LIMITS OF ADAPTIVE OPTICS FOR HIGH-CONTRAST IMAGING

OLIVIER GUYON
Subaru Telescope, 650 North A'ohoku Place, Hilo, HI 96720; guyon@nao.jg
Received 2005 February 23; accepted 2005 April 14

ABSTRACT

The effects of photon noise, aliasing, wave front chromaticity, and scintillation on the point-spread function (PSF) contrast achievable with ground-based adaptive optics (AO) are evaluated for different wave front sensing schemes. I show that a wave front sensor (WFS) based on the Zernike phase contrast technique offers the best sensitivity to photon noise at all spatial frequencies, while the Shack-Hartmann WFS is significantly less sensitive. In AO systems performing wave front sensing in the visible and scientific imaging in the near-IR, the PSF contrast limit is set by the scintillation chromaticity induced by Fresnel propagation through the atmosphere. On an 8 m telescope, the PSF contrast is then limited to $10^{-4}$ to $10^{-5}$ in the central arcsecond. Wave front sensing and scientific imaging should therefore be done at the same wavelength, in which case, on bright sources, PSF contrasts between $10^{-6}$ and $10^{-7}$ can be achieved within 1" on an 8 m telescope in optical/near-IR. The impact of atmospheric turbulence parameters (seeing, wind speed, turbulence profile) on the PSF contrast is quantified. I show that a focal plane wave front sensing scheme offers unique advantages, and I discuss how to implement it. Coronagraphic options are also briefly discussed.

Subject headings: instrumentation: adaptive optics — instrumentation: interferometers — methods: data analysis — techniques: high angular resolution — techniques: interferometric

1. INTRODUCTION

High-contrast imaging of the immediate environment (within a few astronomical units) of nearby stars is critical to the understanding of formation and evolution of planetary systems. The ultimate goal of planetary system studies is to find and characterize planets similar to ours, in the hope that we can find other “habitable” words susceptible of harboring life. Several approaches are currently under development to achieve the high level of contrast required:

1. Nulling interferometry in the mid-IR with a 30 m baseline space interferometer.
2. Visible coronagraphy with a 4–8 m space telescope.
3. Large ground-based telescopes (8–100 m) and high-performance AO systems optimized for bright targets.

While the first two options are targeting Earth-size planets around nearby stars, the ground-based systems have more modest initial goals: planets more massive than Jupiter or young Jupiter mass planets. The plans to build larger (30–100 m) telescopes and the fast progress in high-performance AO systems and coronagraphy, however, open up the possibility of pursuing more ambitious goals. Direct imaging of Earth-size planets is even considered for 100 m diameter telescopes (Gilmozzi 2004; Hawarden et al. 2003).

The contrast detection limit within a PSF is set by photon noise and speckle noise in the image. If only photon noise is considered, the theoretical detection limit for a large telescope (30–100 m) allows relatively easy detection of Earth-size planets around nearby stars (Angel 2003; Hawarden et al. 2003). This, however, seems to be a very optimistic assumption, as current AO systems are all limited by speckle noise (Racine et al. 1999) within the central arcsecond. The detection limit is therefore likely to be driven by how well the speckles can be calibrated/removed. Fast atmospheric speckles average fairly rapidly, but slower evolving speckles are more problematic. Experience acquired on ground-based telescopes has shown that there is no such thing as truly static speckles, as extremely small drifts in the wave front are sufficient to appreciably change the speckle’s intensity: for a $10^{-5}$ speckle (speckle intensity equals $10^{-5}$ of the central star’s intensity) to be stable to within 1%, the corresponding spatial frequency would need to be stable to $2.5 \times 10^{-6}$ wave (4 pm at 1.6 μm).

In noncoronagraphic imaging with current AO systems on large telescopes, the PSF wings are relatively smooth in long exposures: about 10% of the speckle background does not average at about 1” (Fig. 1 in Boccaletti et al. 2003). Through careful PSF calibration, some of the residual speckle structure can be subtracted (Roth et al. 2001), yielding a point-source detection limit more than 10 times fainter than the PSF background level within the central arcsecond. This factor tends to become larger with increasing distance from the PSF core, partially thanks to the chromatic elongation of speckles, which makes the PSF smoother. Differential imaging techniques based on spectral properties of the source could further increase this factor by up to $10^4$ for simultaneous imaging in two bands and $10^5$ for simultaneous imaging in three bands (Marois et al. 2000). Techniques using the coherence properties of speckles have also been proposed (Boccaletti et al. 1998; Guyon 2004).

In this study, performance of an AO system is quantified by the ratio between the light intensity at the point of the PSF considered and the light intensity at the PSF’s center. This quantity is referred to as the PSF contrast in the rest of the paper. The goal of this work is to give limits on this PSF contrast achievable with AO systems and to propose solutions to reach these limits: Which WFS to choose? How to drive the deformable mirror (DM)? Is a coronagraph necessary? If yes, which one? Detection limits for faint companions are significantly harder to predict than PSF contrast for the reasons detailed above and will not be computed in this paper.

The PSF contrast in a photon noise-limited AO system is a function of the ability of the WFS to accurately measure the corresponding spatial frequency in the pupil plane phase. Analytical expressions of the fundamental contrast limits imposed by photon noise and chromaticity of the wave front are derived in § 2. Aliasing effects are discussed in § 3, and solutions to reduce
their impact on the PSF contrast are proposed. In \S\ 4, the sensitivity of common WFSs to photon noise is discussed, and a WFS based on Zernike’s phase contrast is shown to offer optimal sensitivity. Results of \S\S\ 2, 3, and 4 are combined and discussed in \S\ 5 to derive realistic limits to the PSF contrast in ground-based AO systems and identify optimal approaches to detect extrasolar planets. In \S\ 5 it is shown that a focal plane WFS can be especially advantageous, and this option is discussed in more detail in \S\ 6. In \S\ 7 I discuss the need for coronagraphy and the choice of the correct coronagraph to reach the PSF contrast derived in this study.

2. FUNDAMENTAL LIMITS OF WAVE FRONT SENSING FOR HIGH-CONTRAST IMAGING

2.1. Speckles and Wavefronts

Notations used in this work are given in Table 1. In this paper I consider a “perfect” coronagraph: if the entrance pupil of the system had no phase aberrations, the focal plane light intensity would be equal to 0 outside the inner working angle (IWA; smallest angular separation at which the coronagraph can be used to detect a faint companion) of the coronagraph within the spectral bandwidth considered. All observations are made at the zenith: atmospheric dispersion is not taken into account.

The pupil plane complex amplitude is denoted

\[ W(u) = A(u)e^{i\phi(u)}, \]

where \( A(u) \) is the amplitude and \( \phi(u) \) is the phase of the wavefront. The pupil plane phase aberration

\[ \phi(u) = \frac{2\pi h}{\lambda} \cos(2\pi fu + \theta) \]

creates two symmetric images of the central PSF (Malbet et al. 1995):

\[ I(\alpha) = \text{PSF}(\alpha) + \left(\frac{\pi h}{\lambda}\right)^2 \left[ \text{PSF}(\alpha + f\lambda) + \text{PSF}(\alpha - f\lambda) \right], \]

where \( \alpha \) is the imaging wavelength, \( h \ll \lambda \) is the amplitude (in meters) of the sine wave phase aberration of spatial frequency \( f \), and \( \alpha = f\lambda \) is the angular coordinate on the sky. The phases of these speckles are \( \pi/2 - \theta \) and \( \pi/2 + \theta \). Similarly, a multiplicative amplitude error

\[ A(u) = 1 + a \cos(2\pi fu + \theta) \]

creates two symmetric speckles of phases \( -\theta \) and \( \theta \) on either side of the “ideal” image \( \text{PSF}(\alpha) \):

\[ I(\alpha) = \text{PSF}(\alpha) + \left(\frac{a}{2}\right)^2 \left[ \text{PSF}(\alpha + f\lambda) + \text{PSF}(\alpha - f\lambda) \right]. \]

To simplify notations in the rest of the paper, I denote \( f = |f| \), \( u = |u| \), and \( \alpha = |\alpha| \).

1. Photon noise in the WFS.
2. Chromaticity of the optical path length difference (OPD) and amplitude between the WFS wavelength \( \lambda \) and the imaging wavelength \( \lambda_i \).
3. Aliasing. The wave front measurement is corrupted by higher spatial frequency aberrations that propagate into modes that are detected by the WFS.

In this section the first two effects are discussed, while aliasing effects are studied separately in \S\ 3. Table 2 lists the terms computed analytically in this section:

1. \( C_0 \): PSF contrast limit imposed by OPD aberrations in uncorrected atmospheric turbulence.
2. \( C_1 \): PSF contrast limit imposed by amplitude aberrations in uncorrected atmospheric turbulence (scintillation).
3. \( C_2 \): PSF contrast limit imposed by residual OPD aberrations after AO correction. This term is computed analytically in this section as a function of the WFS sensitivity to photon noise \( \beta_p \). Using the technique proposed in Appendix A, \( \beta_p \) is computed for different WFSs in \S\ 4.
4. \( C_3 \): PSF contrast limit imposed by residual amplitude aberrations after AO correction of OPD and amplitude.
5. \( C_4 \): PSF contrast limit imposed by the differential OPD between the WFS and imaging wavelengths. This term is caused by the chromaticity of Fresnel propagation.
6. \( C_5 \): PSF contrast limit imposed by the differential scintillation between the WFS and imaging wavelengths. This term is caused by the chromaticity of Fresnel propagation.
7. $C_c$: PSF contrast limit imposed by the differential OPD between the WFS and imaging wavelengths. This term is caused by the chromaticity of the refraction index of air.

The terms $C_c$ are computed for high levels of correction (Strehl ratio $\approx 1$), which allows simplification of most equations. The final PSF contrast $C_c$ computed as a function of angular separation, is then obtained as follows:

1. No AO correction: $C_c = C_0 + C_1$.
2. AO correction of phase only: $C_c = C_1 + C_2 + C_4 + C_6$.
3. AO correction of phase and amplitude: $C_c = C_2 + C_3 + C_4 + C_5 + C_6$.

2.2. Uncorrected Atmospheric Turbulence ($C_0$ and $C_1$)

In the paraxial approximation, Fresnel propagation of wave of complex amplitude $W(u, 0)$ over a distance $z$ produces a wave $W(u, z)$ described by

$$W(u, z) = W(u, 0) \otimes \exp(i\pi u^2/2z),$$

where $\otimes$ is the convolution operator. This is equivalent to a phase shift of each spatial frequency component of the wave front by

$$d\phi = \pi f^2 z \lambda.$$

A pupil plane complex amplitude

$$W(u, 0) = 1 + i \frac{2\pi h}{\lambda} \sin(2\pi uf + \theta)$$

therefore becomes

$$W(u, z) = 1 + \sin(d\phi) \frac{2\pi h}{\lambda} \sin(2\pi uf + \theta) + i \cos(d\phi) \frac{2\pi h}{\lambda} \sin(2\pi uf + \theta).$$

This equation shows that Fresnel propagation of a pure sine wave phase aberration produces both an amplitude and a phase aberration of identical spatial frequency in the pupil plane. This effect is periodic for each spatial frequency, as $W(u, z + z_T) = W(u, z)$, with $z_T = 2f^2/\lambda$ the Talbot distance (Talbot 1836).

For ground-layer Kolmogorov atmospheric turbulence, the power spectrum of the two-dimensional phase is

$$\phi(f) = \frac{0.023}{r_0^{5/3}} f^{-11/3},$$

where $r_0$ is the Fried parameter. In a telescope pupil of diameter $D$, the power given by a single spatial frequency is obtained by integration of $\phi(f)$ over a two-dimensional domain of width proportional to $1/D$. Through numerical simulations, the corresponding amplitude (in meters) of the sine wave component of spatial frequency $f$ is computed:

$$h(f) = \frac{0.22\lambda_0}{f^{11/6} D_0^{5/6} r_0^{5/6}}$$

where $\lambda_0$ is the wavelength at which $r_0$ is measured.

Taking into account the Fresnel propagation given in equation (9), the following expressions are obtained for the OPD and amplitude components of atmospheric turbulence in equations (2) and (4):

$$h(f) = \frac{0.22\lambda_0}{f^{11/6} D_0^{5/6} r_0^{5/6}} \sqrt{X(f, \lambda)},$$

$$a(f) = \frac{2\pi(0.22\lambda_0)}{\lambda_0 f^{11/6} D_0^{5/6}} \sqrt{Y(f, \lambda)},$$

where

$$X(f, \lambda) = \frac{\int C_n^2(z) \cos^2(\pi f^2 \lambda_0) dz}{\int C_n^2(z) dz},$$

and

$$Y(f, \lambda) = \frac{\int C_n^2(z) \sin^2(\pi f^2 \lambda_0) dz}{\int C_n^2(z) dz} = 1 - X(f, \lambda).$$

Since Fresnel diffraction is chromatic, $X$ and $Y$ are functions of $\lambda_i$. $X$ is the fraction of the atmospheric turbulence that produces phase errors, the remaining part producing amplitude errors (scintillation). For low-altitude turbulence and/or low spatial frequencies, $X \approx 1$: the beam propagation length is too short to allow Fresnel diffraction to transform phase errors in amplitude errors. By combining equations (3) and (12), since $f = \alpha / \lambda_i$, atmospheric phase aberrations produce the following contrast at $\lambda_i$:

$$C_0(\alpha) = \frac{0.484\pi^2 \lambda_0^2 k^5/3 X(\alpha / \lambda_i, \lambda_i)}{\alpha^{11/3} D_0^{2.5/3}}.$$$$C_1(\alpha) = \frac{0.484\pi^2 \lambda_0^2 k^{5/3} Y(\alpha / \lambda_i, \lambda_i)}{\alpha^{11/3} D_0^{2.5/3}}.$$ Since $X + Y = 1$, the combined contribution of phase and amplitude aberrations in the PSF contrast is independent of the turbulence altitude.

2.3. Effect of WFS Photon Noise and Time Lag on the Corrected Phase ($C_2$)

In the Taylor approximation used in this work, atmospheric turbulence is moving in front of the telescope pupil at a speed $v$ (wind speed along the direction $\alpha$ considered). In a closed loop AO system, the corrected amplitude $h_c$ of the spatial frequency considered is the quadratic sum of a component due to time lag and a component due to photon noise (given by eq. [A23]):

$$h_c = \sqrt{[2\pi h(f)vf]^2 + \left(\frac{\lambda}{2\pi}\right)^2 \left(\frac{\beta_p}{\sqrt{1F\pi D^2/4}}\right)^2},$$

where $t$ is the WFS sampling time, $F$ is the source brightness (in photons s$^{-1}$ m$^{-2}$), and $D$ is the telescope diameter.

Parameter $h_c$ is minimal for

$$t_m = \left(\frac{\lambda}{\lambda_0}\right)^{2/3} \frac{0.204\beta_p^{3/2} r_0^{5/3} f^{5/9}}{F^{1/3} v^{2/3} X(\alpha / \lambda_i, \lambda_i)^{1/3}}.$$
The corresponding residual error produces a symmetric pair of speckles (eq. [3]) with a contrast to the central PSF peak:

$$C_2(\alpha) = 2.345 \frac{\lambda^{4/3} \alpha^{2/3} \beta^{4/3} v^{2/3} X(\alpha/\lambda_i, \lambda_i)^{1/3}}{\lambda_{\text{air}}^{13/9} F^{4/3} D^2 \alpha^{5/9} \lambda^{5/9}},$$

where \(\lambda_i\) is the wavelength at which the final image is obtained and might be different from \(\lambda\), the wave front sensing wavelength.

2.4. Effect of WFS Photon Noise and Time Lag on the Corrected Amplitude (\(C_3\))

The light intensity distribution (scintillation) in the pupil plane can be measured and corrected for. In the frozen turbulence flow model adopted in this work, scintillation is an amplitude screen moving in front of the telescope. The effect of photon noise and time lag on the corrected amplitude \(a_c\) can therefore be written

$$a_c = \sqrt{\left[2\alpha(f)v(t)f\right]^2 + \frac{\beta^2}{t^2 F^2 \pi D^2 / 4}}. \quad (21)$$

This equation is identical to equation (18) if \(h_c\) and \(h(f)\) are replaced by \(\lambda a_c / 2\pi\) and \(\lambda a(f)/2\pi\), respectively. The optimal sampling time is therefore

$$t_a = \left(\frac{\lambda}{\lambda_{\text{air}}}\right)^{2/3} \frac{0.204 \beta a^2 \lambda_i^{5/9} f^{5/9}}{F^{1/3} v^{2/3} Y(\alpha/\lambda_i, \lambda_i)^{1/3}}, \quad (22)$$

and the corresponding contrast \(C_3\) is

$$C_3(\alpha) = 2.345 \frac{\lambda^{4/3} \alpha^{2/3} \beta^{4/3} v^{2/3} Y(\alpha/\lambda_i, \lambda_i)^{1/3}}{\lambda_{\text{air}}^{13/9} F^{4/3} D^2 \alpha^{5/9} \lambda^{5/9}}. \quad (23)$$

2.5. Chromaticity of OPD and Scintillation

2.5.1. OPD Chromaticity Produced by Fresnel Propagation (\(C_6\))

Fresnel propagation is chromatic, and the OPD at the telescope pupil is therefore chromatic. When perfectly corrected at one wavelength (the WFS wavelength), the OPD in the imaging wavelength will show a small residual that limits the achievable contrast to

$$C_6(\alpha) = C_0(\alpha) \frac{dX(\alpha/\lambda_i, \lambda_i, \lambda)}{X(\alpha/\lambda_i, \lambda_i)}, \quad (24)$$

where

$$dX(f, \lambda_i, \lambda) = \frac{\int C_2^2(z) \left[\cos(\pi f z^2 \lambda_i) - \cos(\pi f z^2 \lambda)\right]^2 dz}{\int C_2^2(z) dz}.$$

(25)

2.5.2. Scintillation Chromaticity (\(C_5\))

Similarly, Fresnel propagation produces wavelength-dependent intensity variations in the pupil plane. This produces a limit \(C_5\) on the achievable contrast:

$$C_5(\alpha) = \frac{C_1(\alpha) dY(\alpha/\lambda_i, \lambda_i, \lambda)}{Y(\alpha/\lambda_i, \lambda_i)}, \quad (26)$$

where

$$dY(f, \lambda_i, \lambda) = \frac{\int C_2^2(z) \left[\sin(\pi f z^2 \lambda_i) - \sin(\pi f z^2 \lambda)\right]^2 dz}{\int C_2^2(z) dz}.$$

(27)

2.5.3. Chromaticity of the Air Refraction Index (\(C_6\))

The index of refraction of dry air at standard temperature and pressure is wavelength dependent (Edlen 1966):

$$n(\lambda) = 1.0 + \frac{0.2040603}{130 - \lambda^2} + \frac{0.00015997}{38.9 - \lambda^2}.$$

(28)

The corresponding PSF contrast is

$$C_6(\alpha) = C_0(\alpha) \left[\frac{n(\lambda_i) - n(\lambda)}{n(\lambda_i)}\right]^2. \quad (29)$$

3. ALIASING EFFECTS

3.1. WFS Aliasing

The pupil OPD and amplitude aberrations can only be corrected by the AO system below a cutoff spatial frequency \(f_c\) because of limited sampling in the pupil plane DM and/or in the WFS. For the contrast expressions derived in § 2 to be applicable, the measurement accuracy of a pupil plane phase aberration of spatial frequency \(f < f_c\) at the WFS wavelength \(\lambda\) must be limited by photon noise. Unfortunately, measurement of an OPD or amplitude aberration of frequency \(f < f_c\), even in the absence of photon noise, can be corrupted by aliasing: other spatial frequencies (usually above \(f_c\), but not always) can create a WFS signal at frequency \(f\).

The optical part of the WFS (before the detector) does not produce aliasing: a phase aberration at frequency \(f\) only creates an optical signal of frequency \(f/\sin n(\lambda_i)\) in the pupil plane. This signal can be split displacements (for Shack-Hartmann WFS) or intensity modulation (for curvature WFS, for example). Aliasing is therefore produced by the limited sampling of the measurement in the pupil plane.

Two approaches exist to suppress or mitigate aliasing in WFSs:

1. Increasing the WFS spatial sampling. If a good detector (low readout noise and dark current) is used, this solution can be highly successful for all but one of the WFSs considered in this work. The single exception is the Shack-Hartmann WFS, where an increase of spatial sampling in the pupil plane (smaller sub-apertures increases the measurement error on low-order modes due to photon noise (see § 4.1).

2. Preventing spatial frequencies above \(f_c\) to be “seen” by the optical part of the WFS. This can be done by spatial filtering in the focal plane (Poyneer & Macintosh 2004) or by using an anti-aliasing optical filter before the detector (N. Takato 2005, private communication). These solutions reduce aliasing on all pupil plane WFSs and are most effective if the WFS sampling is regular, as is usually the case. Curvature WFSs include a focal plane iris at the vibrating membrane (usually to reduce stray light and sky background), which can be used to reduce aliasing in high Strehl regime. Antialiasing optical filters are routinely used in imaging with CCDs and are often placed immediately before the detector in commercial digital cameras. Antialiasing optical filters with total rejection of high spatial frequencies can be designed (Leger et al. 1997) for monochromatic light, and their performance in
white light is still good: the solution proposed by Leger et al. (1997) reduces aliasing by a factor of 24 with a 20% bandpass.

3.2. Algorithm Used to Compute Aliasing-free DM Control Signals

If the focal plane complex amplitude (or, equivalently, the aliasing-free pupil plane complex amplitude) is perfectly known up to a spatial frequency $f_c$, it is possible to drive a DM to cancel focal plane speckles within a region of the image corresponding to spatial frequencies lower than $f_c$. Malbet et al. (1995) proposed to use a non-linear minimization algorithm to control the DM. In this work, a Gerchberg-Saxton algorithm (Gerchberg & Saxton 1972) is proposed to find the optimal control signals. Since this method is tailored at finding the solution of a problem with constraints on both a function (pupil plane complex amplitude) and its Fourier transform (focal plane complex amplitude), it seems naturally well adapted (Fauchoir et al. 1989). The steps of the algorithm are shown in Figure 1:

1. The region of the focal plane within which the speckles are to be canceled is first chosen. This “diffraction control domain” (DCD) should exclude the central part of the PSF and should not extend beyond the pupil spatial frequency defined by the DM actuator size or WFS sampling. This region can be within a half plane if amplitude errors (scintillation) in the pupil plane are expected, or it can include both sides of the PSF for correction of OPD aberrations only.

2. The focal plane complex amplitude is multiplied by the DCD to represent the complex amplitude that should be canceled by the DM.

3. The two-dimensional complex function computed in step 2 is Fourier transformed to produce the “ideal” pupil plane complex amplitude required to cancel the speckles within the DCD.

4. The pupil plane complex amplitude function is “projected” on the DM: for each actuator of the DM, the phase that best matches the function computed in step 3 is used to update the DM state.

5. Using the updated DM state and the initial measured focal plane complex amplitude, the updated complex amplitude in the focal plane is numerically estimated, and steps 2–5 can then be repeated.

This algorithm converges very rapidly: only a few iterations are required to obtain a high contrast (about $10^{-6}$) if the initial PSF aberrations are low. It is also computationally less greedy than the solution proposed by Malbet et al. (1995) (the computing time is dominated by two Fourier transforms) and can therefore be implemented in a fast closed loop control system.

This algorithm is also very flexible and performs well in nonideal conditions. Nonideal DM characteristics such as irregular actuator shapes, “dead” actuators, or coupled influence functions can easily be included in the algorithm (step 4) with minimal cost in complexity or computation time. For example, Figure 2 shows that a solution yielding good PSF contrast can be found even if the PSF has large aberrations and if the DM has insufficient stroke to fully correct them. This particular example illustrates how the algorithm is able to find a solution for which the diffraction within the DCD is canceled even though the residual phase aberrations in the pupil plane are still large: these aberrations are confined to either high or low spatial frequencies but are very small in the spatial frequency range corresponding to the DCD. It should be noted that in such nonideal conditions, the number of iterations (steps 2–5) required to converge can be quite high (about 100 in this example).

3.3. Closed Loop Operation

In a closed loop AO system, the DM control algorithm proposed in § 3.2 is used to compute frequent but small updates of the DM: the phase function in the pupil plane needed to cancel the speckles is very small. The number of iterations (steps 2–5 of the algorithm) required within the algorithm is therefore small (a single iteration is sufficient). The computing time can consequently be made compatible with kHz update rate on modern computers for systems with up to $10^4$ actuators: on a $128 \times 128$ actuators system (13,000 actuators on the circular pupil) with a focal plane image sampling such that the DCD occupies $256 \times 256$ pixels, the time required for the two fast Fourier transforms is about 1 ms on a modern computer.

The number of photons per focal plane speckle is typically less than 10 per sampling period, and the corresponding relative error on the measured complex amplitude of the speckle due to photon noise is more than 10%. The closed loop performance of the AO correction is therefore not sensitive to small errors introduced by the algorithm described in § 3.2. Even a 5% error in the knowledge of the DM response has a negligible impact on the system performance.

4. WAVE FRONT SENSOR SENSITIVITIES

The sensitivity of a WFS is a quantitative measure of how photon noise affects its measurement of OPD or amplitude. In this section, the sensitivity $\beta_p$ of WFSs for OPD sensing (when only OPD is measured by the WFS) is computed. Parameter $\beta_p$ is used in equation (20) to estimate the contribution $C_2$ of WFS photon noise to the PSF contrast. For each WFS, I show how $C_2$ varies as a function of angular separation and how WFS design parameters affect it. An exact definition of $\beta_p$ and details on how it is computed are given in Appendix A. The sensitivity $\beta_p$ for amplitude sensing is given within the discussion in § 4.7.

The results obtained in this section are only valid for small residual phase variance at the wave front sensing wavelength.

4.1. Shack-Hartmann WFS

In a Shack-Hartmann WFS (SHWFS), the quantities $I_k$ measured are spot displacements. The associated noises for a diffraction-limited spot, in the absence of background light, are (Hardy 1998)

$$\sigma_{I_k} = \frac{0.277\lambda}{d_{sa} \sqrt{N_{sa}}}$$

(30)

for a continuous noiseless detector and

$$\sigma_{I_k} = \frac{0.500\lambda}{d_{sa} \sqrt{N_{sa}}}$$

(31)

for a quad cell detector. In the above equations, $d_{sa}$ is the sub-aperture size and $N_{sa}$ is the number of photons per subaperture. To account for atmospheric turbulence within each cell, $1/d_{sa}$ should be replaced by $(1/d_{sa}^2 + 1/r_0^2)^{1/2}$.

Using the equations detailed in Appendix A, the following results are obtained:

1. For a Shack-Hartmann WFS with a continuous noiseless detector:

$$\beta_p = 0.67 \frac{0.67}{fd_{sa}} \sqrt{1 + \left(\frac{d_{sa}}{r_0}\right)^2}$$

(32)
Fig. 1.—Proposed algorithm used to drive the DM in a closed loop AO system.
Fig. 2.—Example of diffracted light suppression using the algorithm detailed in § 3.2. In this case, the DM stroke is limited to ±0.4 rad (±0.8 rad of phase correction), and the initial pupil phase aberration is about 3 rad from peak to peak. The DM actuators are square shapes, and their influence functions are overlapping (each DM influence function is a square convolved by a Gaussian). In this example, there are 50 actuators across the diameter of the pupil (2000 actuators total).
LIMITS OF AO FOR HIGH-CONTRAST IMAGING

2. For a Shack-Hartmann WFS with noiseless quad cells:

\[
\beta_p = \frac{1.48}{f d_{sa}} \sqrt{1 + \left(\frac{d_{sa}}{r_0}\right)^2}.
\]  

(33)

For both equations, \( f d_{sa} \leq \frac{1}{4} \) (minimum of three lenslets per sine wave period), as lower pupil plane sampling increases \( \beta_p \). For example, with two lenslets per period, if the centers of lenslets coincide with the crests and peaks of the sine wave phase aberration, no signal will be produced by the SHWFS.

With an SHWFS, photometry of the spots can be used to measure amplitude variations in the pupil plane, without altering the accuracy of the phase measurement: the sensitivity \( \beta_p \) is maintained even if OPD and scintillation are measured simultaneously.

The PSF contrast component \( C_2 \) achievable with an SHWFS is shown in Figure 3 for subaperture sizes ranging from 2 to 70 cm. For each subaperture, a continuous noiseless detector was assumed, rather than a less sensitive quad cell.

In the inner region of the PSF, the contrast \( C_2 \) decreases as the \(-17/9\) power of angular separation (eqs. [32] and [20]). No correction is possible beyond the sampling limit of the WFS: the contrast \( C_2 \) reaches a minimum value at this transition point. The contrast \( C_2 \) at small angular separations is independent of the number of subapertures if the subaperture size is larger than the seeing. However, if subapertures are smaller than \( r_0 \), diffraction by each subaperture increases the subaperture’s focal plane spot size and therefore reduces the sensitivity of the WFS. It therefore seems impossible to simultaneously optimize the contrast over a wide range of separations.

To achieve the optimal performance shown in Figure 3, the wave front integration time \( t_b \) needs to be proportional to \( \alpha^{-1/9} \).

4.2. Curvature Wave Front Sensor (CWFS)

In a curvature WFS (Roddier 1988; Roddier et al. 1991), a spherical phase aberration is introduced in the focal plane, which is equivalent, in the pupil plane, to Fresnel propagation. The pupil image is therefore “conjugated” to an altitude that is set by the amplitude of the focal plane phase aberration. Equation (9) shows that Fresnel propagation of a pure sine wave phase aberration produces both an amplitude and a phase aberration of identical spatial frequency in the pupil plane. The curvature WFS therefore transforms phase aberrations into light intensity modulations in the pupil plane.

The WFS measures intensities \( I_k \) in the pupil plane. I assume here that \( N \) such measurements are taken per spatial period:

\[
I_k = \frac{N_{ph}}{N} \left[ 1 + \frac{4\pi A \sin (d\phi)}{2} \sin \left( \frac{2\pi k}{N} + \phi \right) \right],
\]  

(34)

with \( \sigma_{I_k} = (N_{ph}/N)^{1/2} \).

Using the method detailed in Appendix A, the following expression for \( \beta_p \) is obtained:

\[
\beta_p(\alpha) = \sin^{-1} \left( \frac{\pi \delta z \lambda \alpha^2}{\delta z^2} \right),
\]  

(35)

where \( \delta z \) is the conjugation altitude of the pupil plane (in curvature AO systems, two pupil plane images are usually acquired, at conjugation altitudes \(+\delta z\) and \(-\delta z\)). This result is independent of \( N \) for \( N > 2 \).

The PSF contrast component \( C_2 \) achievable with a CWFS is shown in Figure 4. The amount of defocus introduced in the focal plane of the CWFS can be adjusted to tune its sensitivity to an optimal spatial frequency in the pupil plane (for which the term in the sine of eq. [35] is \( \pi/2 \)).

In the inner regions of the PSF, the contrast \( C_2 \) decreases as the \(-29/9\) power of the angular separation (eqs. [35] and [20]), which is significantly steeper than for an SHWFS. This steep increase of wave front error at low spatial frequencies is also referred to as “noise propagation” and is known to be more serious for a CWFS than for an SHWFS. Soon after the contrast reaches a minimum, the defocus distance becomes too large (the sine in eq. [35] becomes close to zero) and no reliable correction can be applied to the wave front. Theoretically, correction of higher spatial frequencies is possible as the sine in equation (35) periodically oscillates between 1 and \(-1\), but this possibility was not considered in Figure 4; in a real CWFS, spectral bandwidth and time evolution of \( \delta z \) (usually closer to a sine wave than a step function) prevent this feature from being usable.

At small angular separations, the wave front integration time \( t_b \) for a CWFS needs to be proportional to \( \alpha^{-7/9} \) to achieve the result shown in Figure 4. This suggests that a curvature WFS can
greatly benefit from a modal control scheme, where the correction speed can be adjusted for each spatial frequency, as opposed to a zonal reconstruction with a fixed integration time. Although the optimal contrast region is narrower in a CWFS than it is with an SHWFS, it also is deeper: at a given separation, a properly tuned CWFS performs better than an SHWFS. This is especially true close to the PSF center, where a “tuned” CWFS can reach a sensitivity $C_1^2 = 1$. Since changing the extrafocal distance in a CWFS is usually very easy, it is in fact possible to continuously move the optimal contrast region between small and large angular separations during an observation. The equivalent technique would be optically more complex in an SHWFS, as the subaperture size would need to be modified.

4.3. Pyramid WFS

The pyramid WFS (Ragazzoni 1996) divides the focal plane into four quadrants, each one being then reimaged in a separate pupil plane. In the geometrical optics approximation, wave front slopes can be measured as contrast between pairs of pupil images. The focal plane point that defines the position of the quadrants (the “center” of the pyramid) can be rapidly rotating around the PSF core to increase linearity and dynamical range, at the expense of sensitivity.

I denote $P_i(x, y)$ the pupil image corresponding to quadrant $i$, as shown in Figure 5, and $P_{\text{ref}}(x, y)$ the pupil image in the absence of a focal plane pyramid. The pyramid WFS can be operated in two ways:

1. Fixed pyramid position. The top of the pyramid (junction point between the four quadrants) is at the center of the PSF core, corresponding to configuration A in Figure 5.

2. Modulation of pyramid position. The top of the pyramid is moving on a circle of radius $r_p$. Configuration B (the central PSF core and one speckle are within the same quadrant) or configuration C (the quadrant containing the PSF core contains either no speckle or both symmetric speckles) can then occur.

4.3.1. Fixed Pyramid Position

I denote $P_0^i$ the pupil image corresponding to quadrant $i$ obtained in configuration A in the absence of phase aberrations for an unobstructed circular pupil. $P_0^i$ consists of a fainter pupil image with bright sharp edges and a significant fraction of the light diffracted outside the geometric pupil. Although the total light in each pupil image $P_0^i$ is one-quarter of the original pupil image, the total light within the geometric pupil (excluding the bright edges) is about 6% of the original pupil. I consider a pupil complex amplitude

$$W(u, 0) = 1 + i \frac{2\pi A}{\lambda} \sin(2\pi uf + \phi), \quad (36)$$

corresponding to a set of two symmetric speckles as shown in Figure 5. In configuration A, these two speckles interfere with $P_0^i$ in quadrants 1 and 3:

$$P_1 = P_0^1 + \sqrt{P_0^1P_{\text{ref}}} \frac{2\pi A}{\lambda} \sin(2\pi uf + \phi), \quad (37)$$

$$P_2 = P_0^2, \quad (38)$$

$$P_3 = P_0^3 - \sqrt{P_0^3P_{\text{ref}}} \frac{2\pi A}{\lambda} \sin(2\pi uf + \phi), \quad (39)$$

$$P_4 = P_0^4. \quad (40)$$

The pyramid WFS therefore directly measures the pupil phase (Véringaud 2004; Véringaud et al. 2005): this is somewhat different from the geometrical optics understanding of this concept, in

![Figure 5](image-url)
which phase slope is measured by the pupil images. Due to the splitting of the focal plane into four zones, reconstruction of the full wave front map requires \( P_1, P_2, P_3, \) and \( P_4 \). For example, pupil images \( P_1 \) and \( P_3 \) are only sensitive to pupil phase spatial frequencies corresponding to zones 1 and 3 of the pyramid.

Using the method detailed in Appendix A,\n\[
\beta_p = \sqrt{2}. \tag{41}
\]

4.3.2. Modulation of Pyramid Position

I consider a motion of the pyramid center on a circle of radius \( r_p > \lambda/d \), with no change in the orientation of the pyramid. As shown in Figure 5, configurations B and C occur as the pyramid moves. In configuration B in Figure 5, within the geometric pupil,
\[
P_1 = P_{\text{ref}} \left[ 1 + \frac{2\pi A}{\lambda} \sin(2\pi u f + \phi) \right], \tag{42}
\]
\[
P_2 = P_3 = P_4 = 0. \tag{43}
\]

In configuration C, \( P_i = P_{\text{ref}} \) for the pyramid zone containing the PSF core and \( P_i = 0 \) for the other pupil images. In a long exposure (longer than the modulation time of the pyramid position),
\[
P_i = P_{\text{ref}} \left\{ f_i^B \left[ 1 \pm \frac{2\pi A}{\lambda} \sin(2\pi u f + \phi) \right] + f_i^C \right\}, \tag{44}
\]

where \( f_i^B \) and \( f_i^C \) are the fraction of the time during which the PSF core is in the pyramid quadrant \( i \) and the configuration is B and C, respectively. The sign of the modulated intensity signal is opposite between \( P_1 \) and \( P_3 \) and between \( P_2 \) and \( P_4 \). Since \( f_1^B + f_3^B = 0.25 \) (the PSF core spends one-quarter of its time on each zone of the pyramid), in a long exposure (longer than the modulation time of the pyramid position),
\[
P_i = \frac{P_{\text{ref}}}{4} \left[ 1 \pm 4f_i^B \frac{2\pi A}{\lambda} \sin(2\pi u f + \phi) \right]. \tag{45}
\]

Each of the four pupil images contains the intensity modulation, but with different signal levels. Since \( f_1^B = f_3^B \) and \( f_2^B = f_4^B \), using the method detailed in Appendix A,
\[
\beta_p = \frac{2\sqrt{2}}{\sqrt{(4f_1^B)^2 + (4f_2^B)^2}}. \tag{46}
\]

Figure 6 shows how \( \beta_p \) varies across the focal plane. It is minimum at the pyramid modulation radius and increases rapidly toward the center of the PSF.

4.3.3. Discussion

Figure 7 shows the contrast component \( C_2 \) for modulated and fixed pyramid WFSs. The sensitivity of the pyramid WFS is better if the pyramid is fixed, and this mode of operation should be preferred in high-contrast AO on bright sources, as the linearity range of the WFS is then not a concern. If the pyramid is fixed, it may be replaced by a “roof top” (a pyramid with only

Fig. 6.—Value of \( \beta_p \) in a modulated pyramid WFS. A two-dimensional map of \( \beta_p \) is shown on the left, and an averaged radial profile is plotted on the right.

Fig. 7.—PSF contrast component \( C_2 \) obtained with a pyramid WFS with and without modulation of the pyramid’s position. \( C_2 \) is plotted for modulation radii ranging from 0.2 to 2". The top line shows the PSF contrast \( C_2 \) corresponding to uncorrected turbulence phase aberration. Parameters used for this simulation are listed in Table 4.
two faces), which would offer the same sensitivity with two pupil images instead of four.

4.4. Mach-Zehnder Pupil Plane Interferometer

In this wave front sensing scheme suggested by Angel (1994), a beam splitter produces two copies of the same wave front. One of the copies is spatially filtered and interferometrically combined with the unfiltered wave front. Wave front phase is transformed into intensity variations in the two pupil images produced by the interferometer. Minimum sensitivity is reached when the first beam splitter is symmetric (50/50 beam splitter), and the interferometer’s OPD is such that the two pupil images have the same brightness when the wave front is perfect. The interferometer’s OPD may be modulated to increase dynamical range.

The WFS measures intensities \( I_k \) in the two pupil planes. I assume here that \( 2N \) such measurements are taken per spatial period (\( N \) measurements per pupil image):

\[
I_k = \frac{N_{ph}}{2N} \left[ 1 \pm \sin \left( \frac{2\pi k}{\lambda} \sin \left( \frac{2\pi k}{N} + \phi \right) \right) \right],
\]

with \( \sigma_k = \left( N_{ph}/2N \right)^{1/2} \), and the sign in front of the sine is different for each pupil image.

Using the method detailed in Appendix A, the following expression for \( \beta_p \) is obtained:

\[
\beta_p = 2.
\]

Figure 8 shows the PSF contrast component \( C_2 \) obtained with a PPMZWFS.

4.5. Focal Plane WFS

In an FPWFS, the amplitude and phase of focal plane speckles are measured by inducing interferences between the focal plane complex amplitude and a set of known “reference waves” (Angel 2003). Optical configurations to produce the reference waves and measure the interferences are discussed in § 6.

The amplitude and phase of a focal plane speckle created by the sine wave pupil phase error are

\[
A_s = \sqrt{N_{ph}} \frac{\pi h}{\lambda},
\]

and

\[
\phi_s = \phi,
\]

where \( x_0 \) and \( y_0 \) are the real and imaginary parts of the speckle

\[
x_s = A_s \cos(\phi_s) = \sqrt{N_{ph}} \frac{x_0}{2},
\]

and

\[
y_s = A_s \sin(\phi_s) = \sqrt{N_{ph}} \frac{y_0}{2},
\]

where \( x_0 \) and \( y_0 \) follow the notations used in Appendix A. Parameters \( x_0 \) and \( y_0 \) are estimated through the measurement of \( N \) intensities:

\[
I_k = \left( \frac{x_k}{\sqrt{N}} + x_s \right)^2 + \left( \frac{y_k}{\sqrt{N}} + y_s \right)^2.
\]

Increasing the number of waves does not lead to better solutions. If the wave front does not contain amplitude variations (no scintillation), the symmetry property of the focal plane speckles allows the use of two speckles to sense a single pupil plane spatial frequency, in which case

\[
\beta_p = \sqrt{2}.
\]

Figure 8 shows the PSF contrast component \( C_2 \) obtained with an FPWFS.

4.6. Sensitivity of an “Ideal” WFS

Using the \( (U, S) \) representation of WFSs given in Appendix B, the unitary matrix \( U \) can be optimized for the wave front sensing
of a pure sine wave phase error of fixed frequency $f$ by finding the smallest possible value of $\beta_p$. Random WFSs can be built and the corresponding value of $\beta_p$ computed using the equations detailed in Appendix A. The quantities $I_0$ measured are light intensities (number of photons), and the associated noise is $I_0^{1/2}$. The numerical simulation results show that the value of $\beta_p$ obtained is then independent of the size of the unitary matrix used to represent WFSs (as long as this matrix is larger than $3 \times 3$):

$$\beta_p = 1. \quad (58)$$

Theoretically, an “optimal” WFS should therefore be able to have a sensitivity to photon noise $\beta_p = 1$. This result does not, however, ensure that a WFS that can satisfy this requirement simultaneously for all spatial frequencies exists, as the above simulation was performed for a single spatial frequency. The results obtained in this work show that only the CWFS reaches this optimal sensitivity, but only for a single value of the spatial frequency. Understanding how the CWFS achieves this result might allow the design of an “ideal” WFS.

When detection is performed in the pupil plane, the goal of the WFS is to transform a phase aberration into a light modulation. I now consider two symmetric focal plane speckles of amplitude $a$ (relative to the central peak amplitude) and phases $\phi_1$ and $\phi_2$ (relative to the phase of the central peak). Fourier transform of this speckle pair yields pupil plane modulation amplitudes of

$$M_a = 2a \cos \left( \frac{\phi_1 + \phi_2}{2} \right) \quad (59)$$

for amplitude and

$$M_p = 2a \sin \left( \frac{\phi_1 + \phi_2}{2} \right) \quad (60)$$

for phase. As described by equation (3), a pupil plane sine wave phase aberration of amplitude $\psi$ (in radians) and phase $\theta$ produces two focal plane symmetric speckles of amplitude $\psi/2$ and phases $\phi_1 = \pi/2 - \theta$ and $\phi_2 = \pi/2 + \theta$. Equations (59) and (60) confirm that, if the phase and amplitude are left unchanged in the focal plane, these two speckles correspond to a pure phase error in the pupil plane ($M_a = 0$). At its optimal spatial frequency, the CWFS adds $\pi/2$ to the phase of each speckle, resulting in $M_a = 2a$ and $M_p = 0$. In this particular case, the two speckles are interfering constructively together in the pupil plane. This level of amplitude modulation is impossible to reach if the two speckles are optically separated (pyramid WFSs, FPWFS) and explains why only the CWFS can be “optimal” ($\beta_p = 1$) for a spatial frequency. Its only drawback is that the pupil plane phase offset is $\pi/2$ only at one value of the angular separation.

In order to build an “optimal” WFS, the focal plane offset would need to be $\pi/2$ at all separations. The most practical solution is to change the phase of the PSF core by $\pi/2$ and $-\pi/2$ alternatively, as shown in Figure 9. In closed loop operation in an AO system, the phase offset does not need to be achromatic but should be approximately $\pm \pi/2$ for minimum sensitivity (the phase offset in a CWFS is also not achromatic).

This WFS was originally developed for microscopy (Zernike 1934) and is named Zernike phase contrast wave front sensor (ZWFS) in this paper. More recently, this WFS has been suggested for ground-based AO systems (Bloemhof & Wallace 2003) because it offers a direct phase measurement (as opposed to wave front slope for SHWFS). The phase mask should be small (size $\approx \lambda/d$) if the contrast needs to be optimized very close to the PSF core. In broadband, a speckle might be outside the phase mask in the red end of the band but inside it in the blue end. The sensitivity $\beta_p$ is then more than 1 in an intermediate transition region between the inner part of the PSF ($\beta$ is infinite) and the outer part of the PSF ($\beta_p = 1$). A Wynne corrector (Wynne 1979; Roddier et al. 1980), which magnifies the pupil by a factor proportional to wavelength, may be used to avoid this effect. This device was originally developed to increase the spectral bandwidth of speckle interferometry and has been successfully used on the sky (Boccaletti et al. 1998).

The ZWFS is highly sensitive ($\beta_p = 1$, everywhere except possibly in the central core of the PSF if the mask is large) and quite achromatic but has limited dynamical range: it is ideal when used after a low-order first-stage AO system.

Figure 8 shows the PSF contrast component $C_2$ obtained with a ZWFS.

4.7. Discussion

Table 3 summarizes the results obtained previously and lists $\beta_p$ and $\beta_a$ for the seven WFSs compared in this study.

4.7.1. Sensitivity for OPD Measurement

For some WFSs (SHWFS, CWFS, and MPPYRWFS) $\beta_p$ reaches its minimum at a given distance from the optical axis and increases closer to the PSF core: these WFSs suffer from the noise propagation effect (low sensitivity to low-order modes due to photon noise). The other WFSs (FPYRWFS, PPMZWFS, and FPWFS) maintain a constant value of $\beta_p$ at all separations: noise propagation is low, and low-order terms can be corrected efficiently at the same time as high-order terms. For most WFSs, correcting for amplitude plus phase instead of only phase increases $\beta_p$, with the exception of the SHWFS, which measures amplitude “for free” (photometry within each subaperture). As shown in § 2, in the photon noise–limited regime, the PSF contrast achievable by an AO system varies as $\beta_p^{4.7}$ (eq. [20]): the differences shown in Figure 10 are therefore important. For example, the contrast can be 1.6 times better in a CWFS-based or ZWFS-based AO system than in an FPWFS-based or FPYRWFS-based system. The “ideal” ZWFS outperforms the PPMZWFS by 1 mag in contrast. In the example shown in Figure 10, the SHWFS, even perfectly tuned for optimal PSF contrast at 0.5, produces a PSF contrast 4.3 worse than a ZWFS. As shown in Figure 11, the sampling frequency required to reach optimal performance on a bright source is especially high at low spatial frequency for the most efficient WFSs.
Noncommon path errors can limit the achievable contrast and are almost unavoidable in pupil plane WFSs (all WFSs except for FPWFS). Very accurate calibration is then required and can be obtained by focal plane phase diversity. The FPWFS is immune to this effect if the wave front sensing and scientific focal planes are shared, which is likely to be the case for visible coronagraphic imaging of extrasolar planets from space (TPF mission). However, on ground-based AO systems, wave front sensing in the visible and scientific imaging in the near-IR are often preferred for scientific and technological (detectors) reasons.

All WFSs studied in this paper have good achromaticity and can be used in broadband light. The SHWFS, CWFS, and FPYRWFS are naturally achromatic, while other WFSs require either achromatic phase shifters (PPMZWFS and ZWFS) or an optical antialiasing filter (OF).

4.7.2. Sensitivity for Scintillation Measurement

The steps to compute $\beta_\alpha$ are not detailed in this work, but comparison with the computation of $\beta_p$ reveals that $\beta_\alpha = 1$ if all the light is used to image the pupil. From this result, $\beta_\alpha$ can be easily estimated for all WFSs considered in this study.

In this work, I choose to adopt $\beta_\alpha = 1$ for all WFSs in subsequent numerical simulations. While this is exact for the SHWFS, which does not optically modify the light intensity in the pupil plane, this is not true for most other WFSs. For example, in the CWFS, $1/\beta_p^2 + 1/\beta_p^3 = 1$ (eqs. [35], [59], and [60]); at the optimal angular separation (defined by $\beta_p = 1$) the CWFS is insensitive to scintillation ($\beta_\alpha = \infty$). If $\beta_p$ is high and $C_3 \gg C_2$, then a fraction of the total flux (or, equivalently, time) needs to be allocated to scintillation sensing, which is performed most efficiently by imaging of the pupil. For example, in the CWFS, a fraction of the time is spent at $dz = 0$ (no defocus in the focal plane). This sharing of the photons increases $C_2$ and decreases $C_3$ until $C_2 + C_3 = 1$ is minimal.

However, as shown in § 5, $C_1 < C_0$ within the central arcsecond: OPD aberrations are stronger than scintillation at low spatial frequencies. Since both terms are moving in front of the telescope with the same speed $v$, the postcorrection scintillation residual $C_3$ can be made comparable to postcorrection OPD residual $C_2$ by allocating a small fraction of the incoming photons to scintillation measurement. The PSF contrasts obtained with the approximation $\beta_\alpha = 1$ are therefore only slightly optimistic within the central arcsecond: $\beta_p$ sets the value of $C_2 + C_3$, not $\beta_\alpha$.

Beyond $\alpha = 1''$, however, $C_0 \approx C_1$, and if a WFS is characterized by $\beta_\alpha = \infty$, $\beta_p = 1$, half of the photons should be allocated to pure scintillation measurement. This would result in $\beta_\alpha = \beta_{ap} = \sqrt{2}$ ($\beta_{ap}$ is the sensitivity when both amplitude and phase are measured by the WFS), which would produce contrasts $C_2$ and $C_3$ equal to $2^{5/3} \approx 1.6$ times the values obtained.
with the optimistic approximation $\beta_d = 1$. The maximum error made by the $\beta_d = 1$ approximation is therefore a factor of 1.6 on $C_2$ and $C_3$ and can only occur at large angular separation ($\alpha > 1''$) with the CWFS and the ZWFS.

5. CONTRAST PERFORMANCE

5.1. Parameters Adopted for Numerical Simulations

Table 4 lists the default parameters adopted in this work for numerical simulations. The atmospheric parameters correspond to conditions frequently encountered atop Mauna Kea, Hawaii. The weights and altitudes of the turbulence layers are derived from four nights of MASS and Scidar measurement atop Mauna Kea (Tokovinin et al. 2005). The photometric zero point of the WFS (corresponding to an equivalent bandpass of 0.1 $\mu m$) is representative of existing WFSs.

Through the paper, some of these parameters are modified to evaluate the contrast performance of a system that departs from this default configuration: wave front sensing and imaging wavelength in § 5.3 and atmospheric parameters in § 5.4. The contrast performance can also easily be derived for telescopes larger than 8 $m$: since the contrast limits $C_0$ to $C_6$ are all proportional to $1/D^2$, the overall contrast for all WFSs is proportional to $1/D^2$.

5.2. Relative Contribution of Contrast Limit Components in Conventional AO

Figure 12 shows the relative contributions of $C_0$, $C_1$, $C_2$, $C_3$, $C_4$, $C_5$, and $C_6$ when an ideal WFS (ZWFS) is used on a bright ($m_v = 5$) star. According to the results obtained in § 4, this wave front sensing scheme is the most sensitive, and other WFSs will show higher values of $C_2$. The main parameters of the simulation are listed in Table 4 and are used throughout this work unless otherwise specified. While the chromaticity of the refraction index of air has a negligible effect on the PSF contrast (component $C_0$), chromatic effects introduced by Fresnel propagation through the atmosphere ($C_2$ and $C_3$) can have a strong impact on the PSF contrast. One very important result from this study is that $C_0$, $C_1$, $C_4$, and $C_5$ are all comparable beyond 2". Similarly, $C_2$ and $C_3$ are comparable beyond about 2". The implications of this result are now discussed separately for AO systems correcting only OPD and AO systems correcting OPD and scintillation. For now, I choose to limit this discussion to AO systems performing wave front sensing in the visible and imaging in the near-IR, as choices of wavelengths are discussed in § 5.3:

1. **OPD correction with AO**.—In an OPD-only AO correction ($C = C_1 + C_2 + C_4 + C_6$), the uncorrected scintillation $C_1$ dominates by far the achievable PSF contrast within the central 2" and limits it to approximately $10^{-4}$ to $2 \times 10^{-4}$ within the central arcsecond. The term due to photon noise and time lag, $C_2$, is much lower, at about $10^{-7}$ for this bright ($m_v = 5$) source. The effect of photon noise would become dominant only for $m_v > 13$ with a high-sensitivity WFS ($\beta_d = 1$). The OPD chromatic term $C_4$ due to Fresnel propagation is small close to the PSF center but is rapidly increasing with angular separation and is comparable to scintillation $C_1$ at 2" and beyond.

2. **OPD+scintillation correction with AO**.—With an AO system correcting both OPD and scintillation ($C = C_1 + C_2 + C_4 + C_5 + C_6$), the term $C_4$ due to the chromaticity of scintillation limits the PSF contrast to slightly better than $10^{-4}$ in the central 2". Scintillation, if measured at $\lambda_i = 0.5 \mu m$, cannot be well corrected for at $\lambda_i \approx 1.6 \mu m$. The improvement to the PSF contrast brought by correction of scintillation with the AO system is quite modest (about a factor of 2). Beyond about 2", $C \approx C_0$, and “classical” AO (wave front sensing in visible, imaging in near-IR), even with scintillation correction, cannot improve the PSF contrast: there is no use to increase the number of elements beyond this limit.

5.3. Choice of Wave Front Sensing and Imaging Wavelengths

The number of photons available for wave front sensing is a function of spectral type and should be maximized to reduce the PSF contrast $C_2$. In this section I consider the number of photons available for wave front sensing to be independent of $\lambda$, which is a good approximation for a spectral type G2 and a fixed spectral bandwidth ($\Delta \lambda/\lambda$ constant).

Figure 13 illustrates how $\lambda$ and $\lambda_i$ affect PSF contrast components $C_0$ to $C_6$.
When \( k = k_i \), chromatic terms \( (C_4, C_5, \text{and} C_6) \) disappear, and the PSF contrast is driven by WFS photon noise through \( C_2 \) and \( C_3 \) (for an AO system correcting OPD and scintillation) or \( C_1 \) (for an AO system correcting only OPD). In all configurations, \( C_6 \) has a negligible impact on PSF contrast. These results are combined in Figure 14, which shows the achievable PSF contrast as a function of imaging wavelength when the WFS wavelength is fixed. An AO system correcting both OPD and scintillation is considered in this figure, with an “ideal” WFS operating in the visible (top panel) or in the near-IR (bottom panel). When \( \lambda = \lambda_i \), the photon noise–driven PSF contrast is between \( 10^{\text{-10}} \) and \( 10^{\text{-6}} \) in the central arcsecond in both cases. The PSF contrast degrades very rapidly as \( \lambda \) becomes different from \( \lambda_i \). In conventional AO, where \( \lambda \approx 0.55 \mu\text{m} \) and imaging is performed in the near-IR, the PSF contrast, dominated by chromatic effects, is limited to a few times \( 10^{\text{-4}} \) in the central arcsecond. When observing a bright source, Figure 14 illustrates that the wave front sensing wavelength should be chosen equal to the imaging wavelength.

This statement is at variance with the common practice of combining visible wave front sensing and near-IR imaging to yield the best possible contrast. Figure 13 does indeed show that this is the optimal choice if chromatic effects are ignored. In current AO systems, the WFS detector and sensitivity are suboptimal (the perfect example is an SHWFS with finite readout noise CCD), and errors are dominated by photon noise except for the brightest sources: in this regime, Figure 13 shows that visible \( \lambda \) and near-IR \( \lambda_i \) are optimal. Moreover, finite number of actuators, lack of aliasing mitigation scheme, and non-common path errors set a limit to the PSF contrast even for bright sources: the effect of these OPD errors on the PSF contrast is mitigated by increasing \( \lambda_i \). For these reasons, most current AO systems do not achieve a \( 10^{\text{-4}} \) PSF contrast in the central arcsecond and are therefore not dominated by chromatic effects. However, an AO system designed to maximize PSF contrast (using an efficient WFS and an aliasing mitigation scheme) would be dominated by chromatic effects if \( \lambda \neq \lambda_i \).

If a visible WFS is used for imaging in the near-IR, the chromatic components \( C_4 \) and \( C_5 \) are fixed phase and amplitude screens moving in front of the telescope’s pupil with speed \( v \). The coherence time of the corresponding speckles is therefore long, unlike the fast residual speckles due to time lag and photon noise (contributions \( C_0 \) and \( C_1 \)). These slow speckles are very detrimental to the final detection limit, as they require long exposure times to average into a smooth continuous background.

5.4. Observing Site

If \( \lambda \neq \lambda_i \) (contrast dominated by chromaticity effect), the PSF contrast \( C \) is dominated by \( C_3 \) as illustrated in Figures 13 and 14. Equation (26) for small angular separations (within the central arcsecond) leads to

\[
C \propto \frac{\pi^2}{\theta_0^{-2/3}}
\]
where

\[ z_2 = \sqrt{\frac{\int C_n^2(z) z^2 \, dz}{\int C_n^2(z) \, dz}} \]  

(62)

The contrast is then independent of wind speed \( v \). In this regime, multiplying \( z_2 \) by 0.6 (40% lower altitude turbulence) is equivalent to improving the seeing by a factor of 2. The high importance of turbulence height on PSF contrast is illustrated in Figure 15, where the turbulence profile given in Table 4 has been scaled in altitude to modify \( z_2 \). All other parameters for this simulation are given by Table 4.

If \( \lambda \approx \lambda_i \) (chromaticity effects are small), the PSF contrast is dominated by \( C_2 \), as shown in Figure 13. At small angular separations,

\[ C \approx r_0^{-5/6} v^{-2/3}. \]  

(63)

The contrast is then independent of the atmospheric turbulence profile.

Figure 16 shows the PSF contrast for the observing conditions listed in Table 5 for \( \lambda = \lambda_i = 0.85 \mu m \) (top panel) and for \( \lambda = 0.55 \mu m, \lambda_i = 1.6 \mu m \) (bottom panel). Parameters not listed in Table 5, such as telescope diameter, were taken from Table 4. Atmospheric conditions A and B in Table 5 were derived from

---

**Fig. 14.**—PSF contrast for a WFS operating at 0.55 (top) and 1.6 \( \mu m \) (bottom) after AO correction of OPD and scintillation. The PSF contrast is shown as a function of angular separation at imaging wavelengths ranging from 0.55 to 2.2 \( \mu m \) in each case.

**Fig. 15.**—Effect of the effective turbulence altitude \( z_2 \) on the PSF contrast. Parameters other than \( z_2 \) for these simulations are given in Table 4.

**Fig. 16.**—PSF contrast for different observing sites. On the top panel, \( \lambda = \lambda_i = 0.85 \mu m \). On the bottom panel, \( \lambda = 0.55 \mu m \) and \( \lambda_i = 1.6 \mu m \). Parameters other than \( r_0, v, \) and \( z_2 \) for these simulations are given in Table 4.
The reference waves can be produced by phase shifting of light extracted from the central part of the PSF in the focal plane (Angel 2003; Codona & Angel 2004). The interference between the speckle cloud and the reference waves is obtained through a Mach-Zehnder–type interferometer with a beam splitter. This approach allows one to create the optimal reference wave for focal plane wave front sensing but requires additional optics.

A slightly less optimal but optically simpler solution is to use the DM to produce the required reference waves. Each reference wave is created by a command sent on the DM: for example, moving a single actuator produces a reference wave of quasi-constant amplitude within the control region and of phase given by the position of the actuator within the pupil. Provided that the influence functions of the DM and the behavior of the coronagraph are known, the complex amplitude of the reference wave in the focal plane can be computed to a good accuracy. In a fast closed loop system, the accuracy with which these reference waves are known does not need to be very high (about 10% accuracy is sufficient) to be photon noise limited.

Although the full optimization of the set of reference waves is beyond the scope of this work, I now show that reference waves well suited to focal plane wave front sensing can be obtained with the DM:

1. **Extent of reference wave in the focal plane.**—The size of a reference wave in the focal plane is set by the size of an actuator of the DM. It is therefore possible to produce reference waves with relatively high amplitude across all of the DCD (the region in the focal plane for which the DM sampling is sufficient to suppress diffracted light).

2. **Ability to set the amplitude of the reference wave.**—Small-amplitude waves can easily be created by moving a single actuator. Increasing the displacement of this actuator will increase the reference wave amplitude up to $1/N_{\text{act}}$ of the peak amplitude of the PSF for an unapodized pupil, where $N_{\text{act}}$ is the total number of actuators in the DM. For a PSF contrast better than $10^{-7}$ and a DM with more than $10^7$ actuators, actuators need to be moved in groups to produce reference waves of sufficient amplitudes.

3. **Ability to produce an achromatic phase in the reference wave.**—A small positive displacement of the center actuator of the DM produces a reference wave of achromatic phase $\pi/2$ in the focal plane ($-\pi/2$ for a negative displacement). The amplitude of this reference wave is, however, chromatic, but this effect will not seriously affect the system performance in a closed loop system, as it is equivalent to multiplying the signal (speckle intensity modulation) by a wavelength-dependent gain of constant sign. The phase of reference waves produced by moving actuators other than the central actuator is also achromatic, provided that a Wynne corrector is used.

Actuators near the edge of the pupil should preferably not be used to produce the focal plane reference waves, as the resulting phase would vary rapidly across the focal plane image.

### 7. Coronagraphy

The contrast limits derived in §§ 4 and 5 assumed that the only sources of scattered light are the wave front errors (phase and amplitude) in the corrected beam. These results only apply to optical systems in which the static diffraction (Airy pattern on a circular aperture) is below the contrast levels derived in this work. In this section I discuss the validity of this approximation and briefly summarize the options available (coronagraphy) to ensure that the limits derived in §§ 4 and 5 can be reached.
equals the Airy pattern diffraction contrast level. Points where the theoretical PSF contrast achievable with a perfect coronagraph fraction for 8, 15, 30, 60, and 120 m diameter telescopes. The circles mark the points where the theoretical PSF contrast achievable with a perfect coronagraph equals the Airy pattern diffraction contrast level.

7.1. Need for Suppression of Airy Pattern

Figure 17 shows that with an efficient WFS, diffraction associated with the Airy pattern is much stronger than light diffracted by residual wave front errors at small angular separations. At some distance from the optical axis, both contributions are equal. Beyond this separation, coronagraphy is not required to reach the contrast level achievable by the AO system. As can be seen in Figure 17, this critical angular separation decreases as the telescope diameter increases. For reasonable-size telescopes (100 m diameter or less), this critical separation is larger than 2", and suppression of the Airy pattern is therefore required.

While light diffracted by residual wave front errors is time variable and might not average nicely with time, the Airy diffraction pattern on the other hand is very stable and can be calibrated accurately. One might therefore wonder if relatively high levels of static diffraction features in the PSF are really detrimental to high-contrast imaging. Static diffraction actually amplifies the time-variable speckles, an effect referred to as “speckle pinning” (Bloemhof et al. 2001; Bloemhof 2003; Aime & Soummer 2004). This is due to the fact that the complex amplitudes of static and dynamic speckles add in the focal plane, and the light intensity $I$ measured is

$$I = (A_s + A_d)^2 = I_s + 2\sqrt{I_sI_d} + I_d,$$  

(64)

where $I_s$ and $I_d$ are the static and dynamic focal plane intensities, respectively, and the phase term between the two contributions has been omitted for simplicity. Assuming that $I_s$ is well known and can be perfectly subtracted, if $I_s > I_d$ (Airy pattern is brighter than C), the “speckle noise” becomes dominated by $2(I_sI_d)^{1/2}$ and is therefore $2(I_sI_d)^{1/2}$ stronger than it would be if $I_s = 0$. In the example considered in Figure 17, at 0.5" separation, adding a coronagraph reduces the “speckle noise” by factors of 52, 26, and 19 on 8, 30, and 60 m telescopes, respectively.

With focal plane wave front sensing, the static diffraction of the telescope pupil can be treated just as atmospheric speckles and can therefore be perfectly cancelled in half of the focal plane with a single DM. The DM phase is chromatic but the pupil intensity is not, the resulting dark region of the PSF is chromatic (Codona & Angel 2004) and does not allow the use of a wide spectral band. Yang & Kostinski (2004), however, found that solutions with very low chromaticity exist with a square telescope pupil, and similar solutions might exist with circular telescope pupils. If successful, this technique could be used instead of a coronagraph in narrowband imaging and would be optically very simple, provided that the DM can produce the required phase functions: the solutions found by Yang & Kostinski (2004) require a large phase slope at the edge of the pupil.

7.2. Coronagraph/WFS Combination

At large angular separation the coronagraph does not need to attenuate the Airy pattern by a large factor (at most a factor of 100 attenuation is required at 0.5" according to Fig. 17), and many suitable coronagraphic options are therefore available.

When observation at small angular separation (a few $\lambda/d$) is required, the choice of the coronagraph is more critical. Coronagraphs with small IWA exist but are often prone to the following:

1. Sensitivity to tip-tilt and low-order modes. This effect is problematic on coronagraphs for which small (less than $\lambda/d$) tip-tilt errors can scatter light outside the IWA (Roddier & Roddier 1997; Rouan et al. 2000; Baudoz et al. 2000). The coronagraphic leaks are especially large if they are combined with a WFS having poor sensitivity to low-order modes, such as a CWFS or a high-order SHWFS. An FPWFS is preferable if the contrast at very small IWA needs to be optimized. Coronagraphs relying on pupil apodization manage to keep this effect small while offering small IWA (Kasdin et al. 2003; Guyon et al. 2005).

2. Chromaticity. Coronagraphs with small focal plane masks are sensitive to the wavelength-dependent PSF scale. Several designs have been proposed to mitigate (Soummer et al. 2003b; Rouan et al. 2000) or solve (Baudoz et al. 2000) this problem.

3. Reduced throughput due to pupil and/or focal plane masks (Kasdin et al. 2003; Kuchner & Spergel 2003; Soummer et al. 2003a), or splitting of light (Baudoz et al. 2000). Lossless apodization can be used to avoid this problem (Guyon et al. 2005).

4. Lower image quality: broader PSF due to apodization (Kasdin et al. 2003) or double images (Baudoz et al. 2000).

These effects can be especially problematic if the WFS is placed after the coronagraph, as would likely be the case in an FPWFS-based system: chromaticity, reduced throughput, and lower image quality would then effectively reduce the WFS signal-to-noise ratio and compromise the achievable contrast ratio.

8. CONCLUSIONS

A thorough comparison of the fundamental contrast limits of AO has shown that visible wave front sensing does not allow high-accuracy correction of near-IR wave front aberrations: chromatic effects then limit the PSF contrast to $10^{-4}$ to $10^{-5}$ within the central arcsecond. Wave front sensing should therefore be performed at the same wavelength as imaging to reach the contrast limit imposed by photon noise (about $10^{-5}$ to $10^{-7}$ in the central arcsecond).

An AO system optimized for high contrast, with wave front sensing and imaging at the same wavelength, can still greatly benefit from a first-stage AO correction with a shorter wavelength (visible) WFS:

1. If residual aberrations are small, an FPWFS can be used efficiently. This WFS offers unique advantages: no noncommon
path errors, no aliasing, and high sensitivity. It therefore appears to be the ideal solution for high-contrast AO.

2. If wave front correction were perfect at the shorter wavelength, the chromatic residuals that the second WFS needs to measure would be small and relatively slow (same speed as the uncorrected turbulence). The PSF contrast achievable in this case could be better than the limits derived in this work, as more photons are used.

This solution is especially attractive since low-noise fast visible detectors exist, while near-IR detectors currently offer lower performance. Theoretically, combining a fast high-sensitivity visible WFS (preferably a ZWFS) with a slower near-IR WFS (preferably an FPWFS) is not as advantageous for red sources as it is for bluer (spectral type G or bluer) sources.

The PSF contrast estimates derived in this paper represent a limit that is hard to reach, as many optimistic hypotheses have been made: observation at zenith, perfect telescope and detectors, perfect DM, high system throughput, favorable atmospheric conditions, perfect coronagraph, bright m_0 = 5 source, no time delay for AO control. On an 8 m telescope, PSF contrast up to 10^{-6} may be reached in the central arcsecond if deviation from these optimistic assumptions is minimal. Even on a 100 m telescope, the corresponding contrast (about 10^{-8}) is 2 orders of magnitude short of what is required to detect an Earth-size planet orbiting a solar-type star. Direct imaging of extrasolar planets is therefore bound to rely heavily on efficient calibration of the speckle noise (through differential imaging techniques, for example).

The author is thankful to the referee, René Racine, for his detailed and thorough look at this work, which led to many useful suggestions.

APPENDIX A

COMPUTING THE SENSITIVITY OF A WFS TO PHOTON NOISE FOR THE MEASUREMENT OF A PURE SINE WAVE PHASE ERROR

In this appendix I define \( \beta_p \), the quantitative measure of sensitivity to photon noise of a WFS, and show how it can be derived. The steps described here were used for each of the WFSs considered in this study to compute \( \beta_p \) for a range of spatial frequencies.

The wave front error given in equation (2) can be represented in a two-dimensional plane coordinate system by

\[
x_0 = \frac{2\pi h}{\lambda} \cos \theta \tag{A1}
\]

and

\[
y_0 = \frac{2\pi h}{\lambda} \sin \theta \tag{A2}
\]

where \( x_0 \) and \( y_0 \) are both in radians.

The WFS produces a set of \( N \)-values \( I_0, I_1, \ldots, I_{N-1} \) (usually light intensities, but can also be centroid positions for an SHWFS) that are used to compute \( x \) and \( y \), the measured values of \( x_0 \) and \( y_0 \). Because of photon noise, \( x \neq x_0 \) and \( y \neq y_0 \). After correction by the DM, the residual sine wave phase aberration at the spatial frequency considered has an amplitude \( A_{\text{res}} \) (in radians):

\[
A_{\text{res}} = \sqrt{(x - x_0)^2 + (y - y_0)^2}. \tag{A3}
\]

In order to compute \( \Sigma \), the residual phase error in the pupil plane in radians rms, the probability distribution \( P(x, y) \) of \( x \) and \( y \) can be used:

\[
\Sigma = \sqrt{\int_{x,y} P(x, y) \left[ (x - x_0)^2 + (y - y_0)^2 \right] \, dx \, dy}. \tag{A4}
\]

\( P(x, y) \) can be written as

\[
P(x, y) = \prod_{k=0}^{N-1} P_k(x, y), \tag{A5}
\]

where \( P_k(x, y) \) is the probability distribution of \( x \) and \( y \) given by the measurement \( I_k \).

I denote \( \sigma^2 \) the variance on the measurement of \( I_k \), for which a Gaussian probability law is assumed. I also assume that \( I_k \) is a linear function of \( x \) and \( y \) around \( (x_0, y_0) \) (this is true in closed loop AO systems, as \( x_0, y_0, x, \) and \( y \) are small):

\[
I_k(x, y) = I_k(x_0, y_0) + \frac{dl_k}{dx} X + \frac{dl_k}{dy} Y, \tag{A6}
\]

where the derivatives are computed in \((x_0, y_0), X = x - x_0, \) and \( Y = y - y_0 \).
In the \((x, y)\)-plane, a measurement \(I_k\) gives a probability \(P_k\) only along one axis (the normalization coefficients of probability distributions are omitted in this section):

\[
P_k(x, y) = \exp \left[ -\frac{(k_xX + k_yY)^2}{2\sigma_k^2} \right],
\]

where \(k_x^2 + k_y^2 = 1\),

\[
k_x = \frac{dI_k}{dx} \left( \frac{dI_k}{dy} \right)^{-1},
\]

and

\[
\sigma_k = \sigma_k \left( \sqrt{\left( \frac{dI_k}{dx} \right)^2 + \left( \frac{dI_k}{dy} \right)^2} \right)^{-1}.
\]

The vector \((k_x, k_y)\) gives the direction along which \(I_k\) is “sensitive.” For example, if \(dI_k/dx = dI_k/dy\), \((k_x, k_y) = (1/\sqrt{2}, 1/\sqrt{2})\), and \(P(x, y)\) is only a function of \(X + Y\). In this example, the probability is constant along a line \(x = \text{const} - y\), as the two partial derivatives of \(I_k\) cancel when moving along this line.

From equation (A5),

\[
P(x, y) = \exp \left[ -\frac{(X^2 \alpha_1 + XY \alpha_2 + Y^2 \alpha_3)}{2} \right],
\]

where

\[
\frac{1}{\alpha_1} = \sum_{k=0}^{N} \frac{k_x^2}{2\sigma_k^2},
\]

\[
\frac{1}{\alpha_2} = \sum_{k=0}^{N} \frac{k_x k_y}{\sigma_k^2},
\]

\[
\frac{1}{\alpha_3} = \sum_{k=0}^{N} \frac{k_y^2}{2\sigma_k^2}.
\]

Equation (A10) is a two-dimensional normal law:

\[
P(x, y) = \exp \left[ -\frac{1}{2(1 - \rho^2)} \left( \frac{X^2}{\sigma_x^2} + \frac{2\rho XY}{\sigma_x \sigma_y} + \frac{Y^2}{\sigma_y^2} \right) \right],
\]

where

\[
\rho = \frac{\sqrt{\alpha_1 \alpha_3}}{2\alpha_2},
\]

\[
\sigma_x = \sqrt{\frac{\alpha_1}{2(1 - \rho^2)}},
\]

\[
\sigma_y = \sqrt{\frac{\alpha_3}{2(1 - \rho^2)}}.
\]

A graphical representation of \(P(x, y)\) is shown in Figure 18. In the coordinate system \((x', y')\), whose axes are aligned with the long and short axes of the ellipse \(P(x, y) = \exp(-1)\),

\[
P(x', y') = \exp \left[ -\frac{1}{2} \left( \frac{x'^2}{\alpha_1^2} + \frac{y'^2}{\alpha_3^2} \right) \right].
\]
where

\[ \lambda_1^2 = \frac{\sigma_x^2 + \sigma_y^2 + \sqrt{\left(\sigma_x^2 - \sigma_y^2\right)^2 + 4\rho^2\sigma_x^2\sigma_y^2}}{2} \]  
\( (A19) \)

\[ \lambda_2^2 = \frac{\sigma_x^2 + \sigma_y^2 - \sqrt{\left(\sigma_x^2 - \sigma_y^2\right)^2 + 4\rho^2\sigma_x^2\sigma_y^2}}{2} \]  
\( (A20) \)

From equations (A4), (A18), (A19), and (A20),

\[ \Sigma = \sqrt{\lambda_1^2 + \lambda_2^2}, \]  
\( \Sigma = \sqrt{\sigma_x^2 + \sigma_y^2}. \]  
\( (A21) \)

\( (A22) \)

\( \Sigma \) can now be used as a quantitative measure of the ability of a WFS to sense a sine wave phase aberration in the pupil plane. For most WFSs, \( \Sigma \) is a function of \( \theta \), in which case I consider \( \Sigma_{\text{max}} \), the maximum value of \( \Sigma \) over all values of \( \theta \). Since the measurement errors are produced by photon noise,

\[ \Sigma = \frac{\beta_p}{\sqrt{N_{\text{ph}}}}, \]  
\( (A23) \)

where \( N_{\text{ph}} \) is the total number of photons available for wave front sensing and \( \beta \) is a function of the WFS. The parameter \( \beta_p \) represents the sensitivity of the WFS to photon noise for the spatial frequency considered.

APPENDIX B

ALGEBRAIC REPRESENTATION OF WFSs

In this appendix I show that any wave front sensor can be represented as a unitary matrix \( U \) and a stochastic matrix \( S_c \).

I denote \( W(u) \) the complex amplitude of the incoming wave front at the position \( u \) in the pupil:

\[ W(u) = A(u)e^{i\phi(u)}, \]  
\( (B1) \)

where \( A(u) \) is the amplitude and \( \phi(u) \) is the phase of the wave front.

In all WFSs, the incoming wave front is estimated from measurements of the intensities (square of the modulus of the complex amplitude) obtained by mutual interferences of parts of the incoming wave fronts. One such intensity, \( I_k \), can be written as

\[ I_k = \left| B_k \right|^2 = \left| \int_{\mathbb{R}^2} f_k(u)W(u)\,du \right|^2, \]  
\( (B2) \)
where \( f_k(u) \) is a complex function of \( u \) and
\[
\forall k, \quad |f_k(u)|^2 < 1. \tag{B3}
\]
The number of such measurements is \( m \). Another constraint is that the incoming light is shared between the different outputs:
\[
r \in \mathcal{P} \implies \sum_{k=1}^{m} |f_k(u)|^2 = 1. \tag{B4}
\]
Finally, the light intensity is conserved between the input (wave front) and the outputs:
\[
\forall A, \quad \sum_{k=1}^{m} I_k = \int_{\mathcal{P}} A(u)^2 \, dr. \tag{B5}
\]

The continuous incoming wave front can be approximated by its values on the points of a fine two-dimensional grid: this representation is accurate up to the spatial frequency defined by the spacing of the grid elements. In this representation, the incoming wave front is a vector \( A \) (representing \( W \) in eq. [B1]), which has as many elements \( A_l \) as there are evaluation points on the pupil. I denote \( n \) the number of such elements. Each function \( f_k(u) \) is represented by a vector \( U_k \) of \( n \) elements \( U_l^k \), \( l = 1 \cdots n \). The measured intensities are a vector \( I \) of \( m \) elements, which is the square of the amplitude of the vector \( B \) (elements \( B_l \)):
\[
I = \begin{bmatrix} I_1 \\
I_2 \\
\vdots \\
I_m \end{bmatrix} = \begin{bmatrix} B_1 \\
B_2 \\
\vdots \\
B_m \end{bmatrix}^2 = |B|^2, \tag{B6}
\]
with
\[
B = UA = \begin{bmatrix} U_1^1 & U_2^1 & \cdots & U_n^1 \\
U_1^2 & U_2^2 & \cdots & U_n^2 \\
\vdots & \vdots & \ddots & \vdots \\
U_1^n & U_2^n & \cdots & U_n^n \end{bmatrix} \begin{bmatrix} A_1 \\
A_2 \\
\vdots \\
A_n \end{bmatrix}. \tag{B7}
\]
The conservation of total light intensity requires
\[
\forall A, \quad \|UA\| = \|A\|. \tag{B8}
\]
Setting \( m = n \), \( U \) is therefore a unitary matrix. In all current WFSs, \( m \) is in fact infinite and the measured quantity is not directly \( I \), but a set of intensities obtained by redistributing the values of \( I \) among a smaller number of variables, which can be represented by a stochastic matrix \( S \), preserving the total flux. For example, in an SHWFS, the intensity measured by a pixel is in fact the integral of the intensity across the pixel.

REFERENCES

Aime, C., & Soummer, R. 2004, ApJ, 612, L85
Angel, J. R. P. 1994, Nature, 368, 203
———. 2003, in ASP Conf. Ser. 294, Scientific Frontiers in Research on Extrasolar Planets, ed. S. Seager & D. Deming (San Francisco: ASP), 543
Baudoz, P., Rabbia, Y., & Gay, J. 2000, A&AS, 141, 319
Bloemhof, E. E. 2003, ApJ, 582, L59
Bloemhof, E. E., Dekany, R. G., Troy, M., & Oppenheimer, B. R. 2001, ApJ, 558, L71
Bloemhof, E. E., & Wallace, J. K. 2003, Proc. SPIE, 5169, 309
Bocacletti, A., Augereau, J.-C., Marchis, F., & Hahn, J. 2003, ApJ, 585, 494
Bocacletti, A., Moutou, C., Labeyrie, A., Kohler, D., & Vakili, F. 1998, A&A, 333, 395
Codona, J., & Angel, R. 2004, ApJ, 604, L117
Edlen, B. 1966, Metrologia, 2, 71
Faucherre, M., Merkle, F., & Vakili, F. 1989, Proc. SPIE, 1130, 138
Gerchberg, R. W., & Saxton, W. O. 1972, Optik, 35, 237
Gilmozzi, R., 2004, Proc. SPIE, 5489, 1
Guyon, O. 2004, ApJ, 615, 562
Guyon, O., Pluzhnik, E. A., Galicher, R., Martinache, F., Ridgway, S. T., & Woodruff, R. A. 2005, ApJ, 622, 744
Hardy, J. 1998, Adaptive Optics for Astronomical Telescopes (Oxford: Oxford Univ. Press)
Hawarden, T. G., et al. 2003, Proc. SPIE, 4840, 299
Kasdin, N. J., Vanderbei, R. J., Spergel, D. N., & Littman, M. G. 2003, ApJ, 582, 1147
Kuchner, M. J., & Spergel, D. N. 2003, ApJ, 594, 617
Lawrence, J. S., Ashley, M. C. B., Tokovinin, A., & Travouillon, T. 2004, Nature, 431, 278
Leger, J. R., Schuler, J., Morphis, N., & Knowlden, R. 1997, Appl. Opt., 36, 4692
Malbet, F., Yu, J. W., & Shao, M. 1995, PASP, 107, 386
Marois, C., Doyon, R., Racine, R., & Nadeau, D. 2000, PASP, 112, 91
Poyneer, L. A., & Macintosh, B. 2004, J. Opt. Soc. Am., 21, 810
Racine, R., & Ellerbroek, B. L. 1995, Proc. SPIE, 2534, 248
Racine, R., Walker, G. A. H., Nadeau, D., Doyon, R., & Marois, C. 1999, PASP, 111, 587
Ragazzoni, R. 1996, J. Mod. Opt., 43, 289
Roddier, C., Roddier, F., Martin, F., Baranne, A., & Brun, R. 1980, J. Opt., 11, 149
Roddier, F. 1988, Appl. Opt., 27, 1223
Roddier, F., Northcott, M., & Graves, J. E. 1991, PASP, 103, 131
Roddier, F., & Roddier, C. 1997, PASP, 109, 815
Roth, K. C., Guyon, O., Chun, M., Jensen, J. B., Jorgensen, I., Rigaut, F., & Walther, D. M. 2001, BAAS, 33, 785
Rouan, D., Riaud, P., Boccaletti, A., Clénet, Y., & Labeyrie, A. 2000, PASP, 112, 1479
Soummer, R., Aime, C., & Falloon, P. E. 2003a, A&A, 397, 1161
Soummer, R., Dohlen, K., & Aime, C. 2003b, A&A, 403, 369

Talbot, W. H. F. 1836, Philos. Mag., 9, 401
Tokovinin, A., Vernin, J., Ziad, A., & Chun, M. 2005, PASP, 117, 395
Vérinaud, C. 2004, Opt. Commun., 233, 27
Vérinaud, C., Le Louarn, M., Korkialaishi, V., & Carhillet, M. 2005, MNRAS, 357, L26
Wynne, C. G. 1979, Opt. Commun., 28, 21
Yang, W., & Kostinski, A. B. 2004, ApJ, 605, 892
Zernike, F. 1934, MNRAS, 94, 377