Comparison of Phase Diversity and Curvature Wavefront Sensing

J.R. Fienup, B.J. Thelen, R.G. Paxman & D.A. Carrara

ERIM International
P.O. Box 134008
Ann Arbor, MI 48113-4008

ABSTRACT

We compare phase diversity and curvature wavefront sensing. Besides having completely different reconstruction algorithms, the two methods measure data in different domains: phase diversity very near to the focal plane, and curvature wavefront sensing far from the focal plane in quasi-pupil planes, which enable real-time computation of the wavefront using analog techniques. By using information-theoretic lower bounds, we show that the price of measuring far from the focal plane is an increased error in estimating the phase. Curvature wavefront sensing is currently operating in the field, but phase diversity should produce superior estimates as real-time computing develops.

Keywords: wavefront sensing, aberrations, phase diversity, adaptive optics

1 INTRODUCTION

Wavefront sensors are essential for obtaining aberration information in order for adaptive optics to correct the aberrations of large telescopes, allowing them to obtain diffraction-limited resolution. The source of the aberrations can be atmospheric turbulence, piston and tilt errors of segmented optics, or mirror-figure error. The aberrations can have time constants of milliseconds, or be fixed in time, or anything in-between. The most commonly used wavefront sensors, Shack-Hartmann and shearing interferometers, employ additional optical hardware that makes direct measurements of wavefront slopes, or spatial derivatives of the phase. Two wavefront sensors more recently developed, phase diversity[1-4] and curvature wavefront sensing,[5-8] make measurements of the light in planes that naturally occur in the optical system, but require sophisticated algorithms to extract an estimate of the aberrations. In its most common configuration, phase diversity (PD) measures images in two planes, one in the nominal focal plane of the telescope and a second slightly defocused plane (typically about one wave of defocus). Alternatively, both images can be in slightly different defocus planes. Curvature wavefront sensing (CWFS) measures two images that are highly out of focus in two planes that are symmetric about the focal plane. Because of the similarity in the measurements made, some members of the optics community have mistakenly assumed that the two approaches are essentially the same. The purposes of this paper are (1) to show that PD and CWFS are indeed two very different wavefront sensors, (2) to compare their performance, and (3) to clarify their respective operating regimes. We will discuss their different measurement regimes and their different reconstruction algorithms in Sections 2 and 3. By using information-theoretic lower bounds, we show the penalty that CWFS must pay for having to operate farther from the focus, in Section 4.

Email: fienup@erim-int.com; Telephone: (734)994-1200 ext. 2500; Fax: (734)994-5704

2 COMPARISON

For this investigation, we are limiting the comparison to a specific application of interest to astronomy: ground-based imaging of unknown and possibly extended objects, using adaptive optics to correct turbulence-induced aberrations. Within this scope, both PD and CWFS have some features in common: (1) they both collect a pair of intensity images at different locations along the optical axis, and (2) in both cases an algorithm is used to estimate a wavefront from the intensity data. In spite of these similarities, PD and CWFS are very different wavefront-sensing modalities. The most striking distinction between the two wavefront sensors is the difference in locations at which the data are collected. PD typically collects data in the focal or caustic region whereas CWFS requires intensity measurements in pupil-like planes. We quantify these regimes of data collection in Section 3. Another difference is that PD requires no *a priori* knowledge of the object and works well with extended objects, whereas CWFS requires that the object is known or is small in extent. The algorithms for the two methods are also quite distinct. Our PD wavefront estimates are found using optimal estimators such as maximum-likelihood or MAP algorithms.[3,9,10] In practice, CWFS uses a Poisson equation solver that can be implemented with bimorph mirrors in a real-time analog processor.[5] We note that estimation-theoretic methods (like those we use with PD) could also be applied to CWFS data. Another difference is that PD provides the option of a joint estimation of the object and the wavefront whereas CWFS yields estimates for the wavefront alone. Finally, although PD is quite mature in the context of image reconstruction, it is less mature than CWFS in its role as a wavefront sensor. Whereas PD has been demonstrated in closed-loop operation in the laboratory[11] and non-real time operation in the field, CWFS is used routinely in an operational telescope system. These distinctions between PD and CWFS are summarized in Table 1.

Table 1: Differences between Curvature and Phase-Diversity Wavefront Sensing.

Feature	Curvature WFS	Phase Diversity
location of data	pupil-plane region, far from focal plane, symmetric about focal plane	focal-plane region, can be asymmetric
object knowledge	assumed known or is small	unknown
object extent	limitations on object extent	no limitations
optical bandwidth	broad-band	broad-band not mature
algorithm	Poisson equation solver, analog computation	nonlinear optimization, digital computation
products	wavefront only	wavefront and reconstructed imagery
maturity of WFS	operational for small sources	closed-loop lab demonstration, non-real time in field

3 DISTANCE FROM FOCAL PLANE

As the distance from the focal plane is a major difference between the two wavefront sensors, we quantify that difference. For PD, typically one image is taken in the nominal focal plane, and the second is taken in a defocus plane, having a quadratic defocus of about $Q = 1$ wave (peak-to-valley), give or take a factor of two. This corresponds to a distance from the focal plane of

$$L = 8Q\lambda(f/D)^2 ,$$ (1)

where L is the extra-focal distance, λ is the center wavelength, f is the effective focal length, and D is the diameter of the telescope aperture. For example, for $D = 4$ m, $f = 120$ m, $\lambda = 0.5$ μm, and $Q = 1$ wave, the extra-focal distance would be $L = 0.36$ cm. Measuring several times farther from focus than this would be detrimental to PD because as Q (or L) increases, the MTF becomes smaller and develops more zeros at the higher spatial frequencies, thereby reducing the contrast of the measured imagery, and decreasing the signal-to-noise ratio.

As described in Ref. 5, one must stay away from the caustic zone where the transport equations assumed by CWFS do not apply, and one must stay far enough away from the focal plane to avoid having the blurred point-spread function smear the wavefront information contained in the intensity patterns which are demagnified images of the pupil. To satisfy these conditions, for the case of aberrations have a Kolmogorov spectrum with correlation distance (Fried parameter) r_o, the extra-focal distance L must satisfy

$$\frac{r_o L}{f} >> \frac{\lambda(f - L)}{r_o} .$$ (2)

In practice $L << f$, and this inequality is approximated well by

$$L >> \frac{\lambda f^2}{r_o^2} .$$ (3)

For example, using the parameters listed above and $r_o = 10$ cm, we have $L >> 72$ cm (or $>>200$ waves of quadratic phase), which is $>>200$ times greater than for PD.

Although CWFS measures in a plane that is physically closer to the focal plane than it is to the pupil plane ($L << f$), the pattern at that distance L has the appearance of a demagnified pupil and is more properly thought of as a pupil-like plane than as an image-like plane.

For CWFS, the extent of the object places further restrictions on L. Consider the version of CWFS currently in use.[5] Let I_1 be the intensity distribution a distance L before the focal plane and I_2 be the 180°-rotated intensity distribution a distance L after the focal plane. In CWFS one computes[5]

$$\frac{I_2 - I_1}{I_2 + I_1} = \frac{f(f - L)}{2L} \left[\frac{\partial W}{\partial n} + P \nabla^2 W \right] ,$$ (4)

where W is the optical path difference, P is the pupil function, ∇^2 is the Laplacian operator, and $\partial W/\partial n$ is the derivative normal to the boundary. One can solve Eq. (4) for W using bimorph mirrors in a real-time analog processor.[5] Figure 1 shows an example. Left-to-right and top-to-bottom, the figure first shows the aperture and the aberration. We chose a small aberration with a Gaussian blob of aberration of amplitude 0.25 waves peak-to-valley in order to illustrate the strength of the CWFS idea. The measurement planes are at $Q = \pm10$ waves of defocus, which is sufficient to show the desired effect for these small aberrations. Next are shown I_1, the pre-flipped I_2, and $I_1 - I_2$, which clearly show the location of the Gaussian blob, because of the Laplacian term in the equation above, demonstrating the principle of the CWFS. Next is shown a small extended object and the same three intensity quantities when convolved with the object. Although the fine fringe structure in the

monochromatic PSF's is washed out, the bright blob at the location of the phase perturbation remains, enabling CWFS to determine the aberration.
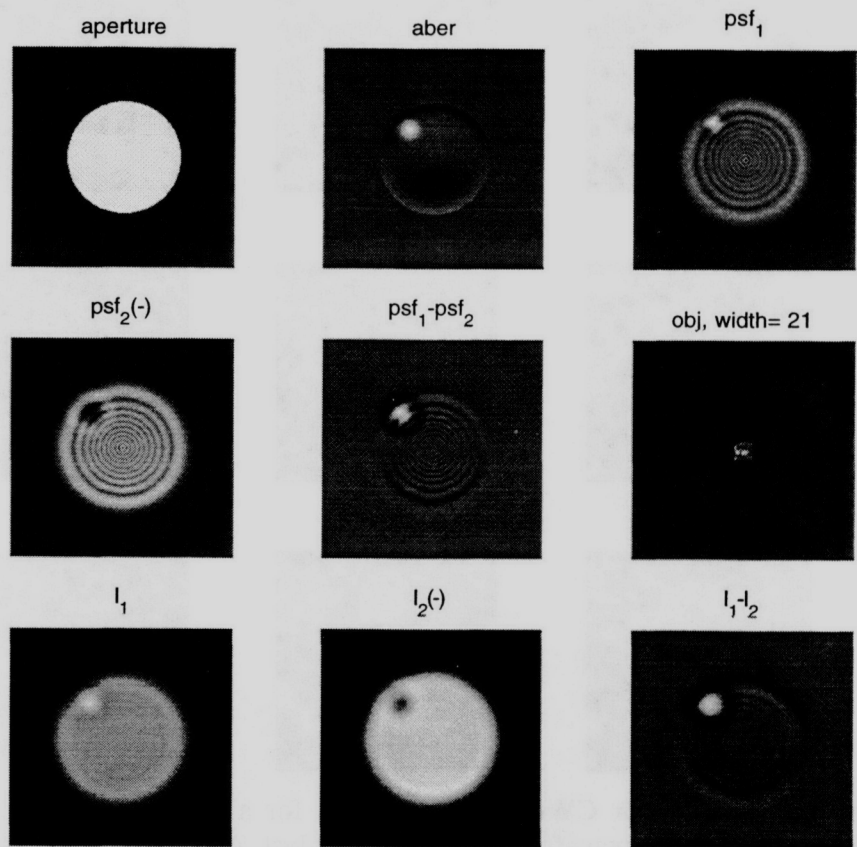

Figure 1. Example quantities for CWFS, showing that the signal does not change much for a small extended object (see text). Array size = 256 pixels.

Figure 2 shows the same thing for a much more extended object. The intensity difference shown in the lower right is dominated not by the aberrations, but by the pattern of the object, destroying the ability of CWFS to reconstruct the aberrations from this information.

In order for an extended object to not destroy the CWFS information, as happened in Figure 2, its image must be much smaller than the scale size of the aberrations in that demagnified pupil plane. That is, if the width of the image of the object is w, then we need

$$w \ll r_o \frac{L}{f} \quad , \quad \text{or} \quad L \gg \frac{wf}{r_o} \quad . \tag{5}$$

For example, for the parameter example discussed earlier, and $w = 40 \, \lambda f/D$ (i.e., 40 resolution elements wide), we need $L \gg 72$ cm, the same distance as required by the aberrations. For a large scene with $L = 1000$ resolution elements, we need $L \gg 18$ m (equivalent to $Q \gg 5000$ waves). For a scene that completely fills the field of view (imaging solar granulation, for example), there does not exist an extra-focal distance large enough for this version of CWFS to work.
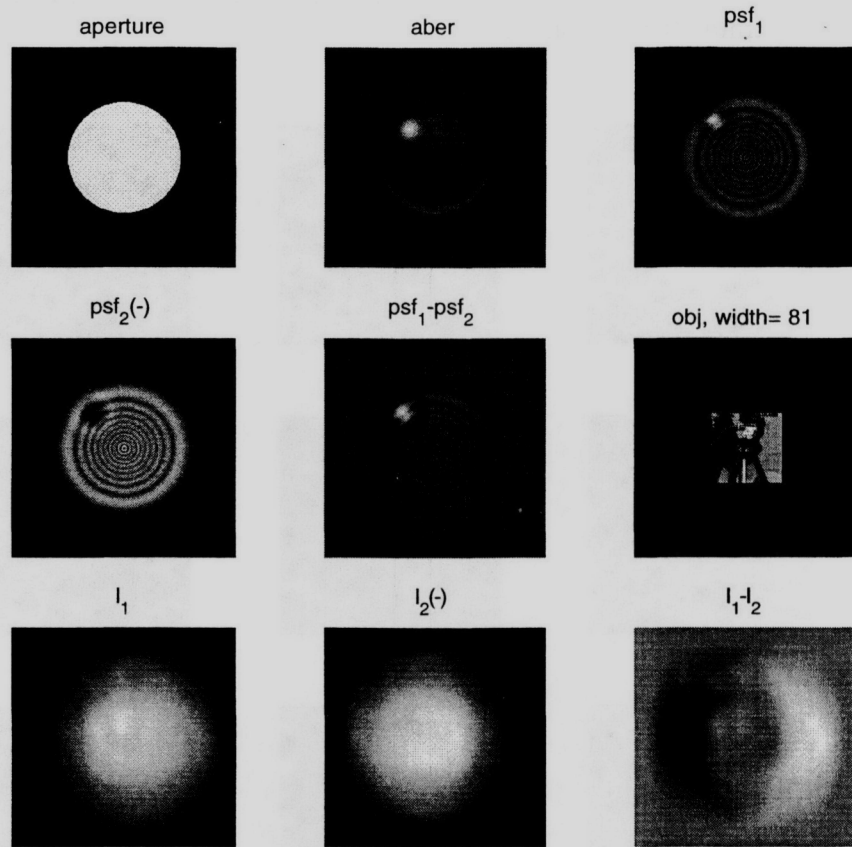
Figure 2. Example quantities for CWFS, showing that for a large extended object the signal is dominated by the object, and the wavefront information is lost. Array size = 256 pixels.

From the analysis above, we see that for imaging with large aberrations or of large scenes, the original CWFS requires very large extra-focal distances.

Since the original CWFS[5] does not work well for very extended objects, a second version was invented.[7] This approach involves two terms, both of them using I_{ft1} and I_{ft2}, the Fourier transforms of I_1 and I_2: one for symmetric aberrations and another for asymmetric aberrations. The latter is[7]

$$S_o(f) = \frac{I_{ft1}I_{ft2}{}^*}{|I_{ft1}I_{ft2}{}^*|} = \exp[i(2\theta + \phi_1 - \phi_2)] \ , \tag{6}$$

where θ is the phase of the Fourier transform of the object and ϕ_i is the phase of the ith OTF (the Fourier transform of the ith PSF). One must first obtain the object's 2θ, then subtract it from the equation above, to obtain $(\phi_1 - \phi_2)$, which is a necessary component to reconstructing the wavefront.[7] One might obtain 2θ by imaging the object through many different realizations of the aberrations, with different values of $(\phi_1 - \phi_2)$, and hope that the $(\phi_1 - \phi_2)$ contribution averages to zero. This would work only if $(\phi_1 - \phi_2) << \pi$.[7] We performed digital simulation experiments to shed light on this requirement, and Figure 3 shows a result, when we assumed the object to be an unresolved point (and so its image is the PSF). Shown are the aperture, the aberration (a Kolmogorov phase error with

$D/r_o = 5$), I_1, I_2, the first OTF magnitude (the MTF, $|I_{ft1}|$), the first OTF phase (ϕ_1), a horizontal cut through ϕ_1 (from d.c. to the edge), the OTF phase difference ($\phi_1 - \phi_2$), and a horizontal cut through ($\phi_1 - \phi_2$). Figure 4 shows the same cut through ($\phi_1 - \phi_2$) for increasing extra-focal distances (± 2, ± 10, and ± 20 waves of defocus). For this range of extra-focal distances, ($\phi_1 - \phi_2$) $<< \pi$ only for the very lowest spatial frequencies, that is, only for the very lowest-order phase errors, and it does not appear to improve with increasing defocus. This does not rule out the possibility that this requirement is met for larger spatial frequencies at extremely large extra-focal distances. Previous simulations[7] showed good results for very low-order phase errors (less than a dozen Zernike coefficients). A more recent variation involving four image planes also has been shown to be effective only for very low-order aberrations.[8]
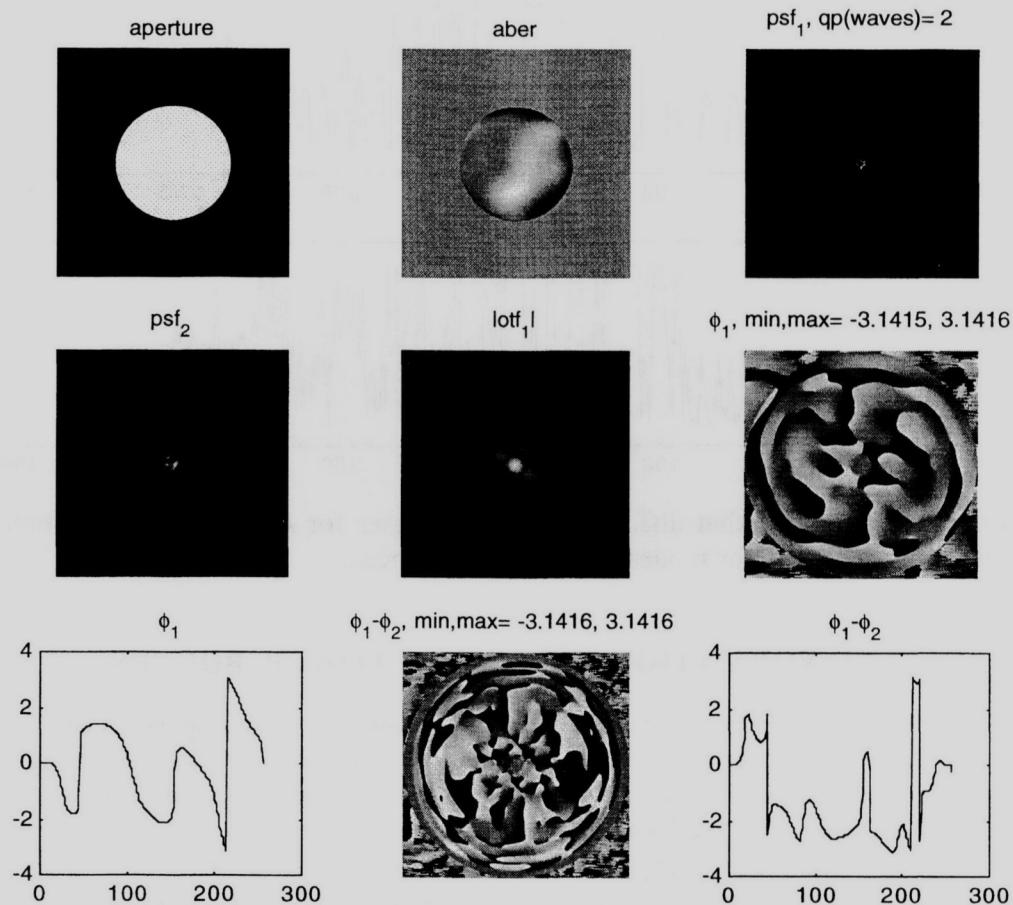

Figure 3. Simulation showing that difference in OTF phases for I_1 and I_2 are not small. Array size = 512 pixels.

In summary, again we see that for imaging with large aberrations or of large scenes, the more recent versions of CWFS require very large extra-focal distances; in Section 4 we quantify the large penalty for measuring so far from focus.
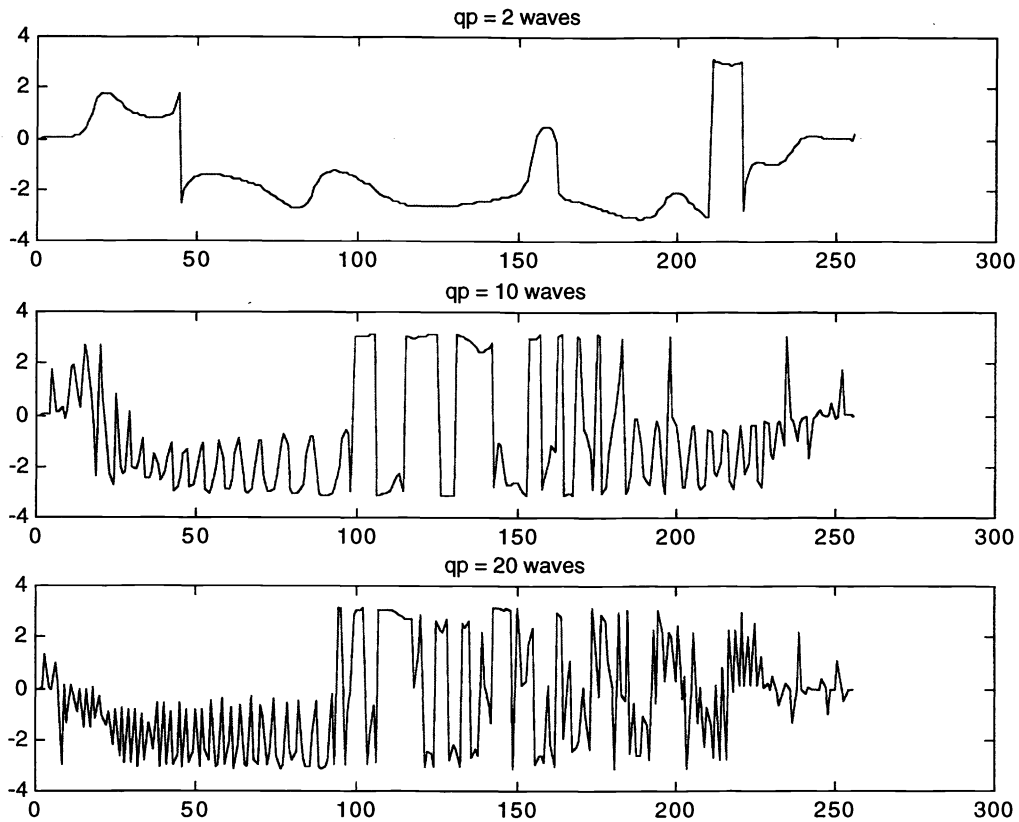
Figure 4. Simulation showing that difference in OTF phases for I_1 and I_2 are not small, even as the extra-focal distance increases, for modest extra-focal distances.

4 INFORMATION-THEORETIC LOWER BOUNDS

Information-theoretic lower bounds provide a natural means for the comparison of PD with CWFS. As background, there is a whole segment of information theory devoted to computing lower bounds on the achievable mean-squared error for general classes of estimators, based on a statistical model for the collected data. The most well-known of these is the Cramer-Rao bound (CRB) which has been used extensively in statistics and signal processing for analyzing the performance of proposed estimators as well as assessing proposed data-collection schemes. By computing CRBs for both PD and CWFS data sets, we can compare these two techniques based on the fundamental content of the data. It should be noted that we are only quantifying the *potential* wavefront estimation accuracy, and that any particular wavefront estimator could be suboptimal and, therefore, fail to achieve the performance bound. Related works on computing CRBs for wavefront estimation include phase retrieval from defocused imagery[12] and averaged CRBs for phase-diverse phase retrieval.[13]

The basic framework for CRBs is a statistical model or probability density function $p(X,\theta)$, where X is a random vector corresponding to the (collected) data and θ is the unknown parameter vector to be estimated. The parameter vector θ can be modeled as either deterministic or random. The

computation of CRBs starts with computing the Fisher information matrix J which is a square matrix with dimension equaling that of the parameter vector θ and is defined by

$$J_{jj'} = E\left(\frac{\partial \log(p(X,\theta))}{\partial \theta_j} \frac{\partial \log(p(X,\theta))}{\partial \theta_{j'}}\right) , \qquad (7)$$

where θ_j is the jth element of θ. The CRB is then specified by the matrix inequality

$$E\left[\left(\widehat{\theta}(X) - \theta\right)\left(\widehat{\theta}(X) - \theta\right)'\right] \geq J^{-1} , \qquad (8)$$

where the prime denotes transpose, $\widehat{\theta}(X)$ is an estimator, and for two square matrices A and B, the inequality $A \geq B$ implies that $A - B$ is a nonnegative definite matrix. The matrix inequality in (8) implies that a lower bound for the (generalized) MSE matrix is given by the inverse of the Fisher information matrix. In particular, this implies that estimates of each of the components of θ have a (scalar) MSE bounded below by the corresponding diagonal element of the inverse Fisher information matrix. Summing this implies the lower bound inequality of

$$E\left[\sum_j \left(\widehat{\theta}_j(X) - \theta\right)^2\right] \geq trace\left(J^{-1}\right) . \qquad (9)$$

There are two different scenarios of interest depending on whether the parameter vector θ is modeled as deterministic with no *a priori* information or is modeled as a random vector with a known probability density function. In the deterministic case the CRB is a lower bound for all unbiased estimators, i.e., estimators $\widehat{\theta}(X)$ which satisfy that

$$E\left(\widehat{\theta}(X)\right) = \theta \quad \text{for all } \theta .$$

In the deterministic case, the CRB is also an approximate lower bound for essentially all "reasonable" estimators when the signal-to-noise ratio is high. In the case of a random parameter vector, the lower bound is referred to as the Bayes CRB and in this case, the bound applies to *all* estimators. It is somewhat intuitive that the (deterministic) CRB will be larger than the Bayes CRB since the model for the latter assumes some additional *a priori* information thereby enabling a more accurate estimation. We will see this intuition confirmed in the CRB results discussed in the next paragraph.

As discussed earlier, one can compare the information content of PD data and data collected for CWFS by computing and comparing CRBs on the MSE of the associated wavefront estimators. Table 2 gives the background assumptions for our CRB results.

The PD imaging model is outlined in Ref. 3. By simply increasing the magnitude of the quadratic defocus term, this imaging model applies to the CWFS scenario as well. The basic assumption is that two images are collected, with symmetric levels of defocus corresponding to $\pm Q$ waves. Small defocus levels (0-3 waves) corresponds to PD while larger defocus levels ($>>10$ waves) corresponds to data that is collected for CWFS. We modeled the unknown aberrations as a linear combination of 52 Zernike basis vectors starting with Z_4, the quadratic, which we scaled to be approximately orthonormal. We modeled the coefficients as being from a Gaussian statistical model corresponding to

Table 2. List of assumptions for the CRBs comparing the performance of PD and CWFS.

| Object	Point
Signal level	10^5 photoelectrons (total both channels)
Read noise level	5 photoelectrons (Gaussian)
Detector sampling	Nyquist
FOV	512 by 512 pixels
Defocus	$\pm(0$ to 200) waves
D/r_o	5
Aberration basis set	Zernikes Z_4-Z_{55}
Aberration type	Kolmogorov

Kolmogorov turbulence with $D/r_o = 5$. The unknown wavefront parameters are the 52 Zernike coefficients, and it is for this that we compute the CRB matrix, assuming a Gaussian model for our data.[14] Invoking the inequality in (9) and taking the square root of the trace of the inverse Fisher information matrix, we have a lower bound on the root MSE (RMSE) of the total aberration estimate. Figure 5(a) shows plots of the deterministic CRB for the RMSE as a function of defocus for each of 10 random aberration realizations assuming that the object was known. Here the x-axis is the level of defocus (the extra-focal distance, in waves of quadratic) and the y-axis is the RMSE CRB on a logarithmic scale. Figure 5(b) is a magnification of the plots in the lower defocus regime. In the same figure, we have also shown the corresponding Bayes CRB, assuming the object is known, which takes into account that the aberration satisfies the Kolmogorov Gaussian model. As expected (as described in the previous paragraph), the Bayes CRB is significantly lower than any of the 10 deterministic CRBs. The general shapes of all of the CRBs are similar. They all have an approximate minimum somewhere in the vicinity of ±1 wave, and all the bounds increase dramatically as the defocus increases from 5 to 200 waves, which is the regime that curvature wavefront sensing would operate. On the basis of these CRB results for the scenario outlined in Table 2, it seems clear that the information content of the data is much higher at lower levels of defocus and that as one moves out past 5-10 waves of defocus, the achievable accuracy of wavefront estimation deteriorates dramatically.

5 CONCLUSIONS

We have demonstrated that phase diversity (PD) and curvature wavefront sensing (CWFS) are very different approaches to wavefront sensing. CWFS makes measurements at extra-focal distances that must be many times greater than that which PD uses. The advantage that this affords CWFS is that the wavefront aberrations are more directly related to the measurements, making it possible to compute the wavefront in real time using bimorph mirrors, which is currently operational. In contrast, near to the focal plane, where PD takes its measurements, the relationship between the aberrations and the measured data is extremely nonlinear, requiring a computationally more demanding nonlinear optimization calculation to determine the aberrations. Owing to improvements in the computational implementation of the PD algorithm and rapid advances in computer speed, closed-loop operation is now possible and will become increasingly less expensive. As shown by the Cramer-Rao lower bounds, the biggest advantages of PD are that its measurements inherently possess much more infor-
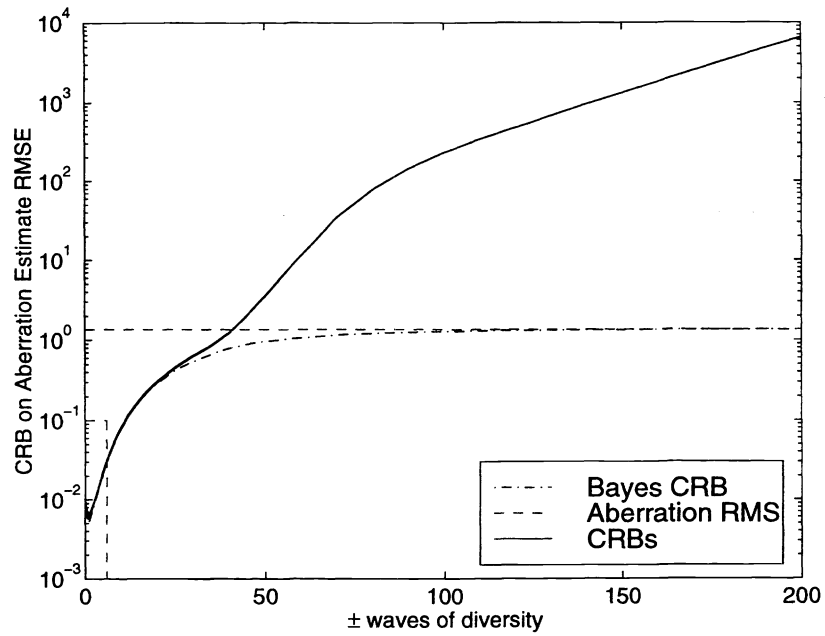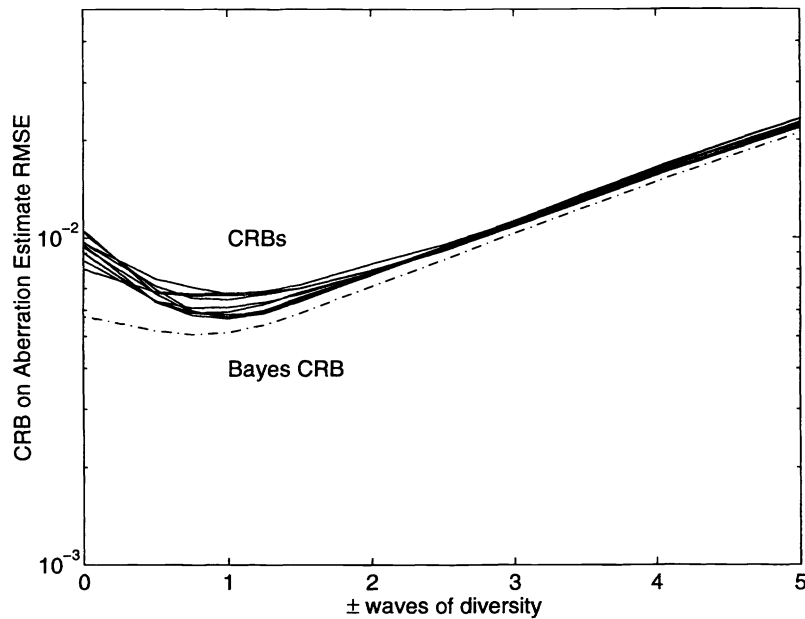
Figure 5. Plots of CRB for aberration RMSE as a function of defocus (extra-focal distance).
(a) The dashed-dotted line is the Bayes CRB, the top set of curves corresponds to CRBs for
10 random aberrations, and the actual root mean-squared (RMS) value of the aberrations
is plotted in the dashed line. (b) Magnification of plots in (a), where the solid lines are the
(deterministic) CRBs for the 10 random aberrations and the dashed-dotted line is the Bayes
CRB that assumes the statistical model for the aberrations.

mation about the aberrations, it can produce aberration estimates with considerably less error, and it can work effectively at substantially lower light levels than CWFS. This translates to increased sky coverage.

ACKNOWLEDGEMENTS

REFERENCES

1. R.A. Gonsalves, "Phase Retrieval and Diversity in Adaptive Optics," Opt. Eng. 21, 829- 832 (1982).

2. R.G. Paxman and J.R. Fienup, "Optical Misalignment Sensing and Image Reconstruction Using Phase Diversity," J. Opt. Soc. Am. A 5, 914-923 (1988).

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

4. R.G. Paxman, J.H. Seldin, M.G. Löfdahl, G.B. Scharmer, and C.U. Keller, "Evaluation of Phase-Diversity Techniques for Solar-Image Restoration," Ap. J. 466, 1087-1099 (1996).

5. F. Roddier, "Curvature Sensing and Compensation: a New Concept in Adaptive Optics," Appl. Opt. 27, 1223-1225 (1988).

6. F. Roddier, "Wavefront Sensing and the Irradiance Transport Equation," Appl. Opt. 29, 1402-1403 (1990).

7. R. Kupke, F. Roddier, and D.L. Mickey, "Curvature-based Wavefront Sensor for use on Extended Patterns," in Adaptive Optics in Astronomy, Proc. SPIE 2201, pp. 519-527 (1994).

8. R. Kupke, F.J. Roddier, and D.L. Mickey, "Wavefront Curvature Sensing on Extended Arbitrary Scenes: Simulation Results," in Adaptive Optical Systems Technologies, Proc. SPIE 3353-41 (Kona, HI, 1998).

9. B.J. Thelen, R.G. Paxman, and J.H. Seldin, "Bayes estimation of wavefronts, fixed aberrations, and object from phase-diverse speckle data," in Digital Image Recovery and Synthesis III, Proc. SPIE 2827-03 (August, 1996).

10. B.L. Ellerbroek, B.J. Thelen, D.J. Lee, D.A. Carrara, and R.G. Paxman, "Experimental comparison of Shack-Hartmann wavefront sensing and phase-diverse phase retrieval," in Adaptive Optics and Applications R.K. Tyson and R.Q. Fugate, eds., Proc. SPIE 3126-38 (July, 1997).

11. R.L. Kendrick, R.M. Bell Jr., G.D. Love, D.S. Acton, and A.L. Duncan, "Closed-loop wavefront correction using phase diversity," in Space Telescopes and Instruments V, Proc. SPIE 3356 (Kona, HI, 1998).

12. J. R. Fienup, J. C. Marron, T. J. Schulz, and J. H. Seldin, "Hubble space telescope characterized by using phase-retrieval algorithms," Appl. Opt., 32, 1747-1767 (1993).

13. D.J. Lee, B.M. Welsh, and M.C. Roggemann, "Cramer-Rao Analysis of Phase Diversity Imaging," in Image Reconstruction and Restoration II, Proc. SPIE 3170 (1997).

14. H. L. Van Trees, Detection, Estimation, and Modulation Theory, Part I, Wiley, New York (1968).