Preprint
UCRL-JC-147058

Fast Wavefront Reconstruction in Large Adaptive Optics Systems Using the Fourier Transform

Lisa A. Poyneer, Donald T. Gavel and James M. Brase

This article was submitted to Astronomical Telescopes and Instrumentation, Waikoloa, HI, August 22-28, 2002

March 13, 2002

U.S. Department of Energy

Lawrence
Livermore
National
Laboratory

Fast wavefront reconstruction in large adaptive optics systems using the fourier transform

Lisa A. Poyneer, Donald T. Gavel and James M. Brase
Lawrence Livermore National Lab, Livermore, CA 94551 USA
poyneer1@llnl.gov, gavel1@llnl.gov, brase1@llnl.gov

March 13, 2002

Abstract

Wavefront Reconstruction using the Fast Fourier Transform and spatial filtering is shown to be computationally tractable and sufficiently accurate for use in large Shack Hartmann-based adaptive optics systems (up to at least 10,000 actuators). This method is significantly faster and can have lower noise propagation than traditional VMM reconstructors. The boundary problem which prevented the accurate reconstruction of phase in circular apertures using square-grid FTs is identified and solved. The methods are adapted for use on the Fried-geometry. Detailed performance analysis of mean squared error and noise propagation for FT methods is presented, using both theory and simulation.

1 Introduction

Current Adaptive Optics (AO) systems use vector-matrix-multiply (VMM) reconstructors to convert gradient measurements to wavefront phase estimates. As the number of actuators n increases, the time to compute the reconstruction using the VMM method scales as $O(n^2)$. The number of actuators involved in AO systems is expected to increase dramatically in the future. In astronomical applications, this is due to both increasing telescope diameters and new higher-resolution applications on existing systems. This increase in size, from hundreds to thousands or tens of thousands of actuators, requires that a faster method for wavefront reconstruction be developed.

A wavefront reconstruction method using the Discrete Fourier Transform (DFT) was suggested by Freischlad and Koilopolous.[1] This method was for square apertures on the Hudgin geometry. A further paper considered small circular apertures.[2] This paper builds on that work in four important ways. First, the circular-aperture case is thoroughly examined. The boundary problem is identified, showing that use of only in-aperture data for circular apertures leads to large errors. Two methods for solving this boundary problem are presented. They both provide perfect reconstruction of sensed modes when no noise is present. Second, the Fourier Transform (FT) methods are adapted to use on the Fried geometry. Third, the performance of these methods, in terms of both speed and reconstruction error, is analyzed. Reconstruction error is analyzed as an estimation problem, which leads to a linear model of system error in response to noise. Theoretical results for small apertures are confirmed by simulation. Finally, the performance of large systems (up to 50,000 actuators) is examined through simulation. By using an FT method presented in this paper, the implementation of a 10,000 actuator system with satisfactory speed and reasonable error performance is feasible given current technology.

This paper is focused on the performance of FT reconstructors using specific discrete models. Therefore it will not directly address phenomena associated with the correction of continuous wavefronts such as branch points or partially obscured subapertures.

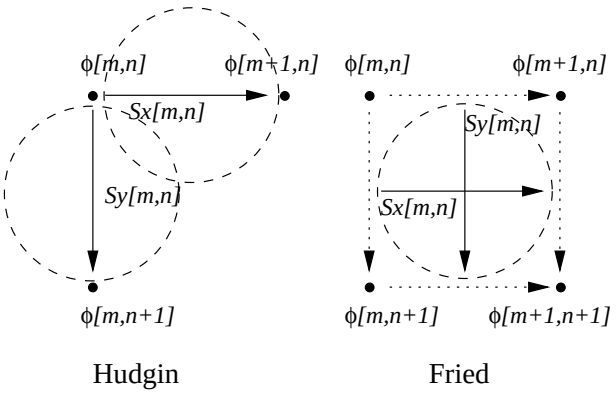

Figure 1: The Hudgin and Fried sensor geometries. The circles represent the Shack-Hartmann wavefront sensor locations. In the Hudgin geometry the gradients are first differences. In the Fried geometry the gradients are the average of the two nearest first differences to the subaperture.

2 The Inverse Spatial Filter

The basic inverse spatial filter is presented here again for reference. Using the Hudgin geometry,[3] the sensor measurements are modeled as the first differences of the wavefront phase. This corresponds to wavefront sensors centered between each pair of points. See Figure 1. The piston-removed (across the aperture) phase $\phi[m,n]$ is an $N \times N$ discrete signal. The gradients $\mathrm{s_x}[m,n]$ and $\mathrm{s_y}[m,n]$ are simply the first differences between adjacent phase points.

$$\mathrm{s_x}[m,n] = \phi[m+1,n] - \phi[m,n] \tag{1}$$

$$\mathrm{s_y}[m,n] = \phi[m,n+1] - \phi[m,n] \tag{2}$$

The DFT is applied to a finite-duration signal, using the assumption that the signal is periodic. The forward transform of a spatial signal of size $N \times N$ is

$$\mathrm{X}[k,l] = \mathcal{F}\{\mathrm{x}[m,n]\} = \frac{1}{N^2}\sum_{p=0}^{N-1}\sum_{q=0}^{N-1}\mathrm{x}[p,q]\exp(-\frac{j2\pi}{N}[kp+lq]) \tag{3}$$

Using the shift property of the DFT, Equations 1 and 2 become

$$\mathrm{S_x}[k,l] = \Phi[k,l](\exp(\frac{j2\pi k}{N}) - 1) \tag{4}$$

$$\mathrm{S_y}[k,l] = \Phi[k,l](\exp(\frac{j2\pi l}{N}) - 1) \tag{5}$$

To get the inverse filter, multiply each of the above equations by the complex conjugate of its exponential term and combine them, solving for $\Phi[k,l]$.

$$\widehat{\Phi}[k,l] = \begin{cases} 0 & k,l = 0 \\ \left[(\exp(-\frac{j2\pi k}{N}) - 1)\mathrm{S_x}[k,l] + (\exp(-\frac{j2\pi l}{N}) - 1)\mathrm{S_y}[k,l]\right] \times & \\ \left[4(\sin^2\frac{\pi k}{N} + \sin^2\frac{\pi l}{N})\right]^{-1} & \text{else} \end{cases} \tag{6}$$

The pole of the filter at $k,l = 0$ is fixed by making that value zero, which sets the DC gain (or piston) of the wavefront phase across the whole square grid to zero. As this mode is disregarded in reconstruction, it does not add any error. Taking the inverse transform produces the estimate $\hat{\phi}[m,n]$.

Freischlad analyzed the performance of this method on a square aperture and on small circular apertures. But a systematic study of the how to use this filter on large circular apertures remained. The following section presents the results of such a study.

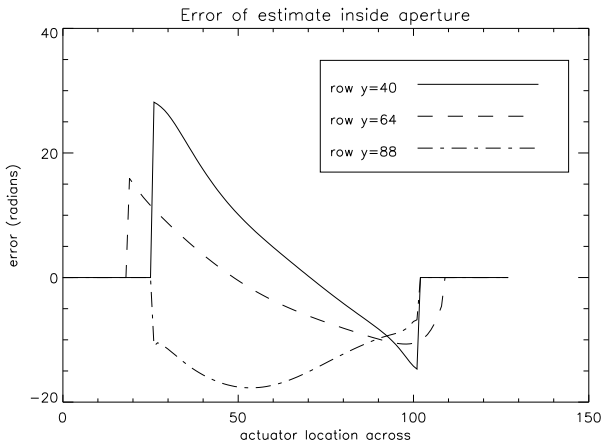

Figure 2: Estimate error from reconstruction of just the gradients inside the aperture, with the rest set to zero. No noise was added so as to clearly isolate the effects of the boundary. The gradients were calculated directly from phase points using Equations 1 and 2. Each curve in the plot is a slice across the aperture along a row of actuators. This simulation was done on a 6376 actuator system, which was 90 actuators in diameter on a 128× 128 grid. The input phase aberration had rms error 1690 nm. The reconstruction had rms error of 1002 nm for this trial. The error spans the aperture and is not easily removed. (Note how the error changes shape and sign from row 40 to row 88.) Using either of the methods presented in Sections 3.1 and 3.2, the reconstruction error was essentially 0 nm.

3 Reconstruction on a circular aperture

The inverse filter was derived for a regular grid of gradient measurements. When dealing with a real AO system (astronomical telescopes in particular), the gradients are typically available only on a circular aperture. The measurement data cannot be simply zero-padded and filtered. Doing so produces huge errors. See Figure 2 for an illustration of these errors. In the first part of this section this boundary problem is identified and explained. Then two methods for altering the gradient data from a circular aperture are presented. These methods produce perfect reconstruction of sensed modes when there is no noise.

There are two key assumptions of the inverse filter derivation which must be satisfied in the ideal case for it to work. The first assumption is that ϕ is spatially periodic. This assumption is necessary for use of the DFT, and it must be maintained for a set of gradient measurements. A check on this condition is that the sum of every row (for x-gradients) or column (for y-gradients) in the $N \times N$ gradient signal sums to zero. The second assumption is based on the modeling of the gradients as first differences. Any closed path of gradients must sum to zero. Both these conditions were acknowledged by Freischlad in his papers.

Zero-padding the gradient measurements is incorrect because in the general case the zero-padded gradient set violates both spatial periodicity and the closed-path loop condition. These inconsistencies manifest themselves in errors that span the aperture. The errors do not become less significant as the aperture size increases.

Proof of this comes from an examination of the Hudgin geometry and reconstruction process. Consider the wavefront phase $\phi[m,n]$ across a square aperture. To obtain a circular aperture, the phase is multiplied by a signal $\mathrm{a}[m,n]$, which is 1 at each point inside the aperture and 0 outside. If the gradients $\mathrm{s_x}[m,n]$ and $\mathrm{s_y}[m,n]$ are calculated directly from this apertured signal, they can be divided into three parts. These are illustrated in Figure 3.

The first part, called the inside gradients $\mathrm{i_x}[m,n]$ and $\mathrm{i_y}[m,n]$, represents the gradients which connect two points inside the aperture. These gradient values are not affected by the aperture $\mathrm{a}[m,n]$. The second part is the boundary gradients which connect one point inside the aperture to one point outside of it. These gradients, called $\mathrm{b_x}[m,n]$ and $\mathrm{b_y}[m,n]$, directly describe the transition from the signal to the zero-phase outside the aperture. The third part connects pairs of points outside the aperture. Note that these outside

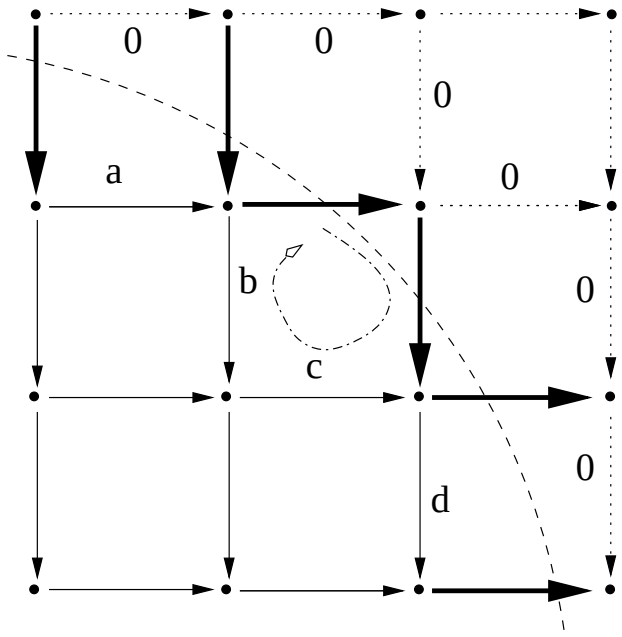

Figure 3: The three types of gradients are shown in the Hudgin geometry at the edge of an aperture. In bold are the boundary gradients b_x and b_y. These connect phase points across the aperture edge. In thin solid lines are the inside gradients i_x and i_y, which can be obtained from measurement. The dotted lines are the outside gradients. Note that a closed loop across the aperture edge does not sum to zero if the boundary gradients are set equal to zero.

gradients are equal to zero everywhere, so they will be ignored.

$$s_x[m,n] = i_x[m,n] + b_x[m,n] \tag{7}$$

$$s_y[m,n] = i_y[m,n] + b_y[m,n] \tag{8}$$

Each of these signals has values for the gradients it describes and is zero for all other values. Because the filtering process is linear, we can consider the results of filtering each portion of $s_x[m,n]$ and $s_y[m,n]$ individually. Filtering the whole $s_x[m,n]$ and $s_y[m,n]$ will produce an exact reconstruction. Taking only the values inside the aperture and zero-padding is equivalent to filtering just $i_x[m,n]$ and $i_y[m,n]$. This means that the error of this estimate is exactly the result of filtering $b_x[m,n]$ and $b_y[m,n]$.

A particularly useful way to examine this error is to calculate the boundary gradients for Zernike modes and then reconstruct them. A few modes, such as piston and focus, produce no error. The others, however, produce complicated error terms that span the aperture. Because the phase aberrations on a circular aperture can be represented via Zernike modes, this means that the total error will be a combination of many different error terms. The actual combination is directly dependent on the phase aberration and can not be removed based on knowledge of the general statistics of the aberration.

Conceptually, the boundary gradients describe the transition of the edges of the wavefront. If they are set to zero, they describe a wavefront where the phase just outside the aperture has the same value as just inside the aperture. The outside gradients also being zero describes a wavefront that has constant phase everywhere outside the aperture. Both these conditions can be true only if the wavefront inside the aperture has a constant value along the edge. Pure piston and pure focus would satisfy this condition. But for an arbitrary wavefront phase aberration, this is almost never the case.

The previous development suggests that a method for obtaining a correct set of gradients involves estimation of the boundary and/or outside gradients. Two different methods are now presented that generate consistent sets of gradients. With no noise, both methods produce perfect reconstruction of all the sensed modes. Furthermore, it will be later shown that the methods require only $O(n)$ operations, preserving the speed advantage of the FT methods.

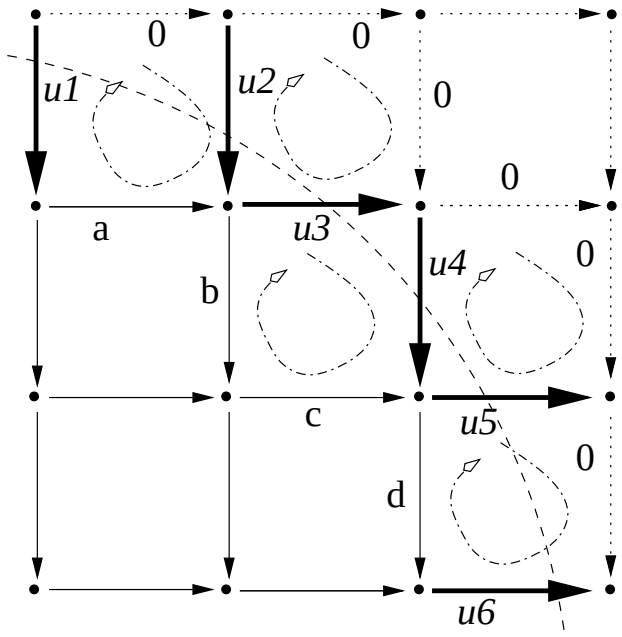

Figure 4: The *Boundary* method. Setting each closed loop across the aperture edge to zero results in an equation relating the unknown boundary gradients to the measured inside gradients and the zeroed outside gradients. In this example, the equations for the boundary gradients u1, u2, etc. are as follows, starting from the upper left corner: $-u1 + u2 = a$, $u2 + u3 = 0$, $u3 + u4 = c + b$, $u4 + u5 = 0$, $u5 - u6 = d$. The complete set of equations for the whole aperture form a linear system, which is then solved to get the estimate of the boundary gradients.

3.1 The *Boundary* Method

The first method is called the *Boundary* method because it estimates the gradients which cross the boundary of the aperture. It follows directly from the development in the previous section of inside, boundary and outside gradients. This process is shown in Figure 4.

Only the inside gradients are known from measurement. The outside gradients can all be set to zero. This leaves the boundary gradients undetermined. A loop continuity equation can be written for each of the two smallest loops which involve a boundary gradient. Setting each of these equations to zero describes a solution which satisfies loop continuity across the whole grid. Using the configuration shown in Figure 4, a partial list of the loop equations is

$$\begin{aligned} -u1 + u2 &= a \\ u2 + u3 &= 0 \\ u3 + u4 &= c + b \\ u4 + u5 &= 0 \\ u5 - u6 &= d \end{aligned} \tag{9}$$

All of these loop continuity equations involving the boundary gradients combine to form a linear system. Where $\mathbf{u}$ is the vector of all boundary gradients and $\mathbf{c}$ is a vector containing sums of measured gradients, the system can be expressed as

$$\mathbf{Mu} = \mathbf{c} \tag{10}$$

The matrix $\mathbf{M}$ is fixed for a given geometry. The vector $\mathbf{c}$ has a fixed combination of gradients, but the value of these gradients depends on the actual measurement. If there is no noise, the system has an infinite number of valid solutions. Each of these solutions represents, in essence, a different piston offset of the aperture from the background of zero phase.

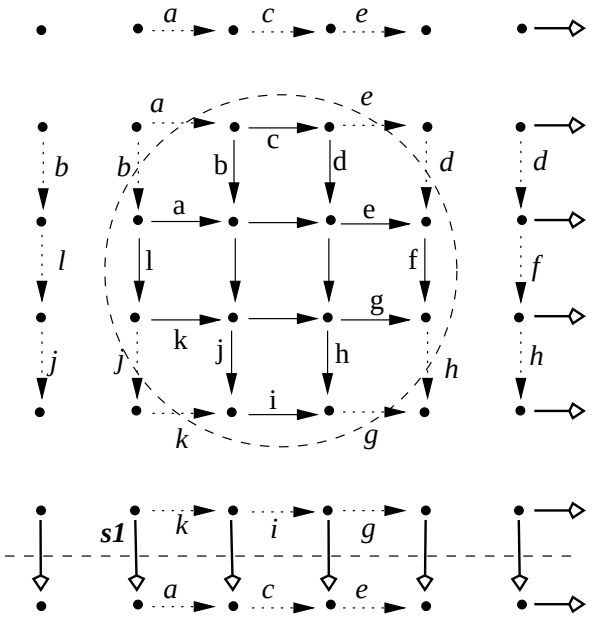

Figure 5: The *Extension* method, shown for $N = 6$. The gradients closest to the aperture edge are extended outside the aperture. For example, gradients a, c and e are each uppermost in their columns and are extended upwards out of the aperture. The unmodified gradients, which are left as zeroes, are not shown in this figure for clarity. The seam gradients are along the right and bottom edges. These gradients 'connect' the spatially periodic copies of the wavefront phase, and they must be set so that every row of the x-gradients and every column of the y-gradients sums to zero. For example, the left-most column in this case must satisfy the equation $s1 = -b - l - j$. Examination of this figure shows how loop continuity is satisfied exactly by the *Extension* method.

When there is noise on the measurements, this system has no exact solution in general. The boundary gradients can instead be estimated using a least-squares projection. This estimation can be solved by such methods as using the pseudoinverse of $\mathbf{M}$. In this case, the boundary gradients are correlated to the noisy measurements, but generally have much higher variance.

3.2 The *Extension* Method

The second method of obtaining a consistent set of gradients is called the *Extension* method because it extends out the gradients from inside the aperture. It is illustrated in Figure 5.

The *Extension* method extends the wavefront shape to outside of the aperture. It does this by a simple method which is based on preserving loop continuity. The x-gradients are extended up and down out of the aperture, while the y-gradients are extended to the left and right. For example, the uppermost x gradient in a given column in the aperture has its value repeated in all the outside gradients above it. Loop continuity is preserved, even where the extended gradients cross each other, such as at the 'corners' of the aperture. All the smallest loops involving these new extended gradients will automatically sum to zero. Each extended gradient is cancelled out in the smallest-loop equation by one with the identical value below or above it.

The final step is to fix the spatial periodicity. The seam gradients are those which connect one copy of the phase signal to the next. Observe in Figure 5 at the bottom how the extended values from the top of the aperture meet those from the bottom. The seam gradients connect them. These seam gradients are set using the spatial periodicity condition by summing up the row or column in which they are in and setting that sum to zero. This exactly satisfies the smallest loop conditions for the seam gradients as well.

The *Extension* method produces a completely consistent set of gradients. These provide perfect reconstruction of the phase when there is no noise, except for the piston. If there is noise, the exact same procedure is done, though loop continuity will not hold on loops involving the seam gradients, just as the boundary

gradients in the *Boundary* method were the best, but not exact, solution when there was noise.

The principles and methods used above can be applied to fill in missing information in general, such as with different aperture shapes or the central obscuration due to the secondary mirror in a large telescope.

4 Adapting to a different sensor geometry

The inverse filter as described previously is specific to the Hudgin geometry. The Fried geometry[4] is frequently used, however, in modeling the behavior of Shack-Hartmann (SH) sensors. It models the gradients that are generated by SH sensors which are centered between the phase points. This allows one sensor to provide both x- and y-gradient measurements. See Figure 1 for an illustration and comparison with the Hudgin geometry. This sensor configuration is common in astronomical AO systems. Its features and implications have to considered if an FT method is used for reconstruction. In this section the inverse filter is derived for the Fried geometry and the *Boundary* and *Extension* methods are adapted to it.

In the Fried geometry the gradient is modeled as the average of the two nearest first differences.

$$\mathrm{s_x}[m,n] = \frac{1}{2}(\phi[m+1,n] - \phi[m,n] + \phi[m+1,n+1] - \phi[m,n+1]) \tag{11}$$

$$\mathrm{s_y}[m,n] = \frac{1}{2}(\phi[m,n+1] - \phi[m,n] + \phi[m+1,n+1] - \phi[m+1,n]) \tag{12}$$

Using the same method as before, the inverse spatial filter to reconstruct the phase can be derived.

$$\widehat{\Phi}[k,l] = \begin{cases} 0 & k,l = 0,\ k,l = N/2 \\ \left[\begin{array}{c} (\exp(-\frac{j2\pi k}{N}) - 1)(\exp(-\frac{j2\pi l}{N}) + 1)\mathrm{S_x}[k,l] + \\ (\exp(-\frac{j2\pi l}{N}) - 1)(\exp(-\frac{j2\pi k}{N}) + 1)\mathrm{S_y}[k,l] \end{array}\right] \times & \\ \left[8(\sin^2\frac{\pi k}{N}\cos^2\frac{\pi l}{N} + \sin^2\frac{\pi l}{N}\cos^2\frac{\pi k}{N})\right]^{-1} & \text{else} \end{cases} \tag{13}$$

This filter has two poles that are zeroed out, one at the piston mode, the other at the highest frequency, or waffle mode. This latter mode is not sensed in the Fried geometry. Nor is it normally controlled for in AO systems. A waffle error will therefore be present, the magnitude of which depends on the amount of waffle in the input phase. In practice, under atmospheric turbulence this waffle component is actually quite small. Rough calculations show that under Komolgorov turbulence, the variance of the wavefront due to waffle is related to the variance of the piston removed wavefront by the following equation. D is the aperture diameter and d is the subaperture diameter.

$$\sigma^2_{\phi_w} \approx 1.33\left(\frac{d}{D}\right)^{\frac{11}{3}}\sigma^2_{\phi} \tag{14}$$

This means that for 12 or more subapertures across the diameter of the telescope, the waffle component is at most 0.01% of the total piston-removed wavefront variance.

4.1 Dealing with a circular aperture

We now know that the values of the gradients outside the aperture must be estimated with some method for proper reconstruction. But in the Fried geometry the gradients do not connect points, so the *Boundary* and *Extension* methods can not be applied directly. A simple invertible linear transform can convert the gradients into two sets that do connect points on the grid. The gradients are now oriented along a different orthogonal basis, in the directions referred to as a and b.[5] See Figure 6.

$$\mathrm{s_a}[m,n] = \mathrm{s_x}[m,n] + \mathrm{s_y}[m,n] \tag{15}$$

$$\mathrm{s_b}[m,n+1] = \mathrm{s_x}[m,n] - \mathrm{s_y}[m,n] \tag{16}$$

With a few minor modifications, such as making the new grids square-shaped, the *Boundary* and *Extension* methods can be applied directly to these two sets in the a and b directions.

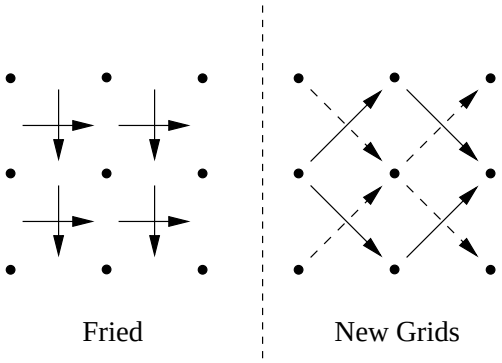

Figure 6: The coordinate transform from the Fried geometry produces two disjoint grids. One grid is connected by the dashed gradients, the other by the solid gradients. Combining these uncoupled grids produces waffle error.

The primary difficulty of using the Fried geometry arises from the recombination two disjoint grids. See Figure 6 for an illustration of the two grids. These grids are independent of each other, which is a direct result of the waffle mode being unsensed and uncontrollable. Processing each grid independently and then recombining them can result in large waffle errors. Fortunately, when no noise is present this waffle error is simple to remove, as it has constant magnitude in the aperture.

When there is noise, matters become more difficult. In practice the *Boundary* and *Extension* methods produce dissimilar noise after filtering. The *Boundary* method was observed to cause waffle-like errors that varied widely across the aperture. The magnitude and sign of the waffle typically changed across the aperture in ways dependent on the noise and specific realization of phase aberration. The noise could not be cleanly removed. The *Extension* method produced more correlated noise on the estimate. It did not normally vary in sign and the magnitude variations were much less severe. The *Extension* method's noise could be removed for the most part in a simple waffle-removal step. This modal removal is discussed in Section 5

4.2 Adapting to an arbitrary sensor geometry

One of the primary concerns with this method is its dependence upon the sensor model. In practice, VMM reconstructors are obtained directly from the AO system. A matrix relating the actuator behavior to sensor measurements is inverted to get the reconstruction matrix. The FT method does not lend itself to such a derivation in general. This method does not convert a system matrix into a reconstruction filter. It instead derives a filter given a regular model of system behavior.

5 Modal removal

As the previous section has demonstrated, it is desirable to discard certain modes from the reconstruction. Piston is a mode which is discarded in the general case. A second mode which is normally removed is waffle. This is because Fried-geometry reconstructions, including FT methods, can introduce large waffle errors in the estimate. In a VMM reconstructor, this modal removal can be built directly into the matrix. Modal removal in an FT-based method is a separate step, but removing a single mode is quick and easy to implement in $O(n)$ steps.

The fastest way to remove a mode from the estimate while spatial filtering would be to identify the frequency coefficients for that mode and zero them. The spatial filter already does this for piston (and waffle in the Fried-geometry case) across the whole square grid. However, zeroing the frequency coefficients is not a good method in practice. This is because modes of interest inside the aperture are not compactly represented in frequency space, so determining the correct amount to remove is non-trivial. Secondly, though the power at high frequencies is low, it is essential for sharp features. Incorrectly removing high frequency components can have deleterious effects on the accuracy of the estimate, especially at the edges.

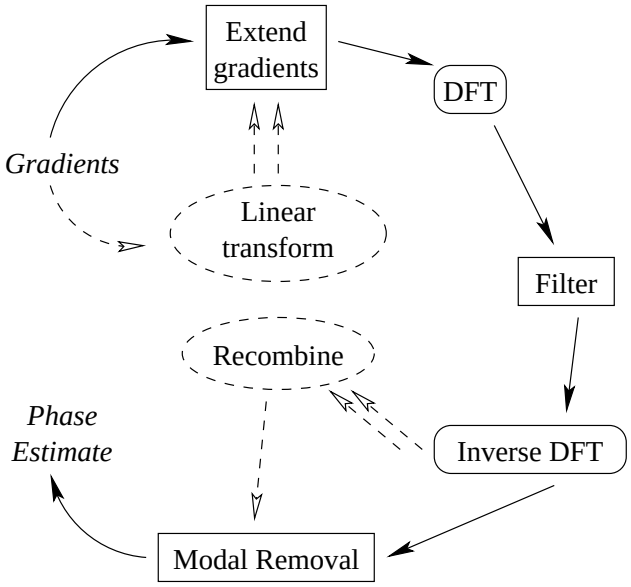

Figure 7: The complete process of FT reconstruction. For the **Grid** method, which is based on the Hudgin geometry, the gradient measurements are first extended. They are then Fourier transformed, filtered, inverse transformed and piston is removed. For the **Center** method, which is based on the Fried geometry, a few more steps are required. These are illustrated by the dashed arrows. The gradients are first converted to the two grids. Then they are extended and filtered. The two results are recombined and then waffle and piston are removed.

Instead, the modal removal process occurs in the spatial domain. The coefficient for the mode is determined by projection. Where $\mathrm{v}[m,n]$ is the mode of concern,

$$c_v = \frac{\sum_{m=0}^{N-1}\sum_{n=0}^{N-1}\hat{\phi}[m,n]\mathrm{v}[m,n]}{\sum_{m=0}^{N-1}\sum_{n=0}^{N-1}\mathrm{v}[m,n]\mathrm{v}[m,n]} \tag{17}$$

Then the estimate with that mode removed is

$$\hat{\phi}_{-v}[m,n] = \hat{\phi}[m,n] - c_v\mathrm{v}[m,n] \tag{18}$$

Care must be taken when removing modes to have an orthogonal basis set. For example, a waffle-mode vector of ± 1 over an odd number of actuators is not orthogonal to the piston vector. If waffle was removed after piston, it would add back a small piston in the aperture.

6 The Complete Methods

The previous sections have identified the boundary problem with circular apertures and presented methods to solve it. Both the Hudgin and Fried geometries have been considered. For the performance analysis, the best overall method for each geometry, as determined by analysis and simulation, is now presented. The complete methods are illustrated as a flow chart in Figure 7

For the Hudgin geometry the method will be called **Grid**, in reference to the gradients connecting points on a grid. Assuming this geometry, the FT reconstruction consists of the following steps. The *Extension* method is applied to the measured gradients. Then the gradients are Fourier transformed and the inverse filter (Equation 6) is applied. The result is inverse transformed and piston is removed to obtain the final estimate.

For the Fried geometry the method will be called **Center** because the gradient measurements are in the center of each subaperture. The gradients are first converted from the x, y directions to the a, b directions

(recall Equation 15). The two grids are made consistent using the *Extension* method. These two grids Fourier transformed and then filtered individually, using Equation 6. The results are inverse-transform and recombined. Waffle and piston are then removed.

These two methods reconstruct all sensed modes perfectly when there is no noise. Next their speed and performance in the presence of noise will be analyzed.

7 Time analysis

The whole purpose of exploring an FT approach to wavefront reconstruction is to exploit the speed the Fast Fourier Transform (FFT). For sizes that are powers of two, this implementation of the DFT is $O(n \log n)$ to compute, where n is the total number of elements. A VMM reconstruction on the same number of elements is $O(n^2)$. Finding a faster way to reconstruct than the VMM is essential for large AO systems.

We must also consider the time cost of the extra processing steps in the FT reconstruction to insure that it remains fast. These extra operations must be at most $O(n \log n)$ so as to not adversely affect performance. For the analysis of the extra processing, the following parameters are used. The width (in phase points) of the square grid is N. The radius of the aperture is R. The number of total elements is $n = N^2$. The number of actuators is approximately $a = \pi R^2$

For the **Grid** method the time analysis is simple. The first step is the *Extension* process. In each dimension, $4R$ gradients are extended to fill out $2RN - a$ other gradients. Then $2N$ seam gradients are set, each by addition of $N - 1$ other gradients. If all of these operations are done serially, this is $2(2RN - a) + 2N(N - 1)$ fundamental operations. For small apertures on large grids, R is at least $N/4$, making the number of above operations be $(3 - \pi/4)N^2 - 2N$. This is $O(n)$. For very large apertures where R is nearly $N/2$, there are $(4 - \pi/2)N^2 - 2N$ operations. This is also $O(n)$.

After the signals are Fourier transformed, the filter is applied by multiplication and addition over the $N \times N$ grid. This is $O(1)$ operations over n elements, or $O(n)$. After the inverse transform, the piston is removed. This removal is an addition of all of the elements and a subtraction of the result from each one, which is again $O(n)$. Therefore the extra processing in the **Grid** method is $O(n)$.

The **Center** method is not so simple. The first step is the linear transform of the gradients. This requires $2n$ additions. Then the *Extension* process is applied to the two separate grids. In this case R varies from $N/(4\sqrt{2})$ to $N/(2\sqrt{2})$. This still results (as above) is $O(n)$ operations. Setting the two seam gradients for two grids is still $O(n)$. The same filtering as in **Grid**, is done, but twice. The final waffle removal is the same amount of computation as the piston removal, namely $O(n)$. So the extra processing in the **Center** method is also $O(n)$, though for a given geometry the **Center** method requires more than twice as much computation.

The processing for both the **Grid** and **Center** methods is $O(n)$. The overall processing time for FT reconstruction is therefore dominated by the actual DFT implementation. For power-of-two sized grids, the FFT is $O(n \log n)$. Using other size grids requires having a fast implementation, such as one based on prime factors.

Not only are the FT reconstructors faster in theory, the implementation of a large system is within computational reach today. A rough estimate of performance requirements illustrates this. For an AO system running in serial computation at 200 frames per second, the entire reconstruction must be done is 5 ms. This means that for the **Grid** method the data extension, two forward FFTs, the filter application and one inverse FFT must be completed in at most 5 ms. A good estimate of FFT time is then 1 ms. For the **Center** method there are twice as many operations, so a single FFT needs to be done in 0.5 ms. A 10,000 actuator system requires a 128×128 element FFT. This can be done in IDL on a 1.7 GHz Pentium 4 in 1.8 ms. This result could be reduced by faster hardware or different implementation.

8 Performance Analysis

It is clear that the FT methods are fast. It must now be shown that they are reasonable accurate, as well. This is particularly important because these methods were not derived or proven to be optimal; they were developed to be fast. By modeling the reconstruction process as an estimation problem, a powerful framework

for performance analysis can be established. In the case of white noise on the gradient measurements, the mean squared error of the reconstruction can be expressed an a linear function of the noise variance. This allows easy comparison of different methods. This section will develop the estimation-problem analysis and discuss the significance of the various error metrics.

8.1 Modeling the measurement and reconstruction processes

The wavefront sensing and reconstruction process can be modeled using vectors and matrices. This model is a simple one - it treats the aberrated wavefront phase as a set of discrete points. It allows for different geometries for gradient measurement (e.g. Hudgin and Fried) to be used, as well as arbitrary reconstruction methods.

The first step in the model is the conversion of the wavefront phase to gradient measurements. The gradients $\mathbf{g}$ are generated from the piston-removed wavefront ϕ via matrix $\mathbf{H}$ under additive noise $\mathbf{n}$. This noise is assumed to be zero-mean and uncorrelated with the measurements and wavefront, with covariance matrix $\mathbf{\Lambda_n}$.

$$\mathbf{g} = \mathbf{H}\phi + \mathbf{n} \tag{19}$$

The estimator function (e.g. reconstructor) can also be expressed as a matrix. If $\mathbf{M}$ is complex, as it is in the case of the FT method, it must be conjugated when it is transposed. However, the development below will omit notation of conjugation for clarity. The actuator commands $\hat{\phi}$ are estimated from measurements $\mathbf{g}$ with the matrix.

$$\hat{\phi} = \mathbf{M}\mathbf{g} \tag{20}$$

Note that this matrix $\mathbf{M}$ can express nearly any linear reconstruction method, whether a standard minimum least-squares (MLS) VMM or the complete **Grid** or **Center** methods. It is required that the reconstructor $\mathbf{M}$ produces an estimate with zero-piston. (If it does not it can be easily converted to do so by multiplication with a piston-removal matrix.) Furthermore, $\mathbf{M}$ is static and does not adaptively change with conditions, (as it would if it depended on the noise variance). This assumption is sufficient for the present analysis of FT reconstructors, though it is not applicable to the more general result that the optimal reconstruction matrix is dependent on the noise distribution.[6]

The error of this estimate is simply the difference between the estimate and the wavefront phase.

$$\varepsilon = \mathbf{M}\mathbf{g} - \phi \tag{21}$$

The bias of the estimate is defined as the expectation of the error of the estimate.

$$\mathbf{b} = \mathrm{E}[\varepsilon] \tag{22}$$

The error variance is then

$$\mathbf{\Lambda}_\varepsilon = \mathrm{E}[(\varepsilon - \mathbf{b})(\varepsilon - \mathbf{b})^T] \tag{23}$$

The mean squared error is also of interest. The mean squared error is a random variable which is the average of the squared error at every point inside the aperture. Therefore we want to deal with its expectation. Recalling that a is the number of actuators,

$$\mathrm{mse} = \frac{\mathrm{E}[\varepsilon^T\varepsilon]}{a} \tag{24}$$

8.2 Modeling the wavefront

8.2.1 Non-random Parameter Estimation

The wavefront has been frequently been modeled as a non-random parameter to be estimated. When ϕ is deterministic the bias and error variance reduce to

$$\mathbf{b} = (\mathbf{M}\mathbf{H} - \mathbf{I})\phi \tag{25}$$

$$\mathbf{\Lambda}_\varepsilon = \mathbf{M}\mathbf{\Lambda_n}\mathbf{M}^T \tag{26}$$

The error of the estimate is dependent on the model chosen to describe the WFS behavior, but the error variance is independent of that model and depends only on the reconstructor and the noise.

In this case the noise propagator metric comes naturally out of the error analysis. The noise propagator, called mse_{np}, is defined as the mean squared error divided by the average variance of the noisy slope measurements. The noise propagator is of dual significance in analyzing the performance of a reconstruction method. For a single configuration of sensors and actuators, it determines how the reconstruction responds to noise. For a group of various configurations of increasing size, the set of their mse_{np}'s can help describe how a system's size affects its response to noise. In the case of white noise of variance σ_n^2, the covariance matrix $\mathbf{\Lambda_n}$ is diagonal with value σ_n^2, which reduces the noise propagator expression to

$$\text{mse}_{np} = \frac{\text{mse}}{\sigma_n^2} = \frac{\text{Trace}(\mathbf{M}\mathbf{M}^T)}{a} \tag{27}$$

This result agrees with standard derivations.[5]

8.2.2 Random Vector Estimation

If the wavefront is assumed to be a random vector, with known mean $\mathbf{m}_\phi$ and covariance matrix $\mathbf{\Lambda}_\phi$, the bias and error variance can also be calculated as above, but with different results.

$$\mathbf{b} = (\mathbf{M}\mathbf{H} - \mathbf{I})\mathbf{m}_\phi \tag{28}$$

$$\mathbf{\Lambda}_\varepsilon = (\mathbf{M}\mathbf{H} - \mathbf{I})\mathbf{\Lambda}_\phi(\mathbf{H}^T\mathbf{M}^T - \mathbf{I}) + \mathbf{M}\mathbf{\Lambda_n}\mathbf{M}^T \tag{29}$$

Assuming that the wavefront phase is zero-mean in time, the estimate is unbiased, regardless of the structure of $\mathbf{M}$. This means that in time, the expected error is zero, though any instance has non-zero error. The mean squared error can be calculated using Equations 24 and 29.

$$\text{mse} = \frac{\text{Trace}\left(\mathbf{\Lambda}_\varepsilon\right)}{a} \tag{30}$$

This result depends on both the reconstructor and the method of gradient generation. This performance metric allows comparative evaluation of different reconstruction methods (by varying $\mathbf{M}$) and of different sensor models (by varying $\mathbf{H}$). This equation shows the importance of considering the impact of noise in the system. It is possible to design an $\mathbf{M}$ such that $\mathbf{M}\mathbf{H} - \mathbf{I} = \mathbf{0}$. However, this also affects the portion of the error due to the noise. The total error may not be minimal in this case.

8.2.3 Wavefront as a random vector

In order to use the performance metrics derived in the previous section, the statistics of the wavefront must be known. These statistics will be different for closed-loop and open-loop performance. Obtaining them is a non-trivial problem. These statistics could potentially be derived from theoretical knowledge of the phenomenon that produces the phase aberrations. This could be atmospheric turbulence, or the heating of optics due to high-power lasers. If the theoretical approach is not possible, the statistics could be estimated from observations of the process or by simulation. Wallner has presented a method[6] for assessing Komolgorov turbulence across an aperture with piston removed. Work by the author in applying this method is still in progress.

8.3 The linear model of mean squared error

The preceding results can be combined to create a linear model reconstruction performance. Expanding Equation 30 produces

$$\text{mse} = \frac{\text{Trace}\left((\mathbf{M}\mathbf{H} - \mathbf{I})\mathbf{\Lambda}_\phi(\mathbf{H}^T\mathbf{M}^T - \mathbf{I})\right) + \text{Trace}\left(\mathbf{M}\mathbf{\Lambda_n}\mathbf{M}^T\right)}{a} \tag{31}$$

The left-hand term of the numerator is the contribution to the mean squared error by the wavefront phase. For a given wavefront phase distribution, this remains fixed. This part of the error will be called mse_ϕ, the

latent error. The right-hand term is the contribution of the noise. This is exactly what the mean squared error is in the non-random parameter case. This depends entirely on the variance of the noise. This part of the error will be called $\text{mse}_n(\mathbf{\Lambda_n})$.

This allows the mean squared error to be broken down into two parts: the latent error mse_ϕ, which is fixed for a given phase aberration model, and $\text{mse}_n(\mathbf{\Lambda_n})$, which depends entirely on the amount of noise. Note that the earlier assumption of a reconstruction matrix independent of noise allows this simplification.

$$\text{mse} = \text{mse}_\phi + \text{mse}_n(\mathbf{\Lambda_n}) \tag{32}$$

Assuming white noise of variance σ_n^2, this further simplifies. Recalling Equation 27 defining the noise propagator, the above expression becomes

$$\text{mse} = \text{mse}_\phi + \sigma_n^2 \text{mse}_{np} \tag{33}$$

This equation describes a line. It says that the expected performance of a reconstructor with white noise on the measured gradients is simply a fixed component (the latent error) and noise component that grows linearly with the noise variance. This allows easy graphical comparison of the performance of various reconstructors under the same wavefront and noise conditions.

9 Performance Results with Hudgin and Fried geometries

Using the metrics defined in the previous section, the performance of the **Grid** and **Center** methods can be analyzed. The most significant result of this analysis is that for a given aperture size, a tradeoff exists between speed and error in FT methods. This is due to the specific grid sizes required by the FFT. This section also presents the noise propagator results for large systems and confirms the linear model of mean squared error.

Keep in mind that this performance analysis is done given the discrete model and specific geometry of each method. In particular, the gradients generated for the simulations are created directly from the Hudgin and Fried geometry equations (see Equations 1, 1, 11 and 12).

9.1 Noise Propagation

The theoretical noise propagator mse_{np} was analyzed for a variety of different aperture sizes for both the **Grid** and **Center** methods. The calculation comes directly from Equation 27. This was tractable for grid width up to 32 actuators across, allowing a circular aperture of 716 actuators. Of most concern, however, are systems with thousands to tens of thousands of actuators. To predict performance in this regime, simulation was relied on. Simulations to estimate the sample-mean of the noise propagator proved to be reasonably accurate compared to what theory predicted (see Figure 8). Simulation was therefore used to predict mse_{np} for larger apertures.

The results of simulation to estimate the noise propagator for large numbers of actuators are shown in Figures 9 and 10. Comparing the plots of the DFT and FFT performance shows that there can be significant performance loss when FFT is used. This is because the FFT grids are of fixed power-of-two sizes. If an aperture is 34 points across, it won't fit into a 32 $\times$ 32 grid, but must be reconstructed on a 64 $\times$ 64 grid. If the DFT were used, it could be reconstructed on a 36 $\times$ 36 instead. As the surrounding grid gets bigger, the noise propagation increases. This observed behavior is confirmed by analysis of the small aperture case. Figure 11 illustrates this for the **Grid** and **Center** methods on a 112 actuator aperture. The increase in noise propagation can be explained as part of the same phenomenon that causes noise propagation to increase with aperture size. Though the number of noisy gradient measurements is fixed, the gradients are duplicated more and more as the grid size is increased. A DFT is a linear combination of these points, which means as more are added the weighting to the noise is increased.

This speed versus error performance tradeoff has important design implications. Ideally, the aperture size will be just right so as to fit into a power-of-two grid. If the system is of an awkward size and cannot be changed, a choice must be made. For example, a system with 38,000 actuators could be calculated on a 220 $\times$ 220 grid using the DFT. This would have noise propagation of about 1.5 in the **Grid** case and 5.0

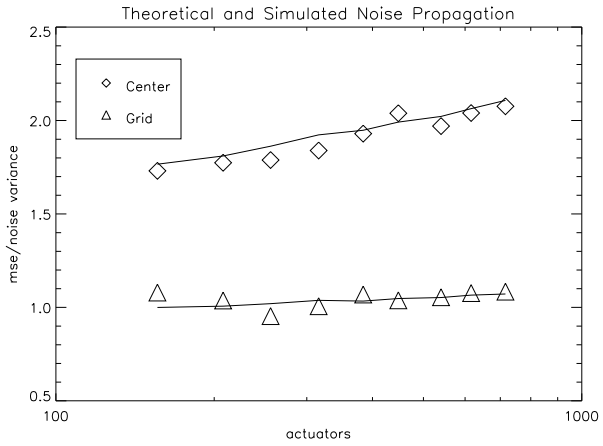

Figure 8: Theoretical and simulation results for noise propagation for the **Grid** and **Center** methods. The solid line is the noise propagator as determined theoretically. The data points are the results of simulation to predict the noise propagator. The simulation converges on the correct solution adequately enough to use it for larger numbers of actuators, the theoretical calculations of which are computationally intractable.

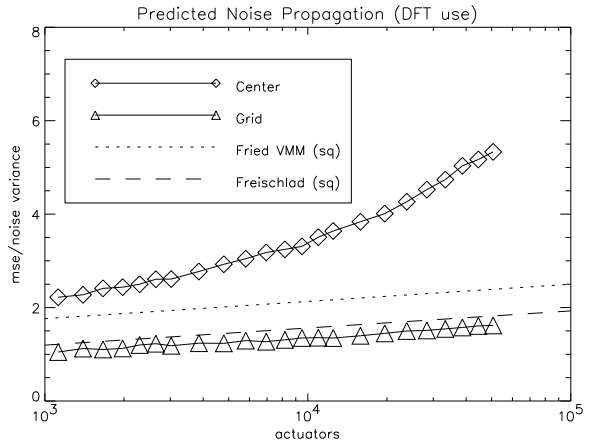

Figure 9: Simulation results for noise propagation for the the **Grid** and **Center** methods. The reconstructions were done on the smallest size grid possible to hold the aperture, and the computations done with an implementation of the DFT. The **Center** method has poor noise propagation for large numbers of actuators. The **Grid** method has excellent characteristics, with mse_{np} having magnitude less than 2 even out to 50,000 actuators. For comparison, the dotted line is the theoretical noise propagator for the Fried-geometry MLS VMM and the dashed line is the noise propagator determined by Freischlad for the FT reconstructor on a square aperture. See Equations 34 and 35

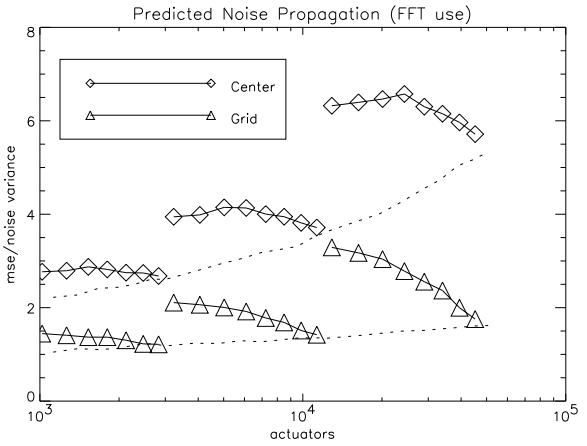

Figure 10: Simulation results for noise propagation for the the **Grid** and **Center** methods. The reconstructions were done on power-of-two sized grids in order to benefit computationally from using the FFT. The DFT results are in dotted lines for comparison. There is a clear performance loss for apertures significantly smaller than the surrounding grid. Only the largest size apertures in a given power-of-two grid approach ideal mse_{np} results.

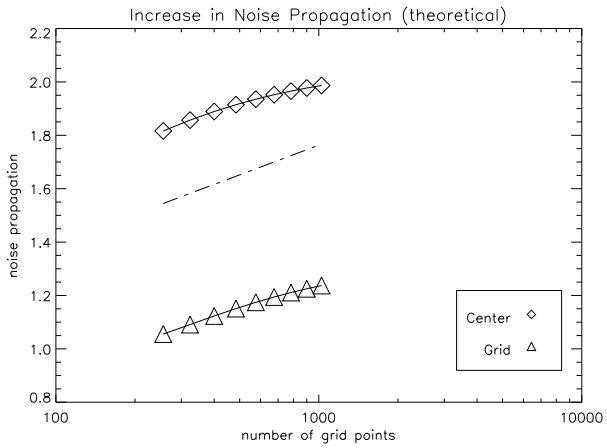

Figure 11: Theoretical results for the increase in noise propagation of a 112 actuator system as the surrounding grid is increased in size. For both methods, increasing the grid size increases the total noise propagation in a regular manner. The dashed line is the noise propagator increase for a square VMM.

in the **Center** case. But this DFT is slower the 256×256 FFT which could also be used. For example, in IDL this DFT takes twice as long to compute as the FFT. Though the FFT is faster, the resulting noise propagation would be increased to about 2.0 in the **Grid** case and 5.9 in the **Center** case. (See Figures 9 and 10.)

In the DFT case, the FT methods have reasonable noise propagation. For comparison, MLS VMM reconstructors have been shown to have an $c + d\ln(a)$ dependence for the noise propagator,[7] where c and d are constants and a is the number of actuators. For a square aperture, mse_{np} for a Fried-geometry least-squares VMM is[4]

$$\text{mse}_{np} = 0.6558 + 0.16028\ln(a) \tag{34}$$

Freischlad determined based on numerical data that mse_{np} for FT reconstruction on a square grid was[2]

$$\text{mse}_{np} = 0.09753 + 0.159156\ln(a) \tag{35}$$

The coefficient d is very close in these two expressions, showing that increasing size has nearly the same effect on performance for both methods. These are plotted against the DFT results in Figure 9.

For the DFT implementation, or for suitably sized apertures using the FFT, the FT methods have favorable noise propagation. The noise propagation for the **Grid** method is much lower than that for a square aperture VMM on the same number of actuators. The **Center** method has higher noise propagation, which starts to be significantly more at around 10,000 actuators.

9.2 Total mean squared error

The second part of the total error is the latent error. As discussed in Section 3, using the *Extension* and *Boundary* methods with the FT filtering produces perfect reconstruction of all sensed wavefront phase modes in the absence of noise. This means for the **Grid** method that the latent error is essentially zero. For the **Center** method the missed waffle mode is extremely small (see Section 4). Compare this with the huge error that occurs without proper processing, as shown in Figure 2.

Simulations confirm the linear model of mean squared error. Simulation and estimation were used to determine both $\mathbf{\Lambda}_\phi$ (which leads to mse_ϕ) and the sample mean of mse_ϕ. Figure 12 shows the results of monte carlo simulations with random realizations of the same wavefront phase profile and varying amounts of white noise on the measurements. For both small and large apertures this linear model of performance was confirmed, with the noise propagator and latent error from simulation agreeing closely with the predicted values.

10 Conclusions

The problem with reconstructing wavefront phase on circular apertures with a square grid has been identified and solved. Though large errors result from zero-padding, the methods presented in this paper enable accurate reconstruction of all sensed modes when no noise is present. These FT methods have been shown to work on the Fried geometry in addition to the Hudgin geometry. The extra processing steps were shown not to increase the order of growth of the FT method execution time. Detailed performance analysis produced a linear model for the mean squared error of reconstruction. The noise propagation of FT methods is reasonable for apertures that nearly fill the square grid, though there exists a tradeoff between speed of performance and reconstruction error when the FFT is used. The above results have been presented for large systems, up to 50,000 actuators. Based on the results in this paper, reconstruction method for a 10,000 actuator system could be realistically implemented using current technology and with adequate performance.

These results are for discrete models which are based on the Hudgin and Fried geometries. How these methods perform in a more continuous domain with SH wavefront sensors and a deformable mirror (DM) is being actively studied by the author. Initial results are very promising, though a more rigorous analysis of SH sensor and DM behavior is necessary. The boundary problem and large errors due to zero-padding still exist and must be corrected. Results indicate that latent error is increased from near zero to reasonable levels due to the SH sensor behavior and fitting error on the DM. Specifically, even without noise, a set of gradients measured with SH sensors will not satisfy loop continuity for either method. Neither the **Grid** or

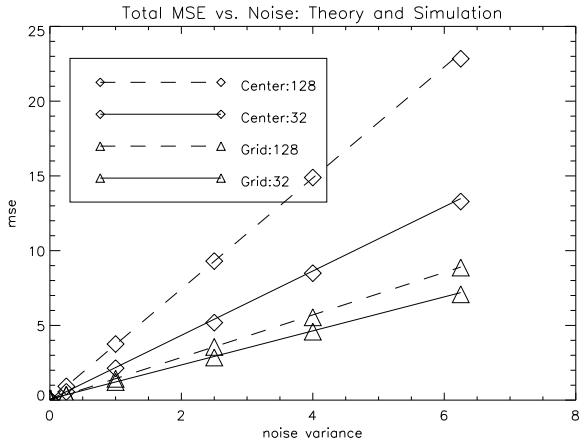

Figure 12: Total mean squared error versus noise variance for two systems. On the a 32×32 grid is a 448 actuator aperture. On the 128×128 grid is a 11,304 actuator aperture. Results for both **Grid** and **Center** methods are shown here. The lines are the predicted performance, based on theoretical or experimental mse_{np} and experimental mse_{ϕ}. The data points are the results of simulation at various levels of noise. The effect of the noise propagator is clearly shown in the different slopes. The latent error is essentially zero.

Center method has been a clear winner in terms of latent error, though both perform as well or better than VMM reconstructors on the same systems. Significantly, the linear model of mean squared error still holds in simulation. The most intriguing area of research is the behavior of DMs, which appear to suppress the noise propagation of FT methods in a highly regular manner.

11 Thanks

Special thanks to Carmen Carrano for all her assistance. This work was performed under the auspices of the U.S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.

References

[1] K. Freischlad and C.L. Koliopoulos. Wavefront reconstruction from noisy slope or difference data using the discrete fourier transform. In *Adaptive Optics*, volume 551, pages 74–80. SPIE, 1985.

[2] K. Freischlad. Wavefront integration from difference data. In *Interferometry: Techniques and Analysis*, volume 1755, pages 212–218. SPIE, 1992.

[3] R. H. Hudgin. Wave-front reconstruction for compensated imaging. *J. Opt. Soc. Am*, 67(3):375–378, 1977.

[4] D. L. Fried. Least-square fitting a wave-front distortion estimate to an array of phase-difference measurements. *J. Opt. Soc. Am*, 76(3):370–5, 1977.

[5] J. Hardy. *Adaptive Optics for Astronomical Telescopes*. Oxford University Press, New York, 1998.

[6] E.P. Wallner. Optimal wave-front correction using slope measurements. *J. Opt. Soc. Am*, 73(12):1771–6, 1983.

[7] R.J. Noll. Phase estimates from slope-type wave-front sensors. *J. Opt. Soc. Am*, 68(1):139–140, 1978.