Phasing the mirror segments of the Keck telescopes: the broadband phasing algorithm

Gary Chanan, Mitchell Troy, Frank Dekens, Scott Michaels, Jerry Nelson, Terry Mast, and David Kirkman

To achieve its full diffraction limit in the infrared, the primary mirror of the Keck telescope (now telescopes) must be properly phased: The steps or piston errors between the individual mirror segments must be reduced to less than 100 nm. We accomplish this with a wave optics variation of the Shack–Hartmann test, in which the signal is not the centroid but rather the degree of coherence of the individual subimages. Using filters with a variety of coherence lengths, we can capture segments with initial piston errors as large as ± 30 μm and reduce these to 30 nm—a dynamic range of 3 orders of magnitude. Segment aberrations contribute substantially to the residual errors of ~75 nm. © 1998 Optical Society of America

OCIS codes: 120.5050, 010.7350, 040.1240.

1. Introduction

One of the greatest challenges associated with the segmented design of the Keck telescopes is the achievement of diffraction-limited images in the infrared. To accomplish this, the individual images from the 36 primary mirror segments must not simply be superposed, but must be superposed coherently; that is, the steps or piston errors between segments must be reduced to a small fraction of a wavelength: The segments must be phased. By contrast, the incoherent superposition of the images, appropriate for optical observations, is a simpler task, and the details of how this is done at Keck are described elsewhere.[1–3]

How accurately must the telescope be phased? At optical wavelengths, phasing is not a concern. Phasing effects are washed out by the atmosphere because the atmospheric coherence diameter r_0 is small (20 cm at a wavelength of 0.5 μm for the 0.5-arc sec seeing that is typical of Keck[4]) compared with the circumscribed segment diameter (1.8 m). However, r_0 scales with wavelength as $\lambda^{6/5}$.[5] Thus, at $\lambda = 3$ μm, r_0 is comparable to a segment diameter, and at $\lambda = 13$ μm it is equal to the full diameter of the telescope. (Keck is designed to be used at wavelengths as long as 20 μm.) This means that at longer wavelengths phase errors can seriously degrade the telescope performance. This is illustrated in Fig. 1, which shows the predicted point response function for a variety of wavelengths (from 0.5 to 10 μm) and rms phase errors (from 0 to ∞). (In this paper we specify phase errors at the glass surface, not at the wave front where they would be a factor of 2 larger.) To generate these point response functions we start with the diffraction pattern for a particular configuration of the primary mirror segments with a Gaussian distribution of phase errors characterized by a standard deviation σ_P and no tip–tilt errors. We then (analytically) average over all possible configurations consistent with this σ_P and finally convolve the result with the seeing profile for r_0 (0.5 μm) = 20 cm. At a wavelength of 0.5 μm, the resolution is limited by the atmosphere and the full-width at half-maximum (see the left-hand column of profiles in Fig. 1) is 0.5 arc sec, independent of phase error. On the other hand, phase errors can have a devastating effect on the resolution at longer wavelengths, as can be seen in the last two columns at the right of the figure. However, we note that it is

G. Chanan is with the Department of Physics and Astronomy, University of California, Irvine, California 92697 and the California Institute of Technology, 105-24, Pasadena, California 91125. M. Troy, F. Dekens, and S. Michaels are with the Department of Physics and Astronomy, University of California, Irvine, California 92697. J. Nelson and T. Mast are with the Lick Observatory, University of California, Santa Cruz, California 95064. D. Kirkman is with the Department of Physics, University of California, San Diego, California 92093.

Received 11 January 1997; revised manuscript received 29 July 1997.

0003-6935/98/010140-16$10.00/0

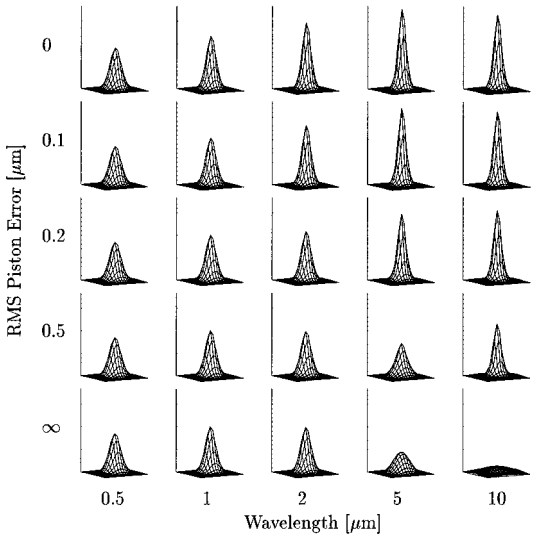

Fig. 1. Theoretical point response functions for the Keck telescope, assuming an atmospheric coherence diameter of r_0 (0.5 μm) = 20 cm for a variety of wavelengths and piston errors of the primary mirror segments. The profiles are averaged over an ensemble of primary mirror configurations that are consistent with the given rms piston error. All plots are to the same scale. FWHM of the images is 0.5 arc sec at $\lambda = 0.5$ μm. Note the devastating effects of phase errors at the longer wavelengths.

difficult to distinguish phase errors of 100 nm from phase errors of 0 at any wavelength; the effects of poor phasing begin to manifest themselves at phase errors of approximately 200 nm. This is essentially the argument that sets the Keck phasing tolerance at $\lesssim$100 nm for normal observing. We note, however, that for planned adaptive optics (and speckle imaging) work, this tolerance may have to be tightened somewhat.[6]

We developed two phasing algorithms, narrow band and broadband, to achieve the above phase specification. Throughout this paper we use these words in a rather specific context. By broadband we mean that we explicitly consider (and exploit) the finite spread in the wavelength of the light. Such finite bandwidth effects can manifest themselves (for large phase errors) even if the bandwidths are small in the sense that $\Delta\lambda \ll \lambda$. Thus in the broadband phasing algorithm we occasionally use a near-infrared filter with a bandwidth of 10 nm, even though in most other contexts this would be thought of as a narrow filter. By contrast, narrow band means that finite bandwidth effects are unimportant, and the diffraction effects of interest can be derived under the assumption that the light is monochromatic.

Before we discuss their development, we briefly compare some of the important properties of narrow-band and broadband phasing. Narrow-band phasing is by far the simpler process to understand; thus we present it first for pedagogic reasons. In principle, narrow-band phasing should be the more accurate procedure; however, it appears to be somewhat less stable, as we show below. As a result, the accuracy we have achieved to date with narrowband phasing is comparable only to what we have obtained with the broadband algorithm. Once these narrow-band stability problems are solved, we hope to run these procedures sequentially, but at the present time, virtually all phasing at Keck is done with the broadband algorithm alone. The main exception—and it is a significant one—is that narrowband phasing provides an important check on the broadband method, particularly with respect to the ring mode distortion caused by dispersion (see Section 7).

In the narrow-band algorithm, the segments are pistoned through small steps that are approximately equal to the desired phasing accuracy, essentially until the zero phase error condition is recognized. This means that the number of steps required is proportional to the initial uncertainty in the phase. On the other hand, in the broadband algorithm the step size can be chosen to match the initial phase uncertainty. Because the accuracy is also proportional to the step size, the procedure can be iterated with the step size for each iteration equal to a fixed fraction of the previous step size. This means that the total number of steps in the broadband procedure increases only as the log of the initial phase uncertainty. This is a profound advantage; it greatly extends the capture range for the broadband algorithm. This in turn eliminates the need to prephase the segments mechanically with a hand-held spherometer—a tedious and time-consuming procedure[7]—to obtain the initial phase uncertainty within approximately a micrometer, as would be required with the narrowband algorithm.

Our central point in this paper is that with the broadband phasing algorithm we can reliably phase the segments of the telescope to a rms piston error of less than 100 nm. This value is sufficiently small that it is virtually indistinguishable from zero for normal observing applications (excluding speckle or adaptive optics). The argument that we can phase to this level is presented in two parts. First we show that the procedure converges rapidly. It is easy to quantify this: We simply execute the phasing procedure and then execute it again. Because the rms difference between the phasing measurements in two successive runs should be $\sqrt{2}$ times larger than the difference between a given measurement and the truth, we define the repeatability to be the rms difference between two successive phasing measurements divided by $\sqrt{2}$. The repeatability of the phasing method is then approximately 30 nm. (We prefer the somewhat more conservative term repeatability to the term accuracy, because, strictly speaking, we are measuring the convergence of a single method, as opposed, say, to measuring our residual errors by means of a second, more accurate method.) That we can repeatedly and reliably put the telescope into the same optical configuration is already a significant result, but we further argue that this is the correct or optimal configuration with regard to segment pistons. The optimality argument is based both on the close agreement between numerical sim-

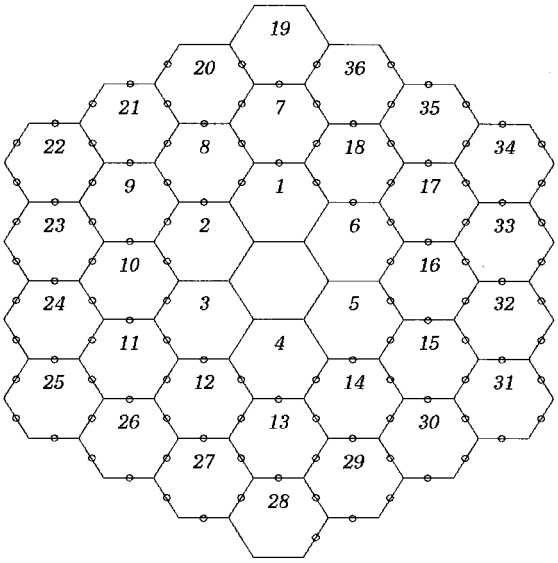

Fig. 2. Geometry of the primary mirror of the Keck telescope showing the 78 circular subapertures that sample the intersegment edges in the phasing procedure. Each segment is 0.9 m on a side. The subapertures are 12 cm in diameter. The 35 peripheral subapertures are used for pupil registration. The six subapertures on the innermost edges (not shown) are obscured by the tertiary mirror and secondary baffles and are not used in the phasing procedure.

ulations and actual results and on the consistency between the solutions found by the narrowband and broadband algorithms, which utilize substantially different diffraction effects and largely independent software.

In the following sections we first briefly describe the Keck telescope itself and then the hardware of the phasing camera, which accomplishes the optical alignment of the telescope. We then examine the diffraction effects associated with out-of-phase segments. Although it is not our principal focus of this paper, we first discuss the narrowband phasing algorithm, because, as noted above, this provides a simple context in which to introduce some of the ideas behind the broadband algorithm. Finally we discuss broadband phasing, including its operational details, systematic effects, and performance.

2. Keck Telescope

The primary mirror of the Keck telescope consists of 36 hexagonal segments that fit together to form a single optical surface with the same area as a circle with a diameter of 10 m (see Fig. 2). Each segment, 90 cm on a side, is controlled in its out-of-plane degrees of freedom by three mechanical actuators on the back of the segment. (The three in-plane degrees of freedom are passively controlled.) There are two capacitive displacement sensors on each of the 84 intersegment edges that sense changes in the local relative heights of the segments. The sensors and actuators are the essential elements of the telescope active control system (ACS), which freezes the relative positions of the segments, stabilizing the structure against gravity, thermal effects, and wind.[8,9] However (and this point is commonly misunderstood), there is no optical reference in the ACS; the ACS stabilizes the sensor readings at their desired values once these are determined—i.e., once the telescope has been aligned optically—but it has no internal way of determining these desired values.[10] The optical alignment task, including tip, tilt, and piston of the segments, falls to the phasing camera system, described in Section 3.

3. Phasing Camera Description

The Keck Observatory phasing camera system[2,11] is a Shack–Hartmann-type wave-front sensor that is permanently mounted at the left-bent Cassegrain focal station of the Keck telescope (one for each telescope). The versatility of this system comes from the ability to place a variety of optical elements at or near the exit pupil. For segment phasing, the key element is an array of 2 mm $\times$ 3 mm prisms, which replaces the usual lenslet array in a traditional Shack–Hartmann camera. This prism array is preceded by a mask at the position of the exit pupil. The mask, at a scale of 1/200 of the primary mirror, defines small circular subapertures (12 cm in diameter referred to the primary) at the center of each of the 84 intersegment edges (shown to scale in Fig. 2). (The six innermost intersegment edges are obscured by the secondary mirror baffles and the tertiary tower; thus only 78 of the edges are actually imaged.) The size of the subapertures is chosen to be significantly smaller than the atmospheric coherence diameter r_0 of approximately 20 cm at a wavelength of 0.5 μm or 27 cm at our shortest operating wavelength of 650 nm. This ensures that the results will be insensitive to atmospheric turbulence in all but the poorest conditions. The optical power for this array is provided by a single downstream objective lens (85 mm, $f/1.2$). This prism array/objective combination is functionally equivalent to an ordinary lenslet array. The advantage of the prisms is much higher image quality (in most lenslet applications one is typically interested only in the first moments of the subimage distribution). In addition, it is easier to generate the required very high focal ratio ($f/140$) with this design, and problems of mask and lenslet registration are virtually eliminated.

The mask–pupil registration is critical to this modified Shack–Hartmann scheme as the subapertures must be aligned accurately with respect to the intersegment edges. We measure this registration (and monitor it with every CCD exposure) by means of an additional subaperture on 35 of the 40 outer segment edges (also shown in Fig. 2). If the mask is registered accurately, each of these peripheral subapertures will be exactly half on and half off the pupil, and the associated 35 subimages will all have precisely the same intensity (i.e., half the intensity of a nonperipheral subimage). For small registration errors, the misalignments in the one rotational and the

two translational coordinates are linear functions of the 35 peripheral spot intensities.[12] We can make these measurements to an accuracy of a few hundredths of a radian in rotation and a few millimeters (referred to the primary mirror) in translation. The entire camera is mounted on an instrument rotator so that the rotation angle can be adjusted; translational adjustments are made with a beam-shearing window on a two-axis mount in the collimated beam. Rotational adjustments are needed only during instrument installation and are therefore done manually. Translational adjustments are needed often because of flexure in the camera and telescope structure with elevation and are made quickly and automatically. A small cross hair (with 30-mm-wide lines, referred to the primary) across each nonperipheral subaperture provides a useful dead band (± 11.5 mm beyond the 7-mm effective intersegment gap) over which the system is insensitive to mask–pupil registration errors.

With a single exposure on a moderately bright star (typically 4th to 7th magnitude), we obtain 78 individual and well-separated subimages or diffraction patterns on the detector (a 1024×1024 Cassini CCD), one subimage for each unobscured intersegment edge. The image scale of 6.77 pixels/arc sec ensures that the subimages are well sampled. (Note that the size of the subimages is set by the 12-cm subaperture diameter, not by the atmospheric coherence diameter because, as noted above, the latter is typically larger by a factor of 2.) The separation of nearest-neighbor subimages on the detector is approximately 60 pixels or 9 arc sec. An efficient computer search algorithm finds the subimages in each CCD image and identifies them with their corresponding intersegment edges, without the need for operator assistance.

4. Single-Subaperture Diffraction

We first focus our attention on a single intersegment subaperture, and for now we consider the light to be monochromatic. The details of the diffraction pattern formed by this subaperture are sensitive to the physical step height between the two segments, as the following analysis shows. We follow the basic approach, and to a large extent the notation, of Born and Wolf.[13]

Let $\boldsymbol{\rho}$, with rectangular coordinates (ξ, η) (in units of length) or circular coordinates (ρ, θ), be the position vector in the subaperture plane and let $\boldsymbol{\omega}$, with rectangular coordinates (x, y) (in radians) or circular coordinates (ω, ψ), be the position vector in the image plane. We consider a circular subaperture of radius a, straddling two segments divided by the line $\eta = 0$; the upper segment ($\eta > 0$) has a piston error of $\delta/2$, and the lower segment ($\eta < 0$) has a piston error of $-\delta/2$. (Thus δ is the physical step height; the corresponding wave-front step height is 2δ.)

In the absence of other aberrations, the complex amplitude in the image plane $\hat{f}(\boldsymbol{\omega}; k\delta)$ is simply the Fourier transform of the (complex) subaperture function $f(\boldsymbol{\rho}; k\delta)$.

$$f(\boldsymbol{\rho}; k\delta) = \begin{cases} \exp(ik\delta) & \eta \geq 0;\ \rho \leq a \\ \exp(-ik\delta) & \eta < 0;\ \rho \leq a, \\ 0 & \rho > a \end{cases} \tag{1}$$

$$\hat{f}(\boldsymbol{\omega}; k\delta) = \frac{1}{\pi a^2} \int_0^{\pi} \int_0^{a} \exp(ik\delta)\exp(ik\boldsymbol{\rho} \cdot \boldsymbol{\omega})\rho d\rho d\theta + \frac{1}{\pi a^2} \int_{-\pi}^{0} \int_0^{a} \exp(-ik\delta)\exp(ik\boldsymbol{\rho} \cdot \boldsymbol{\omega})\rho d\rho d\theta, \tag{2}$$

where $k = 2\pi/\lambda$ and the normalization is chosen such that the on-axis, in-phase intensity is unity (see below). For simplicity and clarity in this discussion, we neglect the relatively small effects that are due to the aperture cross hairs (but these are incorporated into all actual data analysis). It is not hard to show that the imaginary part of $\hat{f}$ vanishes and we have

$$\hat{f}(\boldsymbol{\omega}; k\delta) = \frac{2}{\pi a^2} \int_0^{\pi} \int_0^{a} \cos(k\delta + k\boldsymbol{\rho} \cdot \boldsymbol{\omega})\rho d\rho d\theta. \tag{3}$$

The intensity in the image plane is simply

$$I(\boldsymbol{\omega}; k\delta) = \hat{f}^2(\boldsymbol{\omega}; k\delta). \tag{4}$$

This integral can be handled numerically by two-dimensional fast Fourier transform techniques; however, a modest amount of analysis can make the calculation more efficient and also provide some insight into the problem. We first consider some special cases.

(1) When the segments are in phase, we have $\delta = 0$ and

$$\hat{f}(\boldsymbol{\omega}; 0) = \frac{2}{\pi a^2} \int_0^{\pi} \int_0^{a} \cos(k\boldsymbol{\rho} \cdot \boldsymbol{\omega})\rho d\rho d\theta. \tag{5}$$

The θ integral is independent of ψ. Thus we have the familiar result

$$\hat{f}(\boldsymbol{\omega}; 0) = \frac{2}{a^2} \int_0^{a} J_0(k\rho\omega)\rho d\rho, \tag{6}$$

$$I(\boldsymbol{\omega}; 0) = \left[\frac{2J_1(ka\omega)}{ka\omega}\right]^2. \tag{7}$$

Note that $I(0; 0) \rightarrow 1$ as claimed. The integrated intensity in these units (which is the result for any phase) is

$$I_{\text{tot}} = \int_0^{2\pi} \int_0^{\infty} I(\omega)\omega d\omega d\psi = \frac{4\pi}{k^2 a^2} = \frac{\lambda^2}{\pi a^2}. \tag{8}$$

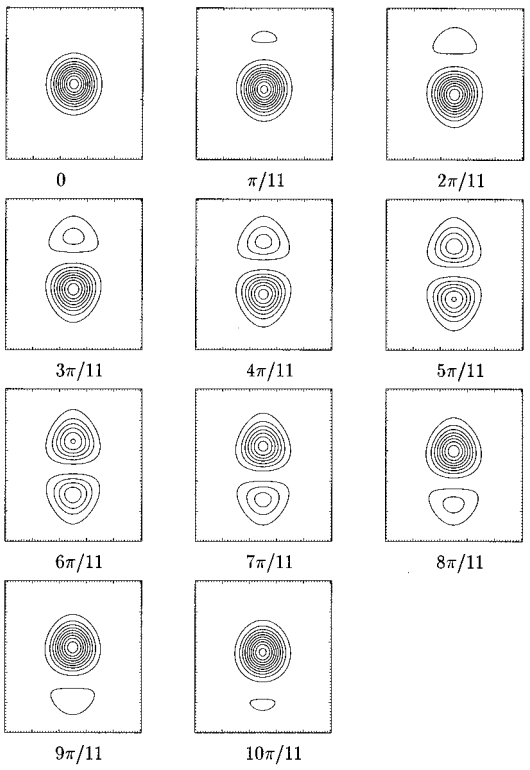

Fig. 3. Theoretical diffraction patterns (monochromatic light) for a split circular subaperture of radius a with a physical step δ between the two halves given by $k\delta = 0, \pi/11, 2\pi/11, \ldots, 10\pi/11$. The boxes are $15/ka$ rad on a side.

(2) The two segments are maximally out of phase when $\delta = \lambda/4$, in which case $k\delta = \pi/2$ and

$$\hat{f}\left(\boldsymbol{\omega}; \frac{\pi}{2}\right) = -\frac{2}{\pi a^2}\int_0^{\pi}\int_0^{a} \sin(k\boldsymbol{\rho}\cdot\boldsymbol{\omega})\rho\,d\rho\,d\theta, \tag{9}$$

$$= \frac{2}{\pi}\int_0^{\pi} \frac{u\cos u - \sin u}{u^2}\,d\theta, \tag{10}$$

where $u = ka\omega\cos(\theta - \psi)$. Despite the superficial similarity between Eqs. (5) and (9), the θ integral here does not readily simplify.

(3) The case of arbitrary δ can be constructed readily from $\hat{f}(\boldsymbol{\omega}; 0)$ and $\hat{f}(\boldsymbol{\omega}; \pi/2)$. We have

$$I(\boldsymbol{\omega}; k\delta) = \left[\cos k\delta\,\hat{f}(\boldsymbol{\omega}; 0) + \sin k\delta\,\hat{f}\left(\boldsymbol{\omega}; \frac{\pi}{2}\right)\right]^2. \tag{11}$$

Figure 3 shows the theoretical diffraction patterns for a sequence of 11 equally spaced values of δ from $k\delta = 0$ to $k\delta = 10\pi/11$. The phase sensitivity is clear. When the two segments are in phase, one obtains the usual circular diffraction pattern (Airy disk). For small $\delta > 0$ the peak location shifts to

$$y = -\frac{16}{3\pi}\frac{\delta}{a}. \tag{12}$$

[To see this, expand Eq. (11) for small δ and small ω.] One can obtain both the sign and the magnitude of Eq. (12) by approximating this stepped subaperture to first order by the best-fit plane. As δ is increased, the original peak continues to shift downward, and a second diffraction peak appears near the top of the subimage. The shift of the original peak, and the relative intensity of the second peak, both increase with increasing δ.

The two peaks become equal at a physical step height of $\lambda/4$, corresponding to an overall path difference of $\lambda/2$ between the two halves of the circular subaperture. As δ continues to increase, the second peak continues to grow at the expense of the first until, at $\lambda/2$ (a path difference of λ), it has replaced and is indistinguishable from the original peak. The separation of the two peaks is weakly dependent on the phase difference, varying from $ka\Delta\omega = 5.14$ as $k\delta \to 0$ to $ka\Delta\omega = 5.29$ for $k\delta = \pi/2$.

At this point, for the first time in this development, we consider explicitly the finite wavelength interval $\Delta\lambda \approx 2\pi\Delta k/k^2$ spanned by the light; this involves the integration of Eq. (11) over k. The finite spread in wavelength becomes significant for the trigonometric factor in Eq. (11) when the condition $\Delta\lambda \ll \lambda^2/2\delta$ is violated. For future reference we abbreviate $\lambda^2/2\Delta\lambda$ as l and refer to it as the coherence length. (The factor of 2 in the denominator is to some extent arbitrary, but is motivated by the fact that the path difference doubles on reflection.) The spread in wavelength becomes significant for the $\hat{f}$ factors when the (generally less restrictive) condition $\Delta\lambda \ll \lambda$ is violated. For simplicity here we assume $\Delta\lambda/\lambda \ll 1$, but make no assumption about the relative size of δ and l. This means that when we integrate Eq. (11) over k, we can approximate the implicitly k-dependent $\hat{f}$ factors by evaluating them at the midpoint of the k interval, but the sin $k\delta$ and cos $k\delta$ factors must be explicitly averaged over k.

We assume a Gaussian bandpass in k:

$$g(k) = \frac{1}{\sqrt{2\pi\sigma_k^2}}\exp\left[-\frac{(k-k_0)^2}{2\sigma_k^2}\right]. \tag{13}$$

Note that, to within a constant of order of unity, σ_k is simply the reciprocal of l. If we interpret Δk as the FWHM of Eq. (13), then $\Delta k = \sqrt{8\ln(2)}\,\sigma_k$, and from the definition of l we obtain

$$\sigma_k = \frac{\pi}{\sqrt{8\ln(2)}}\frac{1}{l} = \frac{1.334}{l}. \tag{14}$$

To perform the k averaging, we multiply the right-hand side of Eq. (11) by $g(k)$ and integrate over k. Extending the integrals to $k = -\infty$ introduces negligible error. We then obtain

$$\langle I(\boldsymbol{\omega}; k\delta)\rangle = \alpha_1\hat{f}^2(\boldsymbol{\omega}; 0) + \alpha_2\hat{f}(\boldsymbol{\omega}; 0)\hat{f}\left(\boldsymbol{\omega}; \frac{\pi}{2}\right) + \alpha_3\hat{f}^2\left(\boldsymbol{\omega}; \frac{\pi}{2}\right), \tag{15}$$

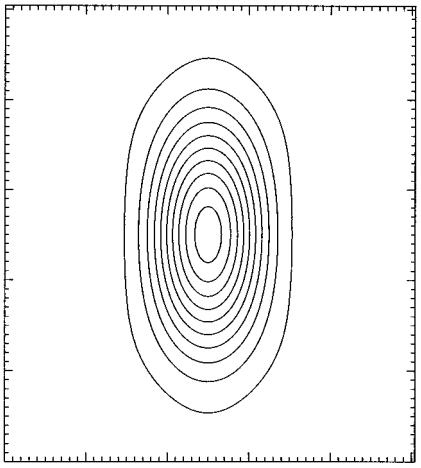

Fig. 4. Theoretical diffraction pattern for a split circular subaperture of radius a with a step between the two halves in the incoherent limit: $\delta \gg \lambda^2/2\Delta\lambda$. As in Fig. 3, the box is $15/ka$ rad on a side.

where the angled brackets denote k averaging and

$$\alpha_1 = \frac{1}{2}[1 + \exp(-2{\sigma_k}^2\delta^2)\cos 2k_0\delta],$$

$$\alpha_2 = \exp(-2{\sigma_k}^2\delta^2)\sin 2k_0\delta,$$

$$\alpha_3 = \frac{1}{2}[1 - \exp(-2{\sigma_k}^2\delta^2)\cos 2k_0\delta]. \qquad (16)$$

This reduces correctly to Eq. (11) as $\sigma_k\delta \to 0$. At the opposite extreme, as $\sigma_k\delta \to \infty$, we obtain

$$\langle I(\boldsymbol{\omega}; \infty)\rangle = \frac{1}{2}\left[I(\boldsymbol{\omega}; 0) + I\left(\boldsymbol{\omega}; \frac{\pi}{2}\right)\right], \qquad (17)$$

i.e., the intensities from the two semicircular subapertures simply add incoherently. This incoherent diffraction pattern is shown in Fig. 4.

The narrow-band algorithm is based on our extracting the phase (δ) information contained in the trigonometric factors of Eq. (11). The broadband algorithm is based on our extracting the phase information contained in the coefficients (α_1, α_2, α_3) of Eq. (15). We now describe these respective algorithms in some detail.

5. Narrow-Band Phasing

Equation (11), which describes the sinusoidal variation in the appearance of the diffraction pattern with $k\delta$, is the basic equation of narrow-band phasing. In principle, given a diffraction pattern (in terms of pixel intensities on a CCD) corresponding to an arbitrary step height (and ignoring the 2π phase ambiguity for now), we could extract δ by means of surface fitting to this functional form. However, a more robust and numerically efficient approach, which accomplishes the same thing, can be formulated in terms of correlation analysis. To this end (for future cross correlation) we generated a set of 11 theoretical diffraction patterns in the form of artificial CCD subimages corresponding to Fig. 3, but of the appropriate resolution and including the effects of the cross hairs. The choice of 11 steps represents a compromise between speed and accuracy; also, the algorithm that reconfigures the primary mirror between steps (see below) can more easily accommodate an odd number of steps. Note that it is not necessary to span a full half-wave, since physical step heights of 0 and $\lambda/2$ are indistinguishable. In principle we should perhaps convolve our theoretical diffraction pattern templates with the expected seeing profile before comparison with the data. However, for a 12-cm aperture and a typical value of r_0 (at 0.5 μm) of 20 cm, the Strehl ratio is between 0.78 and 0.86 for our observing wavelengths of 0.65–0.90 μm. Thus seeing has a relatively small effect on the diffraction pattern and, in the interest of simplicity, we ignore it. This does not bias the measurement of phase, but may result in a slight loss of signal-to-noise ratio.

We can determine the height of an individual edge to the nearest template step by cross correlating the diffraction pattern pixel by pixel against all 11 templates and selecting the template that gives the maximum correlation coefficient. We refer to the number (1 to 11) of the correlation-maximizing template as the correlation index. To be precise, for the (linear) correlation coefficient we use Pearson's r:

$$r = \frac{\Sigma_i(x_i - \bar{x})(y_i - \bar{y})}{[\Sigma_i(x_i - \bar{x})^2]^{1/2}[\Sigma_i(y_i - \bar{y})^2]^{1/2}}. \qquad (18)$$

Here x_i is the intensity in the ith pixel of the image, y_i is the intensity in the corresponding pixel of the template, and the overbars represent averages. The sum is taken over all pixels in a 33×33 pixel box (approximately 5 arc sec on a side) centered on the centroid of the image. (This centering requires reregistering the image with respect to the pixel boundaries, as described below.) For a discussion of the properties of Pearson's r, including a subroutine for calculating it, see Press *et al.*[14] Note that r is independent of the normalization of both the image and the template; it is also unaffected by a constant offset in the pixel counts. This means that no rescaling of the images or background subtraction is necessary. Because we have only 11 template images, there will be a quantization error in the index of the best fitting template, corresponding to step height errors as large as $\pm\lambda/44$ or approximately ± 20 nm. Although such errors are not large, they are also not negligible compared with the ultimate accuracy of the phasing procedure (see below), so it is advantageous to refine the piston estimate by interpolation. In a small neighborhood of its maximum, the correlation index should vary quadratically with the step height error; indeed, numerical calculations show that the correlation index is always within 0.7% of its quadratic approximation when we are within $\lambda/22$ of the true step height. Thus we can refine the phase estimate by a simple quadratic interpolation over this range (i.e., ± 1 unit in the correlation index).

For improved accuracy we use the telescope ACS to

piston the segments so as to shift each intersegment edge step to 11 different positions. (It is convenient but not essential to have the number of measurement positions equal the number of template steps.) Although in principle a single exposure per edge would be sufficient, we find that this multiple-position technique reduces both statistical and systematic errors. The step height measurements are made on all intersegment edges in parallel. To do this we define 11 configurations of the primary mirror for which each edge is shifted to all 11 possible steps (in units of λ: $-5/22$, $-4/22$, $-3/22$, . . . , $3/22$, $4/22$, $5/22$) exactly once each (although different edges pass through step 0, at different times); see Ref. 15 for a unique way to take these measurements. A CCD exposure is recorded for each of the 11 configurations.

For every intersegment edge in every configuration or CCD exposure we extract a 33×33 pixel array containing the diffraction pattern for that edge. Prior to the extraction, the diffraction pattern or subimage must first be reregistered with respect to pixel boundaries so that the centroid coincides exactly with the array center, not just to within the nearest pixel. (We choose the centroid as the origin of the local coordinate system because we do not accurately know the absolute origin.) The intensity of a pixel in the new centroid-centered grid is the weighted sum of the original pixels that it overlaps where each (old) pixel is weighted by its respective spatial contribution. Although the templates are generated with respect to an absolute origin, they are similarly reregistered to facilitate comparison with the data. Each 33×33 data array is then cross correlated against the template sequence, and we derive an interpolated phase as described above. Because there are three possible edge orientations (see Fig. 2), we have templates of the three different orientations stored on disk for this comparison.

Because the diffraction pattern is a periodic function of δ with a period of $\lambda/2$, the conversion from phase (in units of a wavelength) to step height (in units of micrometers) involves the addition of an integer number of half-waves; this unknown integer cannot be extracted from purely monochromatic data. We can resolve this problem of absolute phase ambiguity, corresponding to determining the appropriate integer number of half-waves, within the narrow-band technique by making measurements at two different wavelengths. Details are given in Appendix A. Once the 78 edge heights are determined, they are stored for further analysis. The subsequent analysis is identical in the narrow-band and broadband algorithms; therefore a discussion is deferred to Section 6.

We can readily evaluate the repeatability of the narrow-band phasing algorithm by running the algorithm, sending the resulting piston corrections to the primary mirror, and then repeating the procedure. As noted above, the second set of piston corrections should then be $\sqrt{2}$ times the repeatability of the procedure. From June 1995 to September 1996 there were 19 times when narrow-band phasing was run twice in a row. The mean piston error for the second trial of each pair was 47 nm, corresponding to a repeatability of 33 nm.

As good as the above repeatability is, it is in some sense considerably worse than expected. Because we make multiple measurements of the phase at each edge, we can generate an internal estimate of the phase measurement errors for each of the two wavelengths (prior to the determination of the appropriate integer number of half-waves to be added). In the above 19 pairs of trials, the average internal errors were less than 10 nm. The discrepancy between the 10-nm internal errors and the 33-nm repeatability presumably indicates that, for a few of the 78 edges, we resolved the absolute phase ambiguity incorrectly in one of the two pairs of measurements (see Appendix A for details). This problem may result from the segments drifting slightly between the measurements at the two different wavelengths. We are continuing to investigate this problem, but even at the present level of repeatability, the narrow-band algorithm provides an important check on the broadband technique, as described below.

6. Broadband Phasing

The defining equation of broadband phasing is Eq. (15), which describes how the diffraction pattern washes out as $\sigma_k\delta$, or equivalently δ/l, becomes significant. The length scale of interest for piston errors is now not $\lambda/2$ but l. Therefore imagine a sequence of stepped exposures analogous to that in narrow-band phasing, except that now we select 11 steps spanning not a half-wave but rather roughly the coherence length l. For a specific example, consider a filter with a bandwidth of 10 nm and a central wavelength of 891 nm, corresponding to a coherence length of $l = 40$ μm. On the basis of Monte Carlo calculations described below, we choose a step size of 6 μm (or 150 times the typical narrow-band step size of 40 nm). The full range spanned by these steps is then ± 30 μm with respect to the nominal position.

A typical sequence of real CCD images generated in this manner for a particular intersegment edge is shown in Fig. 5. The first and last steps of the sequence are significantly out of phase, and the diffraction patterns closely resemble the completely incoherent subimage of Fig. 4. As we approach the middle subimages of the sequence, the patterns become more and more coherent. We emphasize that the parameter of interest is the degree of coherence of the subimage (a quantity that we define below), not the precise value of the phase; the latter varies rapidly from subimage to subimage because of the large step size.

We now quantify the degree of coherence of a subimage. One way to do this is suggested by an inspection of Fig. 5. First rotate this figure counterclockwise by 30 deg so that the subimages match the orientation of the templates (Fig. 3). Now consider, for example, the fourth through seventh subimages of Fig. 5 (labeled -12 μm through $+6$ μm). Each of these bears a fair resemblance to one or an-

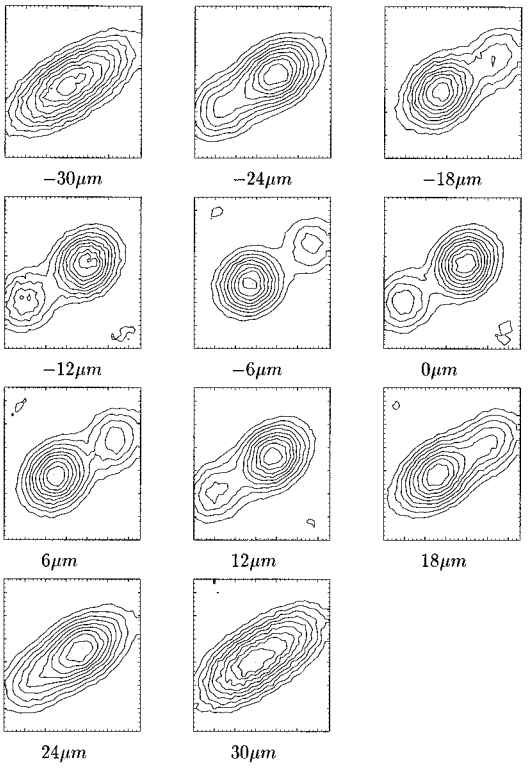

Fig. 5. Typical broadband sequence of CCD subimages. The step size is 6 μm and the coherence length is 40 μm. The edge is in phase between the fourth and fifth subimages (see top panel of Fig. 7). Boxes are 5 arc sec on a side.

other of the template subimages. By contrast, the first and last subimages of Fig. 5 bear little resemblance to any of the template subimages because the details of the diffraction patterns of the former are washed out by incoherence effects. Thus one measure of coherence is to correlate the subimage in question with each of the template subimages and select the maximum correlation coefficient (after interpolation). The coherence of the (monochromatic) template subimages themselves is then unity by definition, as is sensible. However, the coherence of the completely incoherent subimage (Fig. 4) is a disappointingly large 0.88 (determined by numerical integration), so that the dynamic range of the coherence parameter—and, as a result, the signal-to-noise ratio—is uncomfortably small.

We can improve the definition of coherence by noting that there is information contained not only in how closely a subimage resembles the template subimage that it most resembles (the maximum correlation), but also in how little it resembles the subimage that it least resembles. Thus, for example, the (completely coherent) in-phase ($k\delta = 0$) template image has little correlation with the template images with $k\delta \sim \pi/2$, but the incoherent image of Fig. 4 has at least a moderate correlation with all the template subimages. Therefore we define the coherence as the maximum minus minimum correlation coefficient in the above sense. The coherence of the template subimages themselves then ranges from 0.88 to 0.97

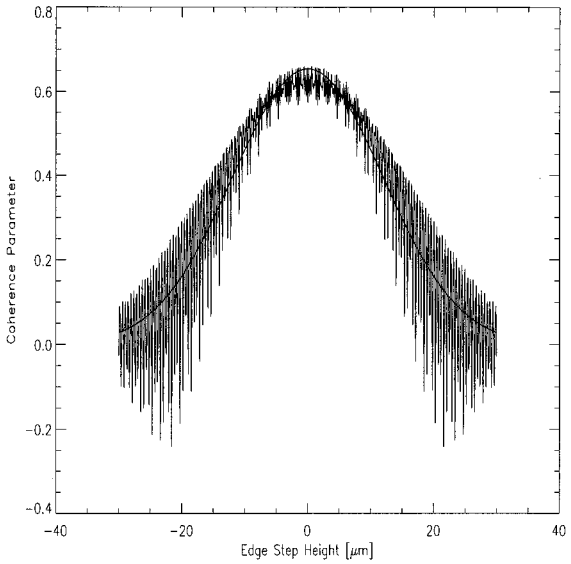

Fig. 6. Theoretical curve of coherence parameter versus edge step height for a filter with a central wavelength of 891 nm, $\Delta\lambda = 10$ nm, and a coherence length of 40 μm. The solid curve is the best-fit Gaussian approximation.

(ideally they would have been unity), whereas the coherence of the completely incoherent subimage is 0.31, where all the above quantities have been determined numerically. Note the large improvement in the dynamic range of the coherence parameter. For convenience, we subtract the latter value of 0.31 from all coherence parameters so that the coherence of a completely incoherent image is zero.

In Fig. 6 we plot the coherence parameter as defined above versus the step height for a series of numerically generated images corresponding to Eq. (15), defined by a filter with a wavelength of 891 nm and a bandwidth of 10 nm. The characteristic horizontal scale of the curve is simply the filter coherence length of $l = 40$ μm. Apart from considerable high-frequency structure, the curve can be well approximated by a Gaussian, as shown in Fig. 6. In fact, in practice we simply use the Gaussian approximation; the high-frequency structure is treated as noise. For our purposes it makes little difference that the origin of the noise is theoretical rather than observational. The actual coherence parameters corresponding to the diffraction patterns in Fig. 5 are shown in the top panel of Fig. 7.

Because the characteristic scale of the coherence function plotted in Fig. 6 is the coherence length of the filter, we can tune the algorithm by selecting $\Delta\lambda$ such that l matches the expected range of piston errors for a given situation. These different coherence lengths define different modes of broadband phasing. We show below that the accuracy of a given mode is roughly proportional to the coherence length. Furthermore, for our standard 11-step sequences, an edge will be measured successfully (and the corresponding segment captured) if its phase lies anywhere within ±5 steps of the nominal position. Thus the capture range also is proportional to the

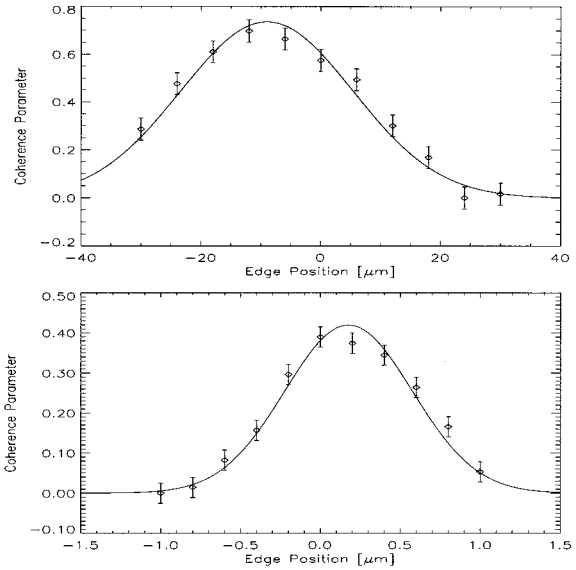

Fig. 7. Typical curves of coherence parameter versus edge position for a filter with a coherence length of 40 μm (Δλ = 10 nm) (top panel) and for a filter with a coherence length of 1.2 μm (Δλ = 200 nm) (bottom panel). The top panel corresponds to the particular sequence of subimages shown in Fig. 5.

coherence length. This means that we can cascade the various broadband phasing modes by choosing the capture range of the n + 1st mode to be equal to the accuracy of the nth mode. As a result the broadband phasing scheme converges rapidly, with the total time required increasing only as the log of the initial piston uncertainty.

In addition to the Δλ = 10-nm filter, we use bandwidths of 30 and 100 nm, or we run unfiltered in which case the effective bandwidth is approximately 200 nm. The central wavelengths are in the range of 850–900 nm, except for the unfiltered case in which the effective wavelength is approximately 700 nm. When we run unfiltered, we choose stars of a consistent spectral type (K0 to K5) to minimize variations in the effective wavelength. The corresponding coherence lengths are l = 40, 12, 3.8, and 1.2 μm. As noted in Section 1, the capture range of ±30 μm that is associated with the l = 40-μm filter completely eliminates the need for mechanically prephasing the segments with a hand-held spherometer, as would have been necessary with the narrow-band algorithm. Indeed, no spherometry was ever done on Keck 2 before successfully phasing the segments.

We refer to the four phasing modes, from least accurate to most accurate (or largest to smallest filter coherence length), as phasing 1000, phasing 300, phasing 100, and phasing 30, respectively, in which the numbers represent (in slightly pessimistic terms) the repeatability of the mode in nanometers, as determined below. The parameters and properties of the various modes are summarized in Table 1. Column 1 gives the mode name (e.g., phasing 1000). Columns 2, 3, and 4 give the filter central wavelength, bandwidth, and coherence length, respectively. Columns 5 and 6 give the step size and capture range for the mode. For completeness, column 7 gives the typical V magnitude of the stars used in the mode. Note that the nominal star brightness decreases as the filter bandwidth increases. This is simply a matter of convenience to keep the count rates and integration times roughly constant in all operational modes.

At this point we want to quantify the errors introduced by fitting the coherence function of Fig. 6 to a Gaussian. We first determined the width (σ_c) of the Gaussian approximation by numerically generating coherence curves that are sufficiently dense with the points to sample adequately the high-frequency structure shown in the figure. We then fitted these to a Gaussian, evaluating the width by the technique of least-squares (or χ^2) parameter estimation.[16] We repeated this for the four different coherence lengths, corresponding to the four different modes. The results are shown in Table 2 in which the mode name is given in column 1 and the theoretical value of σ_c in column 5. Note that σ_c scales with (and is approximately 30% of) the coherence length. We then ran a series of Monte Carlo simulations, one for each phasing mode, in which an edge height was chosen randomly, and then 11 diffraction patterns, corresponding to the 11 measurement steps of the broadband algorithm, were generated numerically. For the simulations to be realistic, the random edge heights for each mode were drawn from a Gaussian probability distribution whose width was roughly the same size as that of the expected distribution of the starting point edge heights for that mode. (The final results, however, are only weakly dependent on the starting point.) The rms values of these starting edge heights are given in column 2 of Table 2. We calculated the edge height by curve fitting (using the above-determined value of σ_c) to obtain the step height corresponding to the maximum of the Gauss-

Table 1. Broadband Phasing Parameters

Mode	Wavelength (nm)	Bandwidth (nm)	Coherence Length (μm)	Step Size (μm)	Capture Range (μm)	Star Magnitude V
Phasing 1000	891	10	40	6	±30	4
Phasing 300	852	30	12	2	±10	5
Phasing 100	870	100	3.8	0.6	±3	6
Phasing 30	700 (eff)	200 (eff)	1.2	0.2	±1	7

Table 2. Comparison of Theoretical and Actual Broadband Phasing Modes

Phasing Mode	Initial Piston Error (nm)	Number of Trials		Gaussian Width		Final Piston Error	
		Theoretical	Experimental	Theoretical (μm)	Experimental (μm)	Theoretical (nm)	Experimental (nm)
Phasing 1000	10,000	1000	12	13.1	14.1	791	741 ± 95
Phasing 300	3000	1000	2	4.02	4.40	212	213 ± 61
Phasing 100	1000	1000	26	1.25	1.40	42	82 ± 5
Phasing 30	300	1000	17	0.30	0.40	22	28 ± 3

ian. The difference between the calculated edge height and the true value used in the simulation is the edge height measurement error. For each mode, the process was repeated 1000 times to obtain good statistics, and the rms edge measurement errors (over these 1000 trials) were determined. We then converted these edge height measurement errors to rms piston errors by multiplying by a factor of 0.686. (This latter factor, which was determined numerically, is less than unity because the 36 segment pistons are overconstrained by the 78 edge height measurements.) The rms piston errors from the Monte Carlo are given in column 7 of Table 2.

Although the performance of the broadband algorithm is discussed in some detail in Section 8, here we give a brief comparison of the performance results of the Monte Carlo and the actual algorithm. There were 57 occasions between June 1995 and January 1997 when we obtained a direct measure of the repeatability of the algorithm, either because one mode was followed directly by a more accurate mode or because one mode was run twice in a row. (When one mode is followed by a more accurate one, the rms phasing command determined by the second mode is approximately equal to the repeatability of the first mode.) The number of repeatability measurements made with each mode is given in column 4 of Table 2.

The real measurements were made in the same manner as were the numerical simulations. That is, the coherence parameter was measured for each edge at 11 discrete points, and we determined the edge height by finding the height that maximized the Gaussian approximation to the coherence parameter. Representative curves of the coherence parameter versus the edge position are shown for the modes corresponding to the longest (l = 40-μm) and shortest (l = 1.2-μm) coherence lengths in the top and bottom panels of Fig. 7. We did not use the theoretical value of the σ_c of the Gaussian; rather we determined it by curve fitting to the empirical data. The empirical values of σ_c are given in column 6 of Table 2. We attribute the (generally small) differences between the theoretical and empirical values of σ_c to differences between the actual and modeled filter bandpasses. (These differences tend to increase with increasing bandwidth.) Finally, the 78 edge height measurements were converted to 36 piston commands by means of the singular value decomposition technique, as described below. The empirical repeatabilities of the various modes of the algorithm are given in column 8 of Table 2. Note that the repeatabilities generally agree quite well with the theoretical predictions of accuracy for each mode. This means that our principal source of error is well understood—it is a direct consequence of the Gaussian approximation. The only exception is the second most accurate mode (phasing 100), in which the errors are almost a factor of 2 larger than expected. This may result from the skewed bandpass of this particular filter, but in any case it is not a serious problem because, as long as all modes converge, it is ultimately the accuracy of the most accurate algorithm that counts. Overall, we believe that the good agreement between both the theoretical and the empirical curve widths, and—more important—between the theoretical and empirical accuracies, constitutes a strong validation of the broadband approach.

After the 78 distinct edge heights are determined, they are tested for adequate signal (the amplitude of the curves in Fig. 7); this typically results in the rejection of only a few ($\lesssim$3) edges. We believe that the most common cause of inadequate signal is aberrations in the segments (see Section 8); these produce local differential tilt errors between the two segments that tend to reduce the signal.[17] The six edges that lie directly underneath the secondary mirror support structure also tend to have a lower signal, presumably from the competing diffraction effects of that structure. (These should not be confused with the six edges that lie directly underneath the secondary baffle, which, as noted above, are completely obscured.) Edges are also missed if they are beyond the capture range of the particular algorithm, although this usually occurs only in the least accurate mode. In the following discussion we assume for simplicity that all 78 edges have been measured successfully.

The 78 measured edge steps δ_j are related to the desired 36 segment piston errors p_i by a system of linear equations of the form

$$p_{i(j)} - p_{i'(j)} = \delta_j; \quad j = 1, \ldots, 78, \tag{19}$$

where $i(j)$ and $i'(j)$ identify the segments associated with the jth edge. Because the absolute (average) phase of the mirror is not of interest, it makes sense to add a constraint to keep it constant, lest the actuators exceed their range:

$$\sum_i p_i = 0; \quad i = 1, \ldots, 36. \tag{20}$$

Equations (19) and (20), which apply to both the broadband and the narrow-band algorithms, constitute a simple linear system of 79 equations in 36 unknowns (p_i). The technique of singular value decomposition provides a powerful, robust, and easily implemented method for solving systems of this type by constructing the pseudoinverse of the matrix (here, 79 rows by 36 columns) that defines the linear system. For a discussion as well as the actual code (which we use essentially without modification), see Press *et al.*[18]

As is generally the case with overdetermined systems of equations, the solution—here a 36-element vector whose negative is the desired piston correction—will not exactly satisfy the original system of equations [with the exception of Eq. (20), which can always be satisfied exactly]. Therefore we can define the residual to the fit for the jth edge, or more succinctly the jth edge residual, as

$$\Delta\delta_j = \delta_j - [p_{i(j)}^* - p_{i'(j)}^*], \tag{21}$$

where $p_{i(j)}^*$ and $p_{i'(j)}^*$ represent the best-fit value of the pistons. There are many potential contributors to the edge residual, including segment aberrations, tilt errors in the segments from imperfect tip–tilt alignment prior to phasing, and seeing. The rms edge residual is a useful figure of merit describing the quality of the fit, and we make use of this below.

Note that, strictly speaking, what we accomplish by the above singular value decomposition is to minimize the edge residual (in a rms sense). If the segments are perfect, this is the same as achieving the minimum piston error, but if the segments are not perfect, the singular value decomposition (edge-residual-minimizing) solution will differ somewhat from the minimum-piston solution. (With an imperfect segment, determination of the segment phase would involve averaging over the surface of the segment.) The rms edge residual itself provides an estimate of this discrepancy. As we show below, although the repeatability of the broadband algorithm is of the order of 30 nm, the accuracy may be worse by a factor of 2 to 3.

Once the desired 36 piston changes are determined, they are communicated to the telescope's ACS. Then we move on to a more accurate mode, or (in the case of the most accurate mode) we repeat the process, to verify the measurements.

As a check we also generated empirical templates instead of the numerically generated ones. That is, we deliberately stepped various segments of the telescope through the same relative phases as are represented in the set of theoretical templates and recorded the corresponding images. We can do this because the telescope ACS is calibrated relatively, though not absolutely (that is, the ACS knows what a micrometer is, but it does not know where the segments are in phase). We determined the phases of the empirical templates by comparing each one to its own reflection about a line parallel to the intersegment edge. The resulting empirical templates were then substituted for the numerically generated ones, and several phasing data sets that had been stored on disk were reanalyzed by correlating against the empirical, as opposed to theoretical, templates. The phasing commands generated from the empirical templates were statistically indistinguishable from those generated from the theoretical ones. This constitutes a strong validation of the overall method; it also shows that the degradation of the images because of seeing effects associated with $a/r_0 \sim 0.4$ are unimportant because the empirical templates include seeing effects but the numerical ones do not. In view of the fact that the results are indistinguishable, we use only theoretical templates because we can modify them easily for different wavelength or pixel size on the detector, without having to take telescope time to prepare new empirical templates.

7. Systematic Effect in Broadband Phasing: Ring Mode

As noted above, the narrow-band phasing algorithm suffers from a problem that limits its accuracy to ~30 nm instead of the expected 10 nm. However, this is still sufficiently accurate to check the broadband phasing algorithm. For example, in our early experience with the phasing-100 mode, the results consistently converged as expected to a rms piston command of $\lesssim$100 nm. Yet when this was followed immediately by a measurement with the narrow-band phasing algorithm, the mirror consistently showed a systematic piston error of almost 300 nm. In addition, this configuration of the mirror possessed a high degree of symmetry—specifically the ring mode—in which the outer rings of the segments are systematically lower than the inner rings. The size of the piston error in the ring mode is roughly proportional to the distance of the ring from the optical axis. An example of the ring mode is shown in Fig. 8.

It has become clear subsequently that the ring-mode offset is caused by dispersion in the prisms inside the phasing camera, which closely simulates the blurring caused by out-of-phase segments. The effect is independent of the step size used in the broadband phasing algorithm and also (perhaps counterintuitively for a dispersion-dependent effect) is independent of the filter bandwidth. The long-term stability of the effect, its symmetry (pure ring mode), its sign (outer segments are low), and its approximate magnitude (300-nm rms) are all explained in terms of the following analysis. The narrow-band phasing algorithm, on the other hand, is not affected by dispersion to the first order. [It is easy to see that narrow-band phasing is not subject to the ring-mode or dispersion bias, because all the relevant equations are independent of $\Delta\lambda$, but we also verified this empirically by checking the phase of the reference beams. We obtain a phase error of 2 nm (rms); this includes all systematic and random effects arising from within the camera (such as dispersion) and thus rules out ring mode at this level.]

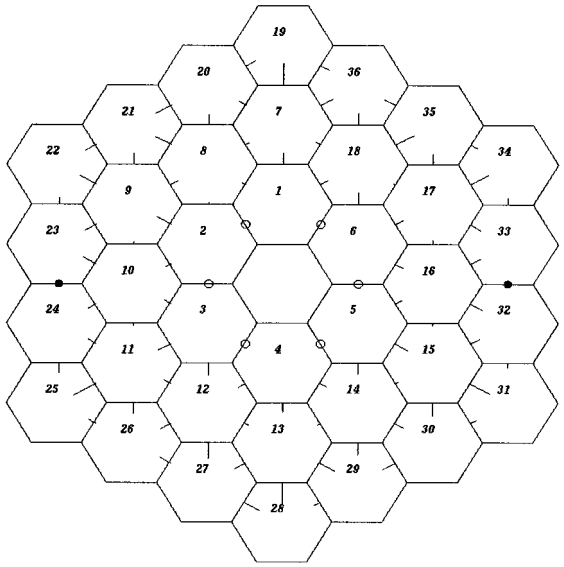

Fig. 8. Uncorrected ring mode in the primary mirror. The ring mode—normally corrected to a high degree in software—is a systematic effect caused by dispersion in the phasing camera optics. The edge height measurements are indicated by line segments drawn perpendicular to the intersegment edges. The length of the line segment is proportional to the edge height; the line is drawn on the lower mirror segment. The ring-mode component here corresponds to a rms edge height of 273 nm. Note that the outer mirror segments are systematically low with respect to the inner segments.

In narrow-band phasing, the signal is the phase ϕ, which is proportional to the edge step height δ:

$$\phi = 2\frac{\delta}{\lambda}\ \text{cycles}. \tag{22}$$

For broadband phasing, the signal is the phase dispersion with wavelength:

$$\Delta\phi = -\frac{2\delta}{\lambda}\frac{\Delta\lambda}{\lambda}, \tag{23}$$

where $\Delta\lambda$ is the filter bandwidth.

Consider two segments that are in phase and that produce the usual circular Airy diffraction pattern in the focal plane, centered at the origin. As the upper segment is slowly pistoned out of the page, the diffraction pattern will appear to shift down the page, but the shorter wavelength components will shift faster, leading to a dispersed, blue-to-the-bottom subimage. Note that the sign of the effect (blue to the bottom) does not change rapidly with step height: It is constant over a range of order λ.

True (material) dispersion in the prisms—for those prisms whose direction of deviation is perpendicular to the corresponding intersegment edge—will closely mimic the above effect, i.e., it will produce a dispersed subimage with the dispersion either parallel or antiparallel to the above direction of phase dispersion. This is the essence of the ring-mode problem: The broadband phasing algorithm seeks to minimize the overall dispersion without regard to whether it is phase dispersion or prism dispersion (because it cannot tell the difference). The algorithm will therefore find a solution in which the phase dispersion is nonzero (and so the mirror is out of phase) so as to cancel out the permanent prism dispersion.

Tracing through the optical system (and noting that the beams exiting the prisms cross in front of the focal plane), we find that if the outer ring of segments is higher than the inner rings, the phase dispersion will produce a subimage that is bluer on the outer edge and redder on the inner edge. If the outer ring is low, the effect is reversed, and the subimage is bluer on the inner edge and redder on the outer edge. Prism dispersion, on the other hand, will always produce a subimage that is bluer on the outer edge and redder on the inner edge. Thus the broadband phasing algorithm will try to cancel out the prism dispersion with a phase dispersion of the opposite sign; i.e., it will drive the outer ring low, as observed.

To estimate the magnitude of the effect, we note that the subimage location in the focal plane is given by

$$x = f(n - 1)\theta, \tag{24}$$

where f is the focal length of the objective lens (located after the prisms), n is the index of refraction of the prism, and θ is the mechanical angle of the prism (not the deviation angle). This produces a variation of the subimage location with wavelength:

$$\Delta x = f\theta\frac{dn}{d\lambda}\Delta\lambda. \tag{25}$$

The spread in the image that is due to the prism dispersion is small compared with the image width $f\lambda/b$, where $b = 596$ μm is the mask subaperture diameter at the camera (at a magnification of 1/200), not at the primary mirror. To be precise, consider the dimensionless ratio β of image dispersion to image width:

$$\beta = \theta b\frac{dn}{d\lambda}\frac{\Delta\lambda}{\lambda}. \tag{26}$$

The rms prism angle is $\theta = 189$ arc min, the dispersion $dn/d\lambda$ of the fused-silica prisms is $-0.0148/\mu$m, and the fractional bandwidths are in the range of 0.01–0.1; this yields β in the range of 0.005–0.05.

The prism dispersion of the subimage in Eq. (25) can be mistaken for the phase dispersion of Eq. (23). Because the functional forms of the respective diffraction patterns for prism dispersion and for phase dispersion are not the same, one might hope that the broadband algorithm can somehow tell the difference, but that is not the case. Because the prism dispersion is small in the sense $\beta \ll 1$, the precise functional form is immaterial; it is virtually inevitable that prism dispersion will cause a first-order (in β) phase shift. We caution, however, that because β is small it does not necessarily imply that the resultant phase shift is small in the sense that $\delta \ll \lambda$.

Before we calculate the size of the dispersion-induced phase shift, which we do both empirically and numerically, we first characterize it in terms of a universal dimensionless constant, which—to a large extent—is independent of the specific parameters of our particular system. Because $\Delta\phi$ and β are both dimensionless and both proportional to $\Delta\lambda$, we can write

$$\Delta\phi = \kappa\beta, \tag{27}$$

where κ is a numerical factor that we expect (but do not assume) to be of order unity. The apparent rms step height associated with the prism dispersion is then

$$\delta = -\frac{1}{2}\,\kappa b\theta\lambda\,\frac{dn}{d\lambda} = -211\kappa \text{ nm}, \tag{28}$$

where we used the rms prism angle and other values of parameters cited above.

We calculated κ numerically by generating diffraction patterns by means of Eq. (15), but shifting them in the focal plane using Eq. (25), for a distribution of wavelengths with $\Delta\lambda = 100$ nm, $\lambda = 870$ nm, (the phasing-100 values), and other values of the parameters appropriate to our system. This calculation was repeated for ten particular values of the (true) step heights near each of nine nominal values of the step height: ± 0.4, ± 0.3, ± 0.2, ± 0.1, and 0 μm, except that we used 100 particular values of the step height near the nominal value of zero for improved statistics. We then extracted the apparent step height from these artificial images by means of the usual coherence analysis and calculated the difference between the inferred or apparent step height and the true step height; this difference is the dispersion-induced phase shift. We found $\delta_{\text{rms}} = 295 \pm 4$ nm, which is equivalent to $\kappa = 1.40 \pm 0.02$. As expected, this difference is independent of the true phase and (also as expected) κ is of order unity.

We also determined κ empirically by producing a ring mode in the primary mirror with the broadband phasing-100 algorithm and then determining the size of the ring-mode component (parameterized as the rms edge height) by analyzing this configuration using the dispersion-independent narrow-band algorithm. The results of this analysis are shown in Fig. 8. The size of the ring-mode component produced in this way is 273 nm, equivalent to $\kappa = 1.29$, or 8% smaller than the numerically determined value. This 8% (or 22-nm) discrepancy may be due to another small source of dispersion in the system or it may be related to the non-Gaussian bandpass of the real filter; in any case, we use the empirical value for the actual corrections.

We now routinely correct the broadband results for the ring mode using the empirically determined value of the dispersion correction—a fixed bias that is removed from each intersegment edge measurement before the singular value decomposition. The correction is the most uncertain for phasing 30 because the wavelength for this mode is the least well defined, but even for this mode we find (by testing phasing 30 against the dispersion-independent narrow-band phasing) that the ring-mode contribution to the rms piston error is only ± 5 nm (after correction).

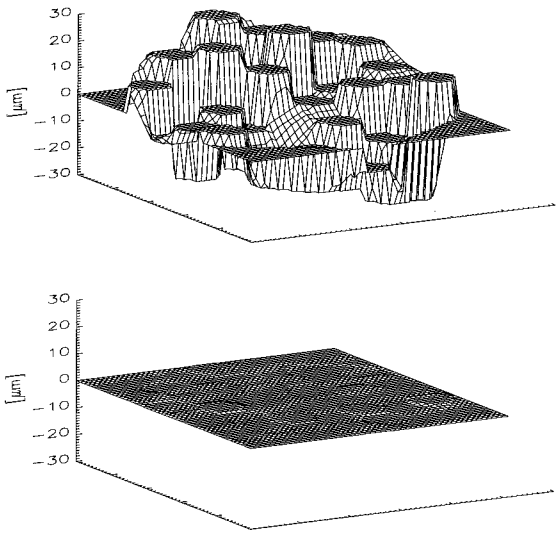

Fig. 9. Surface of the Keck 2 primary mirror before and after phasing. These data represent the first time the procedure was run on this telescope.

8. Performance of Broadband Phasing

We already discussed the average performance of the broadband phasing algorithm when we compared the actual and Monte Carlo results in Section 6. However, it is also instructive to present the unedited results from the first two times that the full phasing procedure was run on Keck 2. The surface of the primary mirror before and after the first complete Keck 2 phasing run is shown in Fig. 9. Details of the first two phasing runs are given in Table 3. Column 1 gives the date. Column 2 gives the mode, i.e., phasing 1000, phasing 300. Column 3 gives the number of edges with adequate signal-to-noise ratio. Column 4 gives the rms piston move executed at the end of a given procedure. Note that the repeatability of the method in a given line is therefore approximately equal to the rms piston move executed in the following line. Column 5 gives the edge residual, i.e., the average (rms) part of the physical step height that cannot be removed by pure pistoning of the segments. Note that executing the first three cycles of the phasing procedure improved the rms piston error by approximately 2 orders of magnitude on both occasions. (The fourth cycle was executed to check the third cycle.) Between the March and June 1996 runs, 21 of the 36 segment figures were adjusted by means of leaf springs or warping harnesses that are attached to the segment support structures. These adjustments typically reduced the segment surface errors from 140 to 40 nm (rms).[19] The improved segment figures manifest themselves in two ways in Table 3. Fewer edges are lost to inadequate signal-

Table 3. Broadband Phasing Performance

Date	Mode	Measured Edges	rms Piston (μm)	Residual (μm)
First Keck 2 attempt				
12 March 1996	Phasing 1000	71	9.469	0.607
14 March 1996	Phasing 300	77	1.384	0.247
14 March 1996	Phasing 100	71	0.147	0.222
14 March 1996	Phasing 100	70	0.140	0.219
Second Keck 2 attempt				
14 June 1996	Phasing 1000	78	7.398	0.795
14 June 1996	Phasing 100	77	0.501	0.109
14 June 1996	Phasing 30	74	0.052	0.102
14 June 1996	Phasing 30	74	0.025	0.099

to-noise ratio, and the edge residual is reduced by a factor of 2, from 220 to 100 nm, in the more accurate procedures. (The residuals are dominated by measurement errors in the less accurate procedures.) Additional adjustment of segments subsequent to June 1996 reduced the Keck 2 edge residual by an additional 30% to 68 nm, or somewhat less than the current 85-nm residual on Keck 1 (where some recent procedural improvements to the segment adjustments have not yet been incorporated).

The fact that the edge residual on Keck 2 went down dramatically as the segment figures were improved constitutes strong, though circumstantial, evidence that the residuals are dominated by segment aberrations. However, is it possible that residual tilt errors left over from segment tip–tilt alignment might be contributing significantly to this residual? To answer this question quantitatively, we examined the (78-element) edge residual vector for a series of 35 phasing-30 runs spanning four months (29 November 1996 to 23 March 1997) on Keck 2, where the segments were restacked between each run. Figure 10 shows the average edge residual versus edge number, where the error bars represent the rms variation on the residual from run to run. The latter is the variable part and is attributable to stacking and measurement errors; the fixed part is attributable to segment aberrations. We expect the fixed and variable parts to be uncorrelated and therefore to add in quadrature. The contribution that is due to stacking and measurement errors is 33 nm, and the contribution that is due to segment aberrations is 59 nm. The latter is clearly dominant. Furthermore, the 59-nm estimate is consistent with the expected segment aberrations: The rms edge height residual should be (and is) within a factor of order unity of the 40-nm segment surface errors. We hope to correlate the segment aberrations with specific edge residuals in the future.

The time required to execute the broadband phasing procedure is dominated by exposure time and CCD readout time. Each exposure is typically 45 s (to average out the effects of atmospheric turbulence), and each CCD readout takes another 11 s. Computation time (for the cross correlation and singular value decomposition) is negligible compared with the exposure and the readout times. The setup time—largely for pupil registration—is approximately 10 min. Once the initial pupil registration is set, the procedure is automatic; the camera automatically adjusts the pointing and keeps the pupil registered without the need for operator intervention. The total time for an 11-exposure cycle is therefore approximately 25–30 min. (This does not count the 15 min required for the prior tip–tilt adjustment of the segments.)

During normal operations the phasing procedure is run approximately once a month. Phasing 1000 is typically used only from a cold start—when segments are installed or exchanged. Phasing 300 is rarely used because the capture range of phasing 100 is comfortably larger than the accuracy of phasing 1000. If no segments are exchanged, the 3-μm cap-

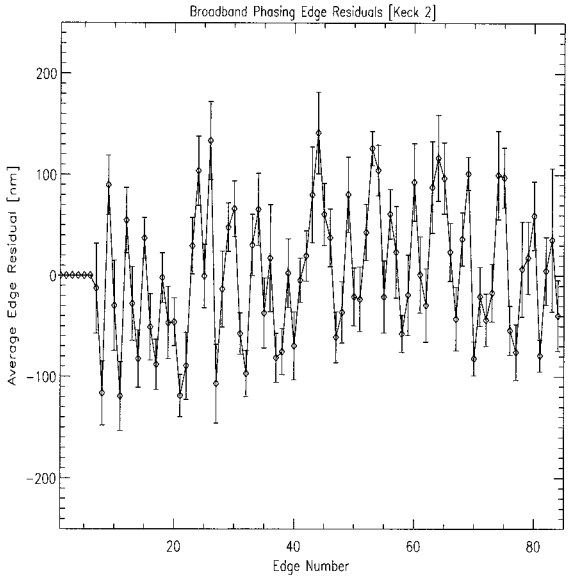

Fig. 10. Residual edge heights (after phasing) for the 78 intersegment edges for a series of 35 phasing-30 runs on Keck 2. Error bars represent the rms variation from run to run (not the error on the mean). The fact that the variations are relatively small shows that random errors in tip–tilt alignment are not the major contributor to the residuals, which are believed to arise from aberrations in the segments themselves.

ture range of phasing 100 is more than sufficient, given the drift rate of the telescope ACS. Phasing 30 is then used to reduce the piston errors further.

We are currently studying the stability of the ACS in the context of phasing. Thus far the data are quite limited, but for a few tests that were on time scales of 10–30 days, during which there were no segment exchanges or other disruptions of the ACS, the required piston corrections were less than 80 nm. Therefore it is likely that in the future even the 1-μm capture range of phasing 30 will prove sufficient.

9. Conclusions

The broadband phasing algorithm repeatedly and reliably puts the telescope into the same configuration with regard to segment phase to within 100 nm or less. The evidence for this is the following:

- As summarized in Tables 2 and 3, the algorithm converges rapidly, with an ultimate repeatability of approximately 30 nm. Indeed, after some initial development work on Keck 1, the algorithm has never failed to converge on Keck 2.
- On a few tests spanning intervals from 10 to 30 days, the phasing solution was shown to be stable at a level of 80 nm or more.
- The edge residuals are stable at a level of 33 nm over time scales of many months (see Fig. 10).

We believe further that the configuration to which we consistently converge is the correct or optimal configuration with regard to phase (again to within the same 100 nm). The evidence for this is the following:

- Detailed Monte Carlo calculations with the broadband algorithm consistently converge to the correct solution, as discussed in Section 6. There are a number of successful consistency checks between the Monte Carlo and the actual algorithm.
- The phasing solution is essentially independent of whether we use empirical or theoretical templates.
- The broadband and narrow-band phasing solutions are consistent. In June 1996 we ran (in order) narrow-band phasing, phasing 30, narrow-band phasing, phasing 30 and narrow-band phasing, and we did not send phasing commands between the five runs. The rms piston difference between the three narrow-band phasing runs and the two phasing-30 runs was 53 nm, or only a little larger than the 43 nm expected on the basis of the quadrature sum of the narrow-band phasing repeatability (33 nm) and the phasing-30 repeatability (28 nm). This is true despite the fact that the two algorithms exploit substantially different diffraction effects and largely independent software.

As noted in the discussion of the singular value decomposition in Section 6, the fact that the edge height residual (68–85 nm) is significantly larger than the repeatability of the algorithm (30 nm) means that the edge-height-minimizing configuration into which we put the primary mirror is not exactly the same as the minimum-piston-error condition. The edge height residual itself provides an estimate of the difference between these two configurations. It is not necessarily clear which is the preferred configuration (although we currently do not have a choice). We hope that further reductions in the segment aberrations in the future will render this point moot.

In summary, the Keck phasing camera, when used with the broadband algorithm, can phase the telescope segments to approximately 30-nm rms piston error with light from moderately bright stars ($V = 4$ to 7). The residuals to the fit of approximately 68 nm (Keck 2) to 85 nm (Keck 1) appear to be dominated by aberrations in the segments themselves. The time required to execute the algorithm grows only logarithmically with the initial piston uncertainty. Phase errors of 30 μm can be reduced by 3 orders of magnitude in approximately 2 h; phase errors of 1 μm can be reduced in as little as 45 min. The phasing procedure requires only modest setup by the operator (principally telescope pointing and pupil registration), and then little or no further intervention is necessary. The algorithm is robust and, after considerable development on Keck 1, it worked the first time (and every subsequent time) on Keck 2.

Current phasing performance is sufficient for diffraction-limited imaging in the infrared, but may need to be improved for speckle and adaptive optics applications.

Appendix A: Resolution of the Absolute Phase Ambiguity in Narrow-Band Phasing

Here we deal with the issue of the half-wave ambiguities associated with narrow-band phasing (a problem that does not arise in broadband phasing). These ambiguities can be eliminated when the narrow-band measurements are repeated in two narrow filters of different wavelengths. The two resulting phases are then reduced to a single step height by a straightforward χ^2 minimization. To be specific, we define $\chi^2(\delta)$, where δ is the step height, by

$$\chi^2(\delta) = \sum_i \frac{\left[\mathrm{frac}\left(\frac{2\delta}{\lambda_i}\right) - \phi_i + m_i\right]^2}{{\sigma_i}^2}. \qquad \text{(A1)}$$

Here, λ_i is the wavelength of the ith filter, frac represents the fractional part of its argument, ϕ_i is the phase in cycles measured in the ith filter ($0 \le \phi_i \le 1$), $m_i = -1$, 0, or $+1$ (chosen so as to minimize χ^2), and σ_i is the uncertainty associated with ϕ_i. The integer m_i is needed to deal with the wrap around of the phase. For example, if $\mathrm{frac}(2\delta/\lambda_i) = 0.01$ and $\phi_i = 0.98$, then $m_i = +1$. (Although we use only two filters, the above equation is valid for larger numbers of filters.) The best-fit value of the step height is the one that minimizes χ^2. (See Press *et al.*[16] for a discussion of parameter estimation by χ^2 minimization, including error analysis and sample routines.) How

to choose these two wavelengths is a nontrivial optimization problem, but a full discussion would take us too far afield here. For our purposes we note that if the wavelengths are too close together, then occasionally the χ^2 minimizing solution will differ from the true value by half of the mean wavelength $(\lambda_1 + \lambda_2)/4$, as is clear from a consideration of the limiting case $\lambda_1 \to \lambda_2$. If the wavelengths are too far apart, occasionally the χ^2 minimizing solution will be in error by

$$\lambda_{12} = \frac{\lambda_1 \lambda_2}{2|\lambda_1 - \lambda_2|}. \tag{A2}$$

To see this, suppose that we have a measure of the phase ϕ_1 and ϕ_2 in each of two different filters, of wavelength λ_1 and λ_2, where we take $\lambda_1 \geq \lambda_2$, and the phase is measured in cycles as above. The phases are related to the step height δ by

$$\delta = (n_1 + \phi_1)\lambda_1/2, \tag{A3}$$

$$\delta = (n_2 + \phi_2)\lambda_2/2, \tag{A4}$$

where the integers n_1 and n_2 are unknown. Solving for ϕ_1 and ϕ_2 and subtracting, we find

$$\delta = \lambda_{12}(\phi_2 - \phi_1 + n_2 - n_1) \tag{A5}$$

with λ_{12} as in Eq. (A2). Thus the step height measurement will be uncertain by an integer multiple of λ_{12}. If λ_1 and λ_2 are sufficiently close together, λ_{12} will be larger than the initial uncertainty in the step height and the ambiguity can be resolved; clearly the problems arise at the other extreme when the two wavelengths are far apart.

Note that if the problem is the former one—errors of half of the mean wavelength—the addition of another measurement in a filter with a wavelength in between the first two will generally not be of much help, but such additional measurements will help the latter problem. Unfortunately, the wavelengths are generally forced closer together than one might hope by the limited bandwidth of the detector or by other optical design considerations. In our case the wavelengths are 651 and 852 nm, and the former effect (errors of 376 nm) tend to be more of a problem than the latter (errors of 1380 nm). In a typical run, approximately two of the 78 edge solutions converge to the wrong value. We suspect that these errors may be triggered when the relative positions of the segments are not held absolutely stable between measurements in the two filters. This is the limiting factor for the convergence of the narrow-band method: The typical rms piston changes by approximately 40 nm between successive runs. This is consistent with Monte Carlo simulations having two errors of half of the mean wavelength or 376 nm.

References and Notes

1. G. A. Chanan, J. E. Nelson, and T. Mast, "Segment alignment for the Keck telescope primary mirror," in *Advanced Technology Optical Telescopes III*, L. D. Barr, ed., Proc. SPIE **628,** 466–470 (1986).
2. G. Chanan, J. Nelson, T. Mast, P. Wizinowich, and B. Schaefer, "The W. M. Keck telescope phasing camera system," in *Instrumentation in Astronomy VIII*, D. L. Crawford and E. R. Craine, eds., Proc. SPIE **2198,** 1139–1150 (1994).
3. G. Chanan, M. Troy, F. Dekens, E. Sirko, and E. Leffert, "Residual misalignments of the Keck telescope primary mirror segments," submitted to SPIE Symposium on Astronomical Telescopes and Instrumentation (1998).
4. J. E. Nelson and P. R. Gillingham, "Overview of the performance of the W. M. Keck Observatory," in *Advanced Technology Optical Telescopes V*, L. M. Stepp, ed., Proc. SPIE **2199,** 82–93 (1994).
5. D. L. Fried, "Optical resolution through a randomly inhomogeneous medium for very long and very short exposures," J. Opt. Soc. Am. **56,** 1372–1379 (1966).
6. G. Chanan, G. Djorgovski, A. Gleckler, S. Kulkarni, T. Mast, C. Max, J. Nelson, and P. Wizinowich, eds., "Adaptive optics for Keck Observatory," Keck Observatory Report 208 (W. M. Keck Library, Kamuela Hawaii, 1996), pp. 3-26 to 3-36.
7. J. Nelson, "Segment mechanical phasing," Keck Observatory Tech. Note 389 (W. M. Keck Library, Kamuela Hawaii, July 1993).
8. J. E. Nelson, T. S. Mast, and S. M. Faber, "The design of the Keck Observatory and telescope," Keck Observatory Report 90 (W. M. Keck Library, Kamuela Hawaii, 1985), pp. 5-1 to 5-44.
9. R. W. Cohen, T. S. Mast, and J. E. Nelson, "Performance of the W. M. Keck telescope active mirror control system," in *Advanced Technology Optical Telescopes V*, L. M. Stepp, ed., Proc. SPIE **2199,** 105–116 (1994).
10. G. Chanan, T. Mast, and J. Nelson, "Keck telescope primary mirror segments: initial alignment and active control," in *Proceedings of the European Southern Observatory Conference on Very Large Telescopes and their Instrumentation*, M.-H. Ulrich, ed. (Garching, Germany, 1988), pp. 421–428.
11. G. A. Chanan, "Design of the Keck Observatory alignment camera," in *Precision Instrument Design*, T. C. Bristow and A. E. Hatheway, eds., Proc. SPIE **1036,** 59–70 (1988).
12. G. A. Chanan, "Mask/pupil registration for the Keck telescope," Keck Observatory Tech. Note 253 (W. M. Keck Library, Kamuela Hawaii, 1988).
13. M. Born and E. Wolf, *Principles of Optics*, 6th ed. (Pergamon, New York, 1969), pp. 395–398.
14. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes: the Art of Scientific Computing* (Cambridge U. Press, New York, 1989), pp. 484–487.
15. T. Mast, J. Nelson, and G. Chanan, "Sampling segment phases: minimal spanning sets of zero-sum triplets," Keck Observatory Tech. Note 208 (W. M. Keck Library, Kamuela Hawaii, 1986).
16. See Ref. 14, pp. 502–515.
17. T. Mast, "Systematic errors in the alignment camera phase measurement," Keck Observatory Tech. Note 299 (W. M. Keck Library, Kamuela Hawaii, 1990).
18. See Ref. 14, pp. 52–64, 515–520.
19. P. L. Wizinowich, T. S. Mast, J. E. Nelson, M. DiVittorio, and G. A. Chanan, "Optical quality of the W. M. Keck telescope," in *Advanced Technology Optical Telescopes V*, L. M. Stepp, ed., Proc. SPIE **2199,** 94–104 (1994).