(2)

AD-A221 707

FINAL TECHNICAL REPORT

OPTICAL WAVEFRONT COMPUTER FOR STABILIZING LARGE SPACE STRUCTURES

SPONSORED BY:	Scientific Officer Office of Naval Research Detachment, Boston 495 Summer Street Boston, MA 02210-2109
TECHNICAL CONTRACT MONITOR:	Mr. William Miceli
CONTRACT NO.	N00014-89-C-0223
TOPIC NO.	89-012
EFFECTIVE DATE OF CONTRACT:	1 September 1989
CONTRACT EXPIRATION DATE:	28 February 1990
NAME OF CONTRACTOR:	Optron Systems, Inc. 3 Preston Court Bedford, MA 01730 (617)275-3100 (617)275-3106 (Fax No.)
PRINCIPAL INVESTIGATOR:	Dr. Ira Farber
SHORT TITLE OF WORK:	Optical Wavefront Computer for Stabilizing Large Space Structures
PERIOD OF WORK COVERED:	1 September 1989 through 28 February 1990

Contents

List of Figures

1 Introduction

This final report describes the theoretical and design analysis performed under a Phase I SBIR contract to investigate a new adaptive optics system architecture for wavefront correction based primarily on optical processing. The highlights of the Phase I accomplishments are:

- The design of a new adaptive optics system architecture incorporating optical rather than electronic processing for wavefront sampling and reconstruction. Unlike conventional adaptive systems incorporating electronic processors, the new optical system exhibits a framing rate that is independent of the sampling and reconstruction resolutions.
- An analysis of requirements and specifications for all major components needed to implement the new optical architecture, including a lenslet array, an image intensifier tube, an optically addressed binary amplitude spatial light modulator, a transmission mask, a computer generated wavefront reconstruction hologram and an optically addressed phase modulator.
- Design of the minimum mean square error phase reconstruction hologram.

Based on the results of this Phase I effort Optron Systems is optimistic about proceeding with a Phase II effort to implement the new architecture.

This final technical report is organized into six sections. The rest of this section contains a brief background on adaptive optics concepts and applications, and it identifies the useful role of the new adaptive architecture in these applications.

Section 2 provides a conceptual explanation of the new adaptive optics architecture.

Section 3 takes a more in-depth look at the various stages of the wavefront detection and reconstruction process, discussing system limitations on bandwidth, light sensitivity, phase dynamic range, and signal to noise ratio.

The final section consists of our conclusions concerning the Phase I research and a summary of our Phase II objectives.

1.1 Background

Adaptive optics systems have been developed for imaging and communication applications where time-varying external factors (e.g. atmospheric turbulence, temperature) limit the usefulness of fixed-geometry optical systems. By compensating for fluctuations in the effective optical pathlength by adjusting the controlling optics, ground-based astronomical imaging systems can achieve better spatial resolution and laser communications systems can minimize energy transmission requirements by reducing beam divergence.

Figure 1 shows the functional blocks of the typical adaptive optics system. It requires a means for sensing the wavefront's phase across the input aperture, a computational unit for calculating the required phase correction signal from the measured phase data, and a spatial light phase modulator (usually a deformable mirror) for impressing the phase correction on the input wavefront. A portion of the corrected wavefront is sampled and fed back to the system, which adaptively calculates the necessary control signals to the deformable mirror to cancel any phase perturbations in the wavefront. This negative feedback structure provides good stability.

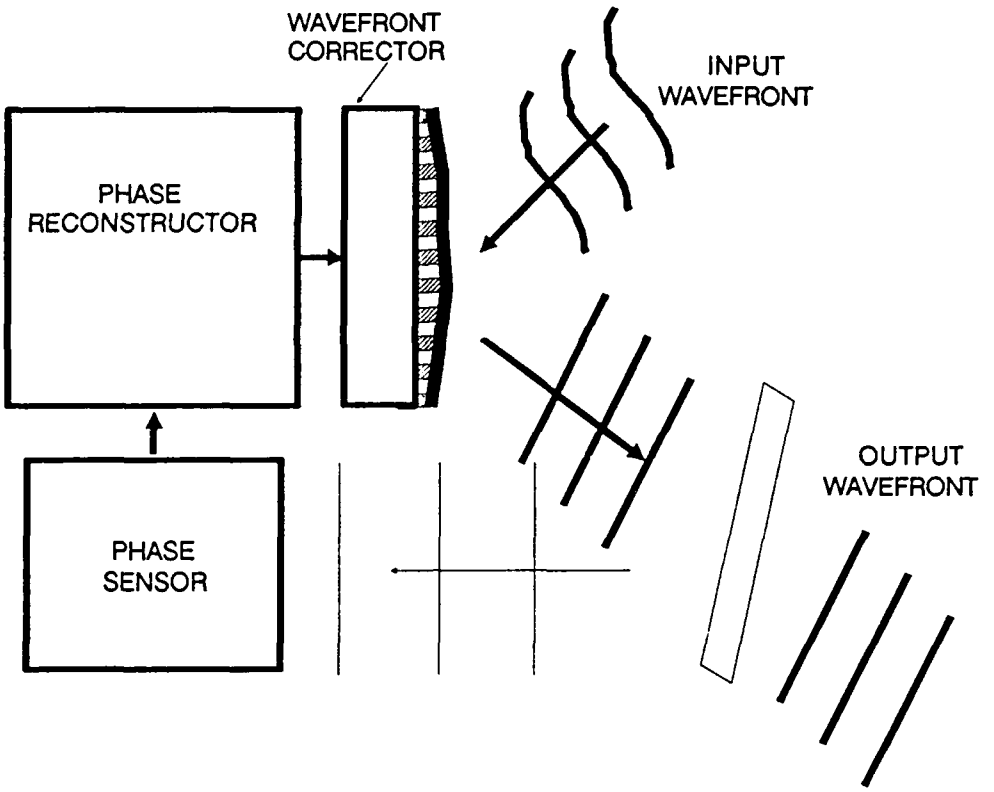


Figure 1: Components of an adaptive optical system for real-time correction of phase disturbances due to atmospheric turbulence.

1.2 Wavefront phase sensing

Two techniques are widely used to measure the phase of the wavefront to be corrected. Both actually sample the slope of the phase rather than measure the phase directly.

Radial AC shearing interferometry

The first method, radial AC shearing interferometry, interferes the incoming wavefront with a displaced version of itself diffracted by a square wave grating [1]. The slope of the wavefront can then be determined from the resulting interferogram. To do this the diffraction grating is rotated, causing the intensity of each point on the interferogram to be modulated at the angular frequency of rotation. The phase angle of each point now encodes the slope of the wavefront at that point. This modulated signal can then be detected and electrically filtered and demodulated to extract the phase angle and thus the wavefront slope at a point.

Early systems used discrete photomultiplier tubes as detectors, each requiring separate power supplies and signal amplifiers. Current detector systems may use CCD arrays mated to image intensifier tubes to provide high photon senstivity while simplifying the complexity of the post-processing electronics. A sensor at Lincoln Laboratories consists of a 64×64 Reticon (photodiode) array capable of running at a 10 kHz frame rate [3].

Hartmann sensors

In a Hartmann sensor the input aperture is divided into an array of subapertures, usually by means of a lenslet array. Provided the subaperture dimension is small compared to the scale of the wavefront disturbances, the wavefront will be approximately planar over the extent of each lenslet, so that the focal plane of the array will consist of an array of focussed spots. The lateral displacement of each spot from the optic axis of the subaperture provides a measure of the slope of that portion of the wavefront. Detection at the focal plane of the Hartmann sensor is generally performed by a high resolution detector array that divides each subaperture into a 2×2 or 4×4 subarray of pixels. Processing is performed on the detected output to find the centroid of each focussed spot, which serves as the position estimate for the slope calculation.

The new architecture that we introduce in section 2 uses the Hartmann sensor concept for wavefront slope sampling.

1.3 Phase reconstruction

Because phase slope rather than phase itself is measured by these sensors, some computation is necessary to reconstruct an estimate of the wavefront phase. Early systems took advantage of the similarity between the phase-from-slope reconstruction problem and the relationship between voltages and currents in a resistive network to create an analog circuit that automatically generated a least squares phase solution as a set of node voltages when injected with currents proportional to the measured slopes. As the power of digital systems has bloomed over the past 15 years, it has become more attractive to place the problem in a linear algebra framework and use a purely digital computation. If the slope measurements are assembled in a vector $\mathbf{g}$ and the phase values to be estimated in a vector ϕ, the relationship between the two can be linearly approximated by the matrix equation

$$\mathbf{g} = \mathbf{A}\,\phi$$

where the matrix $\mathbf{A}$ is determined by the details of the wavefront sensing procedure used. To reverse the procedure, and obtain ϕ from $\mathbf{g}$, we need only take the inverse of $\mathbf{A}$. Unfortunately, this inverse does not usually exist, so some other steps must be taken before a satisfactory solution can be obtained. (The solution is covered in detail in section 3.1). Nevertheless a fixed matrix $\mathbf{B}$ can be calculated that gives the minimum mean square error (MSE) reconstruction via a matrix vector multiplication,

$$\phi = \mathbf{B}\,\mathbf{g} \tag{1}$$

for any slope vector $\mathbf{g}$ [4]. If the wavefront slope is sampled on an $N \times N$ grid, $\mathbf{g}$ and ϕ have $O(N^2)$ elements and $\mathbf{B}$ is an $N^2 \times N^2$ matrix. The number of multiplications needed for the matrix-vector multiplication of 1 is $O(N^4)$, and thus grows quickly as N increases. Standard parallel processing techniques may be employed to attain increased computation rates, but even operation at 1 GigaFLOP (10^9 floating point operations per second) is only fast enough to support real time operation for a 32×32 array. The motivation for this Phase I contract, and the most important innovation of the proposed architecture, is the replacement of the digital electronic matrix computation of ϕ with an optical processing implementation that reconstructs phase quickly and, even more significantly, in an amount of time that is independent of the size of the sampled wavefront array. A variation on this

architecture is also presented wherein the processing time is linearly proportional to array size.

1.4 Wavefront correction devices

Virtually all adaptive optics systems employ some type of segmented or deformable mirror as the wavefront corrector.

Two classes of adaptive optics applications have different requirements: 1) When the mirror is a primary receiving or transmitting mirror (from 0.5m to several meters in diameter) it can provide only limited figure correction, usually less than 1kHz and too slow to correct for atmospheric turbulence; 2) For secondary mirrors within adaptive optics systems, smaller diameters (usually less than 0.25 m) allow faster bandwidths (1 kHz or higher) capable of real-time correction for atmospheric turbulence (1 kHz or higher). For primary mirrors, required phase modulation depths may be as great as several millimeters with an accuracy of $\lambda/20$ and sometimes as high as $\lambda/100$. For secondary mirrors, the required phase modulation depth is less, maybe 2 to 3 wavelengths with similar accuracy requirements, but a faster bandwidth is required.

We are more concerned with the second application - compensation of atmospheric turbulence - because our proposed architecture is directed at relieving the computational burden such systems face as they scale up to large numbers of wavefront sensors and mirror actuators.

Several types of deformable mirrors for atmospheric compensation have been described in the literature [1,2]. Most of the new devices use a thin deformable mirror that is deflected under computer control by an array of discrete actuators pushing or pulling on the back side of the mirror. These actuators are stacks of piezoelectric or electrostrictive materials with attached electrodes. For example, a mirror currently in use at Lincoln Laboratories, designed by Itek Corporation, uses a set of 241 discrete acutators mated to a thin glass facesheet [3]. Phase modulation of $\pm 1.5\mu$ ($\lambda/20$ rms error) is possible with a drive voltage of ± 150 volts.

As a second example of a working system, the European Systems Observatory (ESO) has recently publicized a successful test of an adaptive optics infrared imaging system that uses a deformable mirror [5]. Their system sampled the wavefront on a 5×5 grid and calculated the control voltages that drove the 19 piezoelectric actuators of a deformable mirror.

2 The new system architecture

The previous section provided an overview of conventional approaches to adaptive optics. Laboratory and field experiments have demonstrated the potential of this powerful approach for greatly improving the quality of exoatmospheric ground-based imaging. However, this field is still relatively new and a great deal of progress remains to be made in three areas:

- increasing total aperture size (or equivalently increasing the number of resolvable points on the phase modulating device);
- maintaining tractable computational requirements as the number of actuators and the slope sampling density increases;
- increasing detector quantum efficiency and overall system efficiency.

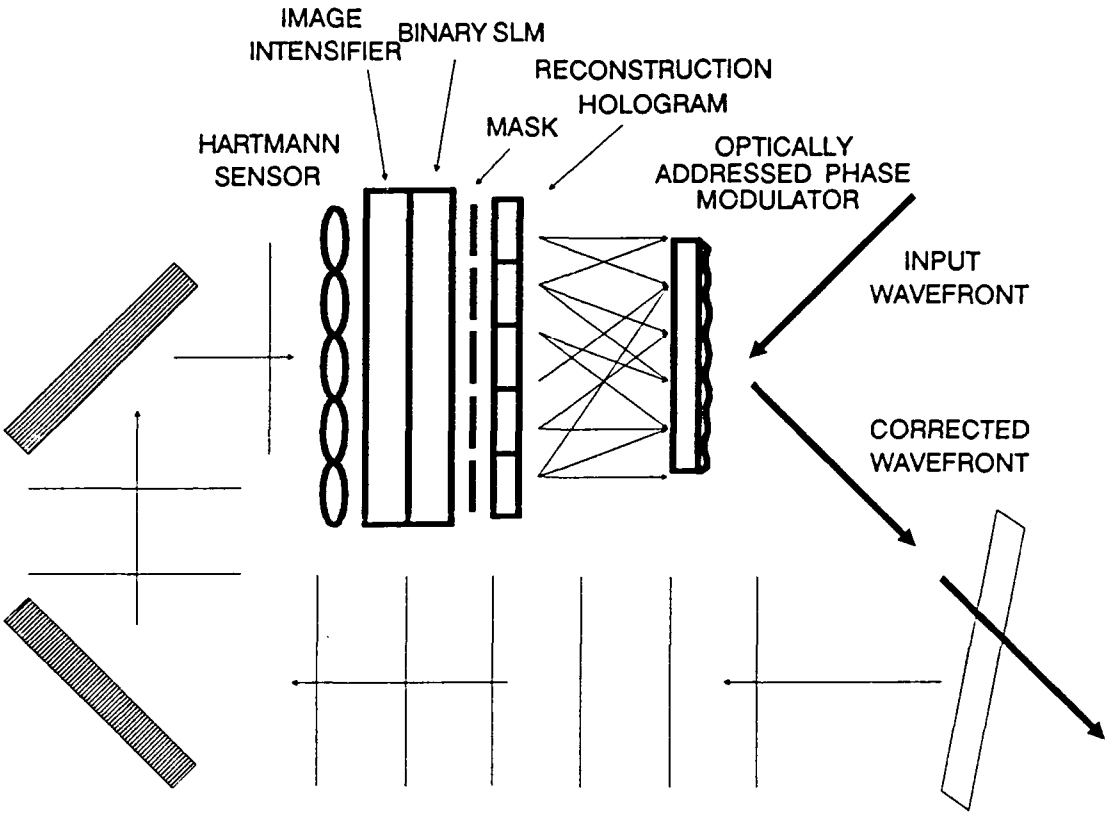


Figure 2: Simplified diagram of the proposed new adaptive optical architecture showing the major components.

Increasing the density of actuators in today's deformable mirrors while maintaining the same modulation depth will be difficult because of space and material limitations. The alternative is to maintain current actuator spacing and increase the diameter of the deformable mirror while adding additional actuator controls. In either case the amount of control electronics will grow linearly with the number of actuators.

At the same time, the computations required to reconstruct the minimum error wavefront phase from the slope measurements increase as the square of the number of controllable phase points.

The system architecture we have developed eliminates the scaling problem for the wavefront reconstruction computation by replacing the digital matrix-vector multiplication with an optical computation performed by a computer-generated hologram in constant time. The other major innovation of this design is a proposal to use a different type of deformable mirror device; one that is optically-addressed and capable of both addition and subtraction input modes.

Our new architecture has the same functional block diagram as was shown in Figure 1, but the implementations of those blocks are quite different than for the conventional approach. Figure 2 is a simplified schematic of the architecture showing the major components. In contrast to the conventional adaptive optics system, the processing peformed by the system of Figure 2 is almost entirely optical - there are no detector arrays and no digital electronics.

The following sections describe the optical processing that occurs at each stage. As an aid to understanding this discussion, an example wavefront is traced through part of the system in Figure 3. The left column describes the progression of processing stages. In the right hand column, the form of the optical signal for the sample wavefront is shown at each stage along the way. The optical signal at the bottom of Figure 3 is used to directly illuminate a hologram. The output of the hologram is a two-dimensional image representing the conjugate of the phase of the sampled wavefront. In this output image, amplitude variations will represent phase variations, as in the output of a phase-contrast or Schlieren optical readout [6]. This image is input to an optically-addressed wavefront corrector, which proceeds to cancel out the existing phase aberrations.

2.1 Wavefront phase sensing

The slope of the wavefront phase is sampled using a conventional Hartmann sensor lenslet array. The focal plane of the Hartmann sensor consists of an $N \times N$ array of focus spots of various displacements from their respective optical axes. The Hartmann sensor is used to encode the local phase slope over each subaperture as a displacement of the focussed spot in the focal plane. A portion of such a sensor is shown in Figure 5. For simplicity, the figure shows the focus spot as being displaced only along the vertical axis. The magnitude of the slope of the wavefront is given by d/f, where d is the displacement of the focussed spot from the optic axis and f is the focal length of the lenslet. Thus the value of d is directly proportional to the slope of the wavefront. The topmost image of Figure 3 represents the image formed at the focal plane of a Hartmann sensor for a hypothetical input wavefront. The x's mark the centers of the subapertures, where the focus spots would occur if the input wavefront were parallel to the lenslet array. The focus spots are displaced in proportion to the local wavefront slope. The intensity variation among the focus spots is due to intensity variation across the input wavefront.

Different sampling geometries using the same lenslet array are possible. Figure 4 shows three that can be incorporated into different variations of the proposed system architecture. The arrangements in Figure 4a,b and c will henceforth be referred to as sampling geometries I, II and III respectively. The dots mark the positions at which phase values will be reconstructed.

For geometry I, the rectangular lenslet array is rotated 45^0. (If the lenslet array is not rotated then the reconstructed phase grid will be.) Alternate rows of lenslets are used to sample the wavefront's slope in the x and y directions. This is indicated in the figure by horizontal and vertical arrows. (Of course each spherical lenslet provides both x and y slope information, but subsequent processing stages only make use of the slope along one axis). Geometry II, shown in 4b, is an extention to geometry I where each lenslet is used to measure slope in both the x and y direction. This geometry is equivalent to the superposition of geometry I with a second copy of itself where the positions of the x and y lenslets has been switched. As a result, the phase can now be reconstructed on a denser grid of points. This geometry also improves the signal to noise ratio of the phase reconstruction.

The third geometry, shown in 4c, is the geometry typically used in conventional adaptive optics systems, where the Hartmann sensor is used with a detector array in the focal plane. In this case the grid is not rotated, and each subaperture samples the x and y slopes coincidentally. Again, the dots represent the grid on which the phase is reconstructed.

Because all three of the sampling configurations of Figure 4 have merit, we will not select

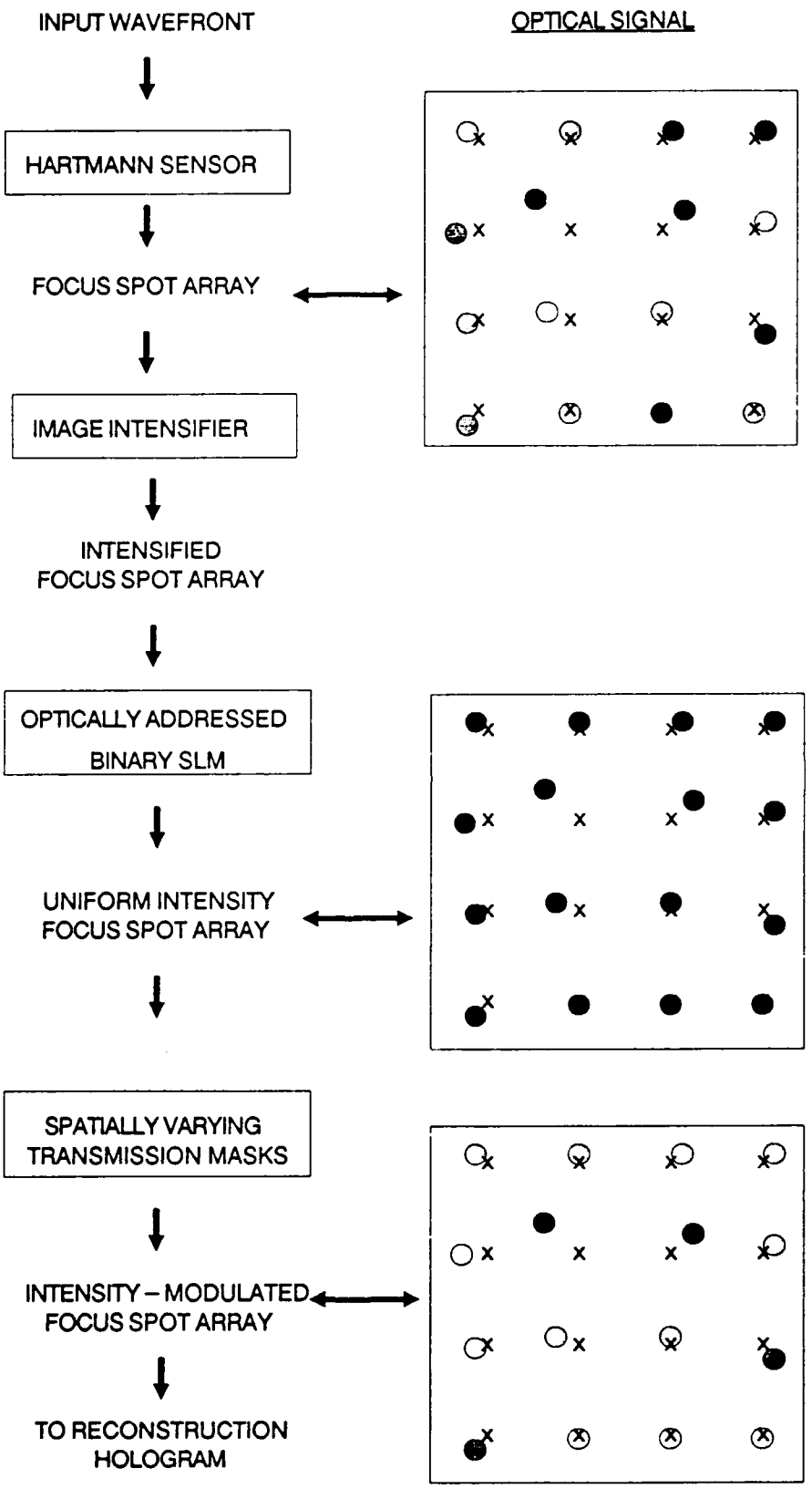
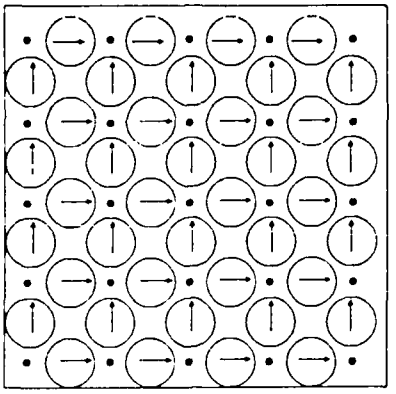
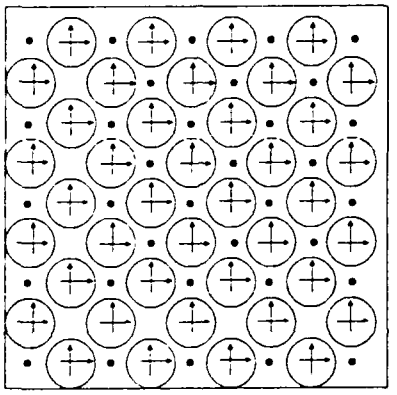


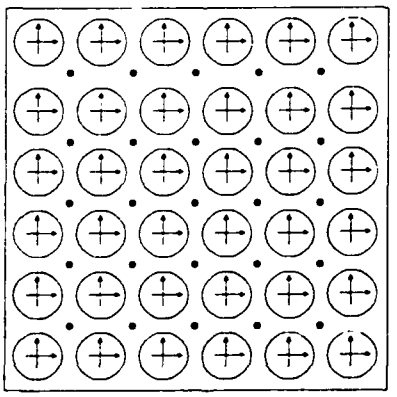
Figure 3: A hypothetical wavefront is traced through successive optical processing stages. The left column, top to bottom, traces the signal through successive optical components, represented by boxes. The right hand column shows the corresponding optical signal after each stage of processing.



(a)



(b)



(c)

Figure 4: Three different sampling geometries utilizing the same 5×5 Hartmann sensor for sampling wavefront slope. Large circles represent the individual lenslets. Arrows specify an x or y slope measurement. Dots represent points at which the phase will be reconstructed.

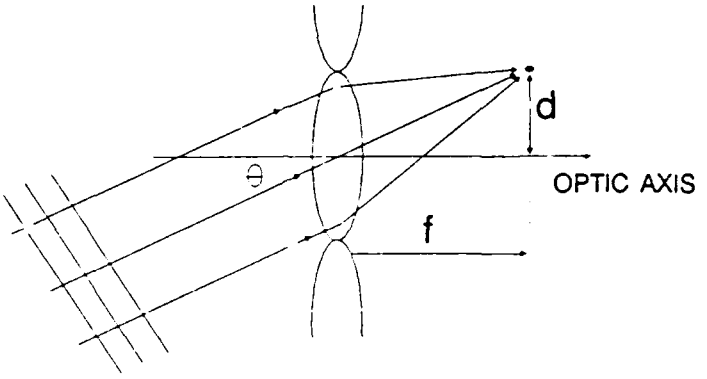


Figure 5: Geometry of the slope sampling function of a single lenslet from a Hartmann sensor array.

one of them and limit our analysis to it alone. Rather, the advantages and disadvantages of each geometry will be made clear when differences exist.

2.2 Light amplification

In astronomical imaging applications, the intensity of the incoming wavefront may be very low. We include at this stage an image intensifier tube in the focal plane of the Hartmann sensor to amplify the focal plane image before proceeding to subsequent optical processing steps. The exact form of the image tube's transfer function is not critical because the intensity of the focal plane image will be binarized by the very next processing stage anyhow.

The output of this stage is an amplified version of the Hartmann sensor focal plane image.

2.3 Intensity binarization/Incoherent to coherent light conversion

A spatial light modulator is used in the next stage to accomplish two tasks at once. First, it serves as an incoherent to coherent image converter. Second, it binarizes the output of the image intensifier, which is the focal plane image from the Hartmann sensor. The former is necessary because the reconstruction hologram at the last stage of processing requires monochromatic, spatially coherent illumination. The output of the image intensifier is unsuitable and must be converted to coherent light. The second task, binarization of the focal plane image, is necessary to remove intensity variations among the focus spots due to intensity variations in the wavefront. Proper correction of phase aberrations in the input wavefront should be done without regard to the amplitude profile of the wavefront.

The requirements on an SLM for this application are: high frame rate (1 kHz or more), high resolution, and binary outputs (no gray scale is needed or desired). Optically addressed ferroelectric liquid crystal modulators are well-suited for this purpose. Typical frame times for ferroelectrics are around 1kHz, with pixel arrays of 256 × 256 and typical contrast ratios of 100:1 or better. We have identified a particular ferroelectric SLM, available from the University of Colorado, that appears promising for use in a Phase II implementation. This SLM uses a layered P-I-N photodiode structure to sense optical input. This design provides much faster operation and better sensitivity than standard photoconductor layers [7]. The

(a) (b)

Figure 6: The transmission masks for modulating the intensity of a focus spot based on its displacement from the optic axis along the (a) x axis or (b) y axis.

SLM is operated in a framed mode in order to supply the SLM with an AC bias voltage. The optical input is time integrated, and periodically reset to zero. The optical readout therefore has a sawtooth character, with its amplitude proportional to the energy of the addressing light. This output waveform is compatible with subsequent stages.

The output of this stage is therefore (ideally) a coherent image of an array of spots of uniform intensity but various displacements from their respective optical axes (Figure 3, second image). No differentiation between x and y slope measurements has yet been made. Section 3 discusses the overall impact due to non-ideal performance of this stage.

2.4 Displacement-to-intensity encoding

This stage is the last before the processed focal plane image is used to illuminate the reconstruction hologram. In the previous stage we removed from the optical signal any amplitude variations due to the original wavefront. In this stage we modulate the intensity of each focus spot according to its displacement. This is the first stage where the focus spots are processed differently depending on whether x, y, or both x and y slope information is to be extracted.

The design of the reconstruction hologram, which is described in detail in a later section, requires that each focus spot have an intensity proportional to the local wavefront slope. The output signal of the previous processing stage has encoded the slope of each focus spot as a displacement from the optical axis. To translate this displacement variation among focus spots into intensity variation, each focus spot is passed through a tiny transmission mask with a spatially varying transmission characteristic. The transmission masks used to translate the x-direction slope and y-direction slope to an intensity modulation are shown in Figures 6a and 6b respectively. The x and y masks are identical but for a 90^0 rotation. They have approximately the same dimensions as a subaperture. Each has a centerline of minimum (ideally zero) transmission, and a transmittance that increases linearly with distance from the centerline. Upon passing through this mask, the intensity of a focus spot will be attenuated so that its intensity upon exit is proportional to the magnitude of its displacement d from the optic axis, which is in turn proportional to the local wavefront slope

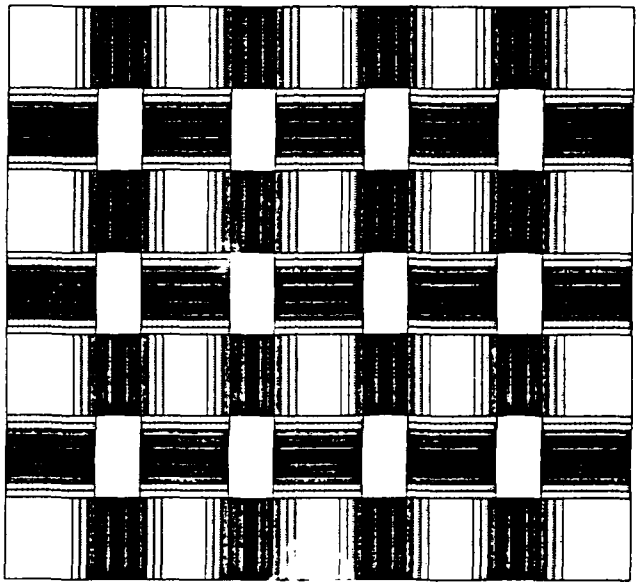


Figure 7: Transmission mask for sampling geometry I, consisting of alternating x and y oriented subaperture masks.

d/f. Thus this mask achieves the necessary intensity encoding. Note that the position of the focus spots is not changed by this step.

Using one of these transmission masks, the intensity of a focus spot may be modulated according to its x displacement or y displacement, but not both. The transmission masks needed for each subaperture are combined into a single large, tiled transmission mask. The mask for sampling geometry I, for example, is shown in Figure 7. Sampling geometries II and III require coincidental measurement of x and y slopes. To do this the input optical signal (i.e. the array of displaced dots) must be replicated and processed in parallel channels (see Figure 8). This does not represent a serious increase in hardware cost or complexity because the only components replicated are the reconstruction hologram and the transmission mask. No additional SLMs or electronics are required, though to maintain the irradiance in each channel at the same level of the single channel system, the power of the laser source has to be doubled. If, on the other hand the beamsplitting of the optical signal into two channels is introduced earlier in the system, say even before the Hartmann sensor, replication of the image intensifier tube and the binary SLM would be necessary.

The division of processing between the two channels can be done in many different ways, all of which work provided the reconstruction hologram is designed accordingly. If the slope measurements needed for geometry II are viewed as the superposition of two complementary versions of geometry I, the transmission masks for the two channels could be as shown in Figure 9a.

As a second possibility, all of the x slope processing could be performed in channel 1 and all of the y processing in channel 2 (Figure 9b). Similar possiblities exist for implementing

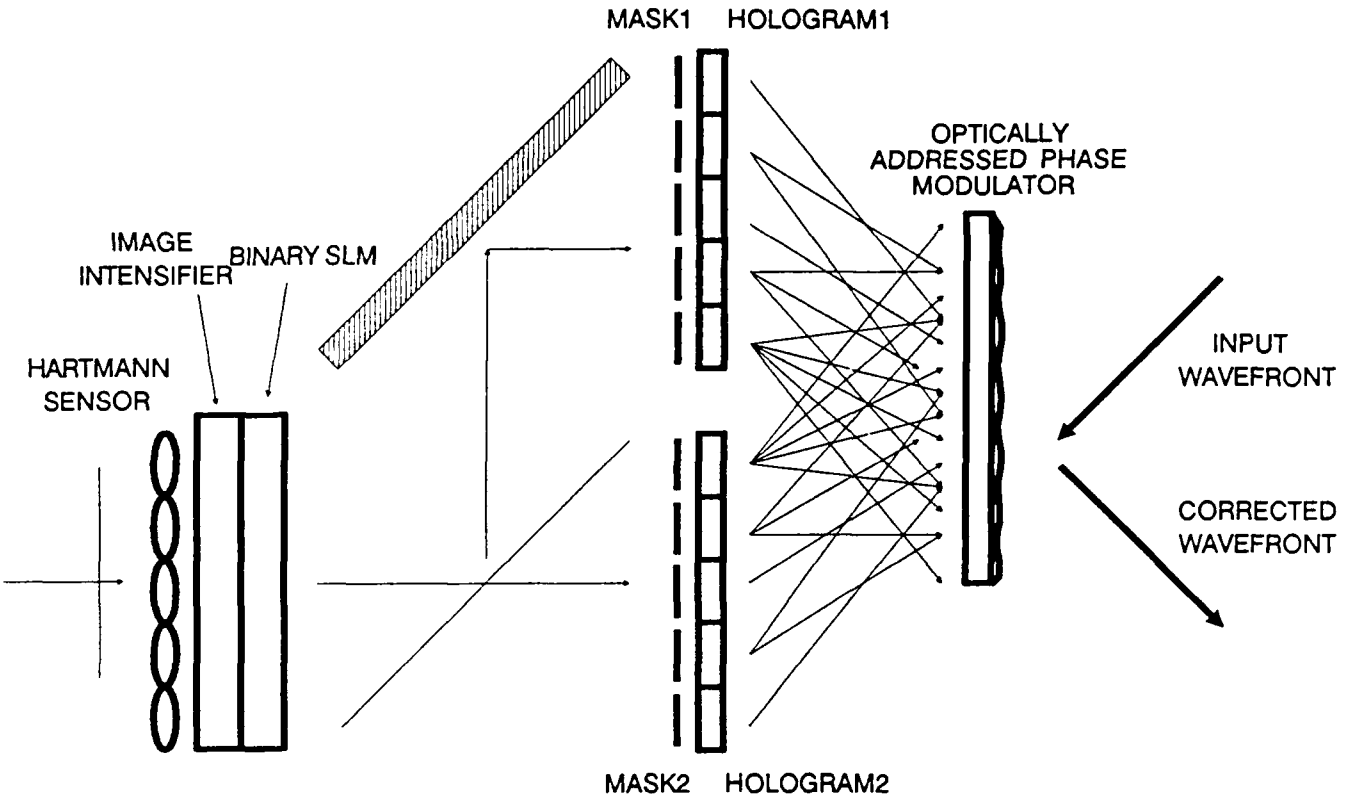


Figure 8: Two channel system needed to implement sampling geometries II and III.

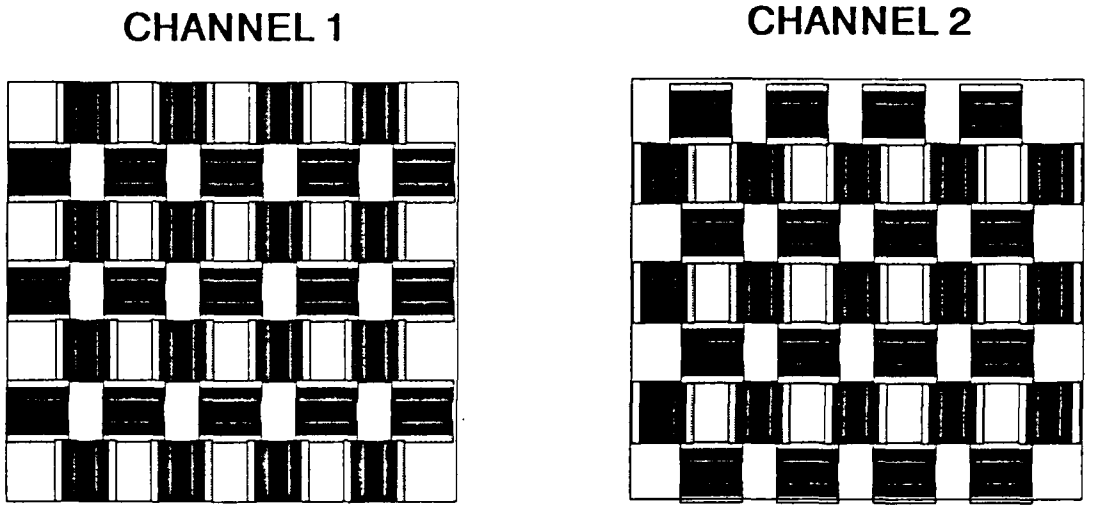
Geometry III using parallel channels.

The example of Figure 3 evidently uses a transmission mask that consists entirely of y axis encoding masks, such as the channel 2 mask in Figure 9b.

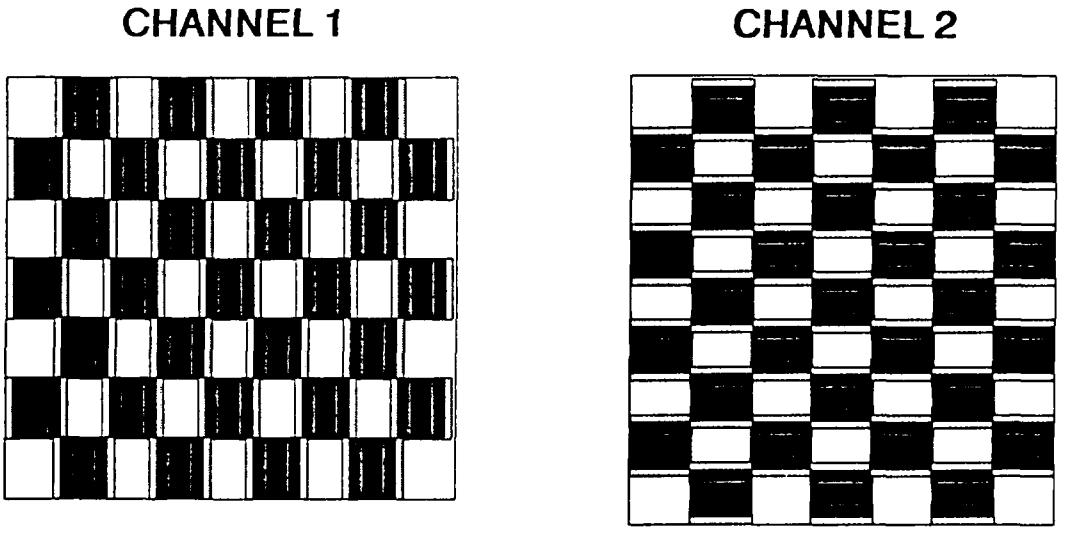
2.5 Illumination of the reconstruction hologram

The function of the reconstruction hologram is to produce a two-dimensional optical image signal whose spatial intensity variations represent the phase variations of the wavefront correction signal. The correction signal is just the conjugate of the wavefront to be corrected. If the wavefront is originally sampled on an $N \times N$ grid, the output of the hologram will essentially be of the same resolution, though the image may be continuous rather than pixelized because the hologram can provide some degree of interpolation between reconstructed phase points. Because it reconstructs the phase of the wavefront from encoded slope data, this hologram is referred to as the reconstruction hologram. The input to the reconstruction hologram (or holograms in the case of a two-channel architecture) is the output of the transmission masks of the previous processing stage. Therefore it is an array of focus spots, each having an intensity and a displacement that are proportional to the local wavefront slope.

The reconstruction hologram, like the Hartmann sensor and the transmission mask, is actually an array of smaller elements, in this case subholograms, each of which processes a single focus spot independently of the others. The desired phase reconstruction image is formed by incoherently superimposing the diffracted outputs of all of the subholograms. The hologram effectively "integrates" the slope measurements to obtain phase. The synthesis and analysis of the reconstruction hologram is covered in detail in section 3.1. For now



(a)



(b)

Figure 9: Two possible sets of transmission masks for channels 1 and 2 (sampling geometry II).

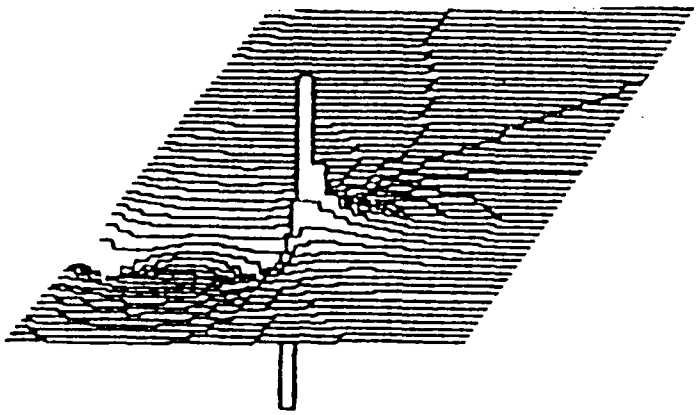


Figure 10: Form of the impulse response for a typical subhologram of the reconstruction hologram.

we simply state that the behavior of each subhologram can be summarized by its impulse function. Each subhologram has a different impulse function, but the form of this function is similar for all of the subholograms. A typical impulse function is shown in Figure 10. It it dominated by a pair of peaks of opposite sign and is symmetric about zero. As we define it, this impulse response represents the output of a particular subhologram when illuminated by a collimated beam of light. It is defined on an $N \times N$ output grid corresponding to the points at which the phase will be reconstructed. If the subhologram is exposed (or generated by computer) using a highly diffusive object beam, it will have the property that its output will be essentially invariant with respect to the position or diameter of the collimated beam used to illuminate it. This is due to the distributed nature of information within such a hologram [8]. Nor does the intensity of the input beam affect the *form* of the impulse response, but, because a hologram is essentially a linear device, the intensity of the subhologram's output will scale in proportion to the intensity of the input. This invariance to input beam position or shape along with its linearity with respect to intensity allows each subhologram to process the intensity information present in a focus spot while ignoring the displacement of the spots. Although it will be seen that the displacement of a spot is used to determine the polarity of the slope measurement, it is not used by the hologram to provide slope magnitude information. The only factor affecting magnitude is the intensity of the focus spot.

The impulse response functions are carefully chosen so that the $N \times N$ array of phase values can be reconstructed by adding the impulse response of each subhologram weighted by the intensity of the focus spot illuminating that subhologram (see section 3.1). The hologram is designed so that the impulse response of each of the subholograms is directed and properly aligned onto a common $N \times N$ output grid in space. Each subhologram's output will indeed be proportional in intensity to the focus spot that illuminates it, so that the sum of all of the subholograms' outputs reconstructs the proper phase within the $N \times N$ output grid.

The complexity of the reconstruction hologram suggests that it should be computer generated. If that is the case, it would be easy and advantageous to eliminate the variable transmission masks and incorporate their function into the design of the hologram.

sign of $h_i(x,y)$	sign of slope g_i	contribution to $\Phi^+(x_0,y_0)$	contribution to $\Phi^-(x_0,y_0)$
+	+	$h_i(x_0,y_0)\cdot g_i$	0
+	−	0	$h_i(x_0,y_0)\cdot -g_i$
−	+	0	$-h_i(x_0,y_0)\cdot g_i$
−	−	$-h_i(x_0,y_0)\cdot -g_i$	0

Table 1: Contributions of subhologram (i,j) to cell (x_0,y_0) in the outputs Φ^+ and Φ^- depends on the sign of the corresponding impulse response $h_i(x_0,y_0)$ evaluated at (x_0,y_0) as well as the sign of the measured slope value g_i.

Negative number representation

We have nearly completed a conceptual description of an adaptive optical processing architecture that can sense wavefront slope and reconstruct an intensity-coded phase signal suitable for wavefront correction, without any detector arrays or digital processors.

The last hurdle is the representation of negative slope values and bipolar impulse response functions. It turns out that all of the impulse response functions, like the one in Figure 10, contain both negative and positive values. In addition, slope values as measured by the Hartmann sensor are bipolar. Because we do not wish to introduce the problems associated with coherent optical subtraction schemes, we have developed an alternate solution for implementing bipolar arithmetic using strictly unipolar signals. It involves modifying the architecture described thus far in two respects.

First, the reconstruction hologram must be redesigned to produce two output images rather than just one: one containing positive contributions and the other consisting of negative contributions to the reconstructed phase. We call these two components Φ^+ and Φ^-, respectively, and their difference gives the desired phase,

$$\Phi(x,y) = \Phi^+(x,y) - \Phi^-(x,y), \qquad 1 <= x,y <= N \tag{2}$$

where the coordinates (x,y) specify a particular cell in the output reconstruction. This alteration involves a simple redesign of the reconstruction hologram.

Each subhologram has a different impulse response function. We denote the impulse response function for subhologram i by $h_i(x,y)$. (Recall that, for some sampling geometries, the subholograms are not arranged in a square grid; hence we use a one-dimensional subscript to avoid confusion. The output phase reconstruction always forms a square array so we use x-y coordinate pairs to refer to specific values). Each $h_i(x,y)$ is a function defined over the extent of the output grid, i.e. for $1 <= x,y, <= N$. We also denote the signed slope value represented by the focus spot in cell i as g_i. Table 1 enumerates all possible sign combinations for an impulse response function $h_i(x,y)$ evaluated at a particular output point (x_0,y_0) and a slope value g_i measured at that subaperture i, and shows what the proper contribution to the Φ^+ and Φ^- outputs should be.

In order that focus spots representing positive and negative slopes may be treated differently, the second modification that needs to be made is the division of each subhologram into two equal size half-holograms. Holograms are divided into left and right or top and bottom halves for x and y slope processing, respectively. Having designating positive x and y directions for slope, the hologram in one half processes its focal spot according to the rules for a positive slope (table 1, rows 1 and 3), while the other half uses the rules for a negative slope value (table 1, rows 2 and 4).

Given the impulse responses $h_i(x,y)$ for all i, the correct output behavior of both halves of every subhologram can be completely determined using the rules of table 1. Section 3.1 shows how to determine the optimal $h_i(x,y)$ for a given sampling geometry.

With these modifications the architecture correctly processes both positive and negative slope values. However, it is now necessary to perform the image subtraction of 2 to obtain the final phase reconstruction. We have analyzed two methods for performing this subtraction: one based on detection and electronic subtraction, and the other based on the use of spatial light modulators that support subtraction of optical inputs. These alternatives are described below.

2.6 Driving the wavefront corrector with the output of the reconstruction hologram

Subtraction mode optically-addressed SLMs

The most attractive approach to combining the positive and negative signal outputs does not use detector arrays to convert the optical signals to electrical ones. Instead it alternately passes one or the other of the optical signals directly to the input of one of a special class of optically addressed SLMs that supports sequential subtraction on its inputs. In this way the input to the SLM is synthesized by a two-step process: first, the SLM is placed in the addition mode, and the positive polarity optical signal is input to the SLM; then, the SLM is switched to subtraction mode, and the negative polarity optical signal is input to the SLM. The net effect is that the SLM will phase modulate the wavefront according to the difference of the two signals, which is the desired wavefront correction. Assuming that such an SLM exists, it could be used in a system like the one in Figure 11. A pair of optical shutters synchronized to the SLM's add/subtract control signal are required to alternately pass the positive and negative component of the output signal. Fast, small aperture electrooptic shutters can be used for this purpose. During one frame time of the binary SLM, first one and then the other shutter would be opened, sequentially passing the signals of opposite polarities to the SLM. Thus the control signals to the shutters would be two square waves oscillating at the system frame rate but 180^0 out of phase with each other. In Figure 11, one shutter is on (solid arrows) while one if off (dotted arrows). Plane mirrors are used to reflect each image onto the common optical input of the SLM.

Optical spatial light modulators capable of sequential subtraction on their inputs are those that convert the optical input signal into an electron charge image using a photocathode. The microchannel spatial light modulator (MSLM) [9] is the best known example of such an SLM. The MSLM is inappropriate for our adaptive optics application because of its slow frame rate and limited phase dynamic range. It is intended primarily for amplitude modulation.

Optron Systems has developed a research prototype of an optically-addressed, pixelized membrane light modulator (Figure 12 as well as several electrically-addressed SLM designs incorporating electron guns, all of which support sequential subtraction on their inputs.

Of course, for electrically addressed systems subtraction is probably best done by purely electronically means prior to addressing the SLM. The problem with electron beam-addressed SLMs is that they have slow frame rates - slower than standard 60Hz video rates. Optically-addressed membrane modulators such as the one being developed at Optron Systems and shown in Figure 12 appear to be the only class of SLMs capable of approaching operation

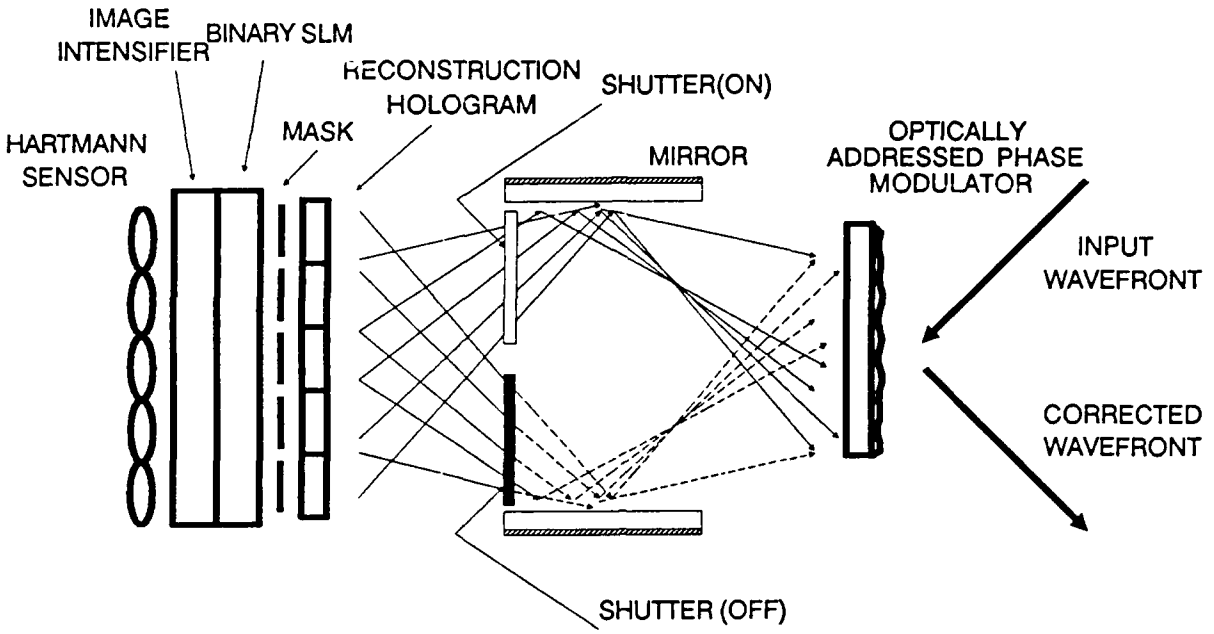


Figure 11: Schematic of an all-optical system that uses a pair of shutters synchronized with the mode control line of an optically-addressed SLM. The shutters alternately pass the positive and negative contributions to the wavefront correction signal while the SLM synchronously toggles between addition and subtraction modes. The feedback path from the corrected wavefront to the wavefront sensor is not shown.

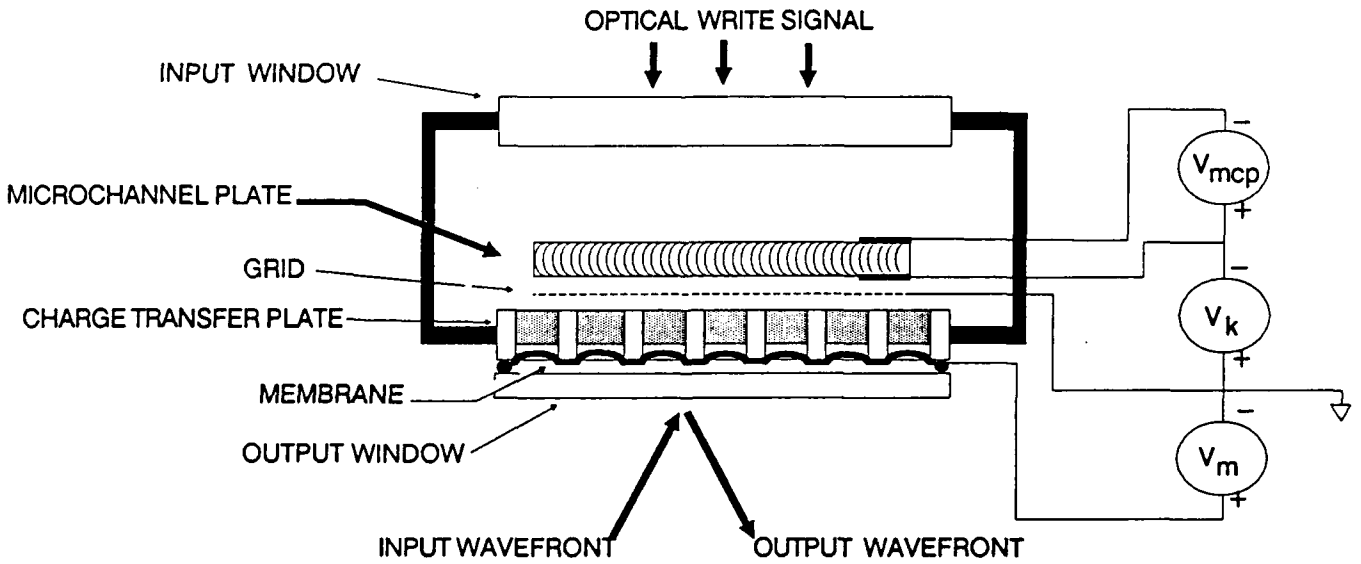


Figure 12: Schematic of Optron Systems' optically addressed membrane light modulator. This SLM can be used as the wavefront correction element in the proposed adaptive optics architecture.

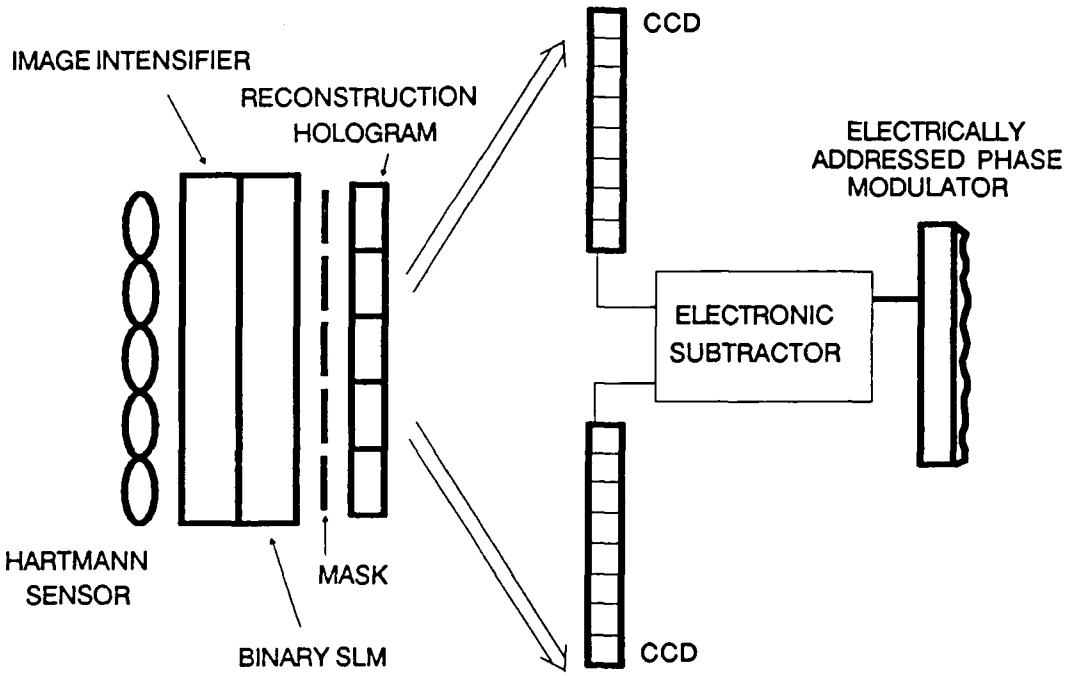


Figure 13: Use of detector arrays and high speed electronics to perform image subtraction and drive an electrically-addressed wavefront correction SLM. The feedback path from the corrected wavefront to the wavefront sensor is not shown.

at the desired 1kHz frame rate. Current research at Optron on the optically addressed membrane light modulator is centered on improving device speed and lifetime.

Electronic subtraction

If a suitable optically addressed wavefront corrector is not available, an alternate configuration uses two area detector arrays to capture the positive and negative polarity output signals, and perform the subtraction electronically (Figure 13).

Since a digital subtraction can be performed in less than 10ns, a array containing 10^5 pixels could be processed at a 1kHz frame rate. This corresponds to a pixel data processing rate of 100Mhz. The limiting factor in such a system will be the rate at which data can be read out from the detector and input to the wavefront corrector SLM, which in this case must be an electrically addressible SLM.

The ideal detector for this application has a fast frame rate, high pixel count and good dynamic range. Aspects of the detector that less critical are: pixel size (though a larger pixel size is more convenient), detector fill factor (since the reconstruction is already pixelized at a low resolution), and light sensitivity and efficiency (since sufficient illumination can be provided by a laser source).

Three detectors occupying different points on the pixel count vs. frame rate tradeoff curve are shown in Table 2.

Only the Reticon arrays can support frame rates of 1kHz or faster. Unfortunately they

Detector type	Pixel count	pixel dimension	frame rate (max)	dynamic range
Thomson TH7852 CCD	288×208	$30\mu \times 19\mu$	125 Hz	3000:1
EG&G Reticon RA6464N	64×64	$\sim 95\mu \times 95\mu$	24.5 kHz	100:1
EG&G Reticon RA2568N	256×256	$28\mu \times 28\mu$	500 Hz	300:1

Table 2: Specifications for candidate detector arrays.

have fewer pixels and less dynamic range than CCD detectors.

A conventional deformable mirror would be used as the wavefront corrector in this type of system, since it provides the necessary speed and accuracy. The resolution and size of the correcting aperture would thus be limited to the relatively small values of current deformable mirrors (see section 1.4).

The frame rate of this type of system will decrease as the pixel count of the detector arrays increases, unless increased parallelism is used to readout the arrays. The second architectural solution has the advantage that its frame rate is completely independent of the size and resolution of the reconstructed phase image.

3 System Analysis

3.1 Least squares wavefront construction

If slope measurements contain noise, the task of finding the best phase correction signal becomes an estimation problem. Since the phase correction signal is simply the conjugate (negation) of the phase of the input wavefront, an equivalent problem is the reconstruction of the phase itself from slope measurements of a wavefront. The problem of reconstructing an estimate of the phase of the wavefront from an array of sampled slope values has been solved in the least squares error sense [4]. The most general solution approach models both the sampled slope values and the reconstructed phase values as vectors, and represents the slope sampling procedure by a transfer matrix. For an $N \times N$ array the slope and phase vectors both have dimensions $N^2 \times 1$ and are denoted $\mathbf{g}$ and $\boldsymbol{\phi}$, respectively. The 1-D phase vector $\boldsymbol{\phi}$ contains the elements of the phase output function Φ using row major ordering. The slope measurement error appearing in the elements of $\mathbf{g}$ is is assumed to be additive Gaussian noise and is also assumed to be uncorrelated between slope measurements. The process of sampling the local phase slope of the wavefront is modeled by taking a discrete approximation of the derivative over $\boldsymbol{\phi}$. If the derivative approximation is linear, this process can be expressed as a matrix operation:

$$\mathbf{g} = \mathbf{A}\,\boldsymbol{\phi}$$

where $\mathbf{A}$ is a constant matrix determined by the sampling geometry. As an example, if $N = 3$ and the sampling geometry of Figure 4 is used, there are $N^2 = 9$ phase points and $2N(N-1)$ sampled slope values (Figure 14). Note that $\mathbf{g}$ consists of the x slope values followed by the y slope values according to the numbering of Figure 14. If the x and y slope values are to be approximated by the differences between adjacent horizontal and vertical

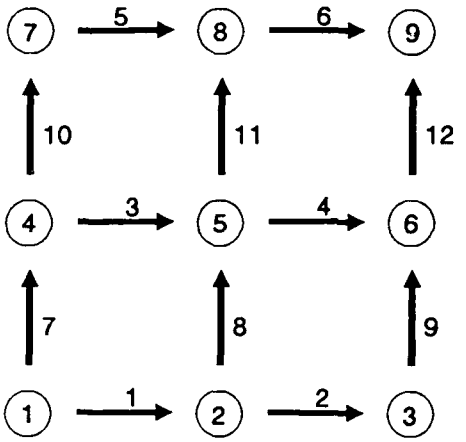


Figure 14: Element numbering for wavefront slope sampling with $N = 3$.

phase values, respectively, the transfer matrix $\mathbf{A}$ is defined as:

$$\mathbf{A} = \begin{bmatrix} -1 & 1 & & & & & & & \\ & -1 & 1 & & & & & & \\ & & & -1 & 1 & & & & \\ & & & & -1 & 1 & & & \\ & & & & & & -1 & 1 & \\ & & & & & & & -1 & 1 \\ -1 & & & 1 & & & & & \\ & -1 & & & 1 & & & & \\ & & -1 & & & 1 & & & \\ & & & -1 & & & 1 & & \\ & & & & -1 & & & 1 & \\ & & & & & -1 & & & 1 \end{bmatrix} \tag{3}$$

We can now express the reconstruction of ϕ from $\mathbf{g}$ using the inverse matrix operation

$$\phi = \mathbf{A}^{-1}\,\mathbf{g} \tag{4}$$

The system of equations represented by (4) is underconstrained by one equation. As a result there is not a single unique solution ($\mathbf{A}$ is singular), but an infinity of solutions differing only by an additive constant. This corresponds to the reality that the phase is determined not absolutely, but relative to some arbitrary origin.

The minimum norm solution $\mathbf{B}$ satisfying

$$\phi = \mathbf{B}\,\mathbf{g} \tag{5}$$

is [4] the Moore-Penrose pseudoinverse $\mathbf{A}^{+}$ defined by

$$\mathbf{B} = \mathbf{A}^{+} = \mathbf{A}^{T}(\mathbf{A}\mathbf{A}^{T})^{-1}\mathbf{g} \tag{6}$$

Thus

$$\phi = \mathbf{A}^{+}\,\mathbf{g} \tag{7}$$

describes the minimum norm least squares error reconstruction, and

$$-\boldsymbol{\phi} = -\mathbf{B}\,\mathbf{g} = -\mathbf{A}^{+}\,\mathbf{g} \tag{8}$$

describes the least squares phase correction image. If we enumerate some of the elements of the components of equation 8, we have

$$-\begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \vdots \\ \phi_K \end{bmatrix} = -\begin{bmatrix} b_{11} & b_{12} & b_{13} & \cdots & b_{1M} \\ b_{21} & b_{22} & b_{23} & \cdots & b_{2M} \\ b_{31} & b_{32} & b_{33} & \cdots & b_{3M} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{K1} & b_{K2} & b_{K3} & \cdots & b_{KM} \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ \vdots \\ g_M \end{bmatrix} \tag{9}$$

This makes it clear that each value of $\boldsymbol{\phi}$ is a linear combination of the slope measurements:

$$-\phi_i = -\sum_{j=1}^{M} b_{ij} g_j \tag{10}$$

where $M = 2N(N-1)$. The function of the reconstruction hologram is to simultaneously perform all N^2 of the weighted sums like (10) and to assemble them in a two-dimensional output array that corresponds to the original wavefront. Viewed as such the hologram is a linear algebra processor, where the weighting coefficients b_{ij} could take on any set of values to specify any desired transfer matrix. Our phase reconstruction problem is merely a special case where the transfer matrix $\mathbf{B}$ is defined by equation 6.

From (9) it can be seen that the N^2 elements of the ith column of $\mathbf{B}$ define the impulse response $h_i(x,y)$ for the ith subaperture. We need only negate the column vector and rearrange the elements into an $N \times N$ matrix to have the form of $h_i(x,y)$. Having done this, we can use the rules given in section 2.5 to determine the exact design of each subhologram. Thus we have shown the process by which the design of the reconstruction hologram is completely determined based on the sampling matrix $\mathbf{A}$.

3.2 Error sources and performance limitations

In this section we identify noise sources that will determine the quality of the phase reconstruction, and factors that limit the frame rate and light sensitivity of the system.

The negative feedback structure of the architecture greatly improves overall performance and relaxes accuracy requirements on the various electrooptic components.

3.2.1 Hartmann sensors

Proper operation of the Hartmann sensing stage assumes that the wavefront being sampled is to a good approximation planar over each subaperture. If this is not the case, the image in the focal plane will not be a well-focussed spot. Reconstruction of phase from slope values by the reconstruction hologram relies on the assumption that all focus spots are of the same size. Because the output of a subhologram is proportional to the *total* intensity of the illuminating focus spot, variations in the size of the focal spot will be confused with variations in the average intensity of the focus spot as determined by the transmission masks. We see this as likely to be one of the largest sources of error in the system.

The maximum wavefront slope measurable by a Hartmann sensor consisting of lenslets of radius r and focal length f is r/f (see Figure 5), which is proportional to the numerical aperture of the lenslets.

3.2.2 Image intensifier

The spatial resolution of the image intensifier (typically 40 lp/mm) should not be a limitation. In addition, because the optical signal is thresholded and binarized in the stage following the intensifier, the exact form of the intensity transfer function is not critical; there is no linearity requirement.

3.2.3 Optically-addressed binary SLM

The frame rate of the binary SLM presents a limit on the overall frame rate of the system. For the photodiode-addressed ferroelectric SLM, frame rates over 1 kHz have been reported [7] with spatial resolutions of 40 lp/mm.

3.2.4 Transmission masks / reconstruction hologram

The simplest design incorporates the transmission mask function into the reconstruction hologram. This eliminates alignment problems and increases light efficiency. The reconstruction hologram will have a limited energy efficiency (perhaps 20%) and some signal-to-noise ratio that will probably determine the overall system signal-to-noise ratio.

3.2.5 Wavefront correction

Two methods of processing the reconstructed phase have been analyzed: one using an optically-addressed SLM operating in addition and subtraction modes, and a second using detector arrays and an electrically-addressed wavefront corrector. For the former, the finite integration time required at the optical input of the SLM will present a limit on the system frame rate. For the latter, fast photodiode detector arrays are available that can frame more than fast enough (20 kHz). Thus the limiting factor will again be the speed of the electrically-addressed SLM. For these devices, resolution may be traded off to achieve higher frame rates.

Once again it is important to point out that the frame rate of the architecture using an optically-addressed SLM is independent of the resolution of the wavefront sensor and wavefront corrector, whereas the frame rate of the system using detector arrays and an electrically-addressed SLM is inversely proportional to the resolution of the wavefront corrector.

4 Conclusions

A novel adaptive optics architecture has been developed that uses optical signal processing for wavefront sensing and phase reconstruction. The main benefit of the new architecture is its scalability - system frame rate is independent of aperture size and resolution. Sources of error and limits on system performance have been identified. The critical parameters of system components have been identified and candidate commercial and research devices

have been identified in most cases. The design of the optimal reconstruction hologram was described in full detail.

A Phase II proposal is being submitted to pursue prototyping and demonstration of the novel adaptive optics architecture described herein.

References

[1] Hardy, J. "Active Optics: A New Technology for the Control of Light," *IEEE Proceedings* **66**, No. 6, 651 (1978).

[2] Everson, J., "New Developments in Deformable Mirror Surface Devices," *Adaptive Optical Components* , SPIE **141**, (1978).

[3] Parenti, R., "Recent Advances in Adaptive Optics Methods and Technology," *Laser Wavefront Control* , SPIE **1000**, 101 (1988).

[4] Herrmann, J. "Least-squares wavefront errors of minimum norm," *J. Opt. Soc. Am.* , **70**, No. 1, 28 (1980).

[5] "Successful tests of adaptive optics," *OE Reports (SPIE)* , March 1990.

[6] Hecht, E. and Zajac, A., *Optics* , Addison-Wesley, Reading, MA, 1974.

[7] Moddel, G., et.al., "High-speed binary optically addressed spatial light modulator," *Appl. Phys. Lett.* **55**, No. 6, 537 (1989).

[8] Hariharan P., *Optical Holography* , Cambridge University Press, Cambridge, England, 1984.

[9] Warde, C. and Thackara, J., "Operating modes of the microchannel spatial light modulator," *Optical Engineering* **22**, No. 6, 695 (1983).