

APPENDIX 4.7.6.A

DYNAMIC RANGE LIMITS ON LARGE SEGMENTED MIRROR TELESCOPES

Report prepared for the National Optical Astronomy Observatory, 2001.

Dynamic Range Limits on Large Segmented Mirror Telescopes: a Report for NOAO

Anand Sivaramakrishnan and Philip E. Hodge
Space Telescope Science Institute
3700 San Martin Drive, Baltimore, MD 21218

ABSTRACT

Segmented primary mirrors have become a mainstream concept for large telescopes of the future. Combining adaptive optics (AO) with segmented ground-based telescopes will enable higher resolution optical and near-infrared imaging than has hitherto been possible. The dynamic range limits of such telescopes then becomes an interesting topic. We develop simple analytic models which complement existing computationally demanding simulations of large segmented mirror geometries. Using these elementary models, we investigate some trade-offs in telescope concepts and their repercussions on dynamic range. Our models provide physical insight into these issues, and suggest directions for large telescopes optimized for high dynamic range. We describe Lyot and phase mask coronagraphs, and discuss their strengths and weaknesses, and outline nulling interferometric methods. We suggest future activities dedicated to understanding the choices that must be made to enable the next generation of large aperture ground-based telescopes to deliver scientifically interesting high dynamic range images. (Fri Oct 5 12:23:33 EDT 2001)

Subject headings: instruments: telescopes --- segmented: coronagraphs -- instruments: miscellaneous -- techniques: coronagraphy

1. INTRODUCTION

A perfect segmented mirror telescope is easily described using the techniques of Fourier optics. We present a simple description of an unobscured, square aperture telescope with square segments to illustrate the essential features of such optical systems. We examine the tradeoff between segment size and dynamic range, especially for coronagraphic applications.

We examine theoretical constraints on partially corrected images in order to understand the nature of high Strehl ratio images. Various approaches to reducing the effect of residual speckle noise in AO imaging are listed.

We discuss coronagraphy on such partially-corrected images, and collate the strengths and weaknesses of Lyot, phase mask and interfero-coronagraphy.

2. THEORY

2.1 The unobstructed aperture PSF

In order to simplify the model, we consider monochromatic imaging in this section. We first develop the aperture illumination function (or aperture function) of a square segmented unobscured aperture of side D . Our approach is to start with an infinite array of segments, and impose aperture boundaries by multiplying by a function that is unity within the aperture and zero elsewhere. We make use of standard Fourier analysis results, which can be found in e.g., Bracewell (1986). Multiplication becomes a convolution in the Fourier domain.

Square segments with side d are placed in a square, 2-dimensional lattice with spacing $b > d$, so the inter-segment gap width is $b - d$. The full aperture of such a telescope is given by

$$A_o = \Pi(x/D_1)\Pi(y/D_1) = \Pi(x/D_1, y/D_1), \quad (2-1)$$

where

$$\begin{aligned} \Pi(x) &= 1 \text{ for } |x| < 1/2, \\ \Pi(x) &= 0 \text{ elsewhere.} \end{aligned} \quad (2-2)$$

Here D_1 denotes the dimensionless quantity D/λ , and x and y are Cartesian coordinates in the aperture plane. The amplitude-spread function (ASF) a of an aperture is the Fourier transform of the aperture function. The point-spread function (PSF) p is the observable property, it is given by aa^* (where $*$ denotes the complex conjugate). A telescope aperture A_o produces an ASF

$$a_o = D_1^2 \text{sinc}(D_1 \xi, D_1 \eta), \quad (2-3)$$

where ξ and η are Cartesian coordinates in the image plane (in units of radians on the sky), and $\text{sinc}(x, y) = \text{sinc}(x)\text{sinc}(y)$.

We describe an infinite array of segments with the aperture function A_i , which we represent as a convolution of individual segment aperture functions $A_s = \Pi(x/d_s, y/d_s)$ with a two-dimensional shah function with periodicity corresponding to the segment spacing b :

$$A_i = \left(1/b_1^2\right) \text{III}(x/b_1, y/b_1) * A_s, \quad (2-4)$$

where $\text{III}(x, y) = \text{III}(x)\text{III}(y)$, $b_\lambda = b/\lambda$, $d_\lambda = d/\lambda$, and $*$ denotes the convolution operation. The shah function is defined as $\text{III}(x) = \sum_{-\infty}^{\infty} \delta(x-n)$. Normalization by $(1/b_1^2)$ is required to ensure unity power in each delta function in the shah function.

The segmented telescope aperture function is therefore

$$A = A_o A_i \quad (2-5)$$

By using the convolution theorem, the ASF of the segmented aperture telescope can be written

$$a = a_o * a_i \quad (2-6)$$

where a_o is given by equation 2-3, and a_i is the Fourier transform of A_i :

$$a_i = \text{III}(b_\lambda \xi, b_\lambda \eta) a_s \quad (2-7)$$

where the individual segment Fourier transform is:

$$a_s = d_1^2 \text{sinc}(d_\lambda \xi, d_\lambda \eta) \quad (2-8)$$

Combining equations 2-3, 2-7, and 2-8 we can write the ASF of the segmented unobscured square aperture telescope with no optical aberrations as:

$$a = D_1^2 d_1^2 \text{sinc}(D_1 \xi, D_1 \eta) * (\text{III}(b_\lambda \xi, b_\lambda \eta) \text{sinc}(d_\lambda \xi, d_\lambda \eta)) \quad (2-9)$$

We now examine the simple model's monochromatic ASF to bring out its principal features. If there are many segments across the full aperture ($D \gg b$), the full-aperture ASF is much narrower than λ/b . The ASF in equation 2-9 is a narrow, full-aperture ASF convolved with a lattice of delta functions, weighted by the single segment ASF, and spaced by the 'segment array resolution element' λ/b determined by the segment spacing. In this case, the function a_o is repeated at each grid point in the reciprocal lattice of the segment lattice, weighted by the function

$$\left(\frac{d_1^2}{b_1^2}\right)^2 \text{sinc}(d_1 n/b_1, d_1 m/b_1) \quad (2-10)$$

where (n, m) are the integer coordinates in the lattice of the ^2III function in image space. The peak field therefore d_1^2/b_1^2 that of an equivalent monolithic aperture.

A feature of this many-segment PSF is that along the axes of symmetry (the two Cartesian axes for the square aperture) the satellite images are a constant fraction $\epsilon^2 = (b-d)^2/b^2$ of the central bright image core, and off the symmetry axes, the satellite images are $(b-d)^4/b^4$ of the core. For a small segment gap, say 1% of the segment spacing, this uniformity holds within a radius of about 10 lattice spacings, or a distance $10\lambda/b$ from the center of the image.

As the segment gap w tends to zero, d/b approaches unity, and the locations of the zeroes of the weighting function 2-10 approach the locations of the delta functions in the lattice. In the limit this reduces to a single central PSF at the origin, without satellite images.

The consequences for high dynamic range imaging and coronagraphy are obvious from these analytical results: larger segments gaps contribute brighter satellite images around the core.

Figure 1 shows the effects of reducing gap size while keeping the number of segments constant, and Figure 2 shows the effect of increasing the number of segments while maintaining a constant inter-segment gap.

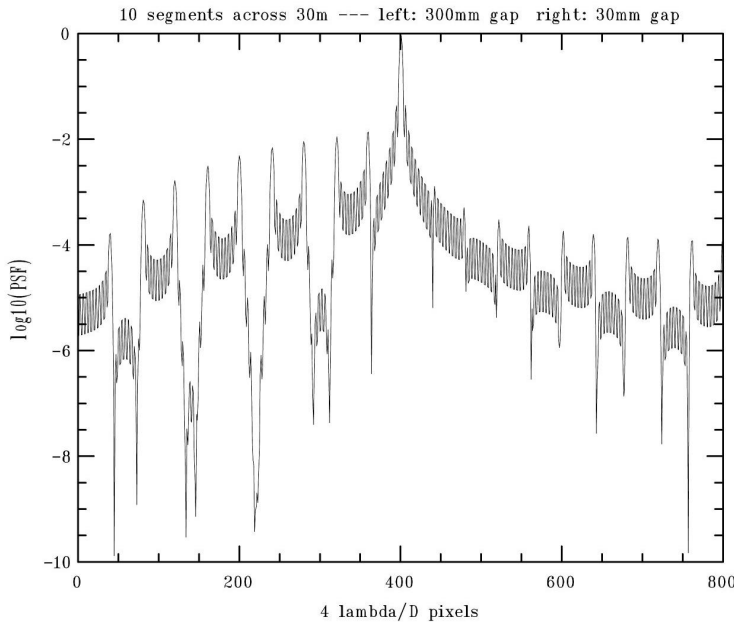
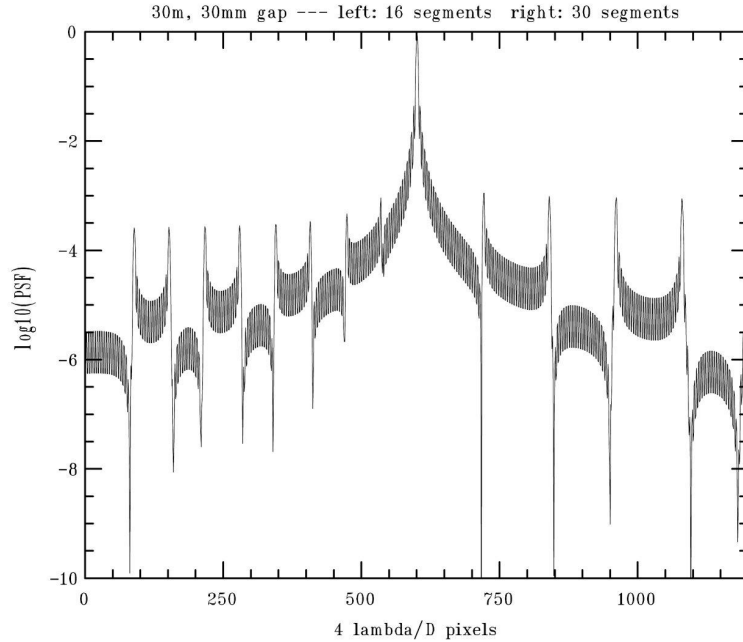


Figure 1: A cut across monochromatic point spread functions of two 30m square aperture configurations, showing the effect of varying the inter-segment gap width. Both configurations have 10 segments across their apertures. In case shown to the left of the central peak (l), the inter-segment gap is 300mm, right of the central peak (r) it is 30mm. Perfectly-corrected PSFs are shown with a spatial sampling of $\lambda/4D$. A common feature of the two PSFs is that there are 10 segments across the full aperture, so both PSF's exhibit the same satellite image spacing of $10\lambda/D$, where D is the full aperture diameter, and λ the wavelength. In (l) the inter-segment gap is 10% of the

segment spacing ($\epsilon = (b-d)/l$, where b is the segment spacing and d the segment size), so a 'beat periodicity' of $1/\epsilon = 10$ satellite image spacings describes the envelope of the PSF. In (r) the inter-segment gap is 1% of the segment spacing, so the envelope of the satellite images has a

period $1/\epsilon = 100$ times that of the satellite spacing. This periodicity is too large to be seen in the plot(b): it results in essentially uniform satellite image brightness around the core. This quantity ϵ is the dominant factor in the brightness ratio between the central image core and its neighboring satellite images. In the square aperture geometry considered here, satellite images along the symmetry axes are a factor of $1/\epsilon^2$ down from the central peak, and the satellites off the symmetry axes are down $1/\epsilon^4$ from the core. Thus in (a) there is a contrast ratio of 10^2 between the core and the first satellite images on the axes, in (b) the same ratio is 10^4 .

Figure 2: A cut across monochromatic point spread functions of two 30m square aperture configurations, showing the effect of varying the segment size. Both configurations have the same inter-segment gap spacing of 30mm. In the case shown to the left of the central peak (l), there are 16 segments across the aperture ($D/b = 16$), right of the central peak (r) has 30 segments across the aperture. Perfectly-corrected PSFs are shown with a spatial sampling of $\lambda/4D$. Case (l), with 16 segments across the aperture, has satellites spaced by $16\lambda/D = 64$ pixels. In this case $\epsilon = 1/67$, so the contrast ratio between the core and the surrounding satellite peaks is $1/\epsilon^2 = 4300$ along the axes (as shown in the figure), and the square of that off the axes. Case (r), with 30 segments across the aperture, has satellites spaced by $30\lambda/D = 120$ pixels. Here $\epsilon = 1/30$, so the contrast ratio between the core and the surrounding satellite peaks is about 900 on the axes and 800000 off the axes.



This description is not dependent on the particular symmetry of the lattice (i.e., square, hexagonal, rectangular, etc.), since it only depends on the fact that the aperture is a regular spatial lattice of segments all the way out to its edges. Kuhn et al. (2001) describes this approach for the particular case of hexagonal symmetry.

2.2. The effect of secondary obscuration

A secondary with a diameter D_s introduces structure on scales of $\lambda/D_{sec} = 1/D_{\lambda_{sec}}$ into the image. It is possible to carry out the above analysis with an aperture function

$$A_o = {}^2\Pi(x/D_l, y/D_l) - {}^2\Pi(x/D_{l_{sec}}, y/D_{l_{sec}}) \quad (2-11)$$

describing the aperture. The presence of a secondary introduces power in the form of the Airy pattern of the secondary. The amount of diffracted power is proportional to the secondary area, and the power is essentially put into an Airy pattern of the secondary shape. It is possible to match the first zero of the secondary's Airy pattern to avoid increasing background light around particular satellite images introduced by the segmenting of the aperture. Kuhn et al. (2001) show a few ways of selecting secondary sizes for a few different multi-mirror designs in order to optimize high dynamic range at particular separations from a bright source.

In section 5.2 we describe the interaction between a secondary obstruction and a diffraction-limited coronagraph.

3. THE ADAPTIVE OPTICS PSF

In order to understand the effect of partial correction of the wavefront by adaptive optics and its effect on the segmented telescope PSF, we first look at the effect of partial correction on the filled aperture ASF (equation 2-3).

3.1. The AO-corrected PSF for arbitrary apertures

The effects of scintillation at wavelengths longer than the I-band are negligible until the AO correction yields Strehl ratios of $\sim 96\%$, which is better performance than large segmented ground-based telescopes are expected to achieve in the next few decades. Angel (1992) derives the Strehl hit of the scintillation due to Kolmogorov-spectrum atmospheric turbulence localized at height h as being

$$0.291(\sqrt{1h}/r_o)^{5/3}. \quad (3-1)$$

If we assume the scale length of the turbulence, r_o (Fried 1966), is of the order of 30cm, and a turbulent layer at $h = 10\text{km}$ dominates the seeing, then at a wavelength of $1\mu\text{m}$ scintillation reduces the Strehl by 4%. This image quality reduction scales as $\lambda^{-7/6}$. We can ignore scintillation at wavelengths longer than about a micron, depending on local atmospheric conditions, unless we are in the realm of extreme adaptive optics.

We represent the effect of atmospheric wavefront aberration by multiplying the aperture function by $A_f = e^{if(x,y)}$. Phase variations induced by the atmosphere (or imperfect figuring) are described by a real wavefront phase function f . We think of A_o in this section as the filled aperture. The aperture function describing an aperture with pure phase disturbances is

$$A_{AO} = A_o A_f \quad (3-2)$$

with a corresponding ASF of $a = a_o * a_f$. For Strehl ratios of about 90% or higher, $A_f \sim 1 + if$, since $f(x, y) \ll 1$. Here we have subtracted of the mean value of f over the full aperture, since adding a constant offset f_o to the phase only multiplies the ASF by e^{if_o} , which has no effect on the PSF. The ASF of the AO-corrected full aperture is then

$$a_{AO} = a_o * (\delta(\xi, \eta) + i\Phi(\xi, \eta)), \quad (3-3)$$

or

$$a_{AO} = a_o + ia_o * \Phi \quad (3-4)$$

with Φ being the Fourier transform of the phase f . The assumption that f is a real function ensures that Φ is a Hermitian function. This does not hold true if there are atmospheric amplitude fluctuations, which would be described by an imaginary component of the phase function in this formalism.

The PSF corresponding to 3-4 is $a_{AO} a_{AO}^*$:

$$p_{AO} = a_o a_o^* + i(a_o^*(a_o * \Phi) - a_o(a_o^* * \Phi^*)) \quad (3-5)$$

to first order in Φ .

The first term, $a_o a_o^*$, is the perfect PSF for the full aperture, which is obviously a real function.

The second term, $i(a_o^*(a_o * \Phi) - a_o(a_o^* * \Phi^*))$, has been discussed in some detail by Bloemhof et al. (2001) in the special case of a symmetric aperture function A_o , where it has a much simpler analytical form. The term contains the imprint of the ASF a_o multiplicatively. Bloemhof et al. (2001) describe the effect of this term as 'speckle-pinning', and show that it results in a speckle pattern that bears the imprint of the Airy pattern on the speckle amplitude, thus 'pinning' the largest amplitude speckles to the first few Airy ring maxima. This high Strehl ratio scenario is significantly different from the concept of a diffraction-limited PSF superimposed on a smooth Gaussian or Moffat style envelope of speckle intensity distributions which might hold at low to moderate Strehls (Racine et al. 1999; Marois et al. 2000). In this pioneering work, Bloemhof et al. (2001) demonstrate the existence of this behaviour in K-band data taken with the Palomar Hale AO system, and suggest ways of using the phenomenon to increase dynamic range of AO imaging.

The theoretical analysis in Bloemhof et al. (2001) is easily extended to arbitrary aperture functions. This first-order speckle-pinning term can be written

$$2\mathbf{r}_{a_o} \mathbf{r}_{\Phi_a} \sin(\mathbf{q}_{a_o} - \mathbf{q}_{\Phi_a}) \quad (3-6)$$

(which is real, though not immediately useful). Since a_o and Φ_a are Hermitian functions (the convolution of two Hermitian functions is itself Hermitian), this term is antisymmetric, because it is the imaginary part of a Hermitian function, so it vanishes at the center of the image, i.e., at the central peak of the PSF for typical apertures. Whatever this term adds to the stellar intensity at some location $s = (\xi, \eta)$ in the image it removes at the diametrically opposite place $(-\xi, -\eta)$ in the PSF. The existence of this antisymmetry could be an aid to reducing high Strehl data. It does not affect the Strehl ratio, and it merely rearranges power within the peaks of the brightest Airy rings. Further work on such analyses is being undertaken by various groups (e.g.,), Riaud et al. (2002), Sivaramakrishnan et al (2002)).

4. CORONAGRAPHY ON MODERATE STREHL IMAGES

Simple, single-channel coronagraphic imaging is advantageous when the halo drops below the Airy rings or other structures produced by the diffraction-limited part of the image, since a coronagraph suppresses only the corrected part of the image. In section 3.1 we showed that an image with a Strehl ratio above $\sim 90\%$ can be decomposed into a diffraction-limited intensity profile, a 'pinned speckle' pattern which follows the diffraction-limited image's amplitude, and a broad speckled halo. Considerable work is under way investigating various ways of combating the deleterious effects of residual speckle on high dynamic range imaging (e.g., Racine et al. (1999), Marois et al. (2000), Bloemhof et al. (2001), Kuhn et al. (2001)), and there has been some progress extending this kind of analysis to Strehl ratios around $\sim 50\%$ (Sivaramakrishnan et al 2002).

The interaction of pupil geometry with various styles of coronagraphic optics is also an important issue. In this section we outline different forms of coronagraphy, and note their applicability to large segmented telescopes in the case of single channel imaging.

All coronagraphs create an image at a focal plane of the telescope, and perform a first operation on that image. This modified initial image is then relayed through a pupil plane (called the Lyot pupil plane in both types of coronagraphs), and a second operation is performed on the incoming wave at that plane before it is finally re-imaged to a focus on a detector or spectrograph slit.

In a Lyot coronagraph, the first operation (on the image) is that the core of the image is blocked by a mask which does not affect the phase. The mask can be hard-edged or graded. A phase mask coronagraph places a mask that affects the phase of the electric field in this image plane rather than the intensity. Part of the core of the image, or two of four quadrants of the image field, can have their field reversed (that is, a phase shift of π introduced). Both coronagraphs use an intensity-only mask in the second operation in the Lyot pupil plane.

Lyot coronagraphy between separations of $\sim 2.5\lambda/D$ and θ_{AO} around a bright object is best suited to Strehl ratios above 80% (Sivaramakrishnan et al. 2001). Here

$$\theta_{AO} = N_{act} \lambda/2D. \quad (4-1)$$

for typical Shack-Hartmann AO systems with N act actuators across the pupil. See the quoted reference for more discussion on this quantity for curvature sensing AO systems. Figure 3 (taken from Sivaramakrishnan et al. (2001)) demonstrates this.

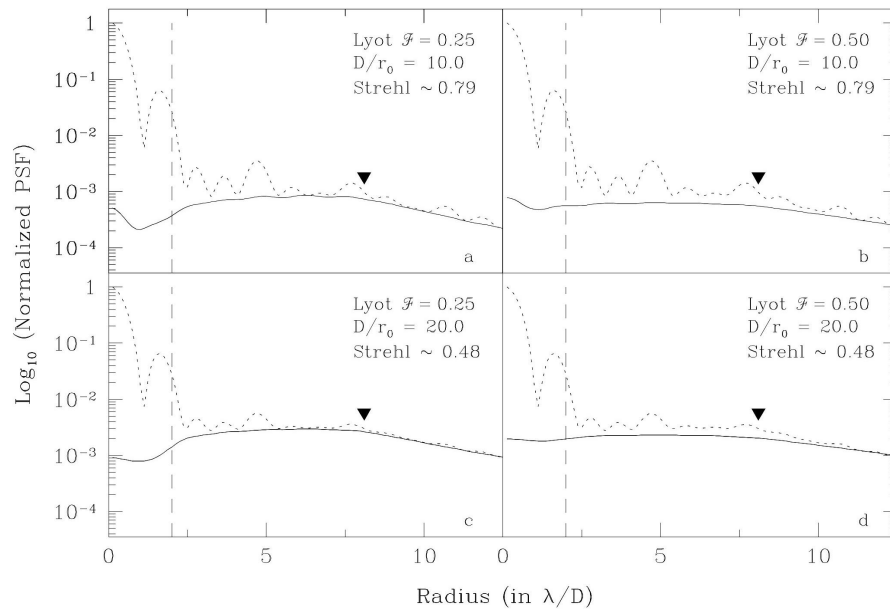


Figure 3: Azimuthally averaged PSF's for the 5m Palomar Hale telescope, with 16 actuators across the pupil diameter (~ 216 actuators within the pupil) and an occulting image stop of $4\lambda/D$ ($0.''27$ at $\lambda = 2.2\mu\text{m}$). Two Lyot stops, the higher throughput one on the left and the lower throughput on the right are shown. The solid line shows the AO-corrected PSF, the dashed line shows the image profile after the coronagraph. The latter has been renormalized to account for the reduction in effective aperture size by the Lyot stops. These are derived from numerical simulations described in Sivaramakrishnan et al. (2001)

Phase mask coronagraphy outside $\sim \lambda/D$ works best with Strehl ratios above $\sim 95\%$ (Roddiier & Roddiier 1997; Rouan et al. 1999). However, using dual channel or multiple band simultaneous imaging (Marois et al. 2000; Kuhn et al. 2001) and some prior knowledge of the nature of the AO-corrected PSF (Bloemhof et al. 2001), it may be possible to extend the applicability of coronagraphy as a high dynamic range technique to lower Strehl ratios to investigate diffuse structure and faint companions within a few diffraction widths of bright stars. Recent data have shown that dual channel polarimetry is an effective antidote to speckle noise (Kuhn et al. 2001).

5. LYOT CORONAGRAPHY ON SEGMENTED APERTURES

Lyot developed a coronagraph (Lyot 1939) to block out the solar disk in order to study the Sun's corona. The advent of adaptive optics made much smaller occulting stops astrophysically interesting (e.g., Golimowski et al. (1992); Nakajima et al. (1995), and many subsequent instruments). The principles of Lyot coronagraphy as applied to ground-based, diffraction-limited imaging are described in detail by Sivaramakrishnan et al. (2001), as well as in several references contained therein.

A telescope well-suited to Lyot coronagraphy has an unobstructed aperture or a small secondary obscuration (~ 10% or less linearly for ground-based AO with Strehl ratios under ~ 90%). This could be at odds with the requirements of wide field imaging. If the primary is segmented, segment gaps should be as small as possible, and, as Kuhn et al. (2001) stress, segments need to retain good figure out to their very edges, and be as large as possible. We quantify the amount of diffracted power from segment gaps in section 5.1.

At moderate Strehl ratios (70%), the benefits produced by AO coronagraphy on Keck are probably not significant enough to make the segment gap diffraction the limiting factor in dynamic range, so it is unlikely that segment gaps and spiders need to be hidden by oversized Lyot stops at these Strehls. While quantitative estimates or experiments are not available, one might expect that as Strehl ratios climb above 90%, segment gap diffraction may start to limit dynamic range. Systematic quantitative investigations of these regimes has not been undertaken.

Diffraction-limited Lyot coronagraphs are robust to guiding errors on the scale of a resolution element, though they achieve this at the cost of search area within a few diffraction widths of the central source. They are also relatively insensitive to asymmetries in the PSF of the telescope.

Their tolerance to wavefront sensing and reconstruction artifacts is still under investigation.

5.1. Interaction between segment size and occulting spot diameter

Here we explain the relation between the number of segments across the pupil diameter and the size of the smallest effective field stop in a Lyot coronagraph.

The number of segments across the aperture determines the spacing of the grid of 'satellite' images (which will be located in the reciprocal lattice of the segment location lattice, to use crystallographic terminology). For example, if there are 20 segments across the primary, satellite images will be located in a grid with an angular spacing $20\lambda/D$ in the square segment case (see section 2.1).

For obstructed apertures, any pupil rotation will have to be matched by rotation of the Lyot stop within the instrument. It is possible to estimate how much energy segment gaps diffract into the pupil. The electric field in the Lyot plane is described by

$$A(x, y) - A(x, y) * W(x, y) \tag{5-1}$$

where A is the aperture illumination function, and W is the Fourier transform of the function w which describes the field mask --- the transmission function of the field mask is written as $1 - w(\xi, \eta)$ (we follow the notation of Sivaramakrishnan et al. (2001) in spirit, although we omit their potentially confusing rescalings of variables in this article). In a traditional Lyot coronagraph, w is a positive function with bounded support, viz., the field stop is a small piece of opaque material. Such a function has a transform W which is infinite in extent (it has unbounded support). If s is some measure of the width of the occulting stop in units of λ/D , the expression 5-1 never quite vanishes everywhere in the Lyot plane, although the energy is mostly restricted to a strip $\sim D/s$ wide around the edges of the pupil. This distance D/s is a critically important length scale in the pupil planes in a Lyot coronagraph. Makidon et al. (2000) showed that in a diffraction-limited Lyot

coronagraph s must be ≥ 4 for optimal performance (tables 2, 3 and 4 in that work). Smaller occulting stops necessitate more severe Lyot stops, so throughput of off-axes sources drops precipitously with decreasing stop size. This is because the light that is not blocked by the occulting spot distributes itself in a band about D/s wide around all aperture edges: most notably the inner edge of a secondary obscuration and the outer edge of the primary. In order to block this light, the Lyot stop occults these two strips. An unobscured aperture can therefore have a somewhat smaller focal plane stop ($s \sim 3$) before losing throughput entirely. Makidon (2001) has developed a systematic method of optimizing the Lyot stop for any particular pupil geometry, given the expected Strehl ratio.

A gap or obstruction L long and T wide ($L/D \sim 1$, $T/D \ll 1$) in the aperture leaves an imprint on the electric field in the Lyot plane: the original gap (i.e., the range of zero values describing the gap in the function $A(x, y)$), less the convolution of the gap shape with the function W , which has unit height, and a linear size of approximately D/s where it is non-negligible.

When the gap length L is $\sim D/s$ or longer, that convolution produces an electric *field* that spreads D/s around the linear obstruction, so it covers an area of $\sim LD/s$. The field over this area is a fraction $\sim sT/D$ of the original field, so the power diffracted by the coronagraph into clear pupil area is of the order of

$$(T/D)^2(LD/s) = s(T/D)(LT) \quad (5-2)$$

per unit incident power per unit area in the entrance aperture. Thus the ratio of spider-diffracted power in clear pupil to incident power is approximately

$$s(T/D)(LT)/D^2 = f_g s T/D, \quad (5-3)$$

where f_g is the fraction of the entrance pupil covered by segment or spider gaps.

We stress that equation 5-3 only holds for segment gaps or spiders longer than D/s . The estimate applies to the fraction of incident energy contained in the diffraction-limited image, i.e., the Strehl ratio S . Numerical estimates using half the sum of the total segment perimeters for L in equation 5-3 can be obtained for particular geometries. This fraction of incident power, $S f_g s (T/D)$, is spread over a scale of sL/D . It becomes a limiting factor in the dynamic range when it compares with the surface brightness of the complement of the Strehl ratio, $(1 - S)$, spread over the halo in a residual speckle pattern, since this is the fraction of incident energy that a coronagraph cannot correct.

Equation 5-1 describing the Lyot field also describes phase mask coronagraphs, a fact which we use in section 6.

5.2. Interaction between secondary obstruction and occulting spot diameter

The field stop of a Lyot coronagraph has some angular diameter, which we characterize in terms of the resolution of the optical system:

$$q_{FS} = sL/D. \quad (5-4)$$

If the secondary diameter D_{sec} is much less than D/s , we can estimate how much power is diffracted by the secondary into clear pupil area in the Lyot plane.

By inspection of the expression 5-1 for the Lyot stop field, and the use of simple geometrical estimates of the convolution $A * W$ in it, The field just outside the secondary obstruction scales as $D_{sec}^2/(D/s)^2 = s^2 D_{sec}^2/D^2$. This is the factor by which field strength is reduced from unity by $A(x, y) * W(x, y)$ in equation 5-1.

The total power scattered into clear pupil is the product of the area of the disk where secondary-diffracted field contributes significant energy and the square of the field strength:

$$(D/s)^2 \cdot (s^4 D_{\text{sec}}^4 / D^4) = s^2 D_{\text{sec}}^2 (D_{\text{sec}} / D)^2 \quad (5-5)$$

for unit incident flux. Without a coronagraph D^2 energy per unit incident flux is placed in the final image. If we neglect Lyot throughput losses, the ratio of secondary-diffracted power to the non-coronagraphic image power is therefore $s^2 D_{\text{sec}}^4 / D^4$. If $s = 4$ and $D_{\text{sec}} = D/16$, this ratio is $1/16^3$. This is the ratio of powers in a perfectly corrected image. With partial AO correction, the secondary still diffracts the perfectly corrected portion of light this way, so at some Strehl ratio S , the amount of power put into clear pupil by the secondary is $S s^2 D_{\text{sec}}^2 (D_{\text{sec}} / D)^2$. How this power compares to the residual speckle in the halo depends on the statistics of the intensity distribution in the halo.

From the above discussion we see that secondary diffraction in a Lyot coronagraph does not matter much until the residual image halo drops to a brightness comparable with the power diffracted by the secondary.

When image quality good enough to worry about oversizing the central obstruction of the Lyot stop, particular science goals and through-put levels will have to be considered separately in numerical experiments to choose a suitable Lyot stop geometry.

We note that these arguments render the optimization of the Lyot stop described in the work of Sivaramakrishnan et al. (2001) a little more pessimistic than necessary for secondary obstructions that are much smaller than D/s . These authors used a uniform width of oversizing on both the outer and inner edges of the annular entrance pupils of the Calypso, Palomar and Gemini telescopes in their numerical experiments. Their results hold for the larger secondary obstruction telescopes, though their stop size recommendations for Gemini might be slightly too cautious for Lyot coronagraphs.

Perhaps the real drawback of secondary obstructions in AO systems delivering 10-90% Strehl images is the introduction of unnecessary unsensed modes in the wavefront reconstructor because of gaps in the wavefront slope data behind the obstruction. In Shack-Hartmann wavefront sensing AO systems, this can result in increased rank deficiency in the AO control matrix, which often translates to non-negligible power near the image core, as the data in e.g., Oppenheimer et al. (2000) show. There may be similar undesirable effects introduced by the extra free edge of the secondary obstruction when solving the boundary value problem in curvature-sensing AO systems.

6. PHASE MASK CORONAGRAPHY

In a phase mask coronagraph the intensity of the image produced by the telescope is left untouched, but a phase mask is inserted to reverse the sign on the electric field vectors over selected areas of this image plane. The principal feature of these phase masks is that they scatter light outside most of the geometrical pupil area in the following pupil plane, where a pupil stop (still known as a Lyot stop) is located. In order to attain this end, such phase masks must contain scale sizes less than λ/D , or no characteristic scale at all (to well below the resolution limit of the optics). A very slow beam is therefore desirable at the first image plane in order to fabricate phase masks within the spatial tolerances required by such instruments. Since Airy disk size is on the order of $f\lambda$ (f being the focal ratio of the imaging beam), occulting spots of order 100 microns require imaging f -ratios of 50-100 or more in the K-band.

The need to effect a sign reversal on parts of the wavefront places limits on the bandpass of phase mask coronagraphs, given current technology. However, efforts are being directed at developing masks that introduce the requisite π phase shift over wider bandpasses than are possible at present (Boccaletti 2001; Helmbrecht 2001). Bandpasses of 20nm at a wavelength of $1\mu\text{m}$ are quoted in the literature. It is hard to predict how fast technology for fabricating wider bandpass phase masks will mature.

There are uncountably many ways of selecting where to reverse the field sign (i.e., introduce a phase shift of π). A choice must be made as to how to arrange these areas by requiring that in the Lyot pupil plane, energy from the on-axis source is segregated enough that it can be blocked by a partially opaque Lyot mask that lets significant amounts of off-axis light through.

Phase mask coronagraphs suppress light within a few resolution elements of the central source, so they reach areas inaccessible to Lyot coronagraphs. Guiding errors on the scale of the diffraction limit of the telescope are damaging to on-axis image suppression. Conversely, off-axis sources that produce images away from the phase-reversal edges of the image plane mask will not be affected much at all by the mask (Rouan et al. 1999). Numerical simulations by Rouan et al. (1999) indicate that 8 magnitudes of dynamic range are achievable with a 3.6m telescope with a 25% central obstruction delivering a Strehl ratio of 80% at a separation of $3\lambda/D$ under realistic conditions, around a 5th magnitude AO guide star.

6.1. Proof of concept for phase mask coronagraphy

The first such instrument design (Roddiier & Roddiier 1997) reverses the electric field sign on about half of the central Airy peak. The optimal phase mask diameter is about 43% of the first Airy ring diameter in the case of an unobstructed aperture. The phase mask size is dependent on the wavelength, so such designs are intrinsically monochromatic or narrow-band unless a special effort is made to address this issue. Roddiier & Roddiier (1997) suggest devising optics to produce a wavelength dependent plate scale in order to increase the bandpass of this design.

6.2. Quadrant phase mask coronagraphy

Rouan et al. (1999) modified the phase mask design by causing a phase reversal over opposite quadrants (centered on the on-axis image). This does away with the need for wavelength-dependent magnification. Phase mask fabrication for this design obviously requires a clean edge between phase shifting and non-phase-shifting quadrants. The tetrad symmetry of the mask limits it to use on telescopes which produce the same or lower (tetrad, dyad or rotational) symmetry in their ideal PSFs, though the mask design can be easily modified to accommodate other symmetries in telescope apertures. Symmetry in the PSF is an essential requirement for the original Roddiier design as well as for quadrant or related hexagonal phase mask coronagraphs. If pupil rotation occurs the coronagraph will have to follow that rotation in order to align the phase mask symmetry with the PSF's symmetry.

The quadrant phase mask multiplies the image field by ${}^2\text{sgn}(\xi, \eta) = \text{sgn}(\xi)\text{sgn}(\eta)$, where $\text{sgn}(x)$ is 1 for $x > 0$ and -1 for $x < 0$. In the Lyot plane this results in a field that is the original aperture function convolved with the transform of ${}^2\text{sgn}(\xi, \eta)$, i.e., $-1/(\pi^2xy)$. The nature of this Lyot plane field $L_{QPMC}(x, y)$ is best seen in the case of a square, unobstructed aperture of side $D_\lambda = D/\lambda$ (pupil coordinates are in units of wavelength in this section). The convolution is antisymmetric, we calculate it for the $x > 0, y > 0$ quadrant here.

$$-\pi^2 L_{QPMC} = \ln(F(x)) \ln(F(y)) \quad (6-1)$$

where F is a real positive function defined by

$$F(x) = (x + D_\lambda/2)/(x - D_\lambda/2) \quad (6-2)$$

when $x \geq D_\lambda/2$, and

$$F(x) = (D_\lambda/2 + x)/(D_\lambda/2 - x) \quad (6-3)$$

when $0 \leq x < D_\lambda/2$.

Along the coordinate axes L_{QPMC} is zero, since it is antisymmetric. As either x or y approach $\pm D_\lambda/2$, (the primary pupil boundary), L_{QPMC} tends logarithmically to ± 8 at locations away from the axes. Its sign depends on the quadrant.

As can be seen in Figure 4, the Lyot plane field intensity decreases rapidly within the pupil for a point source whose image is exactly centered on a perfect quadrant phase mask. The spatial scale of this drop is on the order of wavelengths. This is because there is no inherent spatial scale in function we use to represent the quadrant phase mask.

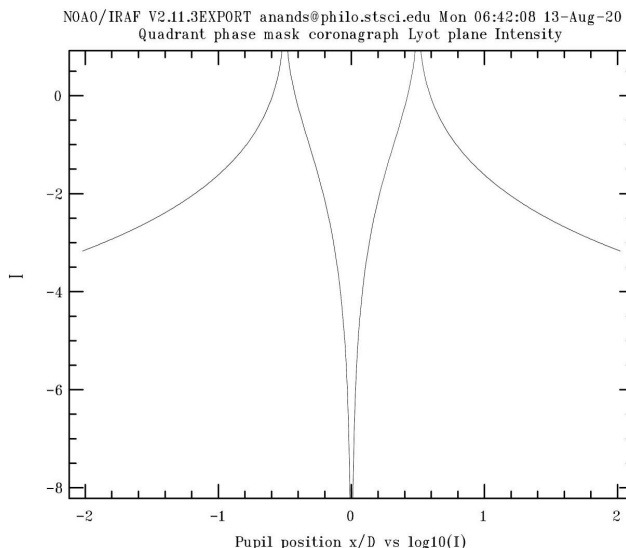
The phase mask function $^2\text{sgn}(\xi, \eta)$ is a distribution or generalized function (Lighthill 1986) rather than a traditional function with finite power and a finite Riemann integral.

Rouan et al. (1999) have proposed a hybrid Lyot--quadrant phase mask coronagraph to improve the design's sensitivity to tilt.

An unobstructed aperture is best for such a design: a central obstruction diffracts on-axis light into the clear area of the reimaged primary in the Lyot plane. An achromatic phase mask is obviously very desirable to reach reasonable survey depths with such an instrument.

The quadrant phase mask coronagraph is well-suited to a high-Strehl off-axis monolithic AO-optimized telescope, or a space telescope. How segment gaps affect the performance of such an instrument is still under investigation, although undoubtedly proponents of the design are addressing this question.

Figure 4: Intensity in the Lyot pupil plane for a quadrant phase mask coronagraph: the pupil intensity across a diameter near the middle of the Lyot pupil plane for a QPMC on a square unobstructed aperture telescope with a 100% Strehl ratio (the field strength is identically zero right across the middle). The entrance pupil is only one wavelength across. This didactic figure shows that the drop in electric field in clear pupil occurs on a spatial scale of wavelengths. The logarithm (base 10) of the intensity is plotted (in arbitrary units) as a function of position across the pupil plane. The square entrance pupil is between ± 0.5 on the abscissa. Most of the energy incident on the aperture gets diffracted by the image plane phase mask to outside the image of the primary in the Lyot plane. On circular apertures Rouan et al. (1999) calculate a Lyot stop diameter that is 95% of the entrance aperture will deliver 10^5 on-axis attenuation. Any aperture boundaries (such as spiders or secondary obstructions) exhibit the same kind of weak logarithmic infinities as are visible at $x = \pm 0.5$. Such infinities are not fatal to the utility of these coronagraphs.



7. NULLING CORONAGRAPHS

Nulling coronagraphs or interfero-coronagraphs split an incoming collimated beam into two equal intensity components, modify one component, and recombine the beams to cause a central source to interfere destructively with itself (e.g., Baudoz et al. (2000)). A pupil flip or rotation is usually the operation performed on one arm. Off-axis sources produce negative and positive (field strength) images as symmetrically-placed satellites of the central nulled source.

Typically the incoming beam is collimated, split, and one arm is sent through a focus whereas the other arm is not. Alternatively, a pupil flip can be accomplished with a cat's eye reflector. Passing through a focus introduces an achromatic π phase shift over the central peak of the Airy pattern, and successive Airy rings suffer more phase shift than just π (Born & Wolf 1993). Recombining the two arms with a second beam-splitter nulls out just the core of the on-axis image. The optical path delays of both arms must be the same --- the instrument operates in 'white light' mode. In order to collect as many photons as possible, two detectors can be placed in each arm of the final beam recombination stage.

There are a few different nulling designs in the literature: Figure 5 shows an all-reflective nulling interferometer that was designed to fit the Gemini South Cassegrain focus (Winsor 2001).

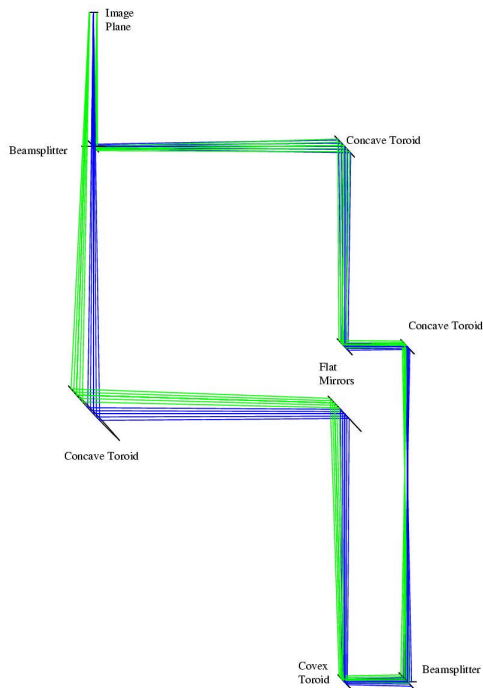


Figure 5: An all-reflective nulling interferometer designed to fit the Gemini South Cassegrain focus (Winsor 2001). The incoming converging beam comes into the bottom right corner, where it is split: the right vertical beam then goes through a focus where the phase changes by π , the left beam does not suffer a phase reversal in its relay, and the beams are then recombined top left at identical f-ratios. On-axis light is shown in blue, off-axis in green.

When the collimated beam passes through focus, the pupil downstream of this focus is flipped with respect to the pupil before focus. This is why the PSF must be symmetric: in point of fact it is the aberration across the collimated beam that must be symmetric. On an alt-az telescope the interfero-coronagraph would have to accommodate pupil rotation somehow.

The optics required by such an instrument are not hard to fabricate for relatively undemanding applications (such as proof of concept), though for extremely high dynamic range imaging such as

terrestrial planet finding from space-based telescopes, systematics start to become uncontrollable. The optical trains of such instruments are inherently more complex than those of phase mask coronagraphs because of considerable non-common-path optics. An alternative approach in very demanding applications is a true nulling imaging coronagraph such as the bandwidth-limited coronagraph (Kuchner & Traub 2001), which eschews non-common-path optics.

8. RECOMMENDATIONS AND FUTURE DIRECTIONS

We summarize the principal points of our work briefly in this section. Following the work of Chanan et al. (2000) and Dekany et al. (2000), we assume that there will be situations when the Strehl ratio of a 30m segmented telescope will be high enough to merit serious thought about high dynamic range imaging. If telescope design decisions are made without considering observing longwards of the K-band, high dynamic range science may be completely precluded *ab initio*.

Whether high dynamic range and wide-field studies can be pursued profitably with the same telescope is a question that springs immediately to mind.

8.1. Understanding the signal-to-noise ratio

A study of the signal-to-noise ratio due to partially corrected wavefronts in 5-90% Strehl regimes will enable a more realistic assessment of the science possible with large segmented mirror telescopes. Over the last few years there has been significant progress in understanding the structure of an AO-corrected PSF, and effort is being spent on developing observational techniques to be used with AO systems. More work on this is still required before an unbiased choice between various large telescope designs can be made. Symmetries in the partially corrected images are being studied (Riaud et al. 2001, Sivaramakrishnan et al. 2002), and ground-based data is being examined with these in mind (Bloemhof et al. 2001).

A quick analysis involving a perfect segmented telescope PSF weighted by its expected Strehl ratio, combined with ad hoc models of the uncorrected part of the image might be a profitable way to study dynamic range limits in a practical way.

The effect of segment phasing error on high dynamic range needs to be quantified with numerical experiments and real data. This error might well dominate the dynamic range achievable with a many-segmented aperture.

8.2. Aperture geometry and segment size and quality

High dynamic range imaging requires as small a secondary obscuration as possible. Telescopes designed for the highest dynamic range applications are often off-axis designs. The secondary obscuration fraction and the segment size impact the smallest usable occulting stop in a Lyot coronagraph. Segment figure needs to be as good as possible out to the segment edge, and segment gaps should be minimized. The choice of secondary diameter should be made only in conjunction with knowledge of the segment size. Many of these issues are described in a recent study (Kuhn et al. 2001). Many of these requirements coincide with thermal IR requirements.

In bandpasses longward of K and L, when Strehl ratios might be expected to approach the 90-95% regimes (e.g., Dekany et al. (2000)), spider diffraction may start to become relevant.

A more quantitative analysis of segment gap diffraction needs to be done to understand the break-even Strehl ratios at which a given segmented aperture design produces equal amounts of uncorrected halo as segment-diffracted light in a coronagraph. In sections 5.1 and 5.2 we considered Lyot coronagraphs: the break-even Strehl ratio might be different for phase mask coronagraphs. Similar analytical or numerical work on phase mask coronagraphs and an investigation of hybrid Lyot/phase mask coronagraphs would enable objective, scientifically-driven choices to be made regarding the design of the new generation of extremely large telescopes.

On a more fundamental note, it is apparent that other telescope configurations dedicated to high dynamic range studies should be studied alongside the segmented filled aperture CELT or GSMT designs before decisions can be made about optimizing a possibly wide field telescope for high dynamic range imaging.

8.3. Key technologies

The ability to align and co-phase multiple segment apertures underlies the entire concept of high dynamic range imaging with a segmented telescope, and is therefore a core technology.

Low noise wavefront sensing and stable wavefront reconstructors are needed for the next step in ground-based AO. Natural guide star AO is often photon-starved, so sensor read noise affects survey limits drastically.

Phase mask fabrications, especially wide band-pass phase masks, are a key technology for the next steps in coronagraphy. Coronagraphy with multiconjugate AO (MCAO) is something that will need study once the practice of MCAO starts to mature, and these systems are built and brought into the mainstream of AO instrumentation.

8.4. Theoretical adaptive optics considerations

Furthering our understanding of the role of a secondary obscuration in wavefront reconstruction is important --- wavefront sensor geometry also enters into this at a fundamental level. While optimised reconstructors may minimize L_1 , L_2 or other measures of the fitting error (using the L^2 norm is the popular minimum variance optimization), they can only guess at unsensed and therefore in principle uncorrectable modes (such as the fundamental waffle mode of the square Fried geometry AO systems) in the wavefront. Directions set by the late Walter Wild need to be followed and his work applied and extended in order to make AO systems operate more effectively.

More study is needed to understand the factors that come into play when moderate Strehl ratio images are delivered to quadrant or other phase mask coronagraphs.

8.5. Practical adaptive optics considerations

Details of the interaction between alt-az telescope pupils and diffraction-limited coronagraphs must be borne in mind when choosing a particular coronagraph design for any given telescope design. Development of wide field MCAO, as well as understanding how to maintain good segment phasing, are of obvious value to high contrast imaging with a segmented telescope.

Extending the range of large segmented telescopes longwards of the K-band eases requirements on phasing and wavefront error, so exploring this avenue is potentially very profitable for high dynamic range on GSMT. Narrow-line observations might be the first possible science observations at mid-IR wavelengths.

9. ACKNOWLEDGEMENTS

We are indebted to R. J. Allen, E. E. Bloemhof, A. Boccaletti, J. E. Krist, J. P. Lloyd, R. B. Makidon, B. R. Oppenheimer and A. T. Tokunaga for much insightful discussion, and we thank J. R. Najita for asking questions that sharpened the analyses undertaken during this study, and B. Gregory for a thorough reading of the manuscript. Numerical work was performed in the Python language, using Python modules developed at STScI (Hsu et al. 2001). This work was funded in part by the NIO office of NOAO, the STScI Director's Discretionary Research Fund, and the STScI Research Programs Office.

REFERENCES

- Angel, J. R., 1994, *Nature*, 368, 203
- Bloemhof, E. E., Dekany, R. G., Troy, M. & Oppenheimer, B. R. 2001, *ApJ Letters*, 558, L71
- Baudoz, P., Rabbia, Y., & Gay, J. 2000, *A&A*, 131, 319
- Boccaletti, A. 2001 private communication
- Born, M. and Wolf, E. 1993 *Principles of Optics*, 6th ed. (CUP; Cambridge)
- Bracewell, R. N. 1986, *The Fourier Transform and its Applications* (McGraw Hill; London)
- Chanan, G. A., Nelson, J. E., Ohara, C. & Sirko, E. 2000, in *Telescope Structures, Enclosures, Controls, Assembly/Integration/Validation and Commissioning*, Sebring & Andersen, eds. *Proceedings of SPIE*, Vol. 4004
- Dekany, R. G., Nelson, J. E., & Bauman, B. 2000, *CELT Reports and Technical Notes*, 4
- Fried, D. L. 1966, *JOSA*, 56, 1372
- Helmbrecht, M. 2001 private communication
- Hsu, J.-C., De La Pena, M. D., Barrett, P. E., Hodge, P. E., Hack, W. J. (2001) *Py-FITS: A Python module for FITS file I/O* (<http://ra.stsci.edu/pyfits/>)
- Golimowski, D. A., Clampin, M., Durrance, S. T., & Barkhouser, R. H. 1992, *Appl. Opt.*, 31, 4405
- Kuchner, M. J., & Traub, W. A. 2001 submitted, *ApJ*
- Kuhn, J. R., Moretto, G., Racine, R., Roddier, F. & Coulter, R. 2001, submitted, *PASP*
- Kuhn, J. R., Potter, D., & Parise, B. 2001 astro-ph/0105239
- Lighthill, M. J. 1958, *Introduction to Fourier Analysis and Generalised Functions* (Cambridge University Press; Cambridge)
- Lyt, B. 1939 *MNRAS*, 99, 580
- Makidon, R. B. 2001 private communication
- Makidon, R. B., Sivaramakrishnan, A., Koresko, C. D., Berkefeld, T., Kuchner, M. J. and Winsor, R. 2000, in *Adaptive Optical Systems Technologies*, P.L. Wizinowich, ed. *Proceedings of SPIE*, Vol. 4007
- Marois, C., Doyon, R., Racine, R. & Nadeau, D. 2000, *PASP*, 112, 91
- Nakajima, T., Oppenheimer, B. R., Kulkarni, S. R., Golimowski, D. A., Matthews, K., & Durrance, S. T. 1995, *Nature*, 378, 463
- Oppenheimer, B. R., Dekany, R. G., Hayward, T. L., Brandl, B., Troy, M., & Bloemhof, E. E. 2000, in *Adaptive Optical Systems Technologies*, Wizinowich, ed. *Proceedings of SPIE*, Vol. 4007
- Racine, R., Walker, G. H., Nadeau, D., Doyon, R. & Marois, C. 1999, *PASP*, 111, 587
- Riaud, P., Boccaletti, A., Rouan, D., Lemarquis, F. & Labeyrie, A. 2001, *PASP* 113, 1145
- Roddier, F. & Roddier, C. 1997, *PASP*, 109, 815
- Rouan, D., Riaud, P., Boccaletti, A., Clenet, Y. & Labeyrie, A. 2000, *PASP*, 112, 1479
- Sivaramakrishnan, A., Hodge, P. E., Makidon, R. B., Bloemhof, E. E. & Lloyd, J. P. 2002, in preparation, *Proceedings of SPIE (Kona)*
- Sivaramakrishnan, A., Koresko, C. D., Makidon, R. B., Berkefeld, T. & Kuchner, M. J. 2001, *ApJ*, 552, 397
- Winsor, R. S. 2001 private communication