Adaptive Wavefront-Correction Algorithms Using Focal-Plane Array Intensity Measurements

R. Mukai,[1] V. A. Vilnrotter,[1] and C.-W. Lau[1]

Earth's atmospheric turbulence induces aberrations in the wavefront of the optical signal from a distant spacecraft. These aberrations cause diffusion and spreading of the signal energy in the receiver's focal plane, potentially resulting in an increase in bit-error rate (BER) by several orders of magnitude when significant background radiation is present, such as during daytime operation. The adaptive optics (AO) algorithms in this article use the focal-plane intensity distribution to estimate aberrations induced by the atmosphere and to correct these aberrations. Successful wavefront correction concentrates the signal energy into a much smaller area in the focal plane, greatly reducing collected background radiation and improving BER performance.

I. Introduction

Existing adaptive optics (AO) systems, which are designed primarily for nighttime astronomical imaging, rely on three critical components necessary for successful imaging [1,4]: (1) a laser guide star, which is an artificial point source of light resembling an 8th magnitude star generated in the upper atmosphere by ionizing sodium atoms with a powerful ground based laser; (2) a wavefront sensor, which measures aberrations in the optical wavefront from the laser guide star; and (3) a tip/tilt mirror for fine-pointing correction. Although most wavefront aberrations are measured using light from the laser guide star, fine-pointing (tip–tilt) correction in present AO systems must be performed by diverting a portion of the signal energy to the tip/tilt correction system since the laser guide star cannot be exactly co-located with the desired source in the sky [1,4]. By measuring wavefront aberrations, existing AO systems determine phase corrections to be applied to a deformable mirror placed in the optical path, compensating for some of the turbulence-induced aberrations. This results in a tightly focused signal spot in the detector plane and enhanced bit-error rate (BER) performance via a corresponding reduction in the collected background energy. Such AO systems have been successfully applied to the problem of astronomical imaging for many years, operating at night when sky background interference is minimal. Deep-space optical communications cannot rely only on nighttime operation, since data transmission usually depends on critical mission considerations, which can occur during day or night. Therefore, the advantages of the focal-plane-array (FPA)-based AO proposed here over existing AO systems are as follows:

---

[1] Communications Architectures and Research and Section.

The research described in this publication was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

(1) Elimination of the cost and complexity of equipment for generating a laser guide star and of the separate wavefront sensor assembly.

(2) No diversion of signal energy away from the detector plane for tip/tilt correction, as required in the existing AO approach. Since a focal-plane array detector is employed, the computation of tip-tilt is accomplished automatically as part of wavefront aberration estimation.

(3) Daytime operation is not a serious impediment to signal-based wavefront correction, since the optical communication system will be designed to meet communications requirements during the day if necessary. Since data communication takes place at a much higher rate than the coherence time of the channel, it is virtually assured that the wavefront correction subsystem has adequate integration time and hence adequate signal energy for estimating wavefront errors during the day. The difficulties of using a guide star during the day are thus avoided.

(4) The proposed system can be upgraded easily. If better stochastic optimization algorithms are discovered in the future, replacing them with newer, more efficient algorithms is a straightforward software or firmware upgrade that avoids the need for expensive optical hardware changes.

In this article, we start by presenting in Section II a model of the system and briefly discussing previous work carried out by Carhart et al. [2] and by the authors [3] in the area of field correction via focal-plane measurement. In Section III, we describe a simulation model of the atmosphere that also was used in previous work [3] and provide details on an algorithm developed at JPL. Conclusions and future directions are given in Section IV.

Before proceeding, we note that this article focuses on the phase *acquisition* problem as opposed to the *tracking* problem. In this context, acquisition refers to computing with good reliability an initial estimate of the phase that will correct for most of the phase distortion in the aperture plane while starting with a focal-plane intensity measurement. This situation exists when the system is first pointed toward the desired spacecraft to receive a laser communications signal. By contrast, tracking refers to the case in which a good phase estimate is already available and the system is only updating its phase estimates to keep up with atmospheric change. This article focuses almost exclusively on acquisition, and tracking will be discussed only briefly.

II. AO Using Focal-Plane Intensity Measurements

Throughout this article, we will use a Fourier optics model to describe the relationship between the optical field at the aperture of the receiver and the field in the focal plane. Since the FPA detector is assumed to measure intensity via photon counting, we will assume that the FPA can measure incoming power, but not phase, at its pixels. This is true of many, but not all, optical FPA receivers. A discussion of a coherent FPA system that is capable of measuring phase is in [4].

Let the complex optical field at the aperture of the receiving telescope be given by $U_a(r, \theta)$ in polar coordinates (due to the circular aperture), and let R_a denote the radius of the telescope aperture. We assume here that $U_a(r, \theta)$ is zero for all $r > R_a$ and that, if a central obscuration exists in the telescope pupil, this obscuration is already present in the function $U_a(r, \theta)$ as well. Let the complex field in the focal plane be denoted $U_f(\rho, \phi)$. Let f denote the focal length of the telescope, and let λ denote the wavelength of the light used to transmit the communications signal. The relationship between $U_a(r, \theta)$ and $U_f(\rho, \phi)$ is given by

$$U_f(\rho,\phi) = \frac{\exp\left(j\frac{\rho^2}{\lambda f}\right)}{j\lambda f}\int_0^R\int_{-\pi}^{\pi} U_a(r,\theta)\exp\left(-j\frac{2\pi\, r\rho}{\lambda f}\cos(\theta-\phi)\right)d\theta\, r\, dr \tag{1}$$

The function $U_f(\rho,\phi)$ is the well-known Airy pattern if $U_a(r,\theta)$ is a perfect plane wave. Unfortunately, atmospheric turbulence, which is modeled using Kolmogorov phase screen throughout this article, will induce random phases in $U_a(r,\theta)$, meaning that $U_a(r,\theta)$ is not an ideal plane wave. This results in spreading of the energy of the focal-plane function $U_f(\rho,\phi)$, which, in turn, results in BER performance reductions of entire orders of magnitude [6,7] when realistic optical filters are employed. Although, in general, atmospheric turbulence introduces random variations in the amplitude of the function $U_a(r,\theta)$, known as scintillation, we have ignored it in our simulations by invoking aperture averaging as discussed in [4,5]. Aperture averaging is valid for the case when the receiving aperture diameter is much greater than the atmospheric coherence length generated by turbulence or, equivalently, when a great many spatial modes are observed simultaneously via a focal-plane detector array. In effect, we are replacing the random field amplitude with its average value, and focusing only on the random phase distortions. The goal of an AO system is to estimate the complex field $U_a(r,\theta)$ over the area of the receiver aperture and to apply the conjugate of the complex function to a correction element (i.e., a spatial phase modulator) in order to optimally correct the incoming wavefront. However, due to the complexity of applying both phase and amplitude corrections to the aperture fields in a practical system, here we shall consider a suboptimum approach that treats the scintillating amplitude as a constant, and concentrate on correcting the phase of the received turbulence-degraded optical fields. Phase estimation is therefore the focus of this article.

A coherent FPA [4] would, assuming appropriate detector spacing and sufficient coverage of the focal plane, allow one to directly sample $U_f(\rho,\phi)$. If this spatial sampling were performed at the Nyquist rate or higher, then one could, at least in principle, reconstruct $U_a(r,\theta)$ with an inverse Fourier transform, after correcting the external quadratic phase term prior to performing the inverse transform operation. This would allow one to determine aperture phases induced by atmospheric turbulence and to apply appropriate corrections. Since the FPA is assumed to measure intensity instead of complex amplitude, we define the intensity function in the focal plane:

$$I_f(\rho,\phi) = |U_f(\rho,\phi)|^2 \tag{2}$$

The difficulty in this case lies in the fact that we do not have access to $U_f(\rho,\phi)$; instead, we have access only to $I_f(\rho,\phi)$. Given that only the magnitudes are available, it is easy to see that by assigning arbitrary phases in the focal plane we could obtain many different aperture phase functions $U_a(r,\theta)$. In general, the inversion problem does not have a unique solution, although atmospheric turbulence may constrain the phase sufficiently to reduce the choices. The key engineering challenge addressed here is that of obtaining phase estimates in the aperture plane even when only the real intensity function $I_f(\rho,\phi)$, and not the complex amplitude $U_f\,(\rho,\phi)$, is available.

The idea of using intensity measurements in the focal plane to control an adaptive optics system has appeared in the literature before. Carhart et al. have published much in this area, and their early paper [2] discusses the use of intensity measurement at a single point in the focal plane of the receiver to control a phase-correction element. This early work had several key features, including a measurement of the focal field intensity at a single point as the measure of goodness to be optimized; a gradient descent method to adjust the phase-correction element in the optical path; and correction of a static phase distortion field.

Unlike most AO systems, the gradient descent system in [2] did not require direct wavefront measurement, resulting in a reduction in overall system complexity. Furthermore, it was shown in [2] that this method does yield very good aperture phase correction, resulting in greatly enhanced image quality in certain experiments.

Building upon the results of [2], in [3] the authors chose to focus on the problem of improving optical communications performance, and BER performance specifically, using an FPA and a gradient descent algorithm. This early work showed that the technique is promising, but there is a serious technical obstacle: gradient descent often requires many successive focal-plane measurements to converge. A gradient descent algorithm that performs one gradient descent iteration per focal-plane measurement may be much too slow to work in a dynamic atmosphere to yield good BER performance [3]. Hence, a critical issue raised in [3] is the need for faster convergence to track phases in a dynamic atmosphere.

The work of Carhart et al. [2] and of the authors [3] previously relied on a new focal-plane measurement for each iteration of the gradient descent algorithm. Given that signals arriving on deep-space communications links may be of very low power, resulting in relatively few photons per symbol, there will be an upper bound on the rate at which updates can be performed. The update rate must not be so fast that too few photons arrive between updates. Otherwise, the algorithm will not have sufficiently reliable intensity information to converge. Another upper bound on update rates comes from the limited adaptation rate of the phase-correction element itself, which may not be able to perform updates at a rate faster than 2 kHz to 10 kHz if relying on mechanical actuators or large-capacitance devices. Hence, the methods in [2] and [3] have an upper bound imposed on their correction update rates. However, there is a necessary lower bound imposed by the need to track a dynamic, changing atmosphere as well, and this requirement for faster convergence and greater processing speeds presents difficulties for some of the previous methods.

III. A Single Intensity-Measurement Approach to Phase Estimation

Given the challenges described at the end of the previous section, the focus needs to be on developing algorithms that estimate the phase vector of the received optical signal at a rate exceeding the coherence time of turbulence, perhaps using a single measurement of the focal-plane intensity distribution [15,16] The system repeatedly tries new phase estimates and, with high probability, acquires the aperture plane phase within a few iterations given adequate signal energy. This allows acquisition to occur within 1 to 2 milliseconds. More importantly, it also allows the algorithm to keep trying many times per atmospheric coherence time until a good solution is found, circumventing the weakness found in the approach in [3]. In Section III.A, we discuss one of the previous approaches to recovering an object from the modulus of its Fourier transform and present two mean-squared error (MSE) cost functions that are minimized. In Section III.B, we present the basics of genetic algorithms. Section III.C presents a "hybrid" algorithm, which is a combination of a genetic algorithm and the classical Gerchberg–Saxton (GS) algorithm. In Section III.D, results using this approach are given, and the groundwork is laid for a better approach in Section IV.

A. Previous Approaches

One of the early approaches to the phase inversion problem based on intensity measurements was the Gerchberg–Saxton algorithm described in [15,16], which is an iterative procedure for obtaining phase estimates. We have already defined $U_f(\rho,\phi)$ and $U_a(r,\theta)$ to be the complex focal-plane and aperture-plane fields, respectively. Their absolute values (moduli) are given by $m_f(\rho,\phi) = |U_f(\rho,\phi)|$ and $m_a(r,\theta) = |U_a(r,\theta)|$. Let the best estimates of the focal-plane and aperture-plane complex fields be denoted $\hat{U}_f(\rho,\phi)$ and $\hat{U}_a(r,\theta)$, and let their moduli be defined as $\hat{m}_f(\rho,\phi) = |\hat{U}_f(\rho,\phi)|$ and $\hat{m}_a(r,\theta) = |\hat{U}_a(r,\theta)|$. Note that the magnitude of the true aperture-plane field is simply the pupil function, which depends only on the telescope design, and not at all on the received field. The Gerchberg–Saxton algorithm now can be summarized below.

(1) Given the current estimate $\hat{U}_a(r,\theta)$ of the aperture-plane complex field, take a Fourier transform to get an estimate $\hat{U}_f(\rho,\phi)$ of the focal-plane field.

(2) Compute the following mean-squared error in the focal plane: $\mathrm{MSE}_{\mathrm{focal},k} = (1/\pi R_a^2)\int_0^{\infty}\int_{-\pi}^{\pi}\left[\hat{m}_f(\rho,\phi) - m_f(\rho,\phi)\right]^2 d\phi\rho d\rho$, where k indicates the current Gerchberg–Saxton iteration.

(3) Take the phases of the focal-plane field estimate $\hat{U}_f(\rho,\phi)$ obtained in step (1) and apply them to the known absolute values $m_f(\rho,\theta)$ obtained from the focal-plane array. This yields the new focal-plane complex field estimate $\hat{U}_{f2}(\rho,\phi)$.

(4) Take the inverse transform of the new focal-plane field estimate $\hat{U}_{f2}(\rho,\phi)$ from step (3) to get an estimate of the aperture field $\hat{U}_{a2}(r,\theta)$.

(5) Compute the following mean-squared error in the aperture plane (here we assume an unobstructed aperture): $\mathrm{MSE}_{\mathrm{aperture},k} = (\sqrt{\pi}R_a)^{-2}\int_{R_a}^{\infty}\int_{-\pi}^{\pi}\left[\hat{m}_{a,k}(r,\theta)\right]^2 r dr\, d\theta$, where, as above, k indicates the current iteration and where $\hat{m}_{a,k}(r,\theta) = |\hat{U}_{a2}(r,\theta)|$.

(6) Apply the known aperture-plane constraint to the field $\hat{U}_{a2}(r,\theta)$ obtained in step (4). Return to step (1).

(7) Continue until some error criterion (or stopping rule) is satisfied. The Gerchberg–Saxton algorithm typically is stopped when one of the MSE errors computed above falls beneath a threshold. In [9], it was shown that, for the case of the Gerchberg–Saxton algorithm, the following relationship holds true: $\mathrm{MSE}_{\mathrm{focal},k+1} \leq \mathrm{MSE}_{\mathrm{aperture},k} \leq \mathrm{MSE}_{\mathrm{focal},k}$.

The initial aperture-field estimate could be generated in several different ways. The simplest model assumes it is an undistorted plane wave. Other computational approaches include randomly generated fields using randomly assigned Zernike coefficients or uniform random phase assignments over the aperture of the object [10].

B. Background on Genetic Algorithms

Genetic algorithms represent one method of solving optimization problems [11,12]. Our presentation of genetic algorithms will start with a high-level overview, and we will slowly move into the details. The application of genetic algorithms to the optimization problem at hand will be discussed in Section III.C, which is meant to be a very brief introduction to genetic algorithms.

Let $f(\mathbf{x})$ be a non-negative scalar cost function of a vector input variable, and let $\mathbf{x}$ be an N-dimensional vector. Our goal is to find $\mathbf{x}^{\mathrm{opt}}$ such that $f(\mathbf{x})$ is minimized. A genetic algorithm (GA) works by maintaining a set of vectors $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_M\}$. The "fittest" vectors, defined by lower cost, survive and reproduce via the crossover operation (to be defined later). The "weakest" vectors, defined by higher cost, are discarded. Vectors that are discarded due to high cost do not reproduce at all. The steps of a genetic algorithm are given below.

(1) Generate a set of initial vectors N-dimensional $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_M\}$. This set is called the *first generation*.

(2) Obtain a set of costs $\{c_i\}$ where $c_i = f(\mathbf{x}_i)$. Lower costs define "better" vectors, while higher costs define "worse" vectors. In our application, cost is equivalent to focal-plane MSE, defined above.

(3) The *keep fraction* $0 \leq \alpha < 1$ defines the fraction of vectors that survives unchanged into the next generation. In other words, the αM lowest MSE vectors are to be kept unchanged for the next generation.

(4) The *discard fraction* $0 < \beta < 1$ defines the fraction of vectors that is eliminated.

(5) After the βM worst vectors are eliminated, there are a total of $(1-\beta)M$ vectors remaining. This set of $(1-\beta)M$ survivors includes the αM lowest MSE elements that survive unchanged into the next generation as a subset. It is necessary to have $(1-\beta) > \alpha$ for this condition to be met.

(6) So far, αM vectors propagate unchanged. In order to keep the population of vectors at a constant size M, it is necessary to generate $(1-\alpha)M$ *new* vectors. Each of the new $(1-\alpha)M$ vectors is generated by performing a *crossover operation* using the $(1-\beta)M$ vectors not eliminated in step (4).

(7) Once the new $(1-\alpha)M$ vectors have been generated, the new set of vectors, called the *next generation*, consists of these new vectors and the αM lowest-cost vectors saved in step (3). Return to step (2) to evaluate costs, and repeat.

The crossover operation described in step (6) generates a new vector as follows:

(1) From the population of the surviving $(1-\beta)M$ vectors, randomly choose two vectors. These will be the parents of the child vector. Call the two randomly selected vectors $\mathbf{p}_1$ and $\mathbf{p}_2$, and call the child vector $\mathbf{c}$.

(2) For each element of $\mathbf{c}$, randomly select the corresponding element of *either* $\mathbf{p}_1$ or $\mathbf{p}_2$. Repeat for each element of $\mathbf{c}$.

(3) Add a small amount of additive white Gaussian noise (AWGN) with a standard deviation σ to the elements of $\mathbf{c}$. This is called "mutation," and it helps the algorithm find new potential solutions in the search space.

The above crossover operation typically will result in $\mathbf{c}$ having about half of its elements from the first parent and half from the second. We note that this is not the only possible crossover operation. One variant of the crossover operation will copy entire segments of either the first or second parent vector to the child, and the boundaries of these segments may be either fixed or random [11,12]. The use of segmentation has the advantage of allowing related portions of the vector $\mathbf{x}$ to be copied to the child together. In many genetic algorithms, α is greater than zero in order to ensure that if a very good (i.e., nearly optimal or optimal) solution is found it will not be discarded. Furthermore, σ in step (3) of the crossover operation is often greater than zero in order to allow the algorithm to test new vectors with new elements as it proceeds.

The genetic algorithm described here was used in conjunction with the Gerchberg–Saxton algorithm as explained in Section III.C.

C. Minimizing MSE

The Gerchberg–Saxton iteration described in Section III.A typically is repeated until $\mathrm{MSE}_{\mathrm{focal}}$ falls beneath a threshold used as a stopping criterion [9]. If the reconstructed aperture image appears to be good, the Gerchberg–Saxton algorithm may be re-started and run another two or three times with different initial conditions, and if all such runs converge to the same image, then the solution found is considered to be unique with a high degree of confidence [9]. It is noted that the Gerchberg–Saxton algorithm is potentially sensitive to initial conditions. The initial inputs used to the Gerchberg–Saxton algorithm are generated randomly. Each initial solution $\mathbf{x}_i = [x_{i,1} \quad x_{i,2} \quad \cdots \quad x_{i,N}]$ is a vector containing N Zernike coefficients. The ith initial phase aperture phase field is given by

$$\hat{U}_{a,i}(r,\theta) = \sum_{n=i}^{N} x_{i,n} z\left(\frac{r}{R_a}, \theta\right) \tag{3}$$

where $z_i(r/R_a, \theta)$ is the ith Zernike polynomial. A total of M such initial phase fields is generated as input to the optimization algorithm, which is a genetic algorithm that works by generating entire sets of initial conditions for the Gerchberg–Saxton algorithm, trying them, and selecting solutions with the smallest focal-plane MSE. The basics of genetic algorithms are thoroughly discussed in [11] and [12], and an application of genetic algorithms to the aperture phase retrieval problem is discussed in [3]. Since the genetic algorithm used here is very similar to the one used in [3], we refer readers to it for more information. A number of key differences between the present genetic algorithm and that in [3] need to be pointed out.

The genetic algorithm in [3] required a new focal-plane intensity measurement for each and every cost function evaluation. Given the large number of cost function evaluations, this algorithm would have been infeasible for use in a real-time AO system [3]. By contrast, the present genetic algorithm requires only one focal-plane intensity measurement. Once a single measurement is taken, the cost function evaluations are all done using the existing measurement without any need for more focal-plane intensity data. Finally, the present GA can be implemented in a large field programmable gate array (FPGA), and there is evidence that it can be made to run in real time based on resource and timing information from the data sheet for the Xilinx Virtex-2 series FPGAs [13] and based on timing and resource requirements for a 32-point fast Fourier transform (FFT) logic core [14].

In this hybrid GA-GS algorithm, the GA provides various test solutions to the GS, which performs the phase-retrieval operation for each initial test vector. The resulting focal-plane MSE is evaluated for each run of the GS, and the fraction α of the best performing initial conditions is kept. A genetic algorithm may stop when either its best solution's MSE falls below a threshold or after a fixed number of iterations. Given that the final goal is a real-time algorithm, we have chosen a fixed number of iterations.

In our simulations, the following parameters were used:

(1) $M = 25$: The first 25 Zernike coefficients were optimized.

(2) $N = 10$: There were 10 solutions per GA iteration.

(3) $\alpha = 0.4$: The best 40 percent (best 4) solutions were always kept.

(4) $\beta = 0.4$: The worst 40 percent (worst 4) solutions were discarded without being used at all in producing new solutions via crossover.

(5) For each solution, Gerchberg–Saxton was iterated 20 times per generation.

(6) The GA was run through only 2 generations, so a total of 40 iterations are needed for each solution.

(7) In all cases, 32×32 two-dimensional FFT and inverse FFT operations were used.

The above will lead some readers to question whether this algorithm can be implemented in real time. Although the present algorithm is a key improvement over the one in [3] in that it works on a single snapshot and thus circumvents the entire need to take one FPA snapshot per iteration, it requires considerable computational power. We argue here that a solution can be generated within less than one millisecond, which is the assumed coherence time for the atmosphere.

Each of the N solutions undergoes 20 Gerchberg–Saxton iterations per GA generation. A Gerchberg–Saxton iteration requires one FFT and one inverse FFT, and we will simply count this as two FFTs per

Gerchberg–Saxton iteration. The result is 40 FFTs per solution × 10 solutions, for 400 FFTs per GA iteration. The GA is only iterated twice, so there are 800 FFTs in total.

Can a 32 × 32 FFT run quickly enough? Basing resource and timing estimates on two documents published by Xilinx corporation, it appears that the answer is "yes" [13,14]. We first note that the 10 solutions are fully independent of each other, and a bank of 10 independent processing modules implemented on a series of FPGAs could be used. If we assume 10 independent processing modules, then a total of 80 FFTs must be run serially on each module. If we assume that a Xilinx Virtex-2 FPGA is used, then a new 32-point one-dimensional FFT can be computed, on average, every 96 clock cycles. Since a 32 × 32 two-dimensional FFT is the true goal, a module should ideally contain 64 FFT cores running in parallel. In the first stage, the first 32 cores could compute along the first dimension in parallel in 96 clock cycles. In another 96 clock cycles, the next 32 cores then could compute along the second dimension in parallel. Thus, a single 32 × 32 two-dimensional FFT could, in principle, be computed in 192 clock cycles, assuming pipelined execution. If we assume the worst case, in which the FFT cores would all have to be restarted each time, it would still take just 258 clock cycles per FFT using 64 cores. Assuming 31 FPGA slices per core, approximately 1984 slices would be needed. Given that a Virtex-2 Pro FPGA has up to 44,000 slices, one could possibly implement the entire design on a single large Virtex-2 Pro. At 258 clock cycles per FFT with 80 serial FFTs, we need 20640 clock cycles per iteration of the GA/Gerchberg–Saxton hybrid algorithm. A rough time estimate for this algorithm would be on the order of 0.1 ms, assuming that FFTs dominate running time and assuming a 200-MHz clock. Of course, the overhead required for MSE computations, comparisons, and GA operations could double this to 0.2 ms. Even so, a computation could, with a proper hardware implementation, be performed in much less than 1 ms using a Virtex-II Pro FPGA, meaning that the algorithm even in this early form is feasible. In Section V, further improvements leading to faster convergence times and fewer iterations will be discussed, and we believe that real-time implementations in affordable FPGA hardware will be feasible.

D. Early Results of $\mathrm{MSE}(\mathbf{x})$ Minimization

Both simulation trials and real-world laboratory data were used to test the above approach. In simulations, the receiving telescope was assumed to have a diameter of 63.5 cm, and the Fried parameter r_0 was assumed to be 15 cm. Random Kolmogorov phase screens with this Fried parameter were generated, and the hybrid genetic algorithm and classical optimization system described in Section III.C was applied. The focal-plane field intensities before and after correction are shown in Figs. 1(a) and 1(b).

An improvement in Strehl ratio by a factor of 2.7 was achieved with the hybrid GA-GS, with the corrected field having a Strehl ratio of 0.77 as shown in Figs. 1(a) and 1(b). This means that peak intensity after correction was 77 percent of that achieved if a perfect plane wave were to replace the corrected incoming light wave. Note also that in addition to a nearly threefold improvement in peak intensity the focal-plane field energy is much better concentrated, which would yield a lower BER in a communications link [8]. By contrast, straight application of the GS algorithm for the same length of time, without an initial search using a repeated application of the hybrid GA-GS algorithm, yields the focal-plane intensity distribution shown in Fig. 1(c). Note that, in this case, a Strehl ratio of 0.24 was obtained, which is substantially smaller than the Strehl ratio attained with the hybrid GA-GS algorithm.

The true aperture-plane phase distribution, the estimated aperture-plane phase distribution, and the residual phase-error distribution are shown in Figs. 2(a), 2(b), and 2(c), respectively, for the case of the hybrid GA-GS algorithm. For this example, the hybrid GA-GS algorithm achieved a focal-plane mean-squared error of 5.7654. The residual phase-error distribution corresponding to the classical GS algorithm is shown in Fig. 2(c). Note that the classical GS algorithm, starting with a phase distribution representing a normally incident plane wave, was unable to estimate the distorted phase front, achieving only a focal-plane mean-squared error of 11.9182 in the same length of time.

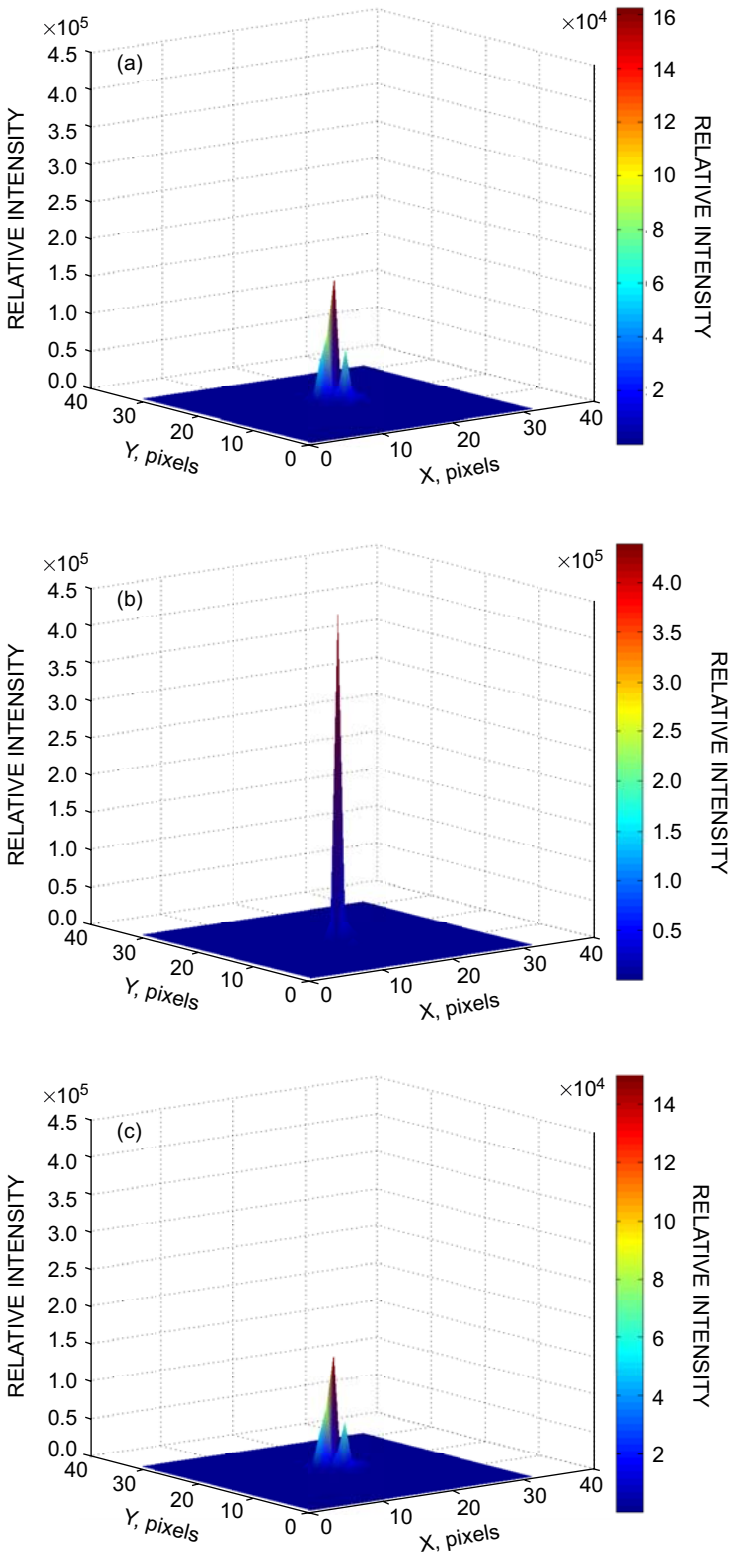

Fig. 1. Focal-plane intensity (a) before correction, (b) after correction using the hybrid GA-GS algorithm, and (c) with the classical GS algorithm.

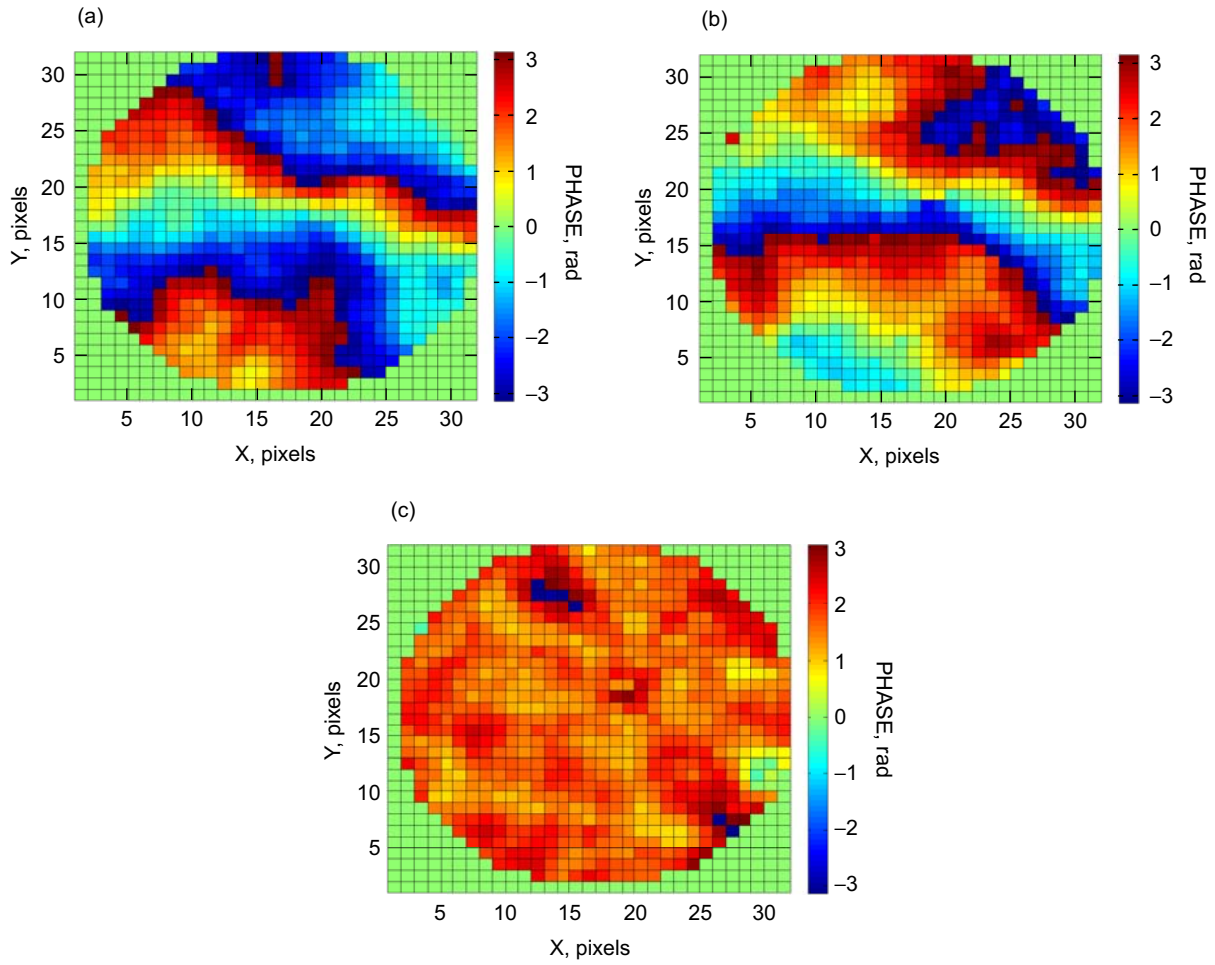

Fig. 2. Illustration of aperture phase correction: (a) true phase distribution, (b) estimated phase distribution, and (c) residual phase error for the hybrid GA-GS algorithm.

Although MSE converges to approximately 12 in Fig. 3(a) for classical GS, the best solution obtained with the hybrid GA-GS algorithm achieves an un-normalized focal-plane MSE of approximately 5.77 in 40 iterations, as shown in Fig. 3(b). Note that this MSE represents two generations of test vectors from the GA; however, the same test vector was selected in the second generation, resulting in a continuous MSE curve. Had a different test vector been selected, there would likely be a noticeable step discontinuity at the 20th iteration of the hybrid GA-GS algorithm.

Simulation results thus show a great deal of promise for the hybrid GA-GS algorithm. Even better, the weakness of previous algorithms [3], the need for excessively high measurement and AO update rates, has been eliminated since this is a single intensity measurement algorithm.

Further verification came from laboratory data gathered by Muñoz Fernández [4] for her doctoral thesis. These data came from a coherent FPA, which takes samples of the complex focal-plane field over time. Using such data, it is possible to obtain estimates of the aperture-plane complex field using an inverse Fourier transform. In our simulations, the focal-plane measurements were used to first obtain $U_f(\rho, \phi)$, after which the absolute value was taken to obtain $m_f(\rho, \phi) = |U_f(\rho, \phi)|$. Next, the inverse Fourier transform relationship was used to obtain $U_a(r, \theta)$ and $m_a(r, \theta) = |U_a(r, \theta)|$. Finally, the intensity measurement $I(\rho, \phi) = |U_f(\rho, \phi)|^2$ was provided to the stochastic optimization algorithm, which was used

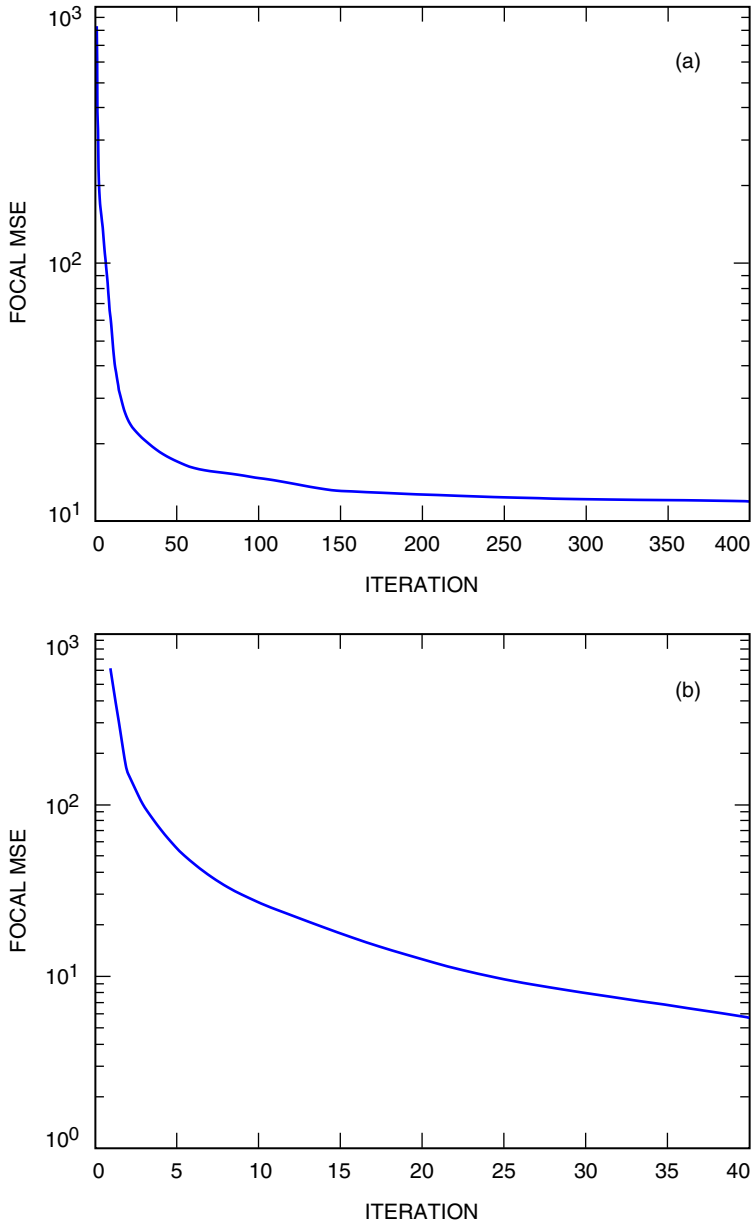

Fig. 3. Focal-plane MSE for (a) the classical GS algorithm and (b) the hybrid GA-GS algorithm.

to estimate the first 25 Zernike coefficients in the expansion. The resulting estimate of the first 25 Zernike coefficients then was used to construct the output estimate of the aperture-plane field.

A comparison of the aperture phases using the inverse Fourier transform of the focal-plane field samples as the baseline compared to the aperture phases estimated via the stochastic optimization method under development is shown in Fig. 4. Although the stochastic optimization method did not yield perfect estimates, it is seen that the estimates still correspond closely.

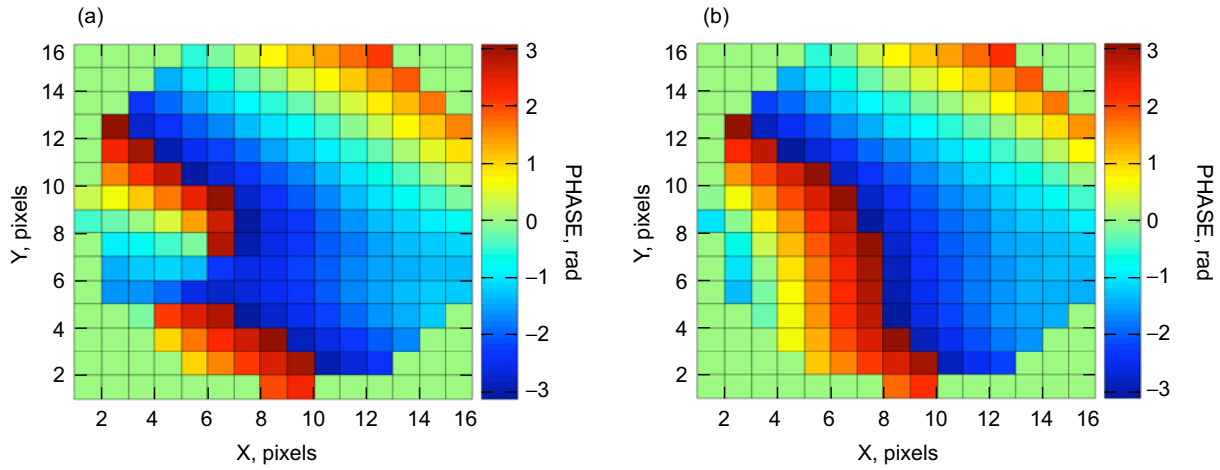

Fig. 4. Results using actual laboratory data: (a) actual phases from the inverse Fourier transform and (b) estimated phases from the hybrid optimization algorithm.

IV. Conclusions and Future Directions

Significant progress has been made in the development of algorithms for AO using FPA-based intensity measurements. In particular, the convergence time weaknesses involving the previous gradient-descent-based approach in [3] have been addressed by using a single intensity-measurement method of aperture phase estimation. Hence, assuming rapid single intensity-measurement phase estimation, it is possible to acquire a good estimate of aperture phase in real time. Even though the solution to the inverse Fourier transform problem is non-unique, the cost function MSE($\mathbf{x}$) was chosen so that any minima found will have a sufficiently high probability of being correct to allow rapid acquisition within several tries.

One challenge facing an actual implementation is the need for computational power, but this challenge is already being addressed. In particular, 16×16 and 32×32 two-dimensional FFTs can be implemented in FPGA hardware, and a large Virtex-2 Pro FPGA could be used to implement the rapid, highly parallel evaluations of the function MSE($\mathbf{x}$) that must be optimized quickly, in real-time, for this algorithm to work. It is likely that a Virtex-2 or Virtex-4 series FPGA would be required due to the need for hardware resources for FFT implementation and due to the need for a high degree of parallelism in executing the FFTs.

Reducing the number of function evaluations often requires reducing the search space. An alternate basis can be used in place of the Zernike polynomials. In particular, if the focal-plane field is expanded using a series of sampling functions, the localized nature of the focal-plane field often will allow a reduced basis to be used in the optimization problem. This leads to faster convergence and fewer function evaluations, significantly reducing the demand on hardware and possibly eliminating the need for the GA altogether, bringing an FPGA-based implementation well within the realm of feasibility. Current work focuses on both a more efficient set of expansion functions for this problem as well as on determining an efficient, highly parallel system design that can be implemented in common FPGA hardware.

Acknowledgment

The authors wish to acknowledge David J. Rochblatt for many hours of helpful technical discussions and advice throughout the course of this work and for suggesting the Gerchberg–Saxton algorithm, a variant of which is the topic of this article.

References

[1] N. Ageorges and C. Dainty, *Laser Guide Star Adaptive Optics for Astronomy*, Boston, Massachusetts: Kluwer Academic Publishers, 2000.

[2] G. W. Carhart, J. C. Ricklin, V. P. Sivokon, and M. A. Vorontsov, "Parallel Perturbation Gradient Descent Algorithm for Adaptive Wavefront Correction," *Proceedings of SPIE*, San Diego, California, pp. 221–227, 1997.

[3] R. Mukai, K. Wilson, and V. Vilnrotter, "Application of Genetic and Gradient Descent Algorithms to Wavefront Compensation for the Deep-Space Optical Communications Receiver," *The Interplanetary Network Progress Report*, vol. 42-161, Jet Propulsion Laboratory, Pasadena, California, pp. 1–21, May 15, 2005. http://ipnpr.jpl.nasa.gov/progress_report/42-161/161U.pdf

[4] M. Muñoz-Fernández, V. Vilnrotter, R. Mukai, and B. Hassibi, "Coherent Optical Array Receiver Experiment: Design, Implementation and BER Performance of a Multichannel Coherent Optical Receiver for PPM Signals under Atmospheric Turbulence," *Proceedings of SPIE, Free-Space Laser Communication Technologies XVIII*, San Jose, California, January 24–25, 2006.

[5] R. M. Gagliardi and S. Karp, *Optical Communications*, New York: John Wiley and Sons, 1976.

[6] F. Roddier, *Adaptive Optics in Astronomy*, Cambridge, United Kingdom: Cambridge University Press, 1999.

[7] V. Vilnrotter, C.-W. Lau, M. Srinivasan, K. Andrews, and R. Mukai, "Optical Array Receiver for Communication Through Atmospheric Turbulence," *Journal of Lightwave Technology*, vol. 23, no. 4, pp. 1664–1675, April 2005.

[8] V. A. Vilnrotter and M. Srinivasan, "Adaptive Detector Arrays for Optical Communications Receivers," *IEEE Transactions on Communications*, vol. 50, pp. 1091–1097, July 2002.

[9] J. R. Fienup, "Reconstruction of an Object from the Modulus of its Fourier Transform," *Optics Letters*, vol. 3, no. 1, pp. 27–29, July 1978.

[10] J. R. Fienup, "Phase Retrieval Algorithms: A Comparison," *Applied Optics*, vol. 21, no. 15, pp. 2758–2769, August 1982.

[11] L. Chambers, ed., *Practical Handbook of Genetic Algorithms: Applications*, vol. I, Boca Raton, Florida: CRC Press, 1995.

[12] L. Chambers, ed., *The Practical Handbook of Genetic Algorithms*, Boca Raton, Florida: Chapman and Hall/CRC Press, 2001.

[13] *Virtex-II Pro and Virtex-II Pro X Platform FPGAs: Complete Data Sheet*, Xilinx, Inc., San Jose, California, October 2005. http://www.xilinx.com/bvdocs/publications/ds083.pdf

[14] *High-Performance 32-Point Complex FFT/IFFT V3.0*, Xilinx, Inc., San Jose, California, March 2002. http://www.xilinx.com/ipcenter/catalog/logicore/docs/vfft32.pdf

[15] R. W. Gerchberg and W. O. Saxton, "A Practical Algorithm for the Determination of Phase from Image and Diffraction Plane Pictures," *Optik*, vol. 35, no. 2, pp. 237–247, 1972.

[16] D. J. Rochblatt, "Systems Analysis for DSN Microwave Antenna Holography," *The Telecommunications and Data Acquisition Progress Report 42-96, October–December 1988*, Jet Propulsion Laboratory, Pasadena, California, pp. 132–157, February 15, 1989.
http://ipnpr.jpl.nasa.gov/progress_report/42-96/96L.PDF