Spatial Filters for Astronomical Interferometry using Discrete Optical Components

Jeffrey A. Meisner[1]
Sterrewacht Leiden

Abstract. Spatial filtering of received optical radiation is used to select a single spatial mode from which interference can be obtained. By constraining the radiation to a single mode, as is done by the antenna in radio interferometry, a definite relationship is established between the initial power in the interfering beams, the strength of the interference signal, and the underlying correlation (fringe visibility), thus providing a means of calibration which is insensitive to variations in the atmosphere. In the case of an infrared interferometer operating at longer wavelengths where the system noise level is dominated by background (thermal) radiation, the spatial filter serves a second purpose: that of rejecting (as much as possible) background radiation responsible for the noise level, while accepting most of the coherent power in a single mode. These criteria are quantified in order to evaluate the performance of proposed spatial filter designs.

While a single-mode waveguide (optical fiber) has excellent properties in relation to both criteria, it may not be the ideal solution in all cases, particularly at longer wavelengths where suitable materials may be unavailable. Instead, concepts for spatial filtering using discrete optical components are analyzed. Although the proposed configuration is highly wavelength dependent, one possibility for a wideband implementation is discussed as well.

1. Introduction

There are two motivations for spatial filtering in optical stellar interferometry. The first has been demonstrated by the FLUOR project, which spatially filters the light from each arm of the interferometer using optical fibers. By accepting only the light in the central mode, the calibration of the interferometer is enhanced, since the observed visibility will be given by the underlying visibility multiplied by the average geometric mean of the powers in the two optical signals, regardless of how much optical power had been coupled into other modes and rejected by the spatial filters. Simultaneous measurement of the photometric level of the two input beams will then provide calibration of the visibility which is insensitive to changes in the magnitude of atmospheric turbulence.

The second purpose of spatial filtering occurs whenever the system noise is not detector limited. In that case, reduction or elimination of light not present in the central mode reduces the system noise level without (greatly) reducing the strength of the interference which is overwhelmingly dominated by interference in the central mode. In the case of source noise limited operation, reduction of the total amount of light provides a decrease in noise according to the square root of the portion of light accepted by the spatial filter. In the case of background limited detection, the advantage is more profound, since a constant amount of background light is present in every mode accepted by the instrument, contributing a noise amplitude, therefore, proportional to the square root of the number of modes contributing to the detector current. This result of spatial filtering will be employed by the MIDI instrument of the VLTI which will measure interference of 10 micron light, a wavelength at which thermal background noise from the warm optics and sky will be the dominant noise source.

In this paper a particular configuration for spatial filtering is proposed (Figure 1) using spatial masks (pinholes) in a confocal train (note that this configuration includes, as a special case, the use of a single pinhole at the image plane of a telescope). The term "confocal" is used in the

[1]Send correspondence to meisner@strw.leidenuniv.nl

same sense that the term is applied to a laser cavity in which the focal points of the focusing elements coincide, *not* in the sense that "focusing" necessarily occurs at those points. In fact the optical configuration shown in Figure 1 is identical to that of a confocal laser cavity unwrapped, that is, with the mirrors replaced by lenses in a unidirectional configuration. As in the analysis of a confocal laser cavity, we likewise decompose the optical field on the basis of Hermite-Gaussian modes, simplifying the analysis of transmission between stages in this periodic structure. If we were to go a step further and identify the pinholes in Figure 1 with the walls of the laser tube (or other aperture stop) we can clearly identify the resulting laser resonator as one which, like the proposed spatial filter, favors transmission (and thus oscillation) of the fundamental 0,0 Hermite-Gaussian mode, the preferential selection of which is thereby intended in both cases.

2. Criteria for Spatial Filtering of Starlight

2.1. Breakdown of incoming light

We wish to decompose an incident wavefront on the basis of Hermite-Gaussian modes as detailed in section 3.1. Consider a plane wave incident on the telescope primary which has only been mildly distorted by the atmosphere (over the extent of the telescope aperture) so that a large part of the optical power can still be found in the mode of the undisturbed plane wave. Then we can view the energy entering the optical system as consisting of the following four components. Without employing a spatial filter we will consider how these components would contribute to the interference signal that would result assuming an optical system adjusted for perfect superposition of the nominal plane wavefronts.

1. The power which corresponds to the mode of a perfect plane wave truncated by the telescope's entrance pupil (note that this "mode" gets reanalyzed into a superposition of Hermite-Gaussian modes of even-even order as discussed below, with most of the power in the 0,0 mode). This component from each telescope will interfere coherently with full amplitude.

2. Power which is diverted, due to distortions in the phase of the plane wave, into even-even modes (other than the 0,0 mode, which would be interpreted as an overall "piston" phase of the signal). Since these are from a phase-only function (assuming that the atmospheric disturbance can be considered near-field) the amplitudes of these components will be in phase quadrature with the amplitudes in those even-even modes due to the coherent plane wave. Thus the interference signal between such components from each telescope will add to the main interference signal in quadrature,[2] and if they are not of a large magnitude, will have essentially no effect on the magnitude of the net interference measured.

3. Power which is diverted, due to distortions in the phase of the plane wave, into modes other than even-even modes. Again, considering the atmosphere to behave as a near-field phase screen, the amplitudes of these contributions will be restricted to (positive or negative) purely imaginary values. Even relaxing the near-field restriction, we would find them to be random complex numbers whose complex mean is necessarily zero (and mutually uncorrelated), but

---

[2]This can be understood as follows. The basis functions of the Hermite-Gaussian modes we are considering are purely real functions, whose phase is 0 or π at all points. If we (without loss of generality) take the nominal plane wave to be at 0 phase, then the overlap of the nominal plane wave with any such even-even basis function will be a (positive or negative) real number (its overlap with odd-even, even-odd, or odd-odd modes will be zero by symmetry).
Now consider the same plane wave after passing through a *weak* phase screen. To first order, the resulting wave can be represented as the sum of the original plane wave plus a component which is purely imaginary (positive or negative) at all points. The overlap of a purely imaginary function with the purely real basis function for any mode (not just even-even modes) will be a purely imaginary number, therefore representing a component in phase quadrature with the contribution to that mode from the plane wave itself. Although this argument also applies to the odd modes discussed in contribution (3), in that case there is a stronger reason for lack of a coherent contribution: that there is no signal initially expected in that mode from the nominal plane wave.

since these modes have no initial power from the plane wave mode, the result of interference from them, again, will be in a random phase with respect to the interference of the plane waves, and, if not large in amplitude, will produce a negligible contribution to the overall measured interference.

4. Power from background illumination, which we expect to have equal energy distributed among all (normalized) modes accepted by the system, and to therefore increase the detector shot noise level, but which is profoundly incoherent (both spatially and temporally) and thus makes no contribution to the interference signal using coherent integration.

2.2. Effect of spatial filtering on visibility calibration

Although contributions (2) and (3) will contribute to the measured total power of the signal (presumably contribution (4), if present, has been removed in a photometric measurement by chopping against the sky background), they do *not* appreciably contribute to the interference signal. Since the interference signal is normalized by dividing it by the geometric mean of the total powers in the interfering beams, we will thus underestimate the true visibility of the source.

To determine the visibility calibration error which will be introduced due to the presence of optical power not in the plane wave mode, we will need to determine the power transmission coefficients of optical modes through the spatial filter. As a rough approximation let us consider that all of the energy of the coherent plane wave component is in the central spatial mode. If 0 designates the central mode (0,0 Hermite-Gaussian mode, as discussed below), and $i > 0$ refer to higher order modes, then, to first order, the interferometer will measure a fringe visibility which is approximately reduced by the factor $1/(1+\epsilon)$ where ϵ is given by:

$$\epsilon = \sum_{i \geq 1} \frac{\rho_i}{\rho_0} \left(\frac{|E_{Ai}|^2 + |E_{Bi}|^2}{|E_{A0}|^2 + |E_{B0}|^2} \right) \tag{1}$$

In this equation ρ_i represents the total power transmission due to only mode i being excited at the input. E_{Ai} and E_{Bi} represents the optical amplitudes in non-central mode i from the A and B telescopes[3], while E_{A0} and E_{B0} are the amplitudes in the central mode excited by the plane wave component. It can be seen that the magnitude of the error ϵ due to energy in mode i is proportional to the power transmission ρ_i for that non-central mode.

Therefore an important figure of merit for the quality of a spatial filter, in regards to visibility calibration enhancement, is given by the relative *rejection factor* ρ_i/ρ_0 for the modes i in which substantial incoherent flux is found. In general, the largest amount of incoherent power will be in the 0,1 and 1,0 "tip-tilt" modes, representing energy robbed from the central mode of the plane wave due to a uniform tilt across the wavefront. Thus the rejection factor for the 1^{st} order modes will be of primary importance. However if that component of wavefront error has been successfully cancelled by an (ideal) tip-tilt correction unit, then the dominant incoherent component will be found in the 2^{nd} order (0,2; 2,0; and 1,1) modes. Therefore in a system in which good tip-tilt correction has been applied, the primary performance criterion would be given by the rejection factors for the 2^{nd} order modes.

2.3. Effect of spatial filtering on signal-to-noise ratio (SNR)

The performance criterion for a spatial filter in regards to enhancement of the detection signal-to-noise ratio, is dependent on the source of system noise. In the case of a detector noise dominated system, there is no gain to be had.

---

[3]This does not include the component in the even-even modes attributable to an undisturbed plane wave, which *does* combine coherently. In a good spatial filter, the even-even modes (which are of order 2, 4, etc.) will be well rejected anyway, so that we are justified in identifying ρ_0 as the approximate throughput of the energy of the plane wave component.

In the case of a source (photon) noise limited system, the gain is modest at best. It can be shown that the increase in the signal-to-noise ratio by introducing a spatial filter, is approximately given by:

$$\frac{SNR^{(f)}}{SNR^{(n)}} \approx \sqrt{\rho_0}\sqrt{\frac{1+\epsilon_n}{1+\epsilon_f}} \tag{2}$$

ϵ_n and ϵ_f are the values of ϵ given by (1) without and with spatial filtering, respectively. The second factor in (2) accounts for the reduction in photon noise from the incoherent component of light that would have otherwise been transmitted to the detector. On the other hand, there is a further *loss* in SNR according to the square root of ρ_0, due to the reduction in the interference itself due to the insertion loss of the filter. Thus a filter which has good rejection factors in order to enhance the visibility calibration, may actually *reduce* the SNR since ρ_0 will typically be reduced well below unity. A substantial increase in SNR only occurs in the case of spatially filtering a signal which has only a small amount of its power in the central mode. This would correspond to the case of a telescope diameter much larger than r_0, which is why restricting the telescope aperture to something of order r_0, is itself a spatial filter which reduces ϵ to something less than one, and this actually increases the SNR (in addition to increasing the calibration of visibility as discussed above).

Now, in the case of a background limited system, the SNR is a strong function of the spatial filter employed. Again, making the approximate assumption that all of the mutually coherent power is in the central mode, it is simple to calculate the resulting SNR with respect to that obtained with a "perfect" spatial filter (one that passes the central mode with unity gain and blocks all others). The signal amplitude in this case simply scales by ρ_0 while the noise amplitude scales according to the square root of the total background flux admitted. Thus:

$$\frac{SNR^{(f)}}{SNR^{(\mathbf{perfect})}} = \rho_0 \left(\sum_{i\geq 0}\rho_i\right)^{-1/2} \tag{3}$$

We can rewrite (3) in terms of the "rejection factors" ρ_i/ρ_0 defined above:

$$\frac{SNR^{(f)}}{SNR^{(\mathbf{perfect})}} = \sqrt{\rho_0} \left(1+\sum_{i\geq 1}\frac{\rho_i}{\rho_0}\right)^{-1/2} \tag{4}$$

Note that in this case, as opposed to the determination of the visibility calibration error due to ϵ defined in (1), there is no weighting resulting from the actual amount of power present in the non-central modes; all modes receive an identical amount of background power. Therefore the effect of the spatial filter on the SNR in this case can be calculated without reference to the magnitude of the atmospheric turbulence.

3. Analysis of the Proposed System

3.1. Hermite-Gaussian mode analysis

While there are many possible orthogonal decompositions of an optical field, there is one particular basis set that will be useful for the analysis of the proposed optical system. Since the system shown in Figure 1 performs fourier transforms on the optical fields in between the masks, we would like to simplify that operation. Thus we choose decomposition into *Hermite-Gaussian* functions, as these are the eigenfunctions of the fourier transform. In two dimensions, the j, k mode in terms

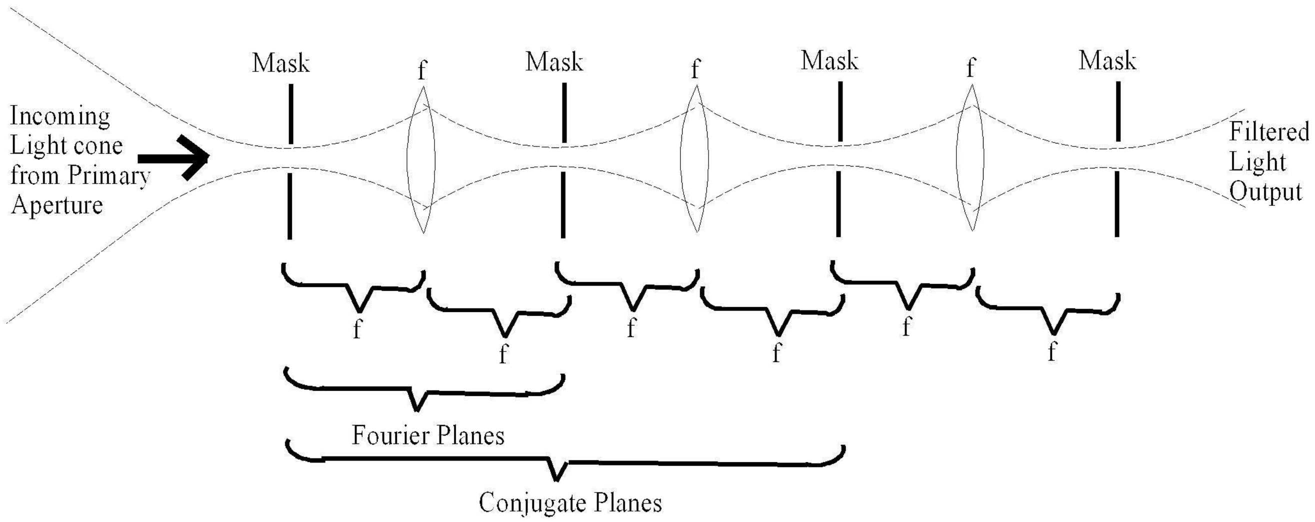

Figure 1. Proposed spatial filter implementation uses multiple masks (pinholes) in successive fourier planes. As is well known, the size of a function times the size of its fourier transform is minimized by the gaussian function (0,0 Hermite-Gaussian mode), which therefore has the least loss in such a system. Higher-order modes suffer greater attenuation, and this discrimination is enhanced as filtering stages are added.

of the *normalized* transverse coordinates[4] x and y, is given by:

$$g_{j,k}(x,y) = g_j(x) \cdot g_k(y) \tag{5}$$

where g_i is the one-dimensional Hermite-Gaussian function of order i. These are orthogonal, and we choose their leading coefficients so that they are also normalized, to satisfy:

$$\int_{-\infty}^{\infty} g_i(x)\, g_j(x)\, dx \;=\; \left\{ \begin{array}{ll} 1 & i = j \\ 0 & i \neq j \end{array} \right. \tag{6}$$

The analysis programs used in this study employ the first five such functions, which are:

$$\begin{array}{rcl} g_0(x) & = & \sqrt[4]{\pi}\, e^{-\frac{x^2}{2}} \\ g_1(x) & = & \sqrt{2}\,\sqrt[4]{\pi}\, e^{-\frac{x^2}{2}}\; x \\ g_2(x) & = & \frac{\sqrt[4]{\pi}}{\sqrt{2}}\, e^{-\frac{x^2}{2}}\; (2x^2 - 1) \\ g_3(x) & = & \frac{\sqrt[4]{\pi}}{\sqrt{3}}\, e^{-\frac{x^2}{2}}\; (2x^3 - 3x) \\ g_4(x) & = & \frac{\sqrt[4]{\pi}}{\sqrt{6}}\, e^{-\frac{x^2}{2}}\; (2x^4 - 6x^2 + \frac{3}{2}) \end{array} \tag{7}$$

Since the two-dimensional modes in (5) are separable in x and y using the one-dimensional Hermite-Gaussian functions (7), an orthonormality condition similar to (6) applies as well. There

[4] The normalized coordinates x and y, are specified in terms of the *physical* coordinates x_{phys} and y_{phys} as follows:

$$x = x_{phys} \frac{\sqrt{2}}{w_0} \quad ; \quad y = y_{phys} \frac{\sqrt{2}}{w_0}$$

where the "beam waist radius" parameter w_0 is, for the optical system depicted in Figure 1, given by:

$$w_0 = \sqrt{\frac{\lambda f}{\pi\, n}}$$

However all calculations will be considered in terms of the normalized spatial coordinates only, with the above relation buried in the definition of the "normalized pinhole radius" R in (11).

is zero overlap between any two non-identical spatial modes j,k and j',k', and unity overlap when $j=j'$ and $k=k'$.

The amplitude of any complex incident optical field $E(x,y)$ in mode j,k is then simply found by the integral:

$$E_{j,k} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_{j,k}(x,y)E(x,y)\,dx\,dy \tag{8}$$

The power in that mode is given by the squared-magnitude $|E_{j,k}|^2$ of its amplitude.

Transmission of optical power present in the j,k mode through a thin optical element (or *mask*) whose transmission is given by $f(x,y)$, will generally dissipate part of that power and redistribute the rest of that power into various output modes j',k'. We can multiply the incident optical field present in the j,k mode (given by its amplitude coefficient $E_{j,k}$ times the basis function $g_{j,k}(x,y)$) by the mask function $f(x,y)$, and analyze this output in the j',k' modes using (8). Thus:

$$E^{(output)}_{j',k'} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_{j',k'}(x,y)\left((E_{j,k}\; g_{j,k}(x,y))\; f(x,y)\right)\,dx\,dy = s_{(j',k'),(j,k)}\cdot E_{j,k} \tag{9}$$

where $s_{(j',k'),(j,k)}$, the *scattering parameter* for that transmission path simply involves the same integral with the strength of the specific mode $E_{j,k}$ not included.

$$s_{(j',k'),(j,k)} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_{j',k'}(x,y)g_{j,k}(x,y)\,f(x,y)\,dx\,dy \tag{10}$$

The analysis in section 3.3. is based on the use of scattering matrices whose elements are computed numerically using (10) with $f(x,y)$ modeling pinholes of various sizes.

3.2. Transmission of Hermite-Gaussian modes

Let us now consider the proposed configuration for a spatial filter consisting of a series of masks in a confocal train as shown in Figure 1. Each lens performs a fourier transform on the optical field from plane m to plane $m+1$. The Hermite-Gaussian modes $g_{j,k}(x,y)$, being eigenmodes of the Fourier transform, are transmitted from one such plane to another intact; only their eigenvalues need be specified. Ignoring the common phase shift due to the speed of light, the propagation constant for transmission between two such planes m and $m+1$, is simply $\jmath^n$ where n is the order of the mode; $n=j+k$ where j and k are the orders in the x and y dimensions respectively. Thus if we removed the masks, we could find the net transmission across 2 such paths (e.g. between conjugate planes) as follows. The even order modes would undergo a phase change of $(\jmath^n)^2=1$, whereas the odd-order modes would encounter a phase change of $(\jmath^n)^2=-1$. Since an odd-order mode, when inverted, reverses its sign at every point, while the inversion of an even-order function is identical to itself, this shows that every mode is spatially inverted between conjugate planes, and thus the entire image is inverted between conjugate planes, as is obvious from ray-tracing.

3.3. Transmission of spatial modes through a single pinhole

By inserting pinhole masks at each plane, we must take into account the scattering matrix defining the transmission of Hermite-Gaussian modes through the mask. These are a strong function of the pinhole size. The normalized pinhole radius R is given by:

$$R = r\sqrt{2\pi n/\lambda f} \tag{11}$$

where r is the physical radius, n is the index of refraction of the transmissive medium (usually 1), λ is the wavelength of light, and f is the confocal parameter equal to the focal length of the lenses, and to the separation between a mask plane and either lens. For instance, with a normalized pinhole radius of 1.5, we obtain the following scattering matrix (this is a symmetric matrix; only the bottom half is shown):

Table of s-parameters for normalized Pinhole radius = 1.50

mode	0, 0	0, 1	0, 2	0, 3	0, 4	1, 0	1, 1	1, 2	1, 3	1, 4	2, 0	2, 1	2, 2	2, 3	2, 4	3, 0	3, 1	3, 2
0, 0	.89																	
0, 1	0	.65																
0, 2	−.16	0	.37															
0, 3	0	−.32	0	.29														
0, 4	−.01	0	−.30	0	.31													
1, 0	0	0	0	0	0	.65												
1, 1	0	0	0	0	0	0	.39											
1, 2	0	0	0	0	0	−.18	0	.22										
1, 3	0	0	0	0	0	0	−.24	0	.20									
1, 4	0	0	0	0	0	.04	0	−.18	0	.19								
2, 0	−.16	0	−.01	0	.04	0	0	0	0	0	.37							
2, 1	0	−.18	0	.05	0	0	0	0	0	0	0	.22						
2, 2	−.01	0	−.10	0	.08	0	0	0	0	0	−.10	0	.14					
2, 3	0	.05	0	−.05	0	0	0	0	0	0	0	−.12	0	.11				
2, 4	.04	0	.08	0	−.08	0	0	0	0	0	.03	0	−.11	0	.11			
3, 0	0	0	0	0	0	−.32	0	.05	0	.01	0	0	0	0	0	.29		
3, 1	0	0	0	0	0	0	−.24	0	.13	0	0	0	0	0	0	0	.20	
3, 2	0	0	0	0	0	.05	0	−.12	0	.10	0	0	0	0	0	−.05	0	.11

We see that the matrix is dominated by zero elements. In fact, for a symmetric mask (such as an on-axis pinhole) the only non-zero elements of the scattering matrix are between modes which are both even or both odd in x, and both even or both odd in y. Thus we can simplify such a system by identifying 4 submatrices containing all the non-zero elements for the even-even modes, the even-odd modes, the odd-even modes, and the odd-odd modes. Any energy launched into one of these groups will remain in the same group (or be dissipated) so we proceed with our analysis as if these were 4 parallel optical systems.

Then, just considering one of the four groups, we can easily find the transfer function from the output of one mask to the input of the following mask. This will simply be a diagonal matrix, since the Hermite-Gaussian functions are eigenmodes of the fourier transform. The diagonal elements will consist of 1 or -1 for the $4n$ and $4n+2$ modes respectively (or, in the case of the even-odd or odd-even modes, the $4n+1$ and $4n+3$ modes, having absorbed the $\jmath$ factor into the overall transmission phase which need no longer be identical between these now-independent groups).

Then to find the *net* transmission from the output of one mask to the output of the next mask, we can multiply the scattering matrix of the mask by the diagonal transmission matrix (on the right). We thus wind up a scattering matrix which goes from one plane to the next similar plane in the system; such a combined matrix is thus no longer symmetric. Using a normalized pinhole size of $R = 1.2$, for instance, we find for the lower order modes these scattering matrices for each of the four non-interacting systems:

Table of s-parameters for normalized Pinhole radius = 1.20 following confocal transmission

- - - - - - Table for EVEN - EVEN modes - - - - - -

modes	0, 0	2, 0	0, 2	4, 0	2, 2	0, 4
0, 0	.7631	.2412	.2413	.0585	.0477	.0585
2, 0	−.2412	−.2241	−.0477	−.1903	−.0720	.0140
0, 2	−.2413	−.0477	−.2240	.0140	−.0720	−.1903
4, 0	.0585	.1903	−.0140	.2107	.0666	−.0317
2, 2	.0477	.0720	.0720	.0666	.0703	.0665
0, 4	.0585	−.0140	.1903	−.0317	.0665	.2106
Power						
Trans.	.7078	.1523	.1523	.0897	.0264	.0896

- - - - - - Table for EVEN - ODD modes - - - - - -

modes	0, 1	2, 1	0, 3	4, 1	2, 3
0, 1	.4219	.1737	.3008	.0782	.1106
2, 1	−.1737	−.1222	−.1106	−.0962	−.0840
0, 3	−.3008	−.1106	−.2499	−.0392	−.0855
4, 1	.0782	.0962	.0392	.0970	.0678
2, 3	.1106	.0840	.0855	.0678	.0694
Power					
Trans.	.3170	.0736	.1741	.0309	.0360

- - - - - - Table for ODD - EVEN modes - - - - - -

modes	1, 0	3, 0	1, 2	3, 2	1, 4
1, 0	.4220	.3009	.1738	.1106	.0783
3, 0	−.3009	−.2499	−.1106	−.0855	−.0392
1, 2	−.1738	−.1106	−.1223	−.0840	−.0962
3, 2	.1106	.0855	.0840	.0694	.0678
1, 4	.0783	.0392	.0962	.0678	.0970
Power					
Trans.	.3172	.1740	.0737	.0360	.0309

- - - - - - Table for ODD - ODD modes - - - - - -

modes	1,1	3,1	1,3
1,1	.1763	.1444	.1444
3,1	−.1444	−.1291	−.1132
1,3	−.1444	−.1132	−.1291
Power Trans.	.07280	.0503	.0503

The sum of the squares of the transmissions from light injected into a single mode i is computed as the "Power Transmission," ρ_i for that mode, and has been tabulated at the bottom of the table for each input mode. As we discussed previously, the figure of merit regarding visibility calibration, the rejection factor for non-central mode i, is given by the ratio of ρ_i to the throughput of the central mode ρ_0. We can see that this stage has a rather poor rejection factor of $.3170/.7078 = .45$ for the 0,1 and 1,0 (tip-tilt) modes. However this number can be improved by using multiple stages of spatial filtering.

3.4. Analysis of a multi-stage spatial filter using pinhole masks

For a spatial filter consisting of a train of such stages in cascade, we consider that the output of one stage goes into the input of the next. Since each stage is identical, the net scattering matrix for a train of M stages can be found by raising each of the 4 scattering matrices to the M^{th} power. For instance, using 5 stages with a larger pinhole diameter of 1.8, we obtain:

Table of s-parameters for normalized Pinhole radius = 1.80 following confocal transmission, 5 stages

- - - - - - Table for EVEN - EVEN modes - - - - - -

modes	0,0	2,0	0,2	4,0	2,2	0,4
0,0	.8309	.0253	−.0253	−.0833	−.0681	−.0834
2,0	−.0253	−.0145	.0077	−.0042	.0010	.0066
0,2	−.0253	.0077	−.0145	.0066	.0010	−.0042
4,0	−.0833	.0042	−.0066	.0118	.0072	.0065
2,2	−.0681	−.0010	−.0010	.0072	.0065	.0072
0,4	−.0834	−.0066	.0042	.0065	.0072	.0118
Power Trans.	.7102	.0010	.0010	.0072	.0048	.0072

- - - - - - Table for EVEN - ODD modes - - - - - -

modes	0,1	2,1	0,3	4,1	2,3
0,1	.2783	.0255	.0612	−.0221	−.0270
2,1	−.0255	−.0030	−.0042	.0016	.0025
0,3	−.0612	−.0042	−.0170	.0057	.0057
4,1	−.0221	−.0016	−.0057	.0020	.0021
2,3	−.0270	−.0025	−.0057	.0021	.0027
Power Trans.	.0830	.0007	.0041	.0005	.0008

- - - - - - Table for ODD - EVEN modes (Similar to EVEN-ODD table, above)

- - - - - - Table for ODD - ODD modes - - - - - -

modes	1,1	3,1	1,3
1,1	.0222	.0076	.0076
3,1	−.0076	−.0030	−.0024
1,3	−.0076	−.0024	−.0030
Power Trans.	.0006	0	0

We have now increased the performance of spatial filtering by using more stages with less filtering in each stage. Thus in this example, the central gaussian mode is transmitted with 71% power transmission (69% output in the same mode, 2% in other even-even modes), almost the same transmission as the previous single stage configuration. However in this case the 0,1 and 1,0 modes each have a net power transmission of only .083, so the rejection factor is now $.083/.71 = .12$, much improved from the previous example (.45). Even more striking, the rejection factor for the 0,2 and 2,0 second order modes has been reduced to .002 (from .21 in the single stage filter).

3.5. Analysis for an incident plane wave

A plane wave incident on the telescope aperture has *most* of its power in the central Gaussian mode, however an accurate analysis should properly break that plane wave into its components. Note that what we are considering at this point is a plane wave truncated by the circular aperture of the telescope objective (the minor effect of the secondary obstruction will be ignored). It should be noted that the entrance pupil of the telescope in fact comprises part of the spatial filter, one that we take for granted. In general it would be necessary to specify the *size* of the

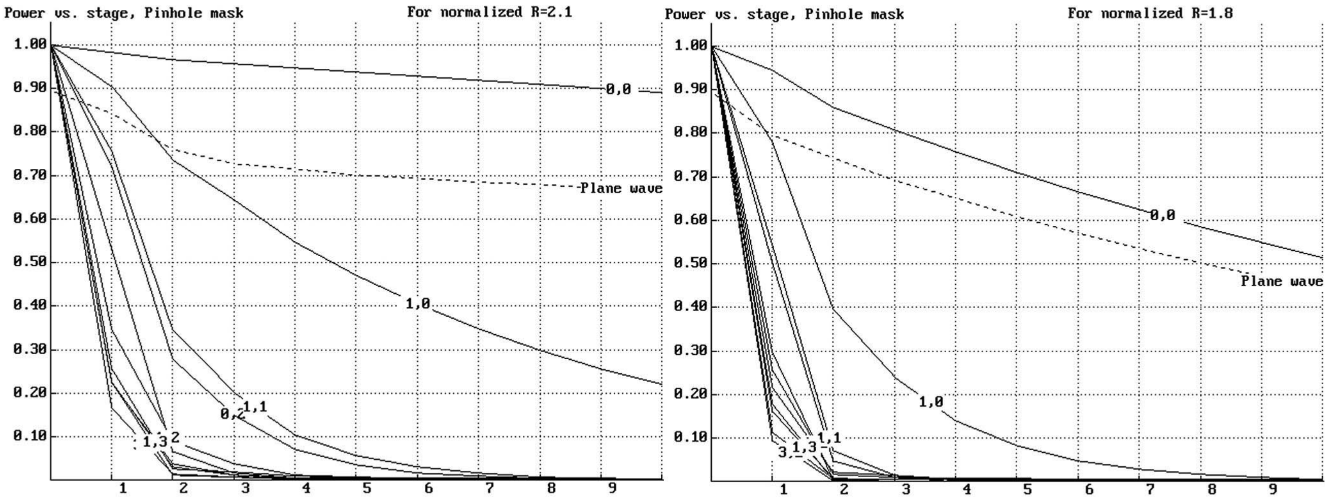

Figure 2. Plot of portion of power which is retained after N stages of the confocal pinhole train spatial filter (Figure 1) for power launched into a single Hermite-Gaussian mode, and for a plane wave input. Note that not all of the initial power of a plane wave is accounted for using the lower-order modes, resulting in a "power transmission" of only 89% at the input stage. Left plot is for a slightly larger pinhole, illustrating the sensitivity of such a system to pinhole size (Figure 3).

circular aperture (or the angle of the light cone) involved, however in the case of a pinhole confocal train corresponding to Figure 1, there is an obvious choice. Since we have chosen a system with identical circular apertures (pinholes) between consecutive fourier planes, we would choose an incoming beam matching that aperture. This is equivalent to saying that in Figure 1, with the first pinhole absent, the telescope aperture would be reimaged coincident with the second pinhole. In other words, the incoming light would have an f-number equal to d/f where d is the physical pinhole diameter and f is the confocal parameter. Therefore the modal decomposition of a plane wave is somewhat a function of the normalized pinhole radius R, but since we are only considering a narrow range of R, the following result will not vary greatly. Note that less of the power would be found in the 0,0 mode *either* using a much smaller *or* a much larger normalized pinhole radius and telescope aperture.

Again using $R = 1.8$, we find the following breakdown of the input plane wave:

Table of amplitude coefficients for Input Wave

Mode	*Amplitude*	*Power*
$0,0$	.89125	.79434
$2,0$	.12693	.01611
$0,2$	.12644	.01598
$4,0$	$-.16090$	.02589
$2,2$	$-.13162$	.01732
$0,4$	$-.16102$	.02592
Total		
Power		0.8956

Note that only 89% of the power is accounted for by the modes we have considered, but it can be conjectured that most of the remaining power which is present in higher order modes would be rejected by any decent spatial filter, thus entailing at most a small error due to this incomplete analysis.

4. Resulting Performance

4.1. Performance in relation to visibility calibration

To find the response of the filter to a plane wave, we would multiply these amplitude coefficients we have found for the plane wave, by the previous scattering matrix describing the transmission of optical signals through the 5 stage filter. In this case we find the output power distributed as follows:

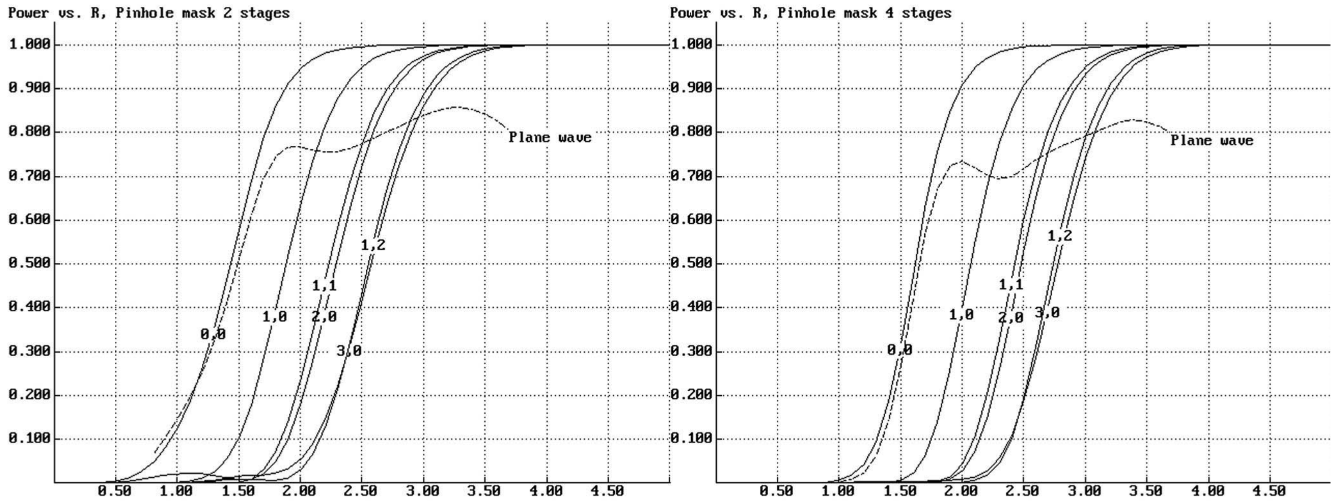

Figure 3. Plot of portion of power retained in a confocal pinhole train spatial filter (Figure 1) for power launched into various Hermite-Gaussian modes as a function of normalized pinhole radius R. Left pane is for a 2-stage system; right pane is for a 4-stage filter. Net power transmission for a matched plane wave is plotted over the region in which over 80% of the plane wave's initial power is accounted for by the modal decomposition; the dip in the curve as $R > 3.3$ is only an artifact of the incomplete analysis of the incoming plane wave.

Table of amplitude coefficients for Resulting Output Wave

Mode	*Amplitude*	*Power*
0, 0	.78275	.61270
2, 0	−.02392	.00057
0, 2	−.02394	.00057
4, 0	−.07847	.00615
2, 2	−.06414	.00411
0, 4	−.07850	.00616
Total Power		0.6303

Thus 63% of the incident power has been transmitted. Again, this compares to 8.3% power transmission for the two 1^{st} order modes, and less than 1% for any higher order mode.

Generally speaking, there will be a trade-off involving the efficiency ρ_0, rejection factors ρ_i/ρ_0, and the number of stages M in the optical train. For instance, in the left pane of Figure 2 the propagation of Hermite-Gaussian modes through each of 10 stages of such a spatial filter is plotted, using a large pinhole size (R=2.1). It can be seen that even after 4 stages, the 0,1 and 1,0 modes retain over half of their initial power. But perhaps we are using a tip-tilt corrector so that those modes are not excited. The second-order modes have already been reduced to a tenth of their initial power after 4 stages. At that point the 3^{rd} and higher order modes have essentially been extinguished, but the central mode retains 95% of its initial power (however the power of a plane wave, plotted with a dashed line, suffers a greater reduction due to the portion of its power present in higher-order modes, as discussed earlier). By using a somewhat smaller pinhole (R=1.8), as shown in the right pane of Figure 2, it is possible to obtain similar rejection of the second order modes in only 2 stages, but now the central mode drops off much faster. The power transmission of modes in a 2-stage and in a 4-stage filter are plotted with respect to normalized pinhole size R in Figure 3.

In any case, an actual determination of the amount of non-coherent signal power remaining after the spatial filter is proportionately dependent upon the initial portion of optical power present in such modes, which is dependent on the ratio of the telescope diameter to r_0 (and the effect of tip-tilt or adaptive optics correction). That analysis is not presently attempted. Instead, we only consider the "rejection factor" ρ_i/ρ_0 which describes the extent to which a visibility calibration error (see equation (1)) due to the presence of an unknown amount of optical power in mode i, will be reduced. Figure 3 suggests that if an effective tip-tilt corrector has been used, then it will be possible to largely reject remaining power in 2^{nd} and higher order modes and still retain 70% throughput for a plane wave, using only a 2-stage filter. However if we are not using active correction for wavefront tilt and instead rely on passive filtering of the 1,0 and 0,1 tip-tilt modes, then a filter with 4 or more stages would be advantageous.

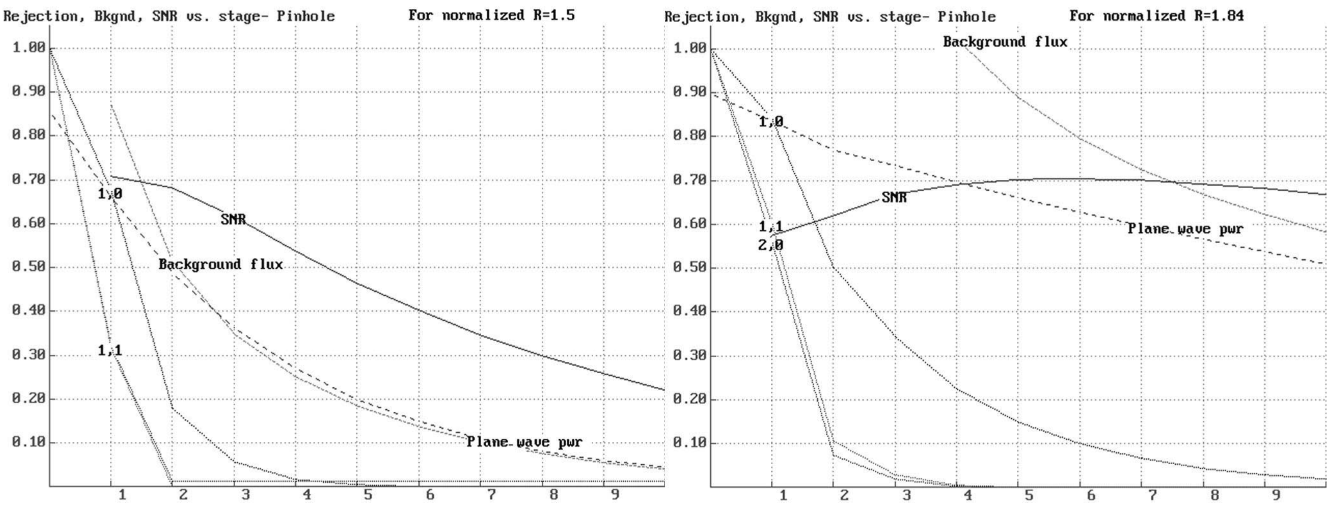

Figure 4. Signal and noise analysis assuming background limited detection, plotted versus stage in a confocal pinhole filter train of up to 10 stages. Resulting SNR (solid line) relative to theoretical maximum SNR, is determined from the power transmission of a plane wave (coarse dashed line) divided by the rms noise level given by the square root of the power transmission of background flux (fine dashed line). Also plotted (dotted lines) are the rejection factors, ρ_i/ρ_0 for the 1^{st} order ("tip-tilt") modes and the 2^{nd} order (defocusing/astigmatism) modes.

4.2. SNR performance in a background limited system

We will now examine the performance of such spatial filters as applied to the case of signal detection where the noise floor is dominated by shot noise contribution from photons received from the sky and/or warm optics. Recall that there is also a potential increase in SNR obtainable in the case of source limited systems according to (2), but that this gain was generally of modest proportions and not simultaneously available in conjunction with the good rejection characteristics required for enhancement of visibility calibration. On the other hand, in a background limited system spatial filtering is absolutely required and is intimately tied to the resulting SNR (note that detection in the image plane of a telescope using a detector array is equivalent to a series of spatial filters determined by the extent of each detector pixel).

Let us see what SNR will be achieved using the proposed configuration in the background limited case. Recall that the performance of the spatial filter in this regard is essentially independent of the magnitude of atmospheric distortion. We will assume that all of the incoming optical power can be found in a plane wave over the telescope aperture; any optical power diverted into other modes will (for all practical purposes) not contribute to the SNR regardless of the spatial filter (this is not strictly true in the case of an image plane detector array in which case there is a small gain from diversity detection of a random speckle pattern). We shall express our results in terms of the SNR reduction compared to an "ideal" spatial filter, that is one that has complete transmission of the undisturbed plane wave mode, and no transmission of all other modes.[5] Note that this criterion is *slightly* different than that proposed in (4) which was based on an incoming wave with a gaussian profile; the optimization point in either case would be virtually the same.

The computer programs used to generate the graphs in Figure 4 were subject to certain approximations. In particular, the background flux level is found by summing the power transmission coefficients of all spatial modes; this is the summation in the second factor of (3). However only 19 of the lower order modes were included in this sum, so that *poor* spatial filters which allowed substantial transmission of higher-order modes (corresponding to the right side of Figure 3 where the pinhole size is too large to attenuate power transmission for the modes plotted), presumably would pass substantial background flux from higher-order modes. Thus calculated background levels exceeding about a third of the maximum possible value of 19 were discounted. An exami-

[5] Although such an ideal filter is not prohibited by the laws of physics, it isn't clear that it even has a realization (for instance, optical fibers which do only accept a single mode cannot be perfectly matched to an incoming plane wave).

nation of the error entailed was performed in the case of a one-stage pinhole filter, in which case there exists an alternative geometrical analysis for calculating the background acceptance flux that does not involve any approximations. The results (not shown) indicated that the limited modal analysis accounted for between 66% and 83% of the true background power for pinhole radii in the range of interest: $.6 < R < 3.2$. This would entail an overestimation of SNR by a factor no greater than $1/\sqrt{.66} = 1.23$. This is really a worst case compared to multi-stage filters in which the higher-order modes that the analysis neglected would generally have very little net power transmission. Likewise, the coherent power of the incident plane wave has over 10% of its power not accounted for in the modes considered, leading to a possible error in the opposite direction. A further comparison using the geometrical analysis in the case of a one-stage filter, also shows fairly good agreement with the method which considers only the lower-order modes. That approximate analysis predicts a peak SNR of .71 (relative to an "ideal" filter) at $R = 1.50$, whereas the exact geometrical analysis predicts a peak SNR of .64 at $R = 1.44$.

In Figure 4 we plot the admitted background flux (expressed in "number of modes"), the portion of transmitted power from a plane wave, and the resulting SNR (compared to the "ideal" spatial filter) at various stages in a multi-stage confocal pinhole filter, similar to the graphs in Figure 2. This time we consider filters having pinhole diameters of $R = 1.50$ and $R = 1.84$. In addition to plotting the SNR and its constituents, we have plotted the rejection factors for the 1^{st} and 2^{nd} order modes; these are identical to the power transmission factors shown in Figure 2 divided by the transmission of the $0,0$ mode. Recall that the rejection factors for the 2^{nd} order modes dominate the visibility calibration error in the case of an incident wave which has undergone perfect "tip-tilt" correction, while the rejection factor for the 1^{st} order modes is crucial in a system lacking a tip-tilt controller.

We observe the following qualitative behavior. Use of "large" pinholes (especially ones larger than shown in the right pane of Figure 4) allows one to increase the SNR by increasing the number of stages, inasmuch as the addition of stages slowly reduces the non-central modes' contribution to the background flux with less than half that much influence on the signal power. Consequently the rejection factors also slowly improve. Of course such filters with many stages may be impractical, and would surely entail other sources of optical power loss! On the other hand, in a filter with small pinholes (such as the left pane of Figure 4) there is almost complete rejection of non-central modes after 2 or 3 stages. Beyond that point, the signal and background power are basically confined to the $0,0$ mode, so that additional stages attenuate both equally resulting in an SNR decline. Interestingly, all (reasonable) possibilities for pinhole size were seen to allow for a peak SNR of around .7 times the ideal SNR, using some number of stages. It can be seen that for $R = 1.50$, it is possible to obtain an SNR close to that number using a 2 or 3 stage filter, and still achieve overwhelming rejection of the 2^{nd} and 1^{st} order modes respectively.

5. Wavelength Dependence

One issue we have ignored up until this point, is the effect of wavelength on the performance characteristics of a spatial filter implemented using pinholes masks. Clearly, a pinhole in such a system will look "larger" at shorter wavelengths, and thus admit a greater portion of the light from the central gaussian mode, the higher order modes, and from a uniform background. This will complicate the implementation of such a device in an interferometric instrument operating over a sizable bandwidth. Since a high-performance interferometer should be wideband in order to accumulate the largest amount of useful information in a given amount of valuable telescope time (and also in order to be able to fringe-track on the dimmest objects), this will tend to be a concern in most practical implementations.

5.1. Effect of varying wavelength on previous analysis

The effect of wavelength variations on the analysis of the spatial filter using pinhole masks in a confocal train, is entirely specified in the definition of the "normalized" pinhole radius R, given by (11). This shows that R grows inversely with the square root of wavelength[6].

Looking at the slopes of the curves in Figure 3, it can be seen that wideband operation, implying a substantial range in R applied to optical energy at different wavelengths, will entail very different characteristics in performance. The normalized pinhole radii used in the two graphs of Figure 4, $R = 1.50$ and $R = 1.84$, could in fact correspond to the long and short wavelength limits of an instrument having a 1.5 : 1 wavelength range. It is possible to select such a filter which will have fairly good SNR performance relative to background illumination over such a wavelength range (and even over a 2 : 1 wavelength range) but the filter's rejection factors will vary tremendously over that range of wavelengths. If non-central mode rejection is a primary performance criterion, it would clearly be desirable to have an achromatic version of this filter for an instrument having such a wavelength coverage.

5.2. A wavelength-independent implementation using FTIR

One class of achromatic implementations for the proposed spatial filter system, involves constructing pinholes whose size would grow according to the square root of wavelength. For instance, this might be implemented in a step-wise manner using concentric filters with different spectral transmissions; additional attention would have to be paid to maintaining a flat phase shift over the effective "pinhole" at any wavelength to avoid defocusing problems. A more elegant solution is hereby presented, employing the principle of *Frustrated Total Internal Reflection* (FTIR).

As is well known, light entering a face of a prism and falling on the hypotenuse of the prism at an angle greater than the so-called critical angle, will be totally reflected at that surface, with no transmission possible. However if a second prism is brought very close (within about a wavelength) of that hypotenuse, that material will receive energy extracted from the "evanescent wave" present at that surface, thus "frustrating" the total internal reflection that would normally occur at that angle of incidence. For a given index of refraction and angle of incidence beyond the critical angle, the transmission coefficient will be a strong function of separation *in wavelengths* between the reflecting surface and the frustrating surface[7]. For instance in Figure 5A, using a material whose index of refraction is 1.414, the FTIR amplitude transmission is plotted versus separation (in wavelengths) for internal angles of incidence of 50°, 60°, and 70°. It can be seen that these curves have somewhat different shapes (as well as different widths).

Only the magnitude of the transmittance is plotted in Figure 5, however there is generally a phase shift as well which is also a function of the separation. However in one special case the phase shift vanishes for all separations. Using a material with a refractive index n, that will occur with the "perfect" incidence angle given by

$$\theta_{perfect} = sin^{-1}\sqrt{\frac{1}{2}\left(1 + \frac{1}{n^2}\right)} \tag{12}$$

---

[6] Although that may seem counter-intuitive, it's fairly obvious that if both the physical radius r_{phys} and the confocal parameter f were scaled together with wavelength, then the normalized pinhole parameter R would not change. However it is because f is physically fixed while λ varies, that the net variation in R, as we have defined it, follows the square root of the change in wavelength.

[7] The following analysis applies only to light in the s polarization. In order for such a system to work properly on randomly polarized light, it appears that it would be necessary to initially separate the polarization components using a polarizing beamsplitter, and run the two components into parallel systems with each striking the hypotenuses of the prisms in the s polarization.

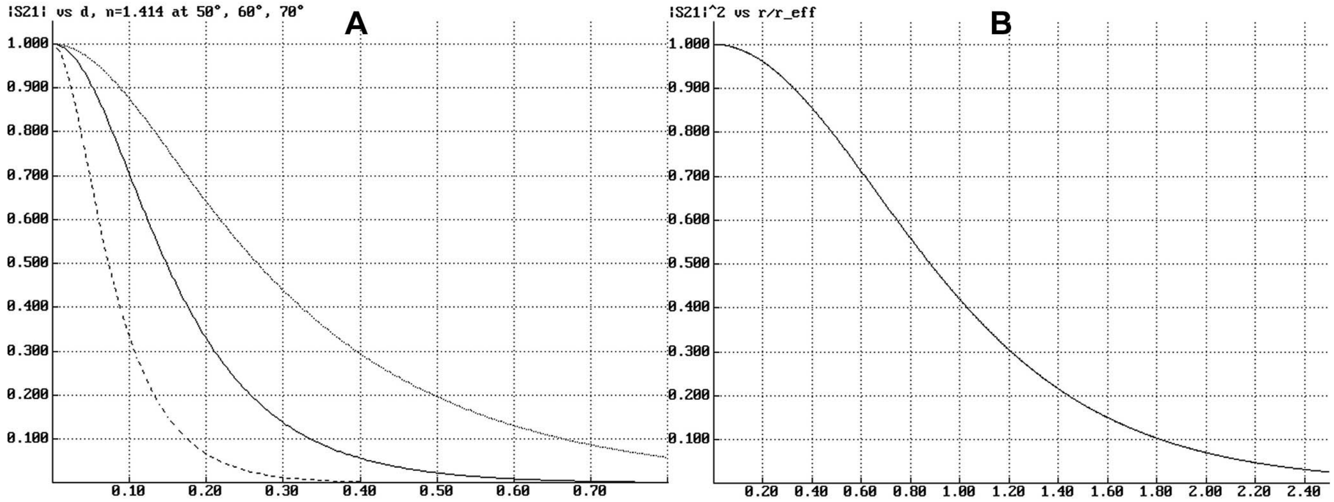

Figure 5. A: Plot of amplitude transmittance versus separation in wavelengths for "Frustrated Total Internal Reflection" (FTIR) using a thin air layer separating surfaces of a material whose refractive index is 1.414 at inclinations of 50°(dotted line), 60°(solid line), and 70°(dashed line). B: Intensity transmittance profile for an aperture generated by FTIR at the "perfect angle" (at which there is no phase shift for the s polarization) with a curved surface as described in the text, plotted in terms of the transverse radial coordinate normalized relative to $r_{eff\,1}$, the point at which the amplitude transmittance falls to $1/cosh(1)$.

Under this condition, the amplitude transmittance as a function of the separation between the surfaces d, takes on the following simple form:

$$S_{21} = \frac{1}{cosh(2\pi n \frac{d}{\lambda})} \tag{13}$$

For instance, in Figure 5A, the middle curve (solid line), corresponding to $\theta = 60°$ meets this condition for the specified index of refraction $n = \sqrt{2}$, and is therefore a plot of (13).

Using this principle, it is possible to make a "pinhole" whose size will vary according to the square root of wavelength using FTIR in which one surface is curved. Of course it will not have the sharp edge of a pinhole, but will follow (13) according to a the radial coordinate $r = \sqrt{x^2 + y^2}$. For a given wavelength λ_1, let us arbitrarily define an "effective radius" $r_{eff\,1}$ as the point at which the argument of the hyperbolic cosine is 1, and thus the point at which the intensity transmission $|S_{21}|^2$ has fallen to .42. Using $r_{eff\,1}$, we can write the transmittance as:

$$S_{21} = \frac{1}{cosh((r/r_{eff\,1})^2)} \tag{14}$$

We can then, as we did for the pinhole case in (11), define a *normalized* mask radius as:

$$R_{FTIR} = r_{eff\,1}\sqrt{2\pi n/\lambda_1 f} \tag{15}$$

If the index of refraction n does not vary (greatly) with wavelength, then it is clear that R_{FTIR} will not be a function of wavelength. This will occur, again, since the physical radius of the aperture thus formed $r_{eff\,1}$, will itself grow according to the square root of wavelength, as we have required.

Note that in order for the aperture thus formed to be circular in the $x-y$ plane (transverse to the propagation direction z), the convex surface of the prism, which is tilted at an angle θ (given by (12)), must actually have a different curvature in the x and y directions. Alternatively, one prism face may be ground as a spherical surface, and the opposing face ground as a cylindrical surface. The resulting intensity transmittance profile relative to the effective radius $r_{eff\,1}$ is plotted in Figure 5B.

Then, as in the case of the pinhole mask confocal train, the characteristics of the device using FTIR masks as shown in Figure 6 will depend solely on the normalized radius R_{FTIR} and the

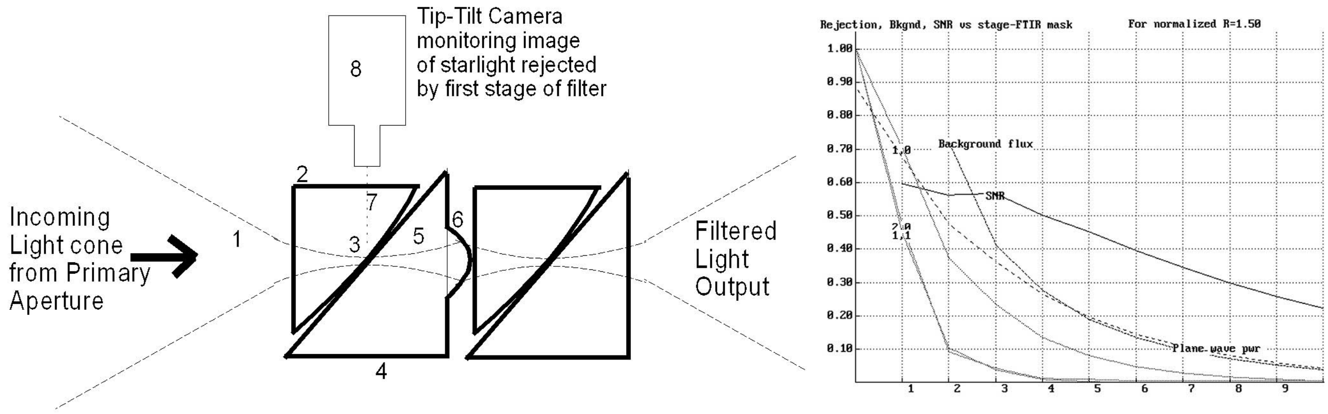

Figure 6. Optical field entering a filter stage (1) enters a prism (2) and focuses onto the hypotenuse (3) at an angle greater than the critical angle. Using FTIR, an aperture following the profile shown in Figure 5B is formed due to the curved surface between it and the second prism (4). The transmitted light (5) is refocused by a lens (6) which feeds the partially filtered beam into the next stage of the system. Light which has been rejected by the first stage of the system is reflected (7) at an angle and contains the image of the stellar object distorted by atmospheric seeing, with the amplitude of the transmitted wave (5) subtracted from it. This image may be used by a camera (8) for observation of image centroid motion to produce feedback for a tip-tilt controller. The performance of such a spatial filter is plotted on the right versus stage (for the choice of normalized $R = 1.50$), which can be compared with a similar plot for pinhole filters in Figure 4.

number of stages M. Because the falloff in transmission shown in Figure 5B is not as sharp as with a pinhole, the rejection of non-central spatial modes will not be as great in a given stage, and more stages will be necessary in order to obtain a given level of performance compared to the pinhole filter train operating at a single wavelength.

Although this is a relatively small effect, it should also be noted that the entire system will never be completely achromatic inasmuch as the effective size of the telescope aperture which truncates the nominal plane wave entering the spatial filter, and itself an element in the overall system, will itself appear to grow (in normalized terms) according to the inverse square root of wavelength. In the same vein, it might be mentioned that the ratio of power diverted from the plane wave mode to non-central modes by the atmosphere will also increase at shorter wavelengths (where the wavefront errors represent a greater portion of a wavelength), so the *net* performance relative to visibility calibration would tend to be poorer at short wavelengths, even with a truly wavelength-independent spatial filter. However it can be seen that with a 3 or 4 stage FTIR filter it is possible to achieve very good rejection of 2^{nd} and higher order modes without greatly compromising SNR in a background limited system, which should ensure good visibility calibration if an effective tip-tilt corrector is in use. The performance of the wavelength-independent FTIR filter is plotted over 10 stages in Figure 6, which can be compared to the somewhat superior performance of the narrow-band pinhole filter previously depicted in Figure 4.

6. Conclusion

Spatial filtering of light may be used in an optical stellar interferometer in order to increase the accuracy of fringe visibility estimation by allowing only light in a single spatial mode, so that the interference signal will be related only to the amount of light accepted by the spatial filter from each beam, regardless of changes in the original strehl ratio of the beams initially entering the system. Secondly, spatial filtering may be used to optimize the signal-to-noise ratio in a background limited instrument, also by accepting light only in the central spatial mode, and rejecting light in other spatial modes which contain background radiation but no coherent signal. In either case the performance of the spatial filter may be characterized by the *rejection factors* ρ_i/ρ_0 for non-central modes, and the throughput ρ_0 for the central mode (or the plane wave throughput). It is found that a visibility calibration error (1) due to optical power diverted into non-central mode i, will be reduced by the rejection factor ρ_i/ρ_0 for that mode. For a background limited system, the SNR in these terms is described in (4).

Although optical fibers may be used for spatial filtering, we wish to consider possibilities for a spatial filter using discrete optical components. The proposed configuration (Figure 1) is analyzed in terms of its transmittance of spatial modes described by Hermite-Gaussian functions (7) since the propagation constants of these modes from one fourier plane to the next is equal to +1 or -1 (times a common phase shift). The performance of such a configuration is plotted in terms of the normalized pinhole size R and number of stages M in Figures 2 - 4.

Although such a system using pinholes is workable for a narrow to medium bandwidth, its performance is quite sensitive to large wavelength variations. One possibility for a wideband version of such a filter is proposed using the principal of "Frustrated Total Internal Reflection" (FTIR), which, if formed using a curved surface at the hypotenuse of a prism, will create an aperture which grows in size according to the square root of wavelength.

7. Bibliography

Buscher, D.; Shaklan, S. "Low-order adaptive optics and single-mode fibers in stellar interferometry," *Proc. SPIE 2201,* pp. 980, 1994

Coude Du Foresto, V.; Ridgway, S.; Mariotti, J.-M. "Deriving object visibilities from interferograms obtained with a fiber stellar interferometer," *A & A Supplement series*, 121, pp. 379, 1997

Meisner, J. "Spatial Filtering with Pinholes for MIDI," *unpublished; posted at:*
http://www.strw.leidenuniv.nl/˜meisner/pinholes.ps

Prasad, S.; Loos, G. "Spatial Filtering of Atmospheric Decorrelation from Wavefronts," *ESO Conference on High-Resolution Imaging by Interferometry II. Ground-Based Interferometry at Visible and Infrared Wavelengths, Proceedings, held in Garching bei Munchen, Germany, October 15-18, 1991.*, pp. 1057, 1992

Shaklan, S. B.; Colavita, M. M.; Shao, M. "Visibility Calibration Using Single-Mode Fibers in a Long-Baseline Interferometer," *ibid.*, pp. 1271, 1992

St-Jacques, D. et al. "Spatial filtering using pinholes at COAST," *Proc. SPIE 3350*, pp. 762, 1998

Yariv, A. *Introduction to optical electronics*, Holt, Rinehart and Winston, U.S., 1971