Hubble Space Telescope characterized by using phase-retrieval algorithms

J. R. Fienup, J. C. Marron, T. J. Schulz, and J. H. Seldin

We describe several results characterizing the Hubble Space Telescope from measured point spread functions by using phase-retrieval algorithms. The Cramer–Rao lower bounds show that point spread functions taken well out of focus result in smaller errors when aberrations are estimated and that, for those images, photon noise is not a limiting factor. Reconstruction experiments with both simulated and real data show that the calculation of wave-front propagation by the retrieval algorithms must be performed with a multiple-plane propagation rather than a simple fast Fourier transform to ensure the high accuracy required. Pupil reconstruction was performed and indicates a misalignment of the optical axis of a camera relay telescope relative to the main telescope. After we accounted for measured spherical aberration in the relay telescope, our estimate of the conic constant of the primary mirror of the HST was −1.0144.

Key words: Hubble Space Telescope, phase retrieval, aberrations.

1. Introduction

Soon after the Hubble Space Telescope (HST) was launched, researchers discovered that its optics were seriously aberrated. Several groups mounted efforts to characterize accurately the aberrations and alignment of the HST. This characterization is important so that we can, first, implement a future optical correction to the aberration; second, know how to align the secondary mirror of the telescope to minimize astigmatism and coma; and third, compute analytically the point spread functions (PSF's) of the system for optimum deblurring of the degraded images currently being collected by the telescope. The latter is important because the PSF changes significantly from one camera to the next and with the position of the field of view, wavelength, and focus setting.

This paper describes elements of a multifaceted program that characterizes the HST performed in the Optical and Infrared Science Laboratory of the Environmental Research Institute of Michigan. Parallel efforts by other groups are described elsewhere in this feature issue of *Applied Optics* and in Refs. 1 and 2.

Most of the characterization approaches were derived from earlier phase-retrieval approaches that have been in use for several years in the applications of image reconstruction with astronomical interferometry[3,4] and of wave-front sensing in electron microscopy[5–7] and optics.[8,9]

In this paper we begin with the more theoretical and then proceed to the specifics of the HST. In Section 2 we calculate Cramer–Rao (CR) lower bounds on the error in estimating the coefficients of a polynomial expansion of the phase-error (aberration) estimate. This proves useful not only for determining how accurate an estimate to expect but also for selecting a focus setting for the telescope that yields a PSF from which the most accurate estimate can be made. In Section 3 we show the results of computer simulating the HST and performing aberration estimates from the simulated data. This also proves useful for determining the accuracy that can be expected and shows the potential sensitivity of the estimate to imperfections in our model of the HST. Included are the effects of multiple-plane (diffraction) propagation and an imprecisely known spatial scale and quadratic phase factor. This section ends with the results of a blind test that was distributed to several groups. Section 4 describes some of the parameters of the HST that are important in the determination of the aberrations by phase retrieval. We calculate sampling requirements that limit the utility of PSF's measured at wavelengths that are too short. We use the *ABCD* matrix method[10] to determine the parameters of the propagation integrals

The authors are with Optical and IR Science Laboratory, Environmental Research Institute of Michigan, Ann Arbor, Michigan 48113.

Received 15 May 1992
0003-6935/93/101747-21$05.00/0.

needed to simulate the optical system. In Section 5 we give the results of processing images taken with the HST. We reconstruct both the unknown phase errors and the pupil function, which also differed from the design. We describe the preliminary results in deconvolving the effects of telescope jitter, which also blurs the images. We verify that the dust artifacts in the images originate from particles on a field flattener. In Section 7 we summarize our results and draw conclusions. Among our conclusions is our estimation that the spherical aberration is somewhat larger than that for which the correction in the replacement cameras is being designed.

2. Lower Bound on Estimating Polynomial Coefficients

The CR bound is an information–theoretical lower bound on the mean-squared error of an unbiased estimate of an unknown parameter. The CR bound for the retrieval of point-by-point phase maps has been computed previously.[9,11] Here we calculate the CR bound on the mean-squared error of estimates of the coefficients of a polynomial expansion of the phase error. We begin by defining a model of the system, derive the lower bound, compute the bound, and draw conclusions.

A. System Model

In this section we employ a simplified model of the HST. The wave front $F(u)$ in the detector plane is assumed to be the Fourier transform of the wave front $f(x)$ in the aperture (pupil) plane, where u and x are each two-dimensional spatial coordinates. The wave front in the pupil is given by

$$f(x) = m(x)\exp[i\theta(x)], \tag{2.1}$$

where $m(x)$ is the binary pupil (aperture function) and $\theta(x)$ is the aberration function (phase errors). Here the aberrations are assumed to be the sum of Zernike polynomials:

$$\theta(x) = \sum_{j=1}^{J} a_j Z_j(x). \tag{2.2}$$

For the HST the commonly used phase polynomials are modified Zernike polynomials orthonormal over an annual aperture of inner radius that is 0.330 times the outer radius. They are given in Table 17 in Appendix A. The eleventh polynomial Z_{11} is the (r^4) spherical aberration term, which is the aberration in the HST of greatest interest; and the fourth, Z_4, is the (r^2) focus term, which can be varied in the optical system by moving the secondary mirror. We define the unknown coefficients as the vector $\mathbf{a} \equiv [a_1, \ldots, a_J]^T$, where $[\ ]^T$ denotes the vector transpose. The measured data $D(u)$ are modeled as independent, Poisson-distributed random variables with an expected value of

$$E[D(u)] = I_a(u). \tag{2.3}$$

Here the intensity, which is given in units of the average number of photons per detector, is

$$I_a(u) = \alpha(u)[F_a(u)F_a{}^*(u) + I_B(u)] + \sigma^2, \tag{2.4}$$

where $\alpha(u)$ is a spatially varying scale factor that is determined by the light level, detector efficiency, nonuniform gain, etc. $I_B(u)$ corresponds to the background intensity, possibly because of the detector dark current, preflash, or background light from the sky, and σ^2 is the variance associated with the CCD readout noise. The subscript a denotes that $f_a(x)$, $F_a(u)$, and $I_a(u)$ are functions of the aberration coefficients $\mathbf{a}$.

Since $D(u)$, the number of photons detected at pixel u, is a Poisson-distributed random variable with an expected value of $I_a(u)$, it has the probability distribution

$$Pr[D(u)] = \frac{[I_a(u)]^{D(u)} \exp[-I_a(u)]}{D(u)!}\cdot \tag{2.5}$$

Furthermore $D(u)$ and $D(v)$ are assumed to be statistically independent for all $u \neq v$. If we denote the entire measurement data as $\{D(u)\}_{u\in U}$, where U is an $N \times N$ array of pixels, then

$$Pr[\{D(u)\}] = \prod_u Pr[D(u)], \tag{2.6}$$

where all products and summations are over $u \in U$. The log-likelihood of the measured data, as a function of the unknown parameters $\mathbf{a}$, is then

$$\begin{aligned} L(\mathbf{a}) &= \ln Pr[\{D(u)\}_u] \\ &= \sum_u \ln Pr[D(u)] \\ &= -\sum_u I_a(u) + \sum_u D(u)\ln I_a(u) - \sum_u \ln D(u)!\,. \end{aligned} \tag{2.7}$$

Omitting terms that do not depend on $\mathbf{a}$, we may rewrite the log-likelihood as

$$L(\mathbf{a}) = -\sum_u I_a(u) + \sum_u D(u)\ln I_a(u). \tag{2.8}$$

B. Cramer–Rao Lower Bound on the Estimation Error

We denote $\hat{\mathbf{a}}$ as an estimator of the unknown parameter vector $\mathbf{a}$. Since $\hat{\mathbf{a}}$ depends on the measured data, it is a random vector. Suppose that the expectation of $\hat{\mathbf{a}}$ is equal to the true parameter vector $\mathbf{a}$. That is, $\hat{\mathbf{a}}$ is an unbiased estimator of $\mathbf{a}$. Now denote the error covariance matrix as

$$\sum_a = E\{(\hat{\mathbf{a}} - \mathbf{a})(\hat{\mathbf{a}} - \mathbf{a})^T\}. \tag{2.9}$$

The CR inequality[12] states that $\Sigma_a - \mathscr{J}^{-1}(\mathbf{a})$ is a nonnegative-definite matrix, where $\mathscr{J}(\mathbf{a})$ is the Fisher information matrix defined as

$$\mathscr{J}(\mathbf{a}) = -E\{\mathscr{J}_o(\mathbf{a})\}, \tag{2.10}$$

and $\mathscr{I}_o(\mathbf{a})$ is the observed Fisher information matrix:

$$[\mathscr{I}_o(\mathbf{a})]_{jk} = \frac{\partial^2 L(\mathbf{a})}{\partial a_j \partial a_k}. \tag{2.11}$$

An important consequence of the CR inequality is that

$$E[(\hat{a}_j - a_j)^2] \geq [\mathscr{I}^{-1}(\mathbf{a})]_{jj}. \tag{2.12}$$

Now the partial derivative of L with respect to $\mathbf{a}_j$ is

$$\frac{\partial L(\mathbf{a})}{\partial a_j} = \sum_u \frac{\partial I_a(u)}{\partial a_j}\left[\frac{D(u)}{I_a(u)} - 1\right]. \tag{2.13}$$

It then follows that

$$\begin{aligned}[\mathscr{I}_o(\mathbf{a})]_{jk} &= \frac{\partial^2 L(\mathbf{a})}{\partial a_j \partial a_k} \\ &= \sum_u \frac{\partial^2 I_a(u)}{\partial a_j \partial a_k}\left[\frac{D(u)}{I_a(u)} - 1\right] \\ &\quad - \sum_u \frac{\partial I_a(u)}{\partial a_j}\frac{\partial I_a(u)}{\partial a_k}\frac{D(u)}{[I_a(u)]^2},\end{aligned} \tag{2.14}$$

$$\begin{aligned}[\mathscr{I}(\mathbf{a})]_{jk} &= -E\{[\mathscr{I}_o(\mathbf{a})]_{jk}\} \\ &= \sum_u \frac{\partial I_a(u)}{\partial a_j}\frac{\partial I_a(u)}{\partial a_k}\frac{1}{I_a(u)},\end{aligned} \tag{2.15}$$

where we have used the fact that $E\{D(u)\} = I_a(u)$. Since

$$\frac{\partial I_a(u)}{\partial a_j} = \alpha(u)\left[\frac{\partial F_a{}^*(u)}{\partial a_j}F_a(u) + \frac{\partial F_a(u)}{\partial a_j}F_a{}^*(u)\right], \tag{2.16}$$

$$\begin{aligned}\frac{\partial F_a(u)}{\partial a_j} &= i\sum_x Z_j(x) f_a(x)\exp(i2\pi ux) \\ &\stackrel{\text{def}}{=} i\mathscr{F}_u[Z_j f_a],\end{aligned} \tag{2.17}$$

where ux is an inner product and $\mathscr{F}_u[\cdot]$ denotes the discrete Fourier transform of its argument evaluated at u, we have that

$$\begin{aligned}\frac{\partial I_a(u)}{\partial a_j} &= \alpha(u)i(-\{\mathscr{F}_u[Z_j f_a]\}F_a(u) + \mathscr{F}_u[Z_j f_a]F_a{}^*(u)) \\ &= -2\alpha(u)\mathrm{Im}\{\mathscr{F}_u[Z_j f_a]F_a{}^*(u)\},\end{aligned} \tag{2.18}$$

$$\begin{aligned}[\mathscr{I}(\mathbf{a})]_{jk} &= 4\sum_u \frac{\alpha^2(u)}{I_a(u)} \\ &\quad \times \mathrm{Im}\{\mathscr{F}_u[Z_j f_a]F_a{}^*(u)\}\mathrm{Im}\{\mathscr{F}_u[Z_k f_a]F_a{}^*(u)\},\end{aligned} \tag{2.19}$$

where $\mathrm{Im}\{\cdot\}$ denotes the imaginary part of its argument.

C. Numerical Results

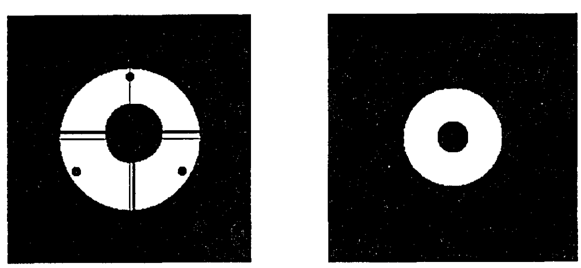
In this section the lower bound on the estimation error is evaluated numerically for some special cases. For these situations $\alpha(u)$ was constant over U and $I_B(u)$ and α^2 were zero. Two pupil functions were used: aperture 1 shown in Fig. 1(a) and aperture 2 shown in Fig. 1(b). The array size is 256×256 with aperture 1 having a diameter of 145 pixels and a central obscuration diameter of 60 pixels and aperture 2 having a diameter of 100 pixels and a central obscuration diameter of 33 pixels. For the computations the value of a_4 was permitted to vary while the remaining coefficients were set to the following values: $a_5 = 0.01, a_6 = -0.01, a_{10} = 0.03, a_{11} = 0.30$, and $a_7 = a_8 = a_9 = a_{22} = 0.00$. In this section the coefficients are in units of radians rms of the wave-front error. Shown in Tables 1 and 2 are the lower bounds on the estimation squared error of each parameter for various values of a_4 and for the two different apertures. The scale factor α was chosen so that on average $N_{\max}$, the largest number of photons in any pixel of each PSF, was 1. To determine the bound for other values of $N_{\max}$, the numbers in the table should be divided by $N_{\max}$. The rms error is given by

$$\sigma_j = \{E[(\hat{a}_j - a_j)^2]/N_{\max}\}^{1/2}. \tag{2.20}$$

As can be seen from Tables 1 and 2, the lower bounds on the estimation errors are significantly lower for the situation in which images are formed well out of focus ($a_4 = 3$) than for situations in which images are formed in focus ($a_4 = 0$) or close to focus ($a_4 = 1$). Note that since the dynamic ranges of the PSF's tend to decrease with increasing $|a_4|$ (which are more out of focus), the total number of photons tends to increase with $|a_4|$ for a fixed $N_{\max}$.

Again it should be stressed that these numbers represent lower bounds for unbiased estimates and in fact may not represent the greatest lower bounds. Hence for an unbiased estimator the actual performance may be far worse than the lower bound, whereas for a biased estimator the actual performance could be better or worse.

The lower bounds were also computed for the situation in which the focus parameter a_4 is known. These results are shown in Tables 3 and 4. As can be seen from these tables the estimation of a_{11} may be



(a) (b)

Fig. 1. Pupil functions used for the numeric computation of CR lower bounds on aberration estimates: (a) aperture 1, (b) aperture 2.

Table 1. Normalized Lower Bounds on $E[(\hat{a}_j - a_j)^2]$ for Aperture 1

j	$a_4 = 0.0$	$a_4 = 1.0$	$a_4 = 3.0$
4	96.460854	40.463634	1.017365
5	0.032564	0.021035	0.000830
6	0.047582	0.034603	0.000560
7	0.026248	0.011118	0.000346
8	0.029077	0.009690	0.000346
9	0.013022	0.018398	0.000415
10	0.013116	0.016309	0.000423
11	4.385449	1.764731	0.045521
22	0.019322	0.007839	0.000197

performed with much greater accuracy when a_4 is known.

As an example of how these tables might be used, suppose that all the coefficients are unknown, as is the case for the HST. This situation is addressed in Tables 1 and 2. Suppose that the $N_{\max}$, the maximum number of detected photons in any pixel, is 16,000 and that no pixels are saturated. Then, when Eq. (2.20) is used, for aperture 1 the lower bounds on the rms error of a_{11} for various focus settings are as follows: 0.0166 for $a_4 = 0.0$, 0.0105 for $a_4 = 1.0$ and 0.0017 for $a_4 = 3.0$. From this we see that, for well-exposed, far out-of-focus PSF's, the CR lower bound allows for acceptably low estimation errors, while the performance for PSF's that are close to focus is marginal at best at this light level.

3. Computer Simulation Results

In this section we explore the accuracy with which phase-error polynomial coefficients can be estimated from HST PSF's by means of the digital simulation of data and reconstruction experiments. We first show the importance of a multiple-plane diffraction model for imaging with the planetary camera (PC) mode of the wide-field/planetary camera (WF/PC) of the HST. Then we demonstrate the effect of imperfectly known system parameters. The section ends with the results of a blind test of our phase-retrieval algorithms.

A. Effect of Multiple-Plane Propagation

Figure 2 shows our model for the optical system for these simulations. The simplified model replaces mirrors with ideal thin lenses and eliminates several elements and folding mirrors. The main telescope, the optical telescope assembly (OTA), consisting of a positive and a negative lens (in reality a concave primary mirror and a convex secondary mirror), forms an image of a star in plane x_2. The image is formed near the surface of a four-faceted mirror in the shape of a pyramid, which divides the image plane into four quadrants. The pyramid facets also have optical power. A PC reimages each quadrant onto a CCD array in plane x_4 by a relay telescope, which is also depicted by two lenses. Only one of the quadrants and relay telescopes is shown in the figure.

The aberrated wave front, $U_1(x_1) = m_1(x_1)\exp[i\theta(x_1)]$, in the input plane of the OTA is spatially limited by the transmittance function $m_1(x_1)$, which includes the effect of the aperture diameter, the obscurations of the secondary mirror and the spiders (struts) holding it in place, and the three pads (bolts) on the primary mirror that hold it in place. On its way through the PC relay telescope, the wave front is multiplied by the transmittance function $m_3(x_3)$, which represents the central obscuration and spiders of the relay telescope in plane x_3.

The relationship between the wave fronts in planes x_1 and x_2 is given by

$$U_2(x_2) = \frac{1}{i\lambda B}\int U_1(x_1)\exp\left[\frac{i\pi}{\lambda B}(Ax_1^2 - 2x_1x_2 + Dx_2^2)\right]dx_1$$
$$= \frac{1}{i\lambda B}\exp\left(\frac{i\pi D}{\lambda B}x_2^2\right)\int\left[U_1(x_1)\exp\left(\frac{i\pi A}{\lambda B}x_1^2\right)\right]$$
$$\times \exp\left(\frac{-i2\pi}{\lambda B}x_1x_2\right)dx_1 \qquad (3.1)$$

and similarly for propagations between other pairs of planes. In the above x_1 and x_2 are both two-

Table 2. Normalized Lower Bounds on $E[(\hat{a}_j - a_j)^2]$ for Aperture 2

j	$a_4 = 0.0$	$a_4 = 1.0$	$a_4 = 3.0$
4	23.568291	43.709743	0.952042
5	0.024185	0.053788	0.000745
6	0.024064	0.048569	0.000090
7	0.013642	0.007164	0.000171
8	0.013959	0.006924	0.000174
9	0.006207	0.029878	0.000153
10	0.006186	0.030008	0.000157
11	0.979072	1.784631	0.039075
22	0.004710	0.008210	0.000181

Table 3. Normalized Lower Bounds on $E[(\hat{a}_j - a_j)^2]$ for Aperture 1 when a_4 is Known

j	$a_4 = 0.0$	$a_4 = 1.0$	$a_4 = 3.0$
4	Known	Known	Known
5	0.032412	0.020977	0.000829
6	0.047395	0.034598	0.000560
7	0.025872	0.010668	0.000345
8	0.028869	0.009541	0.000345
9	0.013019	0.018265	0.000414
10	0.013089	0.016304	0.000423
11	0.025298	0.010237	0.000368
22	0.000003	0.000004	0.000000

Table 4. Normalized Lower Bounds on $E[(\hat{a}_j - a_j)^2]$ for Aperture 2 when a_4 is Known

j	$a_4 = 0.0$	$a_4 = 1.0$	$a_4 = 3.0$
4	Known	Known	Known
5	0.024147	0.053788	0.000745
6	0.024038	0.048569	0.000090
7	0.013642	0.007164	0.000171
8	0.013956	0.006924	0.000174
9	0.006207	0.029878	0.000153
10	0.006185	0.030008	0.000157
11	0.007644	0.007265	0.000318
12	0.000001	0.000006	0.000000

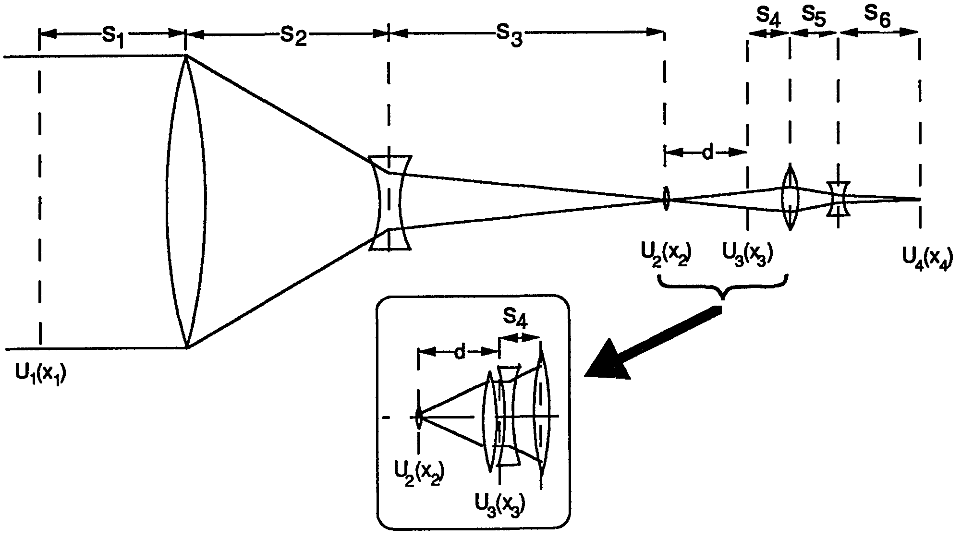


Fig. 2. Simplified thin-lens model of the PC mode of the HST. Plane x_3 contains the obscurations in the PC. A pair of imaginary thin lenses is inserted just before and just after plane x_3 to reduce the size of the FFT required for the digital propagation of a wave front through the system.

dimensional coordinates, x_1x_2 is interpreted as a dot (inner) product, and the integrals are over the interval $(-\infty, \infty)$ in each dimension. The coefficients A, B, and D are determined most conveniently by the $ABCD$ matrix method as described in Section 4. As can be seen from the second line of the equation above, this propagation can be performed with a single Fourier transform with the coefficient of the Fourier kernel being $1/(\lambda B)$.

In the digital computer we approximate the propagation given above by using the discrete Fourier transform:

$$\begin{aligned} U_2(x_2) &= \exp(i\alpha_1 x_2^2)\frac{1}{N}\sum_{x_1=0}^{N-1} U_1(x_1)\exp(i\beta_1 x_1^2) \\ &\quad \times \exp(-i2\pi x_1 x_2/N) \\ &= \exp(i\alpha_1 x_2^2)\mathscr{F}\{U_1(x_1)\exp(i\beta_1 x_1^2)\}, \end{aligned} \tag{3.2}$$

where x_1 and x_2 are now redefined to be integer sample (pixel) numbers for pixels separated by physical distances Δx_1 and Δx_2 for the two respective planes (which are assumed to be the same in each of the two dimensions). The $N \times N$ discrete Fourier transform is computed by using the fast Fourier transform (FFT) algorithm. Comparison of the discrete and continuous transforms reveals the following relationships:

$$\frac{1}{N} = \frac{\Delta x_1 \Delta x_2}{\lambda B} \quad \text{or } \Delta x_2 = \frac{\lambda B}{N\Delta x_1}, \tag{3.3}$$

$$\alpha_1 = \frac{\pi D \Delta x_2^2}{\lambda B}, \tag{3.4}$$

$$\beta_1 = \frac{\pi A \Delta x_1^2}{\lambda B}, \tag{3.5}$$

and similarly for the other propagations.

The requirements on sampling are as follows. To capture the OTA primary mirror, we need $N\Delta x_1 \geq 2.4$ m. To represent the OTA spiders, of 25.6-mm width, we need $\Delta x_1 \leq 25.6$ mm (assuming the gray-level masks as described below). To capture the spiders of 0.834-mm width in the relay telescope, we need $\Delta x_3 \leq 0.834$ mm. To match the CCD pixel spacing in the PC, we need $\Delta x_4 = 0.015$ mm. To capture the entire PSF, we need $N\Delta x_4$ to exceed the sensible diameter of the PSF. The quadratic phase factors determined by the A and D parameters must not be aliased. The difference in a quadratic phase from one pixel to the next should not exceed π radians:

$$\alpha_1\left[\left(\frac{N}{2}\right)^2 - \left(\frac{N}{2} - 1\right)^2\right] = \alpha_1(N-1) \leq \pi, \tag{3.6}$$

and similarly $\beta_1(N-1) \leq \pi$.

It is possible to accomplish the transformation from plane x_1 to plane x_3 with a single FFT. However, to satisfy the requirements above, we found that N of ~ 2000 (2048 for an efficient FFT) would be required. This would be computationally expensive. By propagating from plane x_1 to plane x_2 and then propagating from plane x_2 to plane x_3, we can use much smaller FFT's. Furthermore we kept $\alpha_2(N-1) \leq \pi$ for the second propagation by keeping track of the quadratic phase term $\alpha_2 x_3^2$ analytically for the second and third propagations. This is conceptually identical to inserting a thin lens of focal length d just before plane x_3 and a thin lens of focal length $-d$ just after plane x_3. This pair of imaginary lenses cancel one another in terms of their effect on the optical system but allow the use of FFT's of smaller size in the digital propagations.

We can use a binary mask to represent the obscurations. However, for extremely fine features, such as the spiders, two or more samples across a spider would be necessary to represent it accurately. This would make the sampling requirements more demanding. An alternative is to represent binary obscurations with gray-level transmittance functions. Think of each pixel as a square with the center at the sample value. If there is no obscuration within the square (it is all in the clear aperture), the transmittance of the sample is 1.0, and if the square is totally obscured, the transmittance of the sample is 0.0. For a partially obscured square the transmittance of the sample is the fraction of the square that the clear aperture covers. Transmittance functions $m_i(x_i)$ made in this way should more accurately represent the obscurations than binary masks using the same sampling rate. Functions made in this fashion have the appearance of finer resolution and avoid much of the staircasing effects seen in binary representations. With the computational tricks of using gray-level transmittance functions, the additional propagation, and the imaginary pair of lenses, we were able to use FFT's of size $N = 512$ and even 256 with confidence.

For the model of the HST that uses three propagations including the imaginary pair of lenses at the plane of the PC obscurations (x_3), most of the quadratic phase coefficients, α_i and β_i, are negligibly small, except for those in the first focal plane, in plane x_2. There the sum of all the quadratic phase terms is $2\pi BQx_2{}^2$, where the quadratic coefficient is given by

$$BQ = \frac{1}{2\pi}\left(\alpha_1 - \frac{\pi\Delta x_2{}^2}{\lambda f_{py}} + \beta_2\right) = \left(\frac{D_{12}}{B_{12}} - \frac{1}{f_{py}} + \frac{A_{23}}{B_{23}}\right)\frac{\Delta x_2{}^2}{2\lambda}, \tag{3.7}$$

where α_1, D_{12}, and B_{12} are from the propagation from plane x_1 to plane x_2; β_2, A_{23}, and B_{23} are from the propagation from plane x_2 to plane x_3; and f_{py} is the focal length of the pyramid in plane x_2. The units of BQ are waves per square pixel.

We explored the importance of the multiple-plane propagation model over a simpler single-FFT model of the optical system in a first set of simulation experiments. The values of the system parameters, as indicated in Fig. 2, for these simulations are given in Table 5. (Later work in retrieving the aberrations from HST data used updated parameters, as described in Section 4.) For the purpose of wave-front propagation the OTA secondary obscurations and the primary mirror all can be considered to be in the same plane, making the exact value of S_1 unimportant. In these simulations it was taken to be zero, whereas in Section 4 it was taken to be 4.907 m. The values of A, B, C, and D for the three propagations, computed from the values in Table 5, are given in Table 6. We performed the simulations using $N = 256$ and a wavelength of $\lambda = 0.889$ μm. The sample spacings and array widths for this set of simulations is given in Table 7. The aberration coefficients for the focus and spherical-aberration Zernike polynomials were assumed to be $a_4 = -2.0$ waves and $a_{11} = -0.25$ waves. Simulations were performed for two cases: (1) where OTA and WF/PC obscurations were both in their respective planes and (2) where all obscurations were in a single plane. Case 1 represents the more accurate, multiple-plane propagation simulation of the HST, and case 2 is essentially the single-Fourier-transform approximation that is used in the basic phase-retrieval algorithm. Figures 3(a) and 3(b) show, respectively, the OTA and PC obscurations used for case 1, and Fig. 3(c) shows the composite obscuration mask used for case 2. (Because of the number of FFT's involved, the OTA mask must be rotated by 180° for the multiple-plane propagation compared with the single-plane propagation for one to arrive at a PSF that is not rotated.) The field intensity $|U_4(x_4)|^2$ at the CCD camera after a plane wave is propagated through the HST model for case 1 is shown in Fig. 4(a). This PSF, called $\mathrm{PSF_{both}}$ (where the obscurations are in both planes), is difficultto distinguish visually from the PSF for case 2, $\mathrm{PSF_{OTA}}$ (where all obscurations are in the OTA plane). However, the differences caused by multiple-plane diffraction effects are evident in Fig. 4(b), which shows the difference image ($\mathrm{PSF_{both}} - \mathrm{PSF_{OTA}}$), where each PSF was normalized to have a peak value of unity prior to subtraction. The difference image has a mean value of -0.0002 with a range of $[-0.067, 0.073]$. Another measure of the significance of multiple-plane diffraction effects is the absolute rms error:

$$E = \left[\min_{C} \frac{\sum_{x_4}[C(\mathrm{PSF_{both}}) - \mathrm{PSF_{OTA}}]^2}{\sum_{x_4}[\mathrm{PSF_{OTA}}]^2}\right]^{1/2} = 0.085, \tag{3.8}$$

where the constant C compensates for image-scaling differences. The 8.5% error observed is significant enough to effect the phase-retrieval estimate of the HST aberrations.

Table 5. Parameters of the HST (PC) for Simulations

Symbol	Description	Value (mm)
S_1	OTA input plane to primary	Arbitrary
f_1	OTA primary focal length	5520.0
S_2	OTA element separation	4906.071
f_2	OTA secondary focal length	−679.0
S_3	OTA secondary to pyramid	6406.200
f_{py}	Pyramid focal length	1534.2
d	Pyramid to PC obscuration	895.35
S_4	PC obscuration to primary	234.851
f_3	PC primary focal length	249.840
S_5	PC element separation	234.851
f_4	PC secondary focal length	−112.380
S_6	PC secondary to detector	364.441

Table 6. ABCD Values for Simulations[a]

Propagation	A	B (mm)	C (mm^{-1})	D
$x_1 \to x_2$	-1.3×10^{-8}	57,600	-1.74×10^{-5}	8.22544
$x_2 \to x_3$	1	895.35	−0.0011169	0.0
$x_3 \to x_4$	-3.096×10^{-5}	1078.1	−0.0009276	2.27516

[a]The pyramid is handled separately.

Table 7. Sample Spacings and Array Widths for Simulations (N = 256, λ = 0.889 μm)

	Spacing Δx_i (mm)	Width $N\Delta x_i$ (mm)
1	16.656	4263.9
2	0.01201	3.0743
3	0.25891	66.281
4	0.01446	3.7019

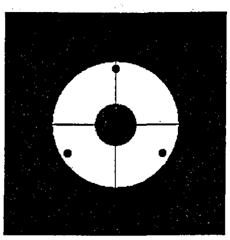
We quantified the significance of multiple-plane diffraction further by performing phase retrieval with single-plane propagation on the two different PSF's, fitting Zernike coefficients through a_{22}. The gradient search algorithms for retrieving the phase coefficients are described in Ref. 13. The algorithms minimize the mean squared error:

$$E = \sum_u W(u)[|G(u)| - |F(u)|]^2, \tag{3.9}$$

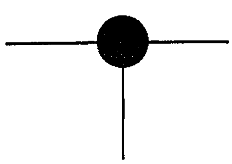
where $|F(u)|$ is the square root of the measured PSF, $|G(u)|$ is the magnitude of the model of the aberrated wave front in the pupil digitally propagated to the PSF plane (the CCD plane), and $W(u)$ is a weighting function (needed for the real data), which is zero wherever there are bad pixels or wherever the signal-to-noise ratio drops too low. When we report results, we give the normalized rms error

$$\text{err} = \left\{\frac{\sum_u W(u)[|G(u)| - |F(u)|]^2}{\sum_u W(u)|F(u)|^2}\right\}^{1/2}. \tag{3.10}$$

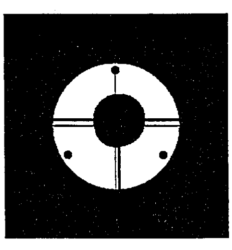
For PSF_{OTA} the single-plane phase-retrieval algorithm yielded the correct Zernike coefficients to within 0.001 waves rms, since the system model used in the phase-retrieval algorithm matches that used to simulate the data. For the PSF_{both} the single-plane phase-retrieval algorithm produced an estimate of -1.977 waves for a_4 (off by 0.023 wave), -0.2545 wave for a_{11} (the magnitude of the spherical aberration was overestimated by 0.0045 wave), and up to 0.004 wave for the other coefficients (which all had true values of 0).



(a)

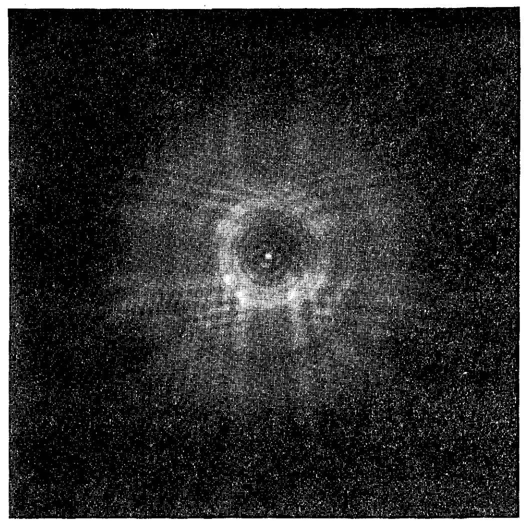


(b)

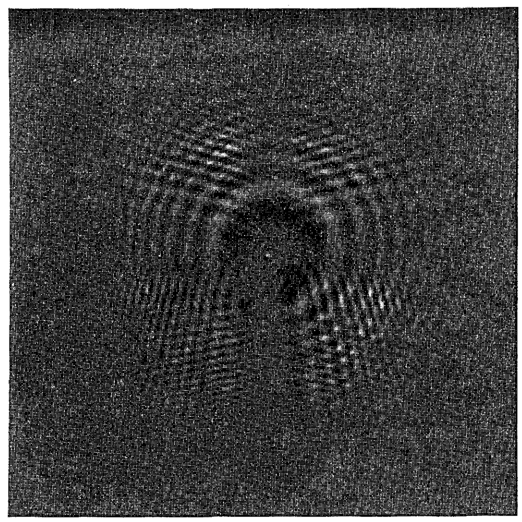


(c)

Fig. 3. Transmittance functions for OTA and PC for simulations: (a) OTA, (b) PC, (c) composite for the single-FFT model.



(a)



(b)

Fig. 4. (a) Simulated PSF. (b) The difference in the PSF's computed using single-FFT and multiple-plane propagation models of HST.

This is not a large error, but the single-plane phase-retrieval algorithm ultimately prevents us from obtaining a more accurate estimate of the coefficients. From Eq. (3.7) and Tables 6 and 7 the value of the quadratic phase factor, which determines the amount by which the OTA obscurations are out of focus at the plane of the PC obscurations, is BQ = 0.00005, and this is the value used for the experiments described above. As discussed below this has approximately the same magnitude but the opposite sign compared with the true value for the HST, which was determined later, as described in Section 5. For this reason an algorithm using a single-plane propagation and a value of BQ with the correct (negative) sign might be expected to underestimate the magnitude of the spherical aberration for HST data.

A. Incorrect Spatial Scale and Obscuration Position

The plate scale refers to the relationship between the arcseconds of the angle in the sky and the number of pixels of the CCD detector. It can be derived from the *ABCD* matrix values describing the optical system, but it can also be determined independently by measuring the distances between the images of stars of known separation. The plate scale also affects the spatial scale in the pupil function that we model, which is related to the spatial scale in the CCD plane by equations of the form $\Delta x_2 = \lambda B/(N\Delta x_1)$. The correct scale factors were known in practice only approximately, and they vary as a function of the location of the PSF on the CCD chip in a way that was not completely characterized at the time of this work. Brewer[14] later performed ray tracing to satisfy this need, as described in Section 4. We determined the sensitivity to the plate scale by simulating data using one spatial scale factor and using an incorrect scale factor in the retrieval algorithm. For these simulations the phase-retrieval algorithm used multiple-plane propagation, and the (incorrectly) assumed scale factors were independently changed for the OTA obscurations and the PC obscurations, or the PC obscurations were translated to simulate an unknown shift of the PC obscuration.

Selected results of these simulation experiments are summarized as follows. If the scales of both sets of obscurations were changed by up to $\pm 3\%$, the error in a_{11} was up to 0.004 wave (assuming that the quadratic phase BQ was correct). As will be seen below, with the real data the sensitivity to the plate scale was considerably greater. When a single parameter was changed, an error in a_{11} of 0.01 wave rms (which would be of considerable concern) was caused by (1) an OTA spatial scale error of +4% or (2) a PC spatial scale error of $\pm 6\%$. A PC shift of ± 2 pixels caused an error in a_{11} of only ± 0.002 wave. In all the cases above the normalized rms error of the fit, given by Eq. (3.10), was in the range of 0.09–0.12. It is unlikely that spatial scale errors exceeding 1% will occur when our best estimates of the actual system parameters are used; therefore spatial scale errors should not ultimately be a limiting factor in estimating a_{11}. However, for determining the other aberrations a spatial scale error of 1% could be significant. In particular a retrieval where the scale of the OTA is too small causes little change in a_{11} but a large change in a_{22}, which could be of concern. It does not appear to be likely that shifts of the PC obscurations will be the limiting factor in estimating a_{11} either as long as we use our best estimates of the actual shifts. However, for determining other aberrations, in particular asymmetric aberrations such as coma, we expect a greater sensitivity to shifts in the PC obscurations.

The results above are for the case of knowing *a priori* the correct value of the quadratic phase factor BQ used in the multiple-plane propagation. If BQ is not known exactly, its value should be optimized along with the plate scale and the aberration coefficients. For the simulations described above, the value of BQ was 0.00005. In another set of simulation experiments a range of values of the plate scale was assumed during different applications of the phase-retrieval algorithm, and multiple values of BQ were used for each case to determine the value of BQ that minimized the fitting error for each of the plate scales. Then, with the optimum values of BQ for each plate scale, the aberrations were reoptimized. The results are given in Table 8. The values of the estimated a_{11} are plotted in Fig. 5 as a function of the plate scale for both cases: when the true value of BQ was used and when the optimum value of BQ was used. The estimated a_{11} varied much more for the optimum value of BQ (which is optimum in the sense that it minimizes the error metric for a given assumed plate scale) than for the true value of BQ. When the optimum value of BQ was used, the estimated a_{11} decreased in magnitude as the value of the plate scale was underestimated.

B. Blind-Test Results

J. Holtzman (Lowell Observatory, Flagstaff, Ariz.) simulated PSF's with realistic amounts of noise and aberrations for a variety of focus settings and distributed the data for blind tests through several phase-retrieval groups.

We processed only the data with a spectral filter corresponding to a wavelength of 889 nm, since at only that wavelength was the Nyquist simulated data Nyquist sampled for the optical fields. The data were crudely filtered by zeroing out the data outside a circle beyond which the signal-to-noise ratio appeared (by eye) to decrease below unity. These results were obtained early in our effort, before we had implemented a weighting function in the phase-retrieval algorithm. All reconstructions reported here used $N \times N = 256 \times 256$ arrays for the FFT's. Limited reconstructions with 512×512 arrays were also performed, and those results differed little from the 256×256 case. The pupil function used was the one circulated in late August 1990 but with the PC secondary obscurations moved 2 pixels in both dimensions (i.e., diagonally). This pupil shift was determined by the iterative transform algorithm.[13] Our results are summarized in Table 9, which shows recovered Zernike coefficients a_j for PSF's simulated for five different (unknown) focus settings, designated *A–E*.

Table 8. Effect of the Plate Scale on Retrieved Zernike Coefficients with Optimized Quadratic Coefficient *BQ*

Root-Mean-Squared Error	Optimum BQ	a_4	a_{11}	a_{22}	Value of Scale s
0.119	0.000117	−2.435	−0.292	−0.0011	3% Too small
0.000	0.00005	−2.500	−0.300	−0.0020	Exact
0.064	0.000031	−2.527	−0.300	−0.0014	1% Too large
0.116	−0.000014	−2.566	−0.304	−0.0023	3% Too large
0.141	−0.000069	−2.568	−0.313	−0.0097	5% Too large

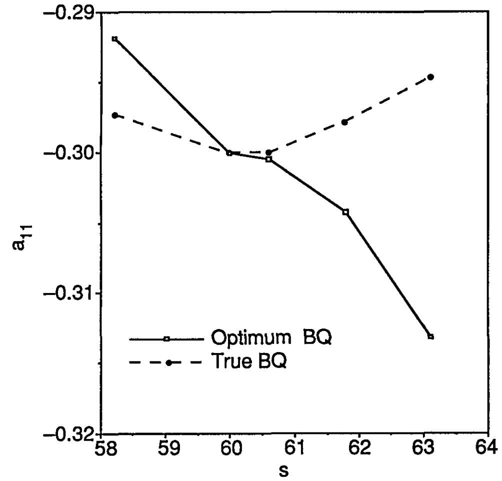


Fig. 5. Estimated spherical aberration a_{11} as a function of assumed spatial scale s for simulated data. The solid curve shows the quadratic phase factor BQ optimized for the given spatial scale; the dashed curve indicates when the true value of BQ is used.

From Table 9 we see that the only coefficients for which the mean value is significantly greater than the standard deviation among the results are (with their mean values) $a_7 = -0.014$, $a_8 = 0.005$, and $a_{11} = -0.227$ (and a_8 differs from 0 by only 2σ, and thus our confidence level in a_8 was not high). All these values agree well with the actual values given in the last column, which were revealed to us at a later date. Note that for the nominally in-focus data, given by image C, the value of spherical aberration a_{11} is farthest from the true value and is substantially underestimated. This is consistent with the results obtained with CR lower bounds and is one of the reasons why we prefer PSF's that are far out of focus.

4. Parameters for the HST

In this section we discuss the specific parameters of the HST that are important to the phase-retrieval algorithms.

A. Sampling Requirement

To avoid aliasing,[15] for an aperture of diameter D, we must sample the detected image intensities with an angular sampling interval (as projected on the sky) ρ_s, which satisfies

$$\rho_s \leq \lambda/(2D), \tag{4.1}$$

whereas for optical fields the sampling requirement is milder:

$$\rho_s \leq \lambda/D. \tag{4.2}$$

The latter requirement is pertinent to the phase-retrieval algorithms that we used, since we must digitally propagate optical fields back and forth within the system. Phase-retrieval algorithms that require only a forward propagation through the system do not have this restriction. Table 10 shows the as-designed sample spacings (in arcseconds and in microradians) of the detector arrays for the various cameras in the HST and the shortest wavelengths for which the optical intensity and optical field are adequately sampled at the detector. For both the wide-field camera (WFC) and the PC, the center wavelength of the narrow-band spectral filter with the longest wavelength is 889 nm. From Table 10 we see that even the optical field at this wavelength is undersampled in the WFC. Consequently we did not use any data from the WFC. There were suitable spectral filters for the faint-object camera (FOC). However, the images from the FOC are quite noisy compared with those from the WFC and PC since the FOC is count-rate limited. For these reasons we restricted our attention to PC images taken through narrow-band filters with wavelengths above 500 nm, which are the most suitable for characterizing the aberrations of the HST.

When a single-FFT model of the optical system is used, if we compute an FFT of the field in the detector plane that has angular sampling interval ρ_s, the resulting field in the aperture plane has a physical width of

$$D_a = \lambda/\rho_s. \tag{4.3}$$

Therefore, for an array in the computer of width N pixels, the scale factor giving the number of pixels per meter of the physical aperture is

$$s = N/D_a = N\rho_s/\lambda, \tag{4.4}$$

where $N\rho_s$ is the angular width (projected onto the sky) of the array in the detector plane over which the

Table 9. Zernike Phase Coefficients Estimated in a Blind Test

j	A	B	C	D	E	s	Mean	Actual
4	−2.089	−1.038	0.004	1.051	2.104	N/A	N/A	various
5	0.002	0.003	−0.004	−0.002	0.001	0.002	0.000	0.001
6	−0.009	−0.004	0.009	−0.010	0.009	0.008	−0.001	0.007
7	−0.010	−0.013	−0.013	−0.015	−0.018	0.003	−0.014	−0.012
8	0.004	0.008	0.002	0.005	0.008	0.002	0.005	0.004
9	0.004	0	−0.001	0.001	0	0.002	0.001	0
10	0.005	−0.002	0	−0.001	0	0.002	0.001	0
11	−0.227	−0.225	−0.209	−0.235	−0.241	0.011	−0.227	−0.236
22	0	0	0.002	0.001	0	0.001	0.001	0
err	0.164	0.209	0.341	0.216	0.171			

Table 10. Sample Spacings for HST Cameras and Wavelengths for Nyquist Sampling

	ρ_s		λ (nm)	
Camera	ρ_s (arcsec)	ρ_s (μrad)	For Intensity	For Field
PC	0.043	0.2085	1000	500
WFC	0.100	0.4848	2333	1667
FOC $f/48$	0.044	0.2133	1023	512
FOC $f/96$	0.022	0.1067	512	256
FOC $f/288$	0.007	0.0339	163	081

FFT is performed. The width D_O in pixels of the aperture of the OTA is given by (2.4 m) s. D_O should be large enough to model accurately the features of the aperture function including the spiders and pads. Since the D_O is ~ 100 times the width of a spider and at least 1 or 2 pixels are needed to represent the width of a spider, we prefer to have $D_O > 100$ or 200 pixels. Table 11 lists the values of s and D_O for all the cameras on the HST, for some of the wavelengths for which there are narrow-band filters, and for $N = 256$.

B. *ABCD* Matrix Calculation of the Propagation Parameters

The most convenient way to determine the parameters of the digital propagation of wave fronts through the optical system is the *ABCD* matrix approach.[10] A paraxial description of any optical system that has no obscurations or vignetting and that may contain multiple elements including lenses, mirrors, and spacings is given by a 2×2 matrix of four values, A, B, C, and D, where

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}. \tag{4.5}$$

Each element of the system is also represented by a 2×2 matrix, and a system described by elements M_1, $M_2, \ldots, M_n$ is given by the matrix product

$$M = M_n \cdots M_2, M_1. \tag{4.6}$$

The propagation of a wave front through the system is then given by Eq. (3.1). Note that only A, B, and D (which together define C) are needed. One method to determine the *ABCD* values for a system would be to compute the matrix product given above, given the parameters of the system that determine the matrix for each element and spacing.

A second approach to determining the *ABCD* values is to use the results of a ray-tracing computer program. A ray at the input plane of the system with height (the distance from the optical axis) y_k and slope v_k is transformed by the system to have height y_k' and slope v_k' at an output plane:

$$\begin{bmatrix} y_k' \\ v_k' \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} y_k \\ v_k \end{bmatrix} \tag{4.7}$$

Table 11. Scale Factors and Aperture Sizes (in Pixels) for Various HST Cameras and Wavelengths, for N = 256[a]

Camera	λ (nm)	N	Scale (m/pixel)	s (pixel/m)	D_O (pixel)
PC	547	256	0.01025	97.58	234.2
PC	631	256	0.01182	84.59	203.0
PC	889	256	0.01666	60.04	144.1
FOC f/48	488	256	0.008937	111.9	268[b]
FOC f/96	488	256	0.01787	55.97	134.3
FOC f/288	488	256	0.05623	17.78	42.7

[a]D_O and s are both doubled for $N = 512$.
[b]Aliased.

or

$$y_k' = Ay_k + Bv_k, \tag{4.8a}$$

$$v_k' = Cy_k + Dv_k. \tag{4.8b}$$

If two different rays for $k = 1$ and 2 are traced through the system, we obtain four such equations in the four unknowns, A, B, C, and D, which have the following solution:

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} y_1' & y_2' \\ v_1' & v_2' \end{bmatrix} \begin{bmatrix} y_1 & y_2 \\ v_1 & v_2 \end{bmatrix}^{-1} \tag{4.9}$$

or more explicitly

$$A = \frac{y_1'v_2 - y_2'v_1}{y_1v_2 - y_2v_1}, \tag{4.10}$$

$$B = \frac{y_1 y_2' - y_2 y_1'}{y_1v_2 - y_2v_1}, \tag{4.11}$$

$$C = \frac{v_1'v_2 - v_2'v_1}{y_1v_2 - y_2v_1}, \tag{4.12}$$

$$D = \frac{y_1v_2' - y_2v_1'}{y_1v_2 - y_2v_1}. \tag{4.13}$$

Values of the ray heights and the slopes for two rays traced by Brewer[14] with a paraxial ray trace with the PAGOS program are given in Table 12. In Ref. 14, y_1, v_1, y_2, and v_2 are denoted as β, b, α, and a. Plane x_2 was taken to be Brewer's surface 18, which is in the vicinity of the image plane at the pyramid, at the apex of the pyramid, after interaction with the optical power of the pyramid. The *ABCD* values that we computed from these rays by using the equations above are shown in Table 13. Included are the values for the case of the optical system including the imaginary pair of lenses immediately before and after plane x_3. These propagations are indicated by $x_2 \to x_3'$ and $x_3' \to x_4$. Note that the coefficient D for $x_2 \to x_3'$ and A for $x_3' \to x_4$, which are proportional to the quadratic phase factors in the propagations, are greatly reduced. The last row is for the entire HST that would be used for a single-FFT propagation through the system. The negative sign of B in this latter case is indicative of a rotation of coordinates in plane x_1, which is necessary for the single-FFT propagation relative to the multiple-plane propagation.

The key parameters desired from the system analysis for use in our 3-FFT model of the HST OTA + PC are (1) the spatial scale factors relating the four planes of interest and (2) the total quadratic phase factor in the plane x_2.

Now we compare the plate scale that we determined by using the *ABCD* approach with that given by differential ray tracing. When the reported pixel size of 15.24 μm (reported earlier as 15.0 μm) at the PC CCD is used, according to this prescription, the

Table 12. Ray Heights and Slopes from Paraxial Ray Tracing[a]

Surface Number, Name	y_1 (mm)	v_1	y_2 (mm)	v_2
(1) Entrance pupil x_1	1200.0	0.0	0.0	0.00083333
(18) Pyramid x_2	−0.2647205	−0.02080727	47.74375	−0.02486843
(22) PC obscuration x_3	−18.89312	−0.02080727	25.47945	−0.02486843
(32) CCD x_4	0.3202301	0.02168971	−59.35954	−0.8977646

[a]Ref. 14.

plate scale at the CCD is given by

$$\frac{15.24\ \mu\text{m/pixel}}{71{,}231.5\ \text{mm}} = 0.2156\ \mu\text{rad/pixel} \times \frac{1\ \text{arcsec}}{4.848\ \mu\text{rad}}$$

$$= 0.04413\ \text{arcsec/pixel}.$$

Brewer[14] gives differential ray-tracing results for the plate scale as a function of the position on the CCD. Figure 6 shows our least-squares fit of his data points to a quadratic surface:

$$\begin{aligned}\text{plate scale (arcsec)} &\\ = 0.044193 &- 0.000016567x - 0.000018200x^2\\ &+ 0.00001690y + 0.000006469xy\\ &- 0.000018315y^2, \end{aligned} \tag{4.14}$$

where x and y are millimeters from the center of the CCD chip. This quadratic fit has a maximum of 0.04420 arcsec at $(x, y) = (-0.385, 0.393\ \text{mm})$. The value of 0.04413 arcsec/pixel differs from the value of 0.04419 arcsec/pixel obtained at the center of the CCD by Brewer's differential ray-tracing method by only 0.14%, which would have a negligible effect on propagation computations. (These compare with the reported original design value of 0.0430 arcsec/pixel, a difference that is quite significant.) The value of 0.04419 arcsec/pixel should be taken to be more accurate since it was obtained by nonparaxial ray tracing. Furthermore Brewer's plate scales are given as a function of the position on the CCD chip, so that they are preferred for the parameters of the propagation integrals, especially for the PSF's measured away from the center of the CCD.

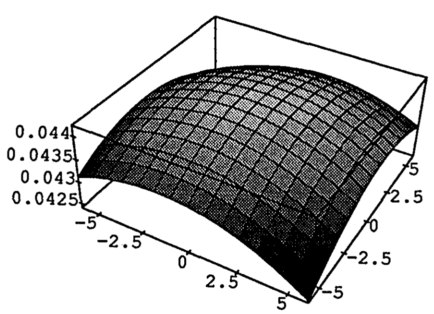


Fig. 6. Quadratic curve fit through the plate scale versus the CCD position. The plate scale is in arc seconds per pixel, and the CCD position is in millimeters. (The data points are from Ref. 14.)

For the PSF's near the center of the CCD, starting with $\Delta x_4 = 0.01524$ mm and working backward, using Eq. (3.3) and the values of B in Table 13, we obtain the spatial scales shown in Table 14, which are evaluated for the particular case of an FFT of length $N = 256$ and $\lambda = 889$ nm. The sample spacing $\Delta x_1 = 16.23$ mm in the input plane is the same whether one or three FFT's is used for the propagation. However, this equality depends on the choice of the locations of the intermediate planes. For example, if plane x_2 is taken to be Brewer's surface 20, which is just 20 mm beyond the apex of the pyramid (closer to the point where the chief ray intersects the pyramid), the multiple-plane propagation requires $\Delta x_1 = 16.04$ mm, a difference of 1.16%, which is significant. These parameters were calculated after the retrieval experiments reported in Section 5 were completed. Fortunately we obtained some of our phase-retrieval results while using parameters that were close to these parameters.

The other important parameter is the total quadratic phase factor in the plane x_2, which is given by Eq. (3.7). However, the optical power of the pyramid is already included in the matrix coefficient D for the propagation $x_1 \rightarrow x_2$, and so we have

$$BQ = \frac{D_{12}\Delta x_2{}^2}{2\lambda B_{12}} + \frac{A_{23}\Delta x_2{}^2}{2\lambda B_{23}}, \tag{4.15}$$

which, when the values in Tables 13 and 14 are used, is 5.04×10^{-5} waves/pixel^2. However, the negative of this number is appropriate because of the sign convention issue discussed next.

C. Sign Conventions for the Phase and Pupil Coordinates

For a period of our effort there was a discrepancy between the sign of the optimum value of the quadratic phase factor BQ used in the multiple-plane diffraction algorithm and that predicted by an analysis of the HST optical system design. We determined that this discrepancy was due to inconsistent sign conventions when the phase of a wave front was described. Two different sign conventions are commonly used for the phase of a wave front. The phase of an expanding spherical wave front is given by

$$\phi(x, y, z) = \frac{2\pi}{\lambda}(x^2 + y^2 + z^2)^{1/2} \tag{4.16}$$

in Goodman[16] and in Born and Wolf[17] but is given by

Table 13. ABCD Matrix Values[a]

Propagation	A	B (mm)	C (mm^{-1})	D
$x_1 \rightarrow x_2$	−0.0002206	57,292.5	-1.73394×10^{-5}	−29.8421
$x_2 \rightarrow x_3$	1.0	895.283	0.0	1.0
$x_2 \rightarrow x_3'$	1.0	895.283	−0.00111696	0.0
$x_3 \rightarrow x_4$	−1.24307	1,113.33	−0.0192194	16.4089
$x_3' \rightarrow x_4$	−0.000474674	1,113.33	−0.000891212	16.4089
$x_1 \rightarrow x_4$	0.000266858	−71,231.5	1.80748×10^{-5}	−1077.32

[a]Propagations involving x_3' include the imaginary pair of lenses before and after the PC obscurations. The pyramid is included in propagation $x_1 \rightarrow x_2$.

the negative of this expression in Siegman.[10] For our analysis we adopted the former positive sign convention. With this convention the phase of a wave front in a given plane is greater at points where the wave front has gone through a greater optical path. Then the transforms are those given in Section 3. We refer to this sign convention as the positive convention and Siegman's as the negative convention.

Our computer simulations use the positive sign convention described above. Other groups characterizing the HST use the negative convention, for which the spherical aberration has a negative sign. To obtain a spherical aberration with a negative sign, we rotated the pupil function 180° with respect to the measured PSF. Therefore we in effect replaced the aberrated pupil function $f(x)$ by $f^*(-x)$. If the Fourier transform of $f(x)$ is $F(u)$, the Fourier transform of $f^*(-x)$ is $F^*(u)$. Since the measured data are $|F(u)|^2 = |F^*(u)|^2$, we can arrive at a consistent solution with a negative spherical aberration despite the use of the positive sign convention. This same effect makes it necessary to negate the value of the quadratic phase factor, BQ, which is discussed above, in the computer software compared with the theory.

To interpret correctly the coefficients other than a_4, a_{11}, and a_{22} in the results given below, it is necessary to establish the orientation of the coordinates. The convention is as follows: the Zernike polynomials we used for Planetary Camera 6 (PC-6) are defined with the x coordinate in the $-v_2$ direction and the y coordinate in the $+v_3$ direction, where v_2 and v_3 are directions defined on the HST. This was determined by comparing our reconstructed pupil functions with the designed pupil functions.[18] In our software x is horizontal, left to right, and y is vertical, top to bottom. Therefore to compare these results with those reported by others using a different convention, the appropriate translation of coefficients is necessary. For multiple-plane propagation, which requires three FFT's to propagate to the image plane, we rotate the entrance pupil by 180° relative to its orientation for the single-FFT propagation case. (Thus the obscurations of the OTA are rotated; however, the obscurations of the PC have the same orientation for both cases.) This is necessary since the first two FFT's serve to rotate the entrance pupil by 180°. However, in our code we do not change the Zernike coefficients to accommodate this effect. Therefore when reporting our Zernike fitting results from multiple-plane propagation, we negate the coefficients of polynomials 2, 3, 7–10, and 16–21 to retain the convention that $x = -v_2$ and $y = +v_3$.

Table 14. Example Evaluation of Spatial Scale Factors for λ = 889 nm and N = 256

Plane	In terms of Δx_{prev}	In terms of Δx_4	Δx_k (mm)
CCD x_4			0.01524
PC obscuration x_3	$\dfrac{\lambda B_{34}}{N \Delta x_4}$	$\dfrac{\lambda B_{34}}{N \Delta x_4}$	0.2537
Pyramid x_2	$\dfrac{\lambda B_{23}}{N \Delta x_3}$	$\dfrac{B_{23} \Delta x_4}{B_{34}}$	0.01226
OTA pupil x_1	$\dfrac{\lambda B_{12}}{N \Delta x_2}$	$\dfrac{\lambda B_{12} B_{34}}{N B_{23} \Delta x_4}$	16.23
OTA pupil x_1 (single-FFT case)		$\dfrac{\lambda B_{14}}{N \Delta x_4}$	16.23

5. Results with HST Data

Considerable effort went into processing data from the HST over a period of several months, during which time the quality of both the data and the phase-retrieval algorithms improved. Since the earlier results are not judged to be as accurate as the later results, we report here only some of the later results from the Hubble Aberration Recovery Program (HARP) 1A and 1B collections. In this section we give the results of phase retrieval, of optimizing over some unknown system parameters, and of pupil reconstruction.

We used the retrieval algorithms described in Ref. 13. Given a description of the pupil function we retrieved the aberrations by using first a gradient search algorithm to estimate a smooth polynomial approximation to the phase error and then an iterative propagation algorithm to retrieve a detailed point-by-point phase map.

However, the optical axis of the WF/PC was not aligned as intended with the OTA, and this resulted in a combined pupil function that was significantly different from the design. Had there been no aberrations, this misalignment would have had little effect on the system performance; but it did have a substantial effect on the attempt to characterize the telescope. Given an estimate of the aberrations, we estimated the pupil function by using an iterative propagation algorithm.[13] This gave the first indication that the optical axes were not aligned as designed. This fact was later confirmed by studies of the variation of the shape of the PSF across the field of view of the WF/PC. The point in the field of view at which the PSF was most symmetric was far from the center of the CCD, where it should have been.

Because the pupil functions are needed to retrieve the phase, and the phase is needed to reconstruct the pupil functions, we estimated both by a bootstrapping procedure. In a first step, with an initial estimate of the pupil function, we retrieved a polynomial phase estimate. Appendix A gives the modified Zernike polynomials that we used. In a second step with this estimate of the phase we reconstructed the position of the PC obscurations to estimate its pupil function. We then repreated these first two steps until no further improvement was made. Usually two sets of these first two steps were necessary. With this improved estimate of the pupil function, we could then use the iterative propagation algorithm to estimate a point-by-point phase map by using the polynomial approximation to the aberrations as the initial estimate. The result of most concern was the coefficient of the 11th Zernike polynomial, which is the spherical aberration that would be corrected in the cameras that would replace the existing cameras in the HST. Initially the goal was to know a_{11} for the HST to be accurate to within 0.01 μm rms of the wave-front error. The relationship between a_{11}, the coefficient of spherical aberration, and the conic constant κ on the primary mirror of the OTA that would produce that aberration is

$$\kappa = -1.0023 + 0.043841 a_{11}. \tag{5.1}$$

A difference in a_{11} of 0.01 μm is equivalent to a difference in κ of 0.00044.

A. Aberration Estimation

The most effort was put into the image designated HARP1A PC-6 F889N_P2, which was taken with PC-6 through the narrow-band filter with a center wavelength of 889 nm. It was the most useful of the HARP1A images. It is well out of focus (focus parameter despace = −260 μm), which is preferred because, without a bright central spike, most of the pixels of the PSF can have a large number of photons without saturating the CCD, i.e., they have a relatively large signal-to-noise ratio. Many other images were available, but, being closer to focus, they were much less suitable for phase retrieval. Our approach was to develop our algorithms while concentrating on a small number of the best images, in which results we could have high confidence, rather than devoting a large amount of effort to a larger number of images of poor quality in which little confidence could be placed.

For these data, by using the single-plane diffraction algorithm, we arrived at a value of $a_{11} = -0.28$ μm rms earlier in the effort. Then later, by using the multiple-plane diffraction algorithm, we obtained $a_{11} = -0.295$ μm, a change of −0.015 μm. Still later, not yet knowing the most accurate estimate of the system parameters, we optimized the error metric over the poorly known system parameters and found the optimized parameters to be the plate scale = 0.0442 arcsec/pixel and the quadratic phase factor $BQ = -0.000054$. Using these optimized parameters, we arrived at $a_{11} = -0.299$ μm rms. Later still, we found from Brewer's ray-tracing results that these optimized parameters were close to the true parameters. Therefore our best estimate of a_{11} from the HARP1A imagery is $a_{11} = -0.299$ μm rms. These results are summarized in Table 15. The first column under P2 is for a single-plane propagation algorithm (only the largest values for a_{12}–a_{21} are shown). Two values for the coefficients are given for the multiple-plane propagation case: the first is for fitting a_2–a_{11} and a_{22}, and the second is for fitting a_2–a_{22}. (Only the value for a_{16} is listed among a_{12}–a_{21} since the other values were negligibly small.) Note that for this real data the magnitude of a_{11}

Table 15. Zernike Coefficients (Micrometer rms Wave-Front Error) for HARP1A images PC-6 F889N_P2 and PC-6 F889N_Q2

j	P2 Single	P2 Multi (11)	P2 Multi (22)	New Parameter P2 Multi (11)	Q2 Single
4	−2.212	−2.227	−2.223	−2.306	0.73
5	−0.018	−0.003	0.006	−0.003	0.06
6	−0.025	0.025	0.026	0.031	−0.02
7	0.004	0.001	0.005	−0.001	0.01
8	0.017	0.010	0.009	0.013	0.06
9	−0.022	−0.020	−0.009	−0.020	−0.00
10	0.002	0.008	0.010	0.005	0.01
11	−0.280	−0.292	−0.295	−0.299	−0.281
12	0.008	(n/a)		(n/a)	0.01
16	−0.009	(n/a)	−0.004	(n/a)	0.01
20	0.006	(n/a)		(n/a)	0.00
22	0.005	0.006	0.007	0.008	0.04
Conic κ =	−1.0146	−1.0151	−1.0152	−1.0154	−1.0146
Root-mean-squared error =	0.1583	0.1352	0.1353	0.1428	0.2508

increased when multiple-plane propagation was used, whereas for the simulated data described in Section 3 the magnitude of a_{11} decreased with the multiple-plane propagation algorithm. The values of a_5–a_{10} changed substantially with the multiple-plane propagation algorithm but stayed below 0.03 μm rms in magnitude; a_{22} varied with the reconstruction, staying below 0.01 μm rms (which, however, could be significant). The column labeled New Parameter shows the values obtained when a larger value of the plate scale (0.0442 arcsec/pixel) and a different value of the quadratic phase coefficient $BQ(-0.000054)$ were used. These values represent what is now thought to be a more accurate system model. Again, all the values of the Zernike coefficients change somewhat. Notably a_{11} increased in magnitude to -0.299, corresponding to the conic constant on the primar mirror of the OTA of -1.0154. The error metric is somewhat larger than that obtained for the old parameters since the pupil shift used in both cases was that optimized for the old parameters. The values other than a_4 and a_{11} are considered to be unreliable at this point, since they change so much depending on the details of the retrieval algorithm for a given data set. These differences show the importance of modeling the system as accurately as possible.

Also shown in the last column of Table 15 is an example of the results for another image, PC-6 F889N_Q2, which was closer to focus and less reliable (as seen from the value of 0.2508 for the rms error of the fit). Since the results for PC-6 F889N_P2 produce a fit that is so much better than the other images, we tend to ignore the results from the other HARP1A images.

There seems to be a trend toward larger magnitudes of a_{11} as the accuracy of the modeling increases. This being the case we suspect that the average value of a_{11} reported by all the groups working on phase-retrieval underestimates the magnitude of the true value.

Point-by-point phase maps have also been reconstructed with the iterative propagation algorithm, which shows the fine structure (sometimes referred to as zones) in the mirror surfaces. However, the reliability of these detailed phase maps has not yet been established and are therefore not included here.

Figure 7 compares a PSF computed from a model of

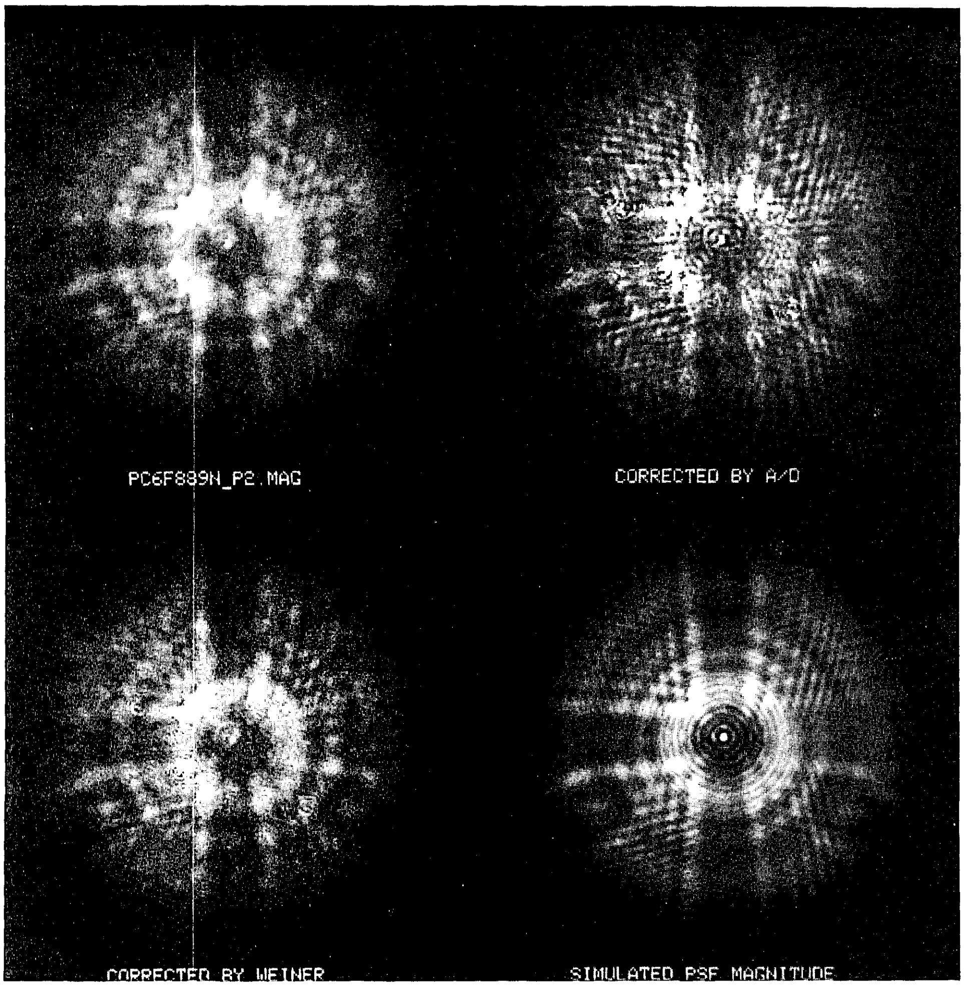


Fig. 7. Measured image PC-6F889N_P2 from HARP1A and the images computed from it. Measured PSF (upper left), the PSF deconvolved with the Ayers/Dainty algorithm (upper right), the PSF deconvolved with the Wiener filter by using jitter data (lower left), and the PSF computed from the model by using a polynomial phase estimate (lower right).

the aberrated system with the measured PSF. The computed PSF is for the case of single-plane propagation and when only the polynomial approximation of the phase error is used. As is discussed in Subsection 5.D, one of the largest factors that causes the computed PSF to differ from the measured PSF is the jitter in the pointing of the telescope, which smears out the measured PSF.

B. Optimizing over System Parameters

Before knowing the correct system parameters (which were later supplied by ray tracing performed by Brewer as described in Section 4), we performed a search over the plate scale and quadratic phase factor that minimized the error metric. Since these two variables were not included in our optimization software, we optimized them in a brute force fashion by selecting many different combinations of their values, by optimizing for the Zernike coefficients for each combination, and by selecting the combination that yielded the smallest error metric. The optimization would have to be redone for any other PSF in a different location in the field of view or in a different wavelength band. (Automatic optimization over these parameters would be preferable.)

The first three columns of values in Table 16 show the effect of the plate scale on the Zernike coefficients obtained from image PC-6 F889N_P2 over a wide range of plate scales. As the plate scale increases (more arcseconds per pixel in the image), the size of the features in the modeled entrance pupil (pixels per meter) increases, and the retrieved values of a_{11} tend to increase in magnitude. The value of the quadratic phase factor BQ that was used in this case was -0.000068, which was found to be optimum by trial and error for a plate scale of 43 arcsec/per pixel. Later ray-tracing results, described in Section 4, predicted a value of BQ of -0.00005, for which the results in the fourth column of values were obtained for the larger plate scale. Note that a decrease in the absolute value of BQ decreased a_{11}.

Table 16. Effect of the Plate Scale on Zernike Coefficients for Image PC-6 F889N_P2 (Micrometer rms Wave-Front Error)[a]

	s (pixels/m)				
	With $BQ = -0.000068$			With $BQ = -0.00005$	With $BQ = -0.000054$
	58.00	60.04	62.00	62.00	61.74
Plate scale (arcsec/pixel)	41.54	43.00	44.40	44.40	44.20
$j = 4$	−2.213	−2.227	−2.311	−2.320	−2.306
5	−0.028	−0.003	−0.001	−0.000	−0.003
6	−0.025	0.025	0.027	0.026	0.031
7	0.004	0.001	0.001	−0.001	−0.001
8	0.021	0.010	0.014	0.014	0.013
9	−0.017	−0.020	−0.023	−0.022	−0.020
10	0.005	0.008	0.002	0.003	0.005
11	−0.291	−0.292	−0.302	−0.299	−0.299
22	−0.001	0.006	0.007	0.008	0.008
Conic κ =	−1.0151	−1.0151	−1.0155	−1.0154	−1.0154
Root-mean-squared err =	0.1545	0.1442	0.1413	0.1415	0.1428

[a]Using multiple-plane propagation, fit coefficients 1–11,22, $N = 256$, PSF weighted by diameter-220 circle, PC obscuration shifted by −4.25, −3.25 pixels.

Figure 8 shows an example of optimizing the plate scale (by the scale parameter s, which represents the number of pixels per meter of the OTA entrance pupil and is proportional to the plate scale). An entire minimization was performed, each with the same parameters except s, which was varied. The points plotted in Fig. 8 represent the final rms error at the end of each minimization. In this case the quadratic phase factor for multiple-plane propagation was $BQ = -0.000054$, which is close to the value predicted by the updated ray-trace model. The error metric has two minima, one at $s = 61.74$, where it has a value of 0.14278 and for which $a_{11} = -0.299$ µm rms, and a second at $s = 62.64$, where it has a value of 0.14276 and for which $a_{11} = -0.301$ µm rms. We do not understand why there would be a double-bottom minimum. When performing many optimizations with different values of BQ, with each of these two values of s, we found that the optimum value of BQ is -0.000059 for $s = 61.74$ and is -0.000075 for $s = 62.64$. The relationship between the plate scale and s is given by $s = \rho_s(N/\lambda)$, where ρ_s is the plate scale in radians (the number of radians separating the pixels). For $\rho_s = 0.0442$ arcsec $= 0.214$ µrad, $N = 256$ pixels, and $\lambda = 0.889$ µm: $s = 61.7$ pixels/m. This value, calculated from the now-known plate scale for the center of the CCD, is in quite good agreement with the first of the two minima that we found by minimizing the fitting error as a function of s. Furthermore we also found later from the optical system design that an appropriate value of BQ is -0.000051 (see Section 4). Therefore both the values of the plate scale and BQ predicted by ray tracing point to the first

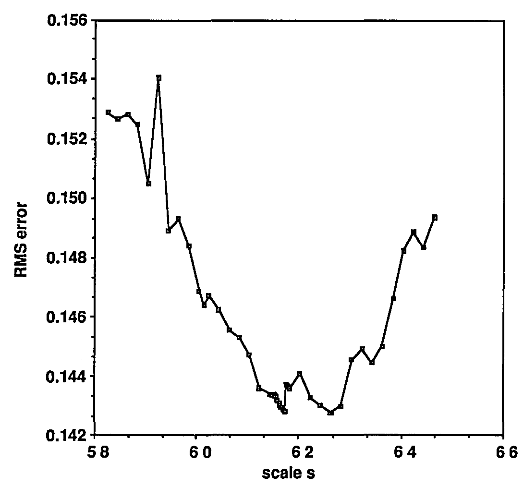


Fig. 8. Optimization of the fitting error as a function of the plate scale. The spatial scale parameter s is proportional to the plate scale.

minimum as being the true value. The results in the last column of Table 16 represent our estimate when we used parameters that are close to the best system parameters that we currently know. In this case $a_{11} = -0.299$ μm rms, which is equivalent to a conic constant of -1.0154.

C. Pupil Reconstruction

Figure 9 shows an example of pupil reconstruction by one iteration of the iterative transform algorithm[13] (ITA) for HARP1B image 33434. In this case the modeled phase error was placed across only the annular aperture of the OTA without any spiders, pads, or PC obscurations. The resulting image shows a clear indication of the location of all the obscurations derived from the measured data: the pads and spiders of the OTA and the spiders and central obscuration of the PC. Based on this picture, we then constructed the modeled obscurations (illustrated together in a single plane) shown in Fig. 10.

We also determined the shift of the PC obscurations by trying different shifts and picking the one that minimized the error metric. When this was done for the PC-6F889N_P2 image, the optimum shift was found to be -4.5 rows and -3.25 columns in pupil space when the array size was $N = 256$. (Fractional-pixel shifts of an array can be computed by Fourier transforming the array, multiplying by an appropriate linear-phase complex exponential, and inverse Fourier transforming). One pixel in this case corresponds to 0.0167 m projected to the primary mirror of the OTA. Therefore the corresponding shift in that plane would be -0.075 m along $-v_2$ and -0.054 m along $+v_3$. This was thought to be a large shift when we consider how close to the center of the chip the image was located (at CCD pixel 531, 425). This is an indication that the optics of PC-6 are not

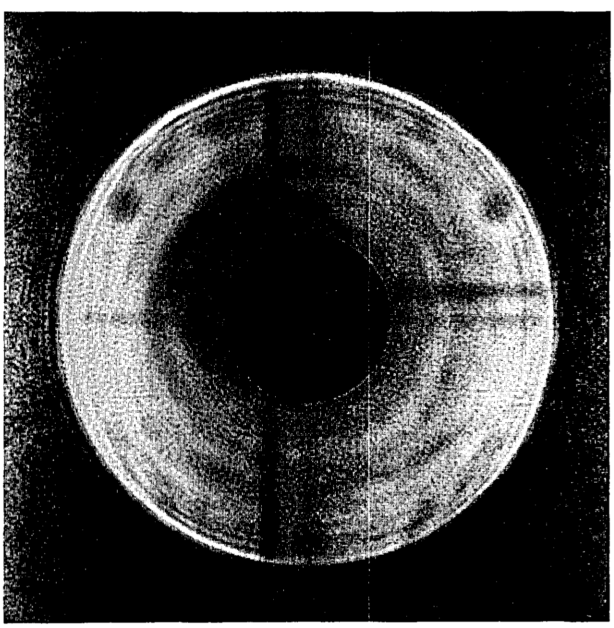


Fig. 9. Pupil function reconstructed by one iteration of the iterative transform algorithm.

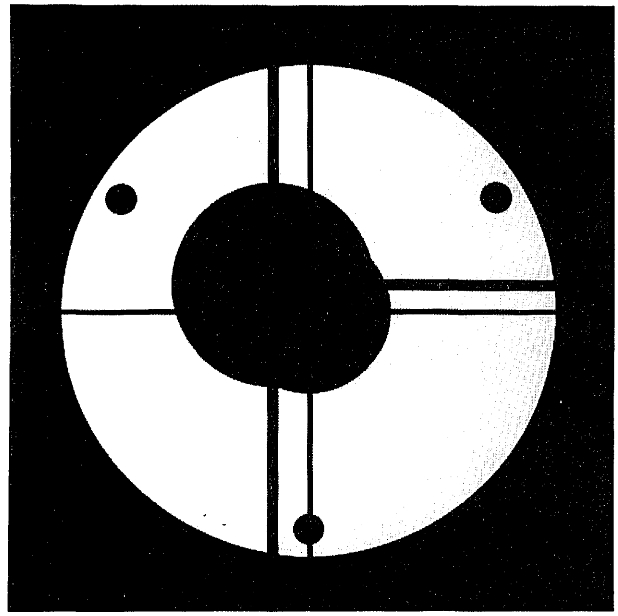


Fig. 10. Model of the pupil function inferred from the reconstructed pupil shown in Fig. 9.

aligned properly with the OTA, causing a pupil shift. This was explored in more depth as described below.

The center of symmetry, i.e., the CCD pixel for which the OTA and WF/PC central obscurations are aligned with one another, was designed to be at pixel 400, 400. The method we used to determine the actual center of symmetry (and thereby infer the alignment of the WF/PC relative to the OTA) is as follows. For each of several images an estimate of the aberrations is put over the 0.33-obscured doughnut-shaped aperture of the OTA. Then one iteration of the ITA is performed, the output of which is darker where there are obscurations in the pupil function. For the case of good-quality narrow-band data, the pads, spiders, and WF/PC central obscuration can be seen clearly, as shown in Fig. 9. We measured the positions of the reconstructed PC spiders to estimate the shift of the PC obscurations relative to the OTA obscurations. This is done for the images taken at different locations in the field of view. Then we perform a least-squares fit of a linear model of the WF/PC obscuration shift to the set of estimated WF/PC obscurations. The least-squares fit is performed for the collection of several images. This yields one equation for the row of the WF/PC shift as a function of (x, y) and another for the column of the WF/PC shift as a function of (x, y). Lastly, these two simultaneous equations are solved for the value of (x, y) for which the row and column WF/PC shift are both zero. This defines the center of symmetry for any given WF/PC channel.

We performed the procedure described above for a collection of seven images from the HARP1B series on PC-6 at a 631-nm wavelength at despace = -90 μm. The result was that the center of symmetry was estimated to be $x = 254.5$, $y = 194.0$ pixels. When we recalculated the center of symmetry leaving out a couple of the images, it changed by ~ 20 pixels. The aberrations for each of the images were not optimized individually, and so with greater care the same procedure could yield an answer with greater accuracy and confidence. Further improvements in the estimates of the PC obscuration shifts could be obtained with the gradient search algorithm described in Ref. 13 by using the derivative with respect to the point-by-point magnitude. Another improvement would be to optimize the shift of a model of the PC obscurations during the phase-retrieval algorithm. Still another improvement would be first to deconvolve the jitter from the images before estimating the shift. Until these improvements are made, the center of symmetry results reported by Roddier and Roddier,[19] who used our approach but performed it more carefully, should be taken to be more accurate than the center of symmetry reported above.

D. Effects of Jitter

We believe that jitter in the data from the HST is one of the largest components of the lack of agreement between our modeled PSF's and measured PSF's. Lacking jitter-free, long-wavelength data we could account for the jitter in the phase-retrieval algorithm, as described below.

The effect of jitter can be seen in Fig. 7, which compares a measured PSF (upper left) with a PSF computed from our polynomial-phase model of the aberrated wave front (lower right). The modeled PSF has a fine-fringe structure that is smeared out in the measured PSF. The modeled PSF does not include the effects of the finite spectral bandwidth, pixelation, or jitter. The spectral bandwidth of the F889 filter is 0.57% (5.1 nm/889 nm according to the *WF/PC Instrument Handbook*) and would not yield this degree of blurring. Furthermore the finite spectral bandwidth causes a blurring that is proportional to the distance from the center of the PSF, which would be negligible near the center of the PSF; consequently the blurring of this PSF that can be seen at its center cannot be a result of the spectral bandwidth. We simulated the effect of pixelization (integration over the area of the CCD pixel) in the modeled PSF and found that it yielded a degree of blurring that is much milder than that seen in the measured PSF; therefore pixelization also cannot explain the loss of the fine-detailed structure in the measured PSF. The jitter, which is known to be of the order of 0.1 arcsec in effective width, or ~ 2 pixels wide, does explain the differences. We model the effect of jitter to be the convolution of the PSF with a jitter spatial density function. The jitter causes a much poorer fit of the model to the measured data than would be the case if there were no jitter (14% rms error being the best fit to date). It can be seen from studies by Lyon *et al.*[20] that jitter could cause errors in a_{11} in the range of 0.01 μm rms (which is the entire error budget). At the very least the jitter causes the level of uncertainty in our results to be much higher than if there were no jitter.

Jitter can be accounted for by two routes. The first is to correct the jitter in the measured data and then operate on that jitter-corrected data with the same phase-retrieval algorithms as before. The second route is to incorporate a model for the jitter directly within the phase-retrieval algorithm.

Any number of deconvolution algorithms can be used to remove the jitter from the measured PSF if the jitter density function is known. The fine-guidance sensors provide a measure of the jitter from which the jitter density function can be computed, but the reliability of that data had not been established. Alternatively we can use a blind deconvolution algorithm to estimate both the PSF and the jitter density function. Figure 7 shows the original measured PSF (upper left), PC-6 F889N_P2, and the PSF deconvolved both by the Ayers/Dainty algorithm[21,22] (upper right) and by the Wiener filtering using the measured jitter data (lower left). The PSF's deconvolved by using the Ayers/Dainty algorithm had significantly greater contrast than those deconvolved by using the measured jitter data. That is, the jitter predicted by the Ayers/Dainty algorithm was significantly greater than that given by the measured jitter data. This is possible since the measured jitter data are undersampled in time relative to some of the high temporal frequencies of the jitter. The two approaches were also combined: the jitter data were used as a starting estimate for the jitter in the Ayers/Dainty algorithm, and then we performed the Lucy algorithm[23] by using the jitter function computed by the Ayers/Dainty algorithm. We obtained several different deconvolutions for a single PSF by using these various techniques, all of which are in good agreement with the measured PSF. This may indicate a lack of uniqueness in deconvolving an unknown or partially unknown jitter function from a PSF for the amount of noise and undersampling present in the data. At this point it is difficult to judge which is the most believable deconvolution result. The additional details in the deconvolved PSF's look quite believable. Some of the fine fringes seen in the simulated PSF (shown in the lower right of Fig. 7), which were washed out in the measured PSF (upper left), became visible in the jitter-deconvolved PSF's (upper right and lower left). The artifacts from dust (presumably on the field flattener in front of the CCD) became much sharper and more visible, as can be seen in Fig. 7.

Phase retrieval with Zernike coefficients was performed on various jitter-deconvolved versions of PC-6F889N_P2 (from Wiener, Lucy, and Ayers/Dainty deconvolutions). For all of them the a_{11} Zernike coefficient obtained for the jitter-deconvolved PSF did not change significantly from that of the original PSF. More extensive simulation studies by Lyon, on the other hand, showed significant changes in the retrieved value of a_{11} when jitter was present.[20] The error in our fit of the modeled PSF to the measured PSF was actually worse for the jitter-deconvolved PSF's than it was for the original PSF. Possibly this is because the deconvolution enhances the noise as well as the fine structure of the measured PSF. This issue deserves further study.

The second route to accounting for jitter is to include a model for the jitter directly within the phase-retrieval algorithm. To that end we derived analytic expressions for gradients of the phase-retrieval error metric with jitter included in the model of the data and for the partial derivative of the error metric with respect to a shift in the WF/PC obscurations. These are given in Appendix B. These analytic gradients were implemented in software within the phase-retrieval algorithm; however, the effort was concluded before this software could be debugged and exercised.

The circular artifacts seen in the PSF's in Fig. 17, which are more pronounced in the deconvolved PSF's, were reportedly a result of dust on a field flattener in front of the CCD. In Appendix C we verify the source of this artifact.

6. Inclusion of the Z_{22} Contribution to Spherical Aberration

It has been argued that since $Z_{22}(r)$ has an r^4 term in it, this contribution to r^4 should be added to the spherical aberration $Z_{11}(r)$ for the purpose of determining what spherical aberration to correct. This should not be done. The r^4 term in $Z_{22}(r)$ is there merely to make $Z_{22}(r)$ orthogonal to $Z_{11}(r)$ and does not represent the real spherical aberration that one would want to correct. The r^4 term in $Z_{22}(r)$ is there only to balance the r^6 term, and it should be included only if the entire $Z_{22}(r)$ function were being corrected (which is not the plan for the replacement of WF/PC). In fact $Z_{11}(r)$ and $Z_{22}(r)$ are orthogonal over a 0.33-obscured annular aperture, not over the actual aperture; nevertheless they are close to orthogonal over the actual aperture, and the coupling between them should be small.

In addition, consider the following. A typical value for the $Z_{22}(r)$ coefficient a_{22} is 0.007 μm rms (see Section 5). When we convert the r^4 term in 0.007 $Z_{22}(r)$ to an equivalent coefficient, a_{11} of $Z_{11}(r)$ yields -0.052 μm rms. If this value were added to an estimate of a_{11} of -0.299 μm rms, the result would be a total estimate of a_{11} of -0.351 μm rms. This large value of spherical aberration is inconsistent with all the other analyses performed on the HST, giving further evidence that the r^4 component of $Z_{22}(r)$ should not be added to our estimate of $Z_{11}(r)$. Again that term should be included only if the entire $Z_{22}(r)$ were being corrected.

7. Conclusions

We applied several new algorithms[13] to retrieving the aberrations of the HST by using the blurred images of stars taken by the telescope on orbit. We also reconstructed pupil functions, which indicated an unexpected shift of the obscurations in PC-6 relative to those in the OTA, which indicates a misalignment of the optical axis of PC-6 relative to the OTA. Our best estimate of the spherical aberration, as specified by the coefficient of the 11th Zernike polynomial a_{11} for the combined OTA and PC-6, is -0.299 μm rms of the wave-front error, which would correspond to a conic constant on the primary mirror of -1.0154. This amount of spherical aberration is larger than some of the earlier results obtained with the difference resulting from a more accurate system model, including multiple-plane propagation, and a larger, more accurate value of the plate scale. For us to determine the aberrations of the primary mirror of the OTA, an adjustment to this prescription is necessary to compensate for the known spherical aberration in PC-6, which the Jet Propulsion Laboratory reports to be equivalent to -0.0010 in the conic constant of the OTA. With this correction the conic constant of the primary mirror of the OTA is estimated to be -1.0144.

The error bars on our estimate are difficult to determine since they depend on systematic errors, such as poorly known parameters of the system, rather than random errors whose standard deviation can be derived. Cramer–Rao lower bounds show that the far out-of-focus PSF's are far more suitable for phase retrieval than the nominally in-focus PSF's. We performed retrieval experiments to show the expected errors caused by imprecise knowledge of the plate scale, a quadratic phase factor, and translation of obscurations in the PC. The jitter also limits the accuracy of our estimate by an unknown amount. We require further analyses to determine the actual error bars. The accuracy of our results to date was not sufficient to give us high confidence in the predictions of the Zernike coefficients (astigmatism, coma, etc.) other than the spherical aberration, of the point-by-point phase maps or of the extent to which there are aberrations in the secondary mirror of the OTA. We believe that the degree of confidence can be increased, and the error bars can be decreased, by several additional refinements, including (1) a more complete accounting for the effects of jitter, (2) an improved estimation of pupil functions, and (3) an automatic optimization over poorly known system parameters.

Portions of this paper were presented in Refs. 24 and 25.

Appendix A: Modified Zernike Polynomials

The polynomials used to describe the aberrations are the modified Zernike polynomials, orthonormal over an annular aperture with an inner radius of 0.330 times the outer radius. They are given in Table 17, which was adapted from Ref. 26 with corrections. Here the two spatial coordinates, x and y, are given explicitly (whereas in the body of this paper x was taken to be a two-dimensional coordinate). The radius $r = (x^2 + y^2)^{1/2}$ is normalized to unity at the outer edge of the aperture.

Table 17. Modified Zernike Polynomials

j	Normalization Factor	Polynomial
2	1.8992573	x
3	1.8992573	y
4	3.8874443	$(r^2 - 0.554450)$
5	2.3137662	$(x^2 - y^2)$
6	2.3137662	$2xy$
7	8.3345629	$x(r^2 - 0.673796)$
8	8.3345629	$y(r^2 - 0.673796)$
9	2.6701691	$x(x^2 - 3y^2)$
10	2.6701691	$y(3x^2 - y^2)$
11	16.895979	$(r^4 - 1.108900r^2 + 0.241243)$
12	12.033645	$(x^2 - y^2)(r^2 - 0.750864)$
13	12.033645	$2xy(r^2 - 0.750864)$
14	2.9851527	$(r^4 - 8x^2y^2)$
15	2.9851527	$4xy(x^2 - y^2)$
16	36.321412	$x(r^4 - 1.230566r^2 + 0.323221)$
17	36.321412	$y(r^4 - 1.230566r^2 + 0.323221)$
18	16.372202	$x(x^2 - 3y^2)(r^2 - 0.800100)$
19	16.372202	$y(3x^2 - y^2)(r^2 - 0.800100)$
20	3.2700486	$x(x^4 - 10x^2y^2 + 5y^4)$
21	3.2700486	$y(5x^4 - 10x^2y^2 + y^4)$
22	74.782446	$(r^6 - 1.663350r^4 + 0.803136r^2 - 0.104406)$

Appendix B: Phase-Retrieval Gradient with Jitter and Pupil Shifts

Expanding on the earlier multiple-plane analysis in Ref. 13, we employ the propagation model

$$G(u) = P[U_1(x_1)], \tag{B1}$$

where $P[U_1(x_1)]$ is a general propagation through a complicated optical system with the multiple planes of diffraction (obscurations), and we replace the error metric

$$E = \sum_u W(u)[|G(u)| - |F(u)|]^2 \tag{B2}$$

with one that includes a jitter density function $J(u)$:

$$E = \sum_u W(u)[|G(u)|_J - |F(u)|]^2, \tag{B3}$$

where

$$|G(u)|_J^2 = |G(u)|^2 * J(u), \tag{B4}$$

$*$ denotes convolution, and $|F(u)|^2$ is a measured PSF degraded by jitter. For our multiple-plane propagation model of the HST, $u = x_4$.

We form the partial derivative of E with respect to a parameter p in the input wave front:

$$\frac{\partial E}{\partial p} = \sum_u W(u)\left[1 - \frac{|F(u)|}{|G(u)|_J}\right]\frac{\partial |G(u)|_J^2}{\partial p}, \tag{B5}$$

where

$$\frac{\partial |G(u)|_J^2}{\partial p} = \sum_{u'} J(u - u')\left[G^*(u')\frac{\partial G(u')}{\partial p} + \text{c.c.}\right], \tag{B6}$$

and where c.c. denotes the complex conjugate of what precedes it. Let

$$G^J(u') = G(u')\sum_u J(u - u')W(u)\left[\frac{|F(u)|}{|G(u)|_J} - 1\right]. \tag{B7}$$

When Eqs. (B6) and (B7) are inserted into Eq. (B5), we have

$$\frac{\partial E}{\partial p} = -\text{Re}\left[\sum_{x_1}\frac{\partial U_1(x_1)}{\partial p}g^{J*}(x_1)\right], \tag{B8}$$

where

$$g^J(x_1) = P^\dagger[G^J(u)] \tag{B9}$$

and $P^\dagger$ is the inverse propagation operator.[13] For a decomposition of the phase into Zernike polynomials with coefficients a_j, we have

$$\frac{\partial E}{\partial a_j} = 2\ \text{Im}\left[\sum_{x_1} U_1(x_1)Z_j(x_1)g^{J*}(x_1)\right], \tag{B10}$$

and for a point-by-point phase map $\theta(x_1)$ we have

$$\frac{\partial E}{\partial \theta(x_1)} = 2\ \text{Im}[U_1(x_1)g^{J*}(x_1)] \tag{B11}$$

and similarly for the other unknown parameters of $U_1(x_1)$.

Next consider an unknown parameter q of the PC obscuration transmittance $m_3(x_3)$. We assume here that there are no aberrations in plane x_3. Then we have, similar to the case in Ref. 13,

$$\begin{aligned}\frac{\partial E}{\partial q} &= -\sum_u G^{J*}(u)\frac{\partial G(u)}{\partial q} + \text{c.c.}\\ &= -\sum_{x_3}\frac{m_3(x_3)}{\partial q}U_3(x_3)\{P_{4\to3}{}^\dagger[G^J(u)]\}^* + \text{c.c.},\end{aligned} \tag{B12}$$

where $U_3(x_3)$ is the propagation of the wave front $U_1(x_1)$ to the x_3 plane, and $P_{4\to3}{}^\dagger$ is the inverse propagation from plane x_4 to x_3.

First, let the parameter q be the transmittance magnitude $m_3(x_3)$ at point x_3. Then we have

$$\frac{\partial E}{\partial m_3(x_3)} = -2\ \text{Re}(U_3(x_3)\{P_{4\to3}{}^\dagger[G^J(u)]\}^*). \tag{B13}$$

Next let the parameter q be the unknown shift x_0 of the PC mask $m_3(x_3 - x_0)$. Letting

$$m_3(x_3) = \sum_{u'} M_3(u')\exp(i2\pi u'x_3/N), \tag{B14}$$

we have

$$\begin{aligned}\frac{\partial m_3(x_3 - x_0)}{\partial x_0} &= \sum_{u'}(-i2\pi u'/N)M_3(u')\\ &\quad\times\exp[i2\pi u'(x_3 - x_0)/N],\end{aligned} \tag{B15}$$

and, about $x_0 = 0$, we have

$$\begin{aligned}\frac{\partial E}{\partial x_0} &= -\sum_{x_3}\sum_{u'}(-i2\pi u'/N)M_3(u') \\ &\quad\times \exp(i2\pi u'x_3/N)U_3(x_3)\{P_{4\to3}{}^{\dagger}[G^J(u)]\}^* + \text{c.c.} \\ &= -2\,\mathrm{Im}\left(\sum_{x_3}\mathscr{F}^{-1}[(2\pi u'/N)M_3(u')] \right.\\ &\quad\left.\times\, U_3(x_3)\{P_{4\to3}{}^{\dagger}[G^J(u)]\}^*\right). \qquad\text{(B16)}\end{aligned}$$

Note that this requires three FFT's for $G(u)$ to be computed, four more FFT's (or two convolutions) for $G^J(u)$ to be computed, and two more FFT's (or one convolution) for each of the two components of the derivative of $m_3(x_3)$ to be computed.

These analytic expressions for the derivatives of the error metric allow for an efficient gradient search algorithm that minimizes the error metric as a function of the unknown parameters, when we assume that the jitter function is known.

Appendix C: Verification of Dust Artifacts

The dust artifacts sharpened during the jitter deconvolution, as can be seen in Fig. 17. If the dust is on the field flattener and moves along with the detector array, it should not be blurred by the jitter. Therefore it would seem that, by deblurring the entire image for the jitter, we would cause the dust artifacts to become more blurred. The fact that the dust artifacts sharpened during jitter deconvolution was contrary to our expectations and caused us to question the hypothesis that the dust was located on the first surface of the field lens. However, we confirmed the hypothesis that the dust is on the field flattener, as described below. A possible explanation of why the dust looks sharper after deblurring for jitter is that the deblurring operation is a high-pass filter that enhances all the edges in the image.

The axial location of the dust causing the artifacts was determined as follows: Assuming an essentially plane wave front at the plane of the dust (which is accurate for the far out-of-focus PSF's) and assuming that the dust acts as a pointlike scatterer, the expected intensity pattern caused by the dust is

$$\begin{aligned}I(x) &= |\exp(i2\pi d/\lambda) + a\exp[i2\pi(r/\lambda + c)]|^2 \\ &= 2 + 2a\cos[2\pi(r-d)/\lambda + 2\pi c], \qquad\text{(C1)}\end{aligned}$$

where the dust is assumed to be at lateral location $x = 0$, d is the distance (which we wish to determine) from the plane of the dust to the plane of the CCD, a is the amplitude of the wave front scattered from the dust, c is an unknown phase constant associated with scattering from the dust, and $r = (d^2 + x^2)^{1/2}$ is the distance from the dust to a given point x on the CCD. The radii of successive peaks and nulls of the resultant concentric-circular fringe patterns are given by

$$x_m = \{2d\lambda[(m + n_0)/2 - c]\}^{1/2}, \qquad\text{(C2)}$$

where n_0, an integer, is a reference fringe number. We obtained this expression by setting the argument of the cosine equal to $2\pi n_0$ and using a Taylor series expansion of r in terms of d and x. For two of the dust artifacts in image PC-6F889N_P2, we measured three successive radii at a null, a peak, and a null to be 28.5, 55.4, and 75.9 μm. There could be a 20% error in these numbers, because the data are given at a spacing of 15.24 μm. Fitting the model for x_m above to these three values, we determined the three unknowns to be $n_0 = 0$, $c = 0.34$ wave, and $d = 2.7$ mm. For comparison, for diffractive propagation purposes, the effective distance from the front surface of the field flattener, at its thinnest point, to the CCD is 1.63 mm + 1.27 mm/1.378 = 2.5 mm, which is in good agreement with the value of d determined from the dust artifacts, considering the uncertainties in estimating the locations of the peaks and nulls of the fringe.

The authors gratefully acknowledge the assistance of A. M. Kowalczyk who wrote much of the software that was necessary for us to obtain the results in this paper and C. A. Bobrowski who exercised some of the software. This work was funded primarily by Jet Propulsion Laboratory under contract JPL 958892 and partially by the Environmental Research Institute of Michigan with internal research and development funds.

References

1. A. G. Tescher, ed., *Applications of Digital Image Processing XIV,* Proc. Soc. Photo-Opt. Instrum. Eng. **1567** (1991).
2. *Space Optics for Astrophysics and Earth and Planetary Remote Sensing,* Vol. 19 of 1991 Technical Digest Series (Optical Society of America, Washington, D.C., 1991).
3. J. R. Fienup, "Phase retrieval algorithms: a comparison," Appl. Opt. **21,** 2758–2769 (1982).
4. J. C. Dainty and J. R. Fienup, "Phase retrieval and image reconstruction for astronomy," in *Image Recovery: Theory and Application,* H. Stark, ed. (Academic, New York, 1987), Chap. 7, pp. 231–275.
5. R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," Optik **35,** 237–246 (1972).
6. W. O. Saxton, *Computer Techniques for Image Processing in Electron Microscopy* (Academic, New York, 1978).
7. D. L. Misell, "A method for the solution of the phase problem in electron microscopy," J. Phys. D **6,** L6–L9 (1973).
8. R. A. Gonsalves, "Phase retrieval from modulus data," J. Opt. Soc. Am. **66,** 961–964 (1976).
9. J. N. Cederquist, J. R. Fienup, C. C. Wackerman, S. R. Robinson, and D. Kryskowski, "Wave-front phase estimation from Fourier intensity measurements," J. Opt. Soc. Am. A **6,** 1020–1026 (1989).
10. A. E. Siegman, *Lasers* (University Science, Mill Valley, Calif., 1986), Chaps. 15 and 20.
11. J. N. Cederquist and C. C. Wackerman, "Phase-retrieval error: a lower bound," J. Opt. Soc. Am. A **4,** 1788–1792 (1987).
12. J. M. Mendel, *Lessons in Digital Estimation Theory* (Prentice-Hall, Englewood Cliffs, N.J., 1987), p. 51.
13. J. R. Fienup, "Phase-retrieval algorithms for a complicated optical system," Appl. Opt. **32,** 1737–1746 (1993).
14. S. Brewer, "WFPC generic prescriptions," Rep. 92-1 (Jet Propulsion Laboratory, Pasadena, Calif., March 1991).
15. R. N. Bracewell, *The Fourier Transform and Its Applications* (McGraw-Hill, New York, 1978).

16. J. W. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, New York, 1968).
17. M. Born and E. Wolf, *Progress in Optics* (Macmillan, New York, 1964).
18. L. Adams, Jet Propulsion Laboratory, Pasadena, Calif. 91109, "WF/PC relay optics spider clocking on OTA primary" (personal communication, 17 August 1990).
19. C. Roddier and F. Roddier, "Reconstruction of the Hubble Space Telescope wave-front distortion from stellar images taken at various focus positions," Final Rep., Contract 958893 (Jet Propulsion Laboratory, Pasadena, Calif., May 1991).
20. R. G. Lyon, P. E. Miller, and A. Grusczak, "Hubble Space Telescope assembly project report, phase retrieval special studies task," Final Rep. PR J14-0013, Contract NAS 8-38494 (Goddard Space Flight Center, NASA, Greenbelt, Md., 4 June 1991).
21. G. R. Ayers and J. C. Dainty, "An iterative blind deconvolution method and its applications," Opt. Lett. **13,** 547–549 (1988).
22. B. L. K. Davey, R. G. Lane, and R. H. T. Bates, "Blind deconvolution of noisy complex-valued image," Opt. Commun. **69,** 353–356 (1989).
23. L. B. Lucy, "An iterative technique for the rectification of observed distributions," Astron. J. **79,** 745–754 (1974).
24. J. R. Fienup, "Phase retrieval for the Hubble Space Telescope using iterative propagation algorithms," in *Applications of Digital Image Processing XIV,* A. G. Tescher, ed., Proc. Soc. Photo-Opt. Instrum. Eng. **1567,** 327–332 (1991).
25. J. R. Fienup, "Hubble Space Telescope aberrations and alignment determined by phase retrieval algorithms," in *Space Optics for Astrophysics and Earth and Planetary Remote Sensing,* Vol. 19 of 1991 OSA Technical Digest Series (Optical Society of America, Washington, D.C., 1991), pp. 19–21.
26. C. Burrows, *Hubble Space Telescope Optical Telescope Assembly Handbook* (Space Telescope Science Institute, Baltimore, Md., May 1990).