, pp. 61–62.
15. J.B. Breckinridge and H.J. Wood, *Appl. Optics* **32**, 1677–1680 (1993).