Application of Genetic and Gradient Descent Algorithms to Wavefront Compensation for the Deep-Space Optical Communications Receiver

R. Mukai,[1] K. Wilson,[1] and V. Vilnrotter[1]

Present adaptive optics systems use a wavefront sensor to detect phase errors in the incoming wavefront. Knowledge of these phase errors then is used to correct the incoming wavefront, reducing image distortion. However, these systems require that a portion of the incoming light be diverted to the wavefront sensor and away from the main receiver's focal plane, which results in a loss of signal power in the receiver in optical communication applications. Accordingly, it is desirable to develop a system that relies entirely on the focal-plane intensity distribution as opposed to a separate wavefront sensor in order to detect and correct wavefront errors. Two common stochastic optimization techniques—genetic algorithms and gradient descent algorithms—are evaluated in this article.[2] *Although these algorithms are promising, further work is necessary to enable them to be used in practical adaptive optics systems due to their slow convergence speeds relative to the rate of change in the atmosphere.*

I. Introduction

Optical communications systems for space applications face the problem of distortion of the optical beam as it passes through Earth's atmosphere. Turbulent eddies in the atmosphere result in random phase errors in the wavefront at the aperture of the receiving telescope. This leads to severe scattering of the received signal energy in the telescope's focal plane [12,14–16] and can result in degraded bit-error rate (BER) performance in the presence of a background such as a bright daytime sky. Correction of the phase front thus is highly desirable in order to achieve minimum bit-error rates.

---

[1] Communications Architectures and Research Section.

[2] If the incoming signal is integrated over a large number of incoming signal pulses, the resulting average will be very close to the intensity distribution given by wave optics. The instantaneous signal, however, is not equal to this average since this is an array of photon-counting detectors, and photon-counting effects would have to be taken into account for shorter integration intervals. Here, we assume that integration intervals are sufficiently long to provide an accurate average intensity distribution in the focal plane.

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.

Adaptive optics systems today operate by using a wavefront sensor to detect phase errors caused by atmospheric turbulence [12]. Once the phase errors are detected by the wavefront sensor, appropriate corrections are applied to a deformable mirror and to a tip/tilt mirror to stabilize the focused spot of received energy on the detector, leading to improved optical performance. When applied to optical communications, adaptive optics requires that photons be routed away from the main detector to the wavefront sensor. One approach to mitigate the amount of signal loss is to implement a laser guide star to correct the higher-order atmosphere-induced aberrations. Ground-based Rayleigh laser guide stars do not correct for atmospheric tip/tilt, and some of the received signal power still will need to be directed to a quadrant detector to stabilize the location of the downlink signal in the focal plane, thereby enabling a reduction in receiver field of view. This loss of signal energy will impact the signal-to-noise ratio (SNR) and the performance of the system and must be considered in the end-to-end link design. An approach that is less complex and that does not require the diversion of energy from the communications detector is of particular interest to deep-space optical communications.

A direct approach not requiring the use of laser guide stars or of tip/tilt mirrors would be to use the information taken from the distribution of received signal power in the focal plane to perform the required wavefront corrections. Information on the atmosphere-induced wavefront aberrations can be gleaned from an array of intensity sensors in the focal plane. Although there is not a one-to-one correspondence between phase distortions in the aperture plane and intensity distribution in the focal plane, perturbing the phases in the aperture plane allows one to gather enough information from the intensity distribution in the focal plane to correct wavefront errors. It is conceivable that this information can be used to correct the wavefront without the need for a wavefront sensor. This would enable adaptive optics (AO) systems to operate without the signal losses imposed by wavefront sensors. This concept was illustrated for image correction by Carhart et al. [3] and by Givéon et al. [9,10]. Previous work by Weyrauch et al. [20] has led to the development of gradient descent methods for optical communications, but no bit-error results were given. Instead, emphasis was placed on optimizing cost or fitness functions related to communication link quality. The algorithms presented in this article are a first step toward feasible algorithms for real-time adaptive optics systems that would rely entirely on information from the focal plane without the need for a wavefront sensor.

In this article, the objective is to simulate and evaluate the reception of a 64-PPM (pulse-position modulation) optical link at the 1064-nanometer wavelength. The detector is a focal-plane array located at the focus of a receiving telescope. The evaluation metric is the bit-error rate. The detector array considered was a 128×128 pixel array. In the absence of atmospheric turbulence and aberrations by optical components, the spatial distribution of the photons in the focal plane is an ideal Airy pattern [12]. Atmospheric turbulence introduces wavefront errors in the beam and blurs the size of the focused spot.

Previously published works by Vilnrotter and Srinivasan [15,16] describe the selection of the optimal subset of focal-plane array detection elements for PPM pulse detection with minimal bit-error probabilities. Their optimal solution approach requires a specific number of receiving elements in the focal plane to keep bit-error probabilities at a minimum for the given focal-plane distribution. Increasing the number of elements results in not only the capture of signal photons but also in the capture of increasing numbers of background photons. This results in reduced BER performance [15,16]. The ideal Airy pattern, which corresponds to diffraction-limited performance, yields the lowest overall bit-error rates for a given background level and for a given signal pulse strength. Accordingly, we seek to use adaptive optics to bring the received pattern in the focal plane as close to the diffraction-limited Airy pattern as possible and thereby minimize the BER.

Adaptive optics techniques for improving images are not new, and they have been used to improve the resolution of large Earth-based astronomical telescopes achieving near-diffraction-limited imaging performance in the infrared [12]. In the diffraction limit, the radius of the focused spot at wavelength λ in the focal plane of a telescope of diameter D and focal length f is given by [12]

$$w_0 = 1.22\frac{\lambda f}{D} \tag{1}$$

However, atmospheric turbulence-induced wavefront aberration limits the resolution of large telescopes to that of a telescope of diameter r_0 Fried coherence length [12]. This has serious consequences for both astronomical imaging and optical communications. As an example, for $r_0 = 7.0$ cm, $D = 1.0$ m, and a wavelength of 1064 nm, the focused spot diameter increases by a factor of approximately 14. This increase in spot size is equivalent to an increase in field of view. For daytime operations, this will result in an increase in background noise and a concomitant decrease in BER performance [15,16].

II. Model of the System

There are two planes of interest in our model of the receiving telescope: the aperture plane and the focal plane. A received signal is incident on the telescope aperture plane. The telescope's optical system relays the signal to the focal plane, where it is detected. A focal plane array (FPA) consisting of a 128×128 square grid of intensity detectors located at the telescope focus receives the signal. The coordinate systems of these two planes are shown in Fig. 1.

We begin with a model of the receiving telescope, which is treated in this article using the Fourier optics approach of Goodman [11]. Assuming an aperture-plane field given by $U_a(r,\theta)$ and a focal-plane field given by $U_f(\rho,\phi)$, we can write [11]

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

where λ and f are as previously defined and

$k \triangleq \dfrac{2\pi}{\lambda}$: wave number

(r,θ) : polar coordinates in the aperture plane

(ρ,ϕ) : polar coordinates in the focal plane

$R = \dfrac{D}{2}$: radius of the telescope's aperture

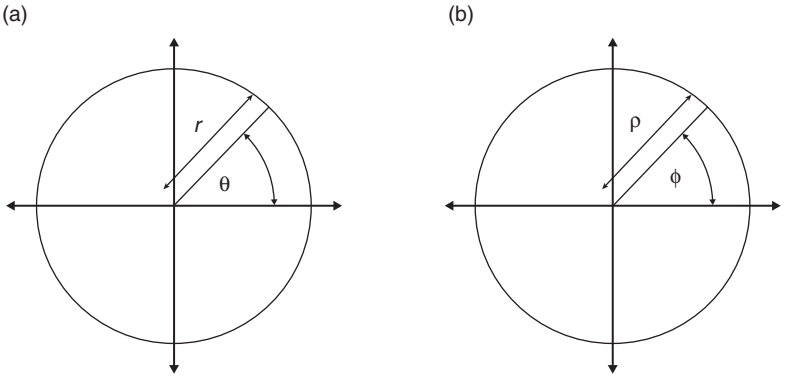

Fig. 1. Coordinate systems: (a) aperture plane and (b) focal plane.

For a uniform plane wave arriving at the telescope along the telescope's axis, such that $U_a(r,\theta) = 1$, the resulting focal-plane field is the ideal Airy pattern. Under realistic conditions, the atmosphere causes phase distortions. These distortions result in changes in the effective path length along the z-axis, the axis of the telescope. We denoted these by $d(r,\theta)$. The resulting aperture-plane field can be written

$$\begin{aligned} U_a(r,\theta) = & \exp\left(j\frac{2\pi}{\lambda}d(r,\theta)\right) \\ = & \exp\left(jkd(r,\theta)\right) \end{aligned} \tag{3}$$

Substituting Eq. (3) into Eq. (2) yields

$$U_f(\rho,\phi) = \frac{\exp\left(j\frac{k}{2f}\rho^2\right)}{j\lambda f}\int_0^R\int_{-\pi}^{\pi}\exp\left(jkd(r,\theta)\right)\exp\left(-j\frac{k}{f}r\rho\cos(\theta-\phi)\right)rd\theta dr \tag{4}$$

Equation (4) is a two-dimensional Fourier transform that is scaled by the wavelength and telescope focal length.

The goal of adaptive optics is to introduce compensating phase shifts in the pupil plane to correct $d(r,\theta)$ [12]. Although there are a variety of approaches that can be used to generate the compensating phase shifts, including liquid crystal micro-electro-mechanical systems (MEMS) mirrors and deformable mirrors [12], there are two basic approaches to measuring the required phase correction. These are [12]

(1) Use a wavefront sensor. This involves direct measurement of the wavefront aberrations, enabling compensation to be applied to cancel the function $d(r,\theta)$ in Eq. (3).

(2) Measure the intensity distribution in the focal plane and deduce the phase distortions, although there is not a one-to-one correspondence between the measured intensity distribution and the wavefront distortion. There are many signal processing approaches available here that enable the AO system to rely entirely on focal-plane information. A number of these approaches are discussed in [12], and a parallel gradient descent approach is described in [3] and expanded upon in [4,5,17–20]. We investigate this approach and describe it in detail in this article.

It is useful to decompose the wavefront path error function $d(r,\theta)$ into a complete orthonormal set of functions. The most common basis set used in optics for this purpose is that of the Zernike polynomial [2,12]. One benefit of using the Zernike polynomials is that of faster, more efficient convergence. By perturbing the Zernike coefficient in the Zernike expansion of the wavefront error $d(r,\theta)$ instead of perturbing individual actuators, one can achieve faster, more efficient convergence in a gradient-descent-based optimization algorithm [3].

III. Algorithms for Adaptive Optics

In this section, we begin with a discussion of the basics of genetic algorithms. We then proceed to describe an application of genetic algorithms to the problem of controlling an adaptive optics system and give a description of our algorithm for doing so. The simulation setup used to test genetic algorithms is presented next, and this section ends with a set of simulation results.

A. Basic Genetic Algorithms

Genetic algorithms represent one method of solving optimization problems [6,7]. Let Φ denote an instance of a given optimization problem, and let $\Sigma = \{\overrightarrow{\sigma}_1(k), \overrightarrow{\sigma}_2(k), \cdots, \overrightarrow{\sigma}_N(k)\}$ denote a set of solutions to the problem Φ at time k. In an optimization problem, we can define a measure of goodness, or fitness, for each of the solutions in Σ. Let these fitness values be given by $f\left(\overrightarrow{\sigma}_n(k)\right)$. Then we can use the function f to order the solutions in $\sum$ according to their quality.

There are many types of genetic algorithms in existence, and we refer the reader to [6,7] among others for more information. The genetic algorithm used in this article works as follows:

(1) Define a "keep fraction" $0 < \alpha < 1$. In each iteration of the algorithm, the $N\alpha$ fittest solutions are kept for the next generation.

(2) Define a "kill fraction" $0 < \beta < 1$. In each iteration of the algorithm, the $N\beta$ least-fit solutions are discarded completely.

(3) The remaining solutions are replaced with new solutions. These are the result of "crossover combination" of the remaining $(1-\beta)N$ solutions [1]. The $N\alpha$ best solutions will propagate unchanged, and there are $(1-\alpha)N$ solutions to be replaced with the "offspring" of the $(1-\beta)N$ solutions not eliminated in Step (2).

(4) Once the new solutions are generated, they are again sorted by fitness, and the process is repeated.

The above approach is not the only possible approach to genetic algorithms, and so far we have been vague in our description of the term "crossover combination." The use of crossover combination will be described in greater detail in the next section. A thorough discussion of genetic algorithms and their many variations is given in two texts edited by Chambers [6,7].

B. Genetic Algorithms for Adaptive Optics

The use of genetic algorithms in adaptive optics is discussed here. Recall Eq. (4) and that, from the previous discussion, energy is spread out in the focal-plane field due to optical phase distortions. The objective of adaptive optics is to use deformable and tip/tilt mirrors to correct the wavefront and effectively reduce $d(r,\theta)$ to zero. Practically, it is not possible to make $d(r,\theta)$ zero. However, we can use corrective optics to minimize its root-mean-square (rms) value and bring the system closer to diffraction-limited operation.

The first step in our approach is to expand $d(r,\theta)$ over the Zernike polynomials. Using the coefficients of the expansion, we apply the appropriate corrections to the deformable and tip/tilt mirrors to reduce the wavefront phase error [3]. Expanding upon the genetic algorithm described in the previous section, we now present our basic genetic algorithm:

(1) Each solution $\overrightarrow{\sigma}_n(k) \in \Sigma$ is a vector of Zernike coefficients that describe the deformation to be applied to the deformable mirror.

(2) The fitness function, $f\left(\overrightarrow{\sigma}_n(k)\right)$, to be applied in each case could be defined either as the Strehl ratio, as the minimum number of detector elements capturing a certain percentage of focal-field energy, or as a bit-error rate. Since each vector of coefficients $\overrightarrow{\sigma}_n(k)$ will result in a given field distribution in the focal plane, it can be argued that each one has a fitness for a given phase distortion in the aperture plane and that this can be influenced by applying a set of commands to the deformable mirror in the pupil plane.

(3) The fitness $f\left(\overrightarrow{\sigma}_n(k)\right)$ of each solution is determined by applying that solution to the deformable mirror. Once that is done, a Strehl ratio or other measure of fitness can be taken. This is done once for each solution $\overrightarrow{\sigma}_n(k) \in \Sigma$.

(4) The solutions then are sorted from the least fit to the most fit. The $N\alpha$ most-fit coefficient sets are kept, and the $N\beta$ least-fit sets are eliminated from the solution pool.

(5) The surviving $(1-\beta)N$ solutions are modified to generate the $(1-\alpha)N$ new solutions to be propagated to the next generation.

The method of generating the $(1-\alpha)N$ child solutions can vary. The crossover combination method, variants of which are used in genetic algorithms [1], is described in detail:

(1) Each solution $\overrightarrow{\sigma}_n(k)$ is a vector of M Zernike coefficients. In other words, $\overrightarrow{\sigma}_n(k) = \left[z_{n,1}(k)\ \ z_{n,2}(k) \cdots z_{n,M}(k)\right]$, where $z_{n,i}$ is the ith Zernike coefficient of the nth solution.

(2) We randomly select $\overrightarrow{\sigma}_{n_1}(k)$ and $\overrightarrow{\sigma}_{n_2}(k)$ as the "parents" of the new child solution.

(3) For each element of the child solution, randomly choose the corresponding element of one of the parent solutions. By setting the probability of choosing each parent to one-half, we ensure that on average a child solution contains half of its elements from the first parent and half from the second parent.

(4) Add a small amount of Gaussian noise to each element of the child solution. This represents "mutation."

The genetic algorithm presented here has the following features:

(1) It always keeps the best solutions. If an optimal or nearly optimal solution is found by chance, then that solution will not be lost. The best solution always survives to the next generation.

(2) It always discards the worst solutions in order to direct the search toward increasingly "fit" solutions.

(3) It uses a "two-parent" method of creating new solutions. We can increase the likelihood that the best features of the parent solutions will, at some point, combine. For example, if $\overrightarrow{\sigma}_{n_1}(k)$ has correct coefficients in the first half of its vector and $\overrightarrow{\sigma}_{n_2}(k)$ contains correct coefficients in its second half, this recombination procedure is likely to produce offspring solutions that contain many features of both.

(4) It includes random mutation. It prevents the system from getting stuck if none of its coefficients is good.

The primary difficult with applying either this algorithm or any other genetic algorithm to the problem of adaptive optics is that phase changes in the aperture plane are rapid, with significant changes occurring on a time scale of milliseconds [12]. A genetic algorithm (GA) therefore needs to be able to adapt its solution vectors quickly in response to atmospheric changes. Fitness evaluation is performed by applying each of the solutions to the deformable mirror (DM) and measuring the intensity in the focal plane. It is important that the solutions obtained be of sufficiently good quality to ensure that the BER will not increase during the process of testing and adaptation. Hence, one of the main obstacles to the use of genetic algorithms in adaptive optics is the need to adapt to change rapidly.

The number N of solutions per generation is crucial to GA performance. Most of our simulations were run with $N = 80$, but this yields poor performance when correcting 400 Zernike coefficients. It took approximately one week to perform a test with $N = 800$, but results were far more impressive. Both cases are shown in Section III.D.

C. Simulation Setup

The simulation model makes use of the "frozen atmosphere" assumption. In this model, a fixed Kolmogorov phase screen is used to simulate the phase errors caused by the atmosphere. This screen is called a frozen atmosphere because the atmosphere is considered to consist of discrete stationary states that remain constant over a time period τ_o. The justification for the use of a frozen atmosphere model comes from the following statement from Hardy: "It is assumed that the turbulence structure does not change over time scales of less than a few seconds, so that temporal effects are entirely due to transport by the wind."[3] In our approach, we use this concept to bound the time over which the simulation must converge. A dynamic atmosphere is thus simulated by moving the phase distortion across the aperture D with a speed D/τ_0 [12].

The objective function $f\left(\vec{\sigma}_n(k)\right)$ used in our simulations is the number of pixels containing 80 percent of the energy of the focal-plane distribution. This metric is suboptimal from a BER perspective in optical communications, as will be demonstrated by the simulation results of this section. Other metrics that can be optimized include the Strehl ratio and the set of metrics defined by Vilnrotter and Srinivasan [15,16] to minimize the BER. Alternatively, one also could attempt to minimize the BER directly. Our choice of 80 percent enclosed energy was motivated by several factors:

(1) The BER that would result from a given focal-plane intensity distribution is difficult to compute in practice, and long simulation runs would be required to estimate it. This metric was not directly optimized for this reason. As a compromise, we sought to find a metric whose optimization would yield nearly optimal BER performance.

(2) In our initial analysis, we found that optimizing the Strehl ratio did not yield as low a BER as optimizing the number of pixels containing 80 percent of the enclosed energy on the 128×128 pixel detector array.

As previously stated, the objective function $f\left(\vec{\sigma}_n(k)\right)$ must be evaluated after each change is applied to the aperture plane phase. This implies that the rate at which the AO algorithm runs is inherently limited by the maximum update rate of the deformable mirror or other phase-correction element. Nevertheless, these algorithms offer strong evidence that the focal-plane intensity distribution contains enough information to permit correction of most phase errors without the need for a wavefront sensor. This result is corroborated by image correction results obtained by Carhart et al. [3] and by Givéon et al. [9,10].

In the results that follow, we consider two approaches for computing the BER. The first method is based on the work by Vilnrotter and Srinivasan [15,16] and computes an optimal subset of FPA pixels to be used as a single detector. We call this the "optimal subset" method. It is optimal in that it yields the lowest achievable BER when the system is required to use the sum of the outputs of a fixed subset of pixels as its decision metric for PPM symbol decisions. The optimal solution given in [14–16] involves the use of multiplicative weights in computing the optimal decision metric. This method is not used here due to the complexity of computing the optimal weights. The second approach simply takes those pixels that contain 80 percent of the enclosed energy of the focal-plane array signal distribution and uses those as a single detector. The second method thus uses a suboptimal subset of pixels as a single detector, yielding bit-error rates that are greater than those achieved using the first method. We will refer to these methods as the "optimal subset method" and the "80 percent method" in the remainder of this article.

The parameters in this simulation are given in Table 1. It is assumed that we have an ideal intensity detector, and no attempt is made to model physical effects such as quantum efficiency that would be present in an avalanche photo-diode (APD) array or other effects that would be present in realistic detectors. The genetic algorithm is used to minimize the number of pixels containing 80 percent

[3] J. W. Hardy, *Adaptive Optics for Astronomical Telescopes*, New York: Oxford University Press, Chapter 9, p. 316, 1998.

Table 1. Parameters used in simulation.

Parameter	Value
Telescope aperture diameter	1.00 m
Telescope center obscuration diameter	0.10 m
Telescope F number	75.80
Fried parameter	0.07 m
Wavelength	1064.00 nm
FPA pixel size (for 128×28 FPA)	40.30 μm

of the enclosed energy. Once the genetic algorithm has computed the appropriate phase corrections, these corrections are applied. The resulting BER curves are computed using the optimal subset method of Vilnrotter and Srinivasan [15,16]. No attempt was made to directly minimize the number of pixels in the optimal subset, and this is reserved for future research. Since genetic algorithms are slow to converge, this approach is used only in the still atmosphere simulation. Correction in the presence of movement of the phase screen is considered only in the gradient descent case.

D. Genetic Algorithm Results

We begin with simulations that cover correction of only the first 100 Zernike coefficients in the expansion of the function $d(r, \theta)$. We begin by defining some important terms:

(1) Exact 100: In this case, we use perfect knowledge of the first 100 Zernike coefficients to remove all 100 of them from $d(r, \theta)$.

(2) GA 100: In this case, the genetic algorithm is used to estimate and correct the first 100 Zernike terms. Here, there were $N = 80$ test solutions per generation.

(3) GA 400: In this case, the genetic algorithm is used to estimate and correct the first 400 Zernike terms. Here, there were $N = 80$ test solutions per generation.

(4) GA 400 800: In this case, the genetic algorithm is used to estimate and correct the first 400 Zernike terms. Here, there were $N = 800$ test solutions per generation. Since this took approximately one week to run, only one BER curve was generated.

(5) No AO: This refers to the uncorrected case.

(6) Diffraction limited: This refers to the case of a perfectly planar wavefront truncated by the finite aperture.

(7) Guide Star AO – 11 Across: This refers to an AO system that uses a laser guide star and performs the best achievable correction with 11 actuators across a deformable mirror. The number of actuators across the deformable mirror is the main limitation here.

(8) Guide Star AO – 21 Across: This refers to an AO system that uses a laser guide star but now has a mirror with 21 actuators across the aperture.

(9) In all simulations, we assumed 3.14×10^{-2} background photons per diffraction-limited spot area per slot, an assumption corresponding to a spectral irradiance of 100 microwatts per square centimeter-nanometer-stearadian, which corresponds to sunlight scattering from optics [14].

(10) The error bars shown on the BER curves in Figs. 2 and 3 were generated in the following manner. A total of 10 random Kolmogorov phase screens were generated using the same parameters, and the genetic algorithm was used to correct wavefront errors from each of the 10 phase screens. The 10 phase screens were individually analyzed to determine the BER, resulting in a total of 10 BER curves. Each curve shown is the median BER curve from 10 trials, and the error bars are based on the lowest and highest BER values from 10 trials.

It should be noted that an AO system with K actuators across the deformable mirror is roughly equivalent to a system that achieves ideal correction of the first K^2 Zernike coefficients. For example, an AO system with a laser guide star and 11 actuators across can achieve performance comparable to exact correction of the first 100 to 120 Zernike coefficients. Similarly, an AO system with a laser guide star and 21 actuators across can achieve performance comparable to exact correction of the first 400 to 500 Zernike coefficients.

Figures 2 and 3 illustrate performance of the genetic algorithm for the 100- and 400-coefficient cases, respectively. We note that, since the genetic algorithm was designed to optimize the 80 percent enclosed energy metric instead of the size of the optimal subset metric, it does not yield the optimal subset for BER performance in this case.

An expanded analysis using 400 Zernike coefficients is shown in Figure 3. Performance does not reach the "Exact 400" curve in all cases, as shown by the median BER curve labeled "GA 400." This is most likely due to the size of the search space and the possible presence of local minima. The genetic algorithm requires a larger population and/or a longer search time to converge on the optimal solution for 400 coefficients. With 800 solutions per generation for the correction of 400 Zernike coefficients (the

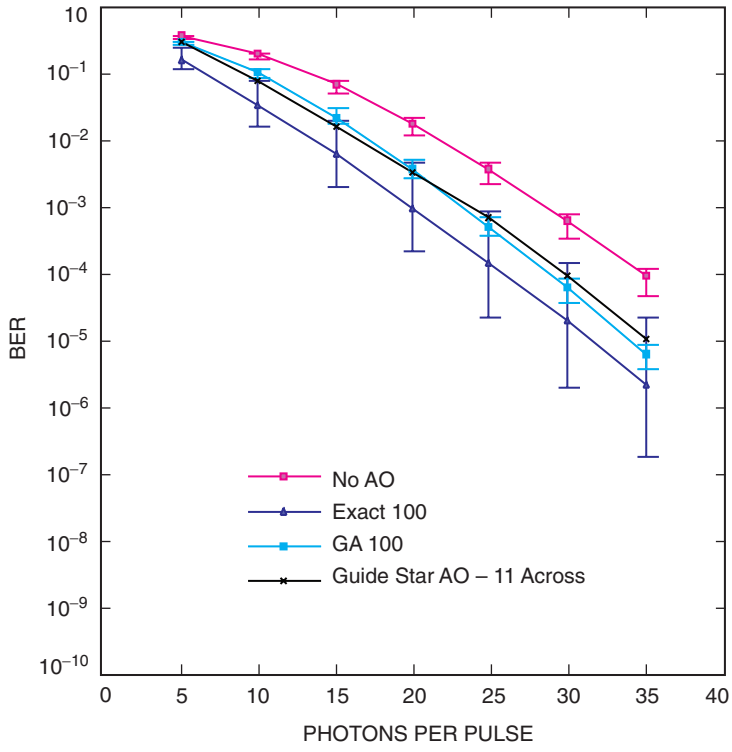

Fig. 2. BER curves for correction of 100 Zernike coefficients with no wind.

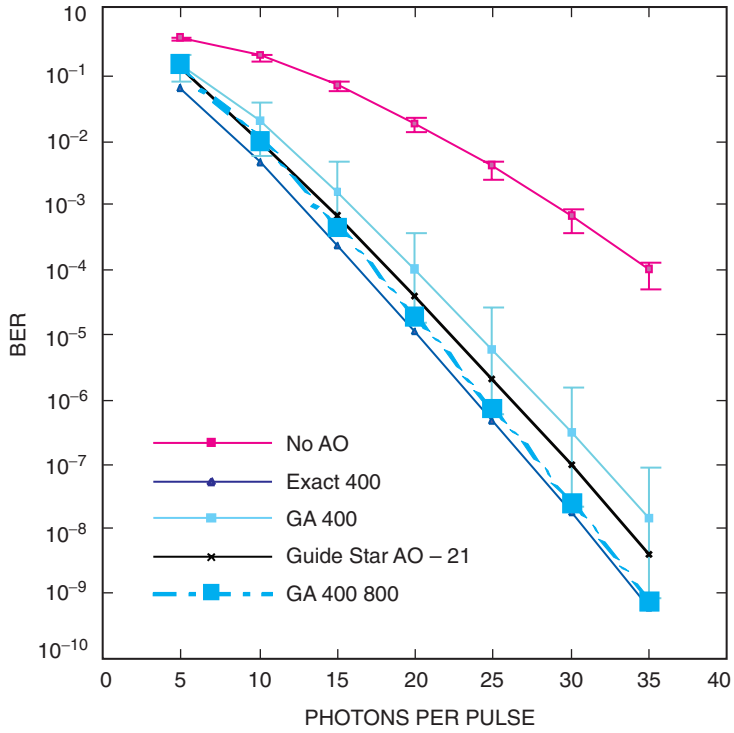

Fig. 3. BER curves for correction of 400 Zernike coefficients with no wind.

"GA 400 800" curve with $N = 800$), the GA achieved BER performance almost identical to "Exact 400," meaning that it converged to a nearly optimal solution. It also took over one week of computer time to generate a single BER curve for such a powerful GA. Using just 80 solutions per generation, we obtained the "GA 400" curve (with $N = 80$), which yields suboptimal performance. Due to the high complexity of this algorithm, the decision was made to use a parallel stochastic gradient search method instead [3]. The convergence time performance of the GA using 800 solutions per generation is shown in Fig. 4, and a quick comparison with convergence curves for the gradient descent algorithm of Section IV (Figs. 7 and 8) shows that, while gradient descent algorithms have promise, genetic algorithms do not. The genetic algorithm's performance with $N = 80$ is similar to that of an AO system with a guide star and 17 actuators across the deformable mirror, which is comparable to correcting the first 250 to 300 coefficients.

Although the genetic algorithm is not a candidate for implementation due to its slow convergence speed, illustrated in Fig. 4, the performance shown in Fig. 3 yields information on the problem at hand:

(1) The use of 800 solutions per generation rather than 80 leads to reliable convergence, which indicates that searching a greater portion of the search space improves the likelihood of convergence.

(2) The error bars for the GA 40 case, based on the minimum and maximum BER values, touch the GA 400 800 line and come very close to the Exact 400 line. This suggests that even with just 80 solutions per generation the algorithm may run into a global minimum in the cost function, which is the number of pixels containing 80 percent of the energy in the focal plane. However, the failure of the algorithm to consistently reach this very good level of performance would suggest that there may be large numbers of local minima in the search space, and a genetic algorithm would need a large number of solutions per generation in order to have a high probability of reaching the global minimum, which is the optimal solution.

(3) The error bars for the GA 400 case, indicating the difference between best-case and worst-case performance over 10 trials, are very wide, spanning 1 to 2 orders of magnitude in BER. This would suggest that there is considerable variation among the local minima encountered and that the quality of these minima may vary significantly.

(4) This suggests, but does *not* prove, that the search space is characterized by large numbers of local minima of highly variable quality. In most cases, a genetic algorithm with too few solutions per generation will tend to settle on these local minima, but a sufficiently broad search normally yields the global minimum.

We suspect, but have not proven, that the search space of this problem thus is one characterized by large numbers of local minima. This will have implications for the development of gradient descent algorithms and, in particular, for the performance of the parallel gradient descent method described by Carhart et al. [3–5], and these algorithms are likely to require modification to get out of local minima.

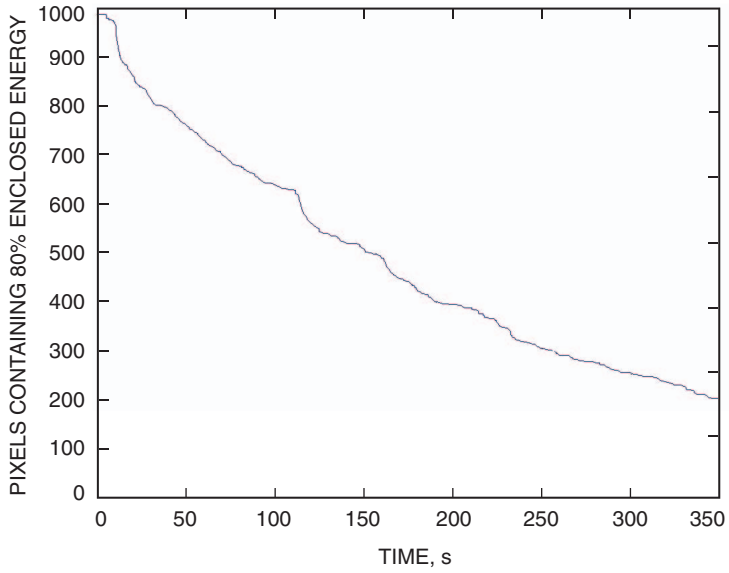

Fig. 4. Convergence sample from the GA 400 800 run assuming a 10-kHz update rate. The GA still has not converged after 350 seconds. Final convergence generally takes at least 500 seconds.

IV. Gradient Descent Method

The gradient descent algorithm is a faster method of converging to an optimized solution. This method has been used for improving image resolution and for communications link tests [3–5,9,10,17–20]. In this article, we use it with a focal-plane array detector to improve the performance of the optical communications link.

Given a vector of Zernike coefficients,

$$\vec{\sigma}_n(k) = \left[z_{n,1}(k)\ z_{n,2}(k)\ \cdots\ z_{n,M}(k) \right] \tag{5}$$

and a cost function,

$$f\left(\vec{\sigma}_n(k)\right) \tag{6}$$

we seek to estimate

$$\frac{\partial f\left(\vec{\sigma}_n(k)\right)}{\partial z_{n,m}} \quad m = 1, 2, \cdots, M \tag{7}$$

The estimates so computed are used to generate an estimate of the gradient vector:

$$\nabla_{\vec{\sigma}_n} f\left(\vec{\sigma}_n(k)\right) = \left[\frac{\partial f\left(\vec{\sigma}_n(k)\right)}{\partial z_{n,1}} \; \frac{\partial f\left(\vec{\sigma}_n(k)\right)}{\partial z_{n,2}} \; \cdots \; \frac{\partial f\left(\vec{\sigma}_n(k)\right)}{\partial z_{n,M}}\right] \tag{8}$$

Denoting our best estimate of the gradient vector by $\hat{\nabla}_{\vec{\sigma}_n} f\left(\vec{\sigma}_n(k)\right)$, we perform the following update:

$$\vec{\sigma}_n(k+1) = \vec{\sigma}_n(k) - \alpha \hat{\nabla}_{\vec{\sigma}_n} f\left(\vec{\sigma}_n(k)\right) \tag{9}$$

if $f\left(\vec{\sigma}_n(k)\right)$ is a cost function we seek to minimize. The parameter α is a learning rate between 0 and 1.

The gradient descent method of Eq. (9) requires us to estimate the derivative in Eq. (7). We can estimate the derivative using the following procedure:

(1) Perturb the mth coefficient $z_{n,m}$ by an amount $\Delta z_{n,m}$. Observe the change in the objective function $f\left(\vec{\sigma}_n(k)\right)$.

(2) Compute the quantity

$$\frac{\partial \hat{f}\left(\vec{\sigma}_n(k)\right)}{\partial z_{n,m}} = \frac{f\left(\vec{\sigma}_n(k)|_{z_m = z_{n,m} + \Delta z_{n,m}}\right) - f\left(\vec{\sigma}_n(k)|_{z_m = z_{n,m}}\right)}{\Delta z_{n,m}} \tag{10}$$

as our best estimate of the derivative.

(3) Repeat the above procedure for all M Zernike coefficients.

We see that computing a single estimate of the gradient vector $\nabla_{\vec{\sigma}_n} f\left(\vec{\sigma}_n(k)\right)$ requires a total of M perturbations of the deformable mirror or other adaptive optical element in the system. This is the primary performance limitation of this gradient descent method. Nevertheless, this does represent a form of directed search, which is still faster and more efficient than the genetic algorithm method.

The form of gradient descent given above, while conceptually simple, yields a convergence time proportional to M, the number of coefficients to correct. An even more efficient algorithm, the parallel gradient descent algorithm given by Carhart et al. in [3], achieves convergence times proportional to $\sqrt{M}$. This algorithm thus yields significantly faster convergence. It can be described as follows. Select a small positive number δ to be used as the size of the perturbation. Generate a random perturbation vector:

$$\Delta\vec{\sigma}_n(k) = \left[\Delta z_{n,1}(k) \; \Delta z_{n,2}(k) \; \cdots \; \Delta z_{n,M}(k)\right] \tag{11}$$

where

$$\Delta z_{n,m} = \pm\delta \tag{12}$$

with equal probability.

Denote the resulting change in the objective function by

$$\Delta f_n(k) = f\big(\overrightarrow{\sigma}_n(k) + \Delta\overrightarrow{\sigma}_n(k)\big) - f\big(\overrightarrow{\sigma}_n(k)\big) \tag{13}$$

Let μ denote a small, positive learning-rate parameter. The update equation for gradient-descent-based minimization is

$$\overrightarrow{\sigma}_n(k+1) = \overrightarrow{\sigma}_n(k) - \mu\Delta f_n \Delta\overrightarrow{\sigma}_n(k) \tag{14}$$

This algorithm was used by Carhart et al. for image sharpening [3], and the metric being optimized was one related to image quality. Here we are instead minimizing the number of pixels (FPA elements) receiving 80 percent of the enclosed energy.

Although the convergence time of the gradient descent algorithm is proportional to $\sqrt{M}$, it is also linearly proportional to the update rate η of the deformable mirror or other phase-correcting element. For example, if η is increased by a factor of 10, then the gradient algorithm can perform 10 times as many iterations per second, resulting in a factor of 10 reduction in convergence time. However, integration time is inversely proportional to the update rate, and there is a trade-off between input SNR to the gradient descent algorithm and the need for a reasonably fast update rate. Accordingly, this algorithm's performance in a real-world system is directly affected by the update rate of the phase-correcting element (i.e., deformable mirror or other element). A key advantage of gradient descent over genetic algorithms is computational simplicity; it is possible for a modern computer to reach update rates in the megahertz range, although present deformable mirror (DM) technology does not permit such rapid updates.

A. Simulation Model and Gradient Descent Results

The gradient descent algorithm is faster than the genetic algorithm, and it yields similar BER performance. In all simulations, we assumed 3.14×10^{-2} background photons per diffraction-limited spot. We begin by presenting BER curves for the correction of 100 Zernike coefficients. Here is a list of important terms:

(1) No AO: This refers to the uncorrected case.

(2) Exact M: This refers to the removal of the first M Zernike coefficients in the Zernike expansion of $d(r, \theta)$. For example, "Exact 100" refers to the complete removal of the first 100 Zernike coefficients in the expansion of $d(r, \theta)$.

(3) Grad M: In this case, the gradient descent algorithm is used to estimate and to correct the first M Zernike coefficients.

(4) Guide Star AO - K Across: This refers to an AO system that uses a Rayleigh laser guide star in conjunction with a deformable mirror with K actuators across.

(5) Diffraction Limited: This refers to the case of a perfectly planar wavefront truncated by the receiving telescope aperture.

In all simulations, we assumed 3.14×10^{-2} background photons per diffraction-limited spot area, an assumption corresponding to a spectral irradiance of 100 microwatts per square centimeter-nanometer-stearadian [14]. As in the genetic algorithm (GA) case, we also computed BER curves using the median

from 10 different Kolmogorov phase screen realizations, and the error bars reflect the lowest and highest BER values from these 10 realizations.

The results for 100 coefficients are illustrated in Fig. 5, and gradient descent yields performance similar to that of "Exact 100" correction.

For 400 coefficients, Fig. 6 shows that gradient descent without wind achieves performance similar to that of current AO with a laser guide star and just 13 actuators across the aperture, while the GA achieved performance comparable to guide star AO with 17 actuators across. As stated in the discussion of Section III.D on genetic algorithms, there is some evidence for the existence of many local minima in the search space, and there is a possibility the gradient descent algorithm is finding suboptimal local minima in many cases.

1. Gradient Descent Performance in the Presence of Wind. The gradient descent method's convergence, while swifter than that of the genetic algorithm, is too slow at a 10-kHz update rate but sufficiently fast at a 2-MHz update rate if a sufficient number of signal photons are available. An illustration of this effect is shown in Fig. 7 for 100 Zernike coefficients and in Fig. 8 for 20 Zernike coefficients. Using the frozen atmosphere model, the atmosphere changes significantly in just 0.01 s when moving at 7 m/s. The phases over any given aperture point are only correlated over 7 cm, so the algorithm must be capable of converging in a time less than 0.01 s if the atmosphere moves at 7 m/s. Observed turbulence-weighted daytime wind velocities are on the order of 30 m/s [8], resulting in a required convergence time of ∼2 ms. Let "convergence time" be defined as the time required for the gradient descent algorithm to bring the cost function to within 10 percent of the minimum value that will be achieved over the course of operation. The convergence time for a gradient descent algorithm correcting 100 coefficients is on the order of 0.05 s. Convergence is rapid enough only for about 20 to 50 coefficients with an update rate of 10 kHz.

In these simulations, 10-kHz and 2-MHz update rates were assumed. No noise is included in the analysis at this time. A trade study between update rates and SNR at the input to the gradient descent algorithm will be a topic of future study. The results shown here help to illustrate the properties of the gradient descent algorithm and suggest directions for future research. The results for a 10-kHz update rate and 1-m/s wind are presented in Fig. 9.

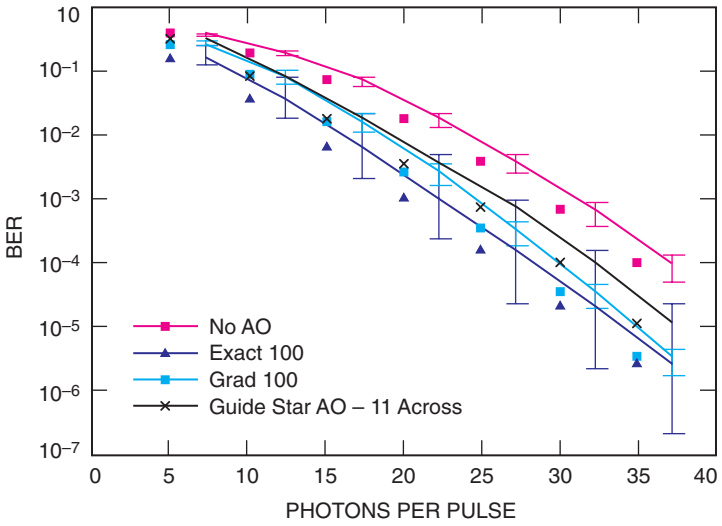

Fig. 5. BER curves for correction of 100 Zernike coefficients with no wind.

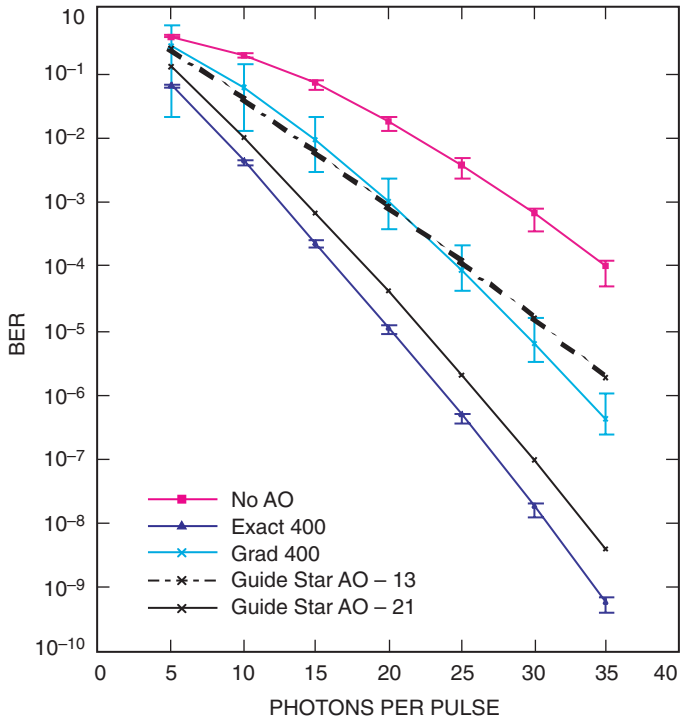

Fig. 6. BER curves for correction of 400 Zernike coefficients with no wind.

From Fig. 9, we note that some improvements in BER performance are achieved with a 10-kHz update rate in the presence of slow 1-m/s wind. We are using the modified parallel gradient descent algorithm of Carhart et al. [3], which has a convergence speed proportional to $\sqrt{M}$ instead of M, where M is the number of Zernike coefficients being corrected, yet convergence speed with a 10-kHz update rate is insufficient for 5-m/s wind, as will be shown in Fig. 10. For this reason, no attempts were made to correct more than 100 Zernike coefficients in the presence of wind during the course of this study. At 5-m/s wind speed, the present gradient algorithm yields a BER that is nearly the same as that of an uncorrected system for 100 coefficients. We state again that the gradient descent method used here seeks to minimize the number of pixels containing 80 percent of the enclosed energy, and the metric being minimized is not the optimal metric. There is a trade-off between correcting a large number of coefficients (i.e., 100), which makes convergence time in the presence of wind too slow, and correcting too few coefficients (i.e., 25), which means we are not correcting a sufficient number of coefficients to have a significant impact on BER.

With fewer coefficients to correct, the algorithm converges more rapidly. One difficulty in a dynamic atmosphere is converging rapidly enough to keep up with changes. The frozen atmosphere model used here, even with a wind speed of 5 m/s, poses a significant challenge to gradient-descent-based AO at 10 kHz, as shown in Fig. 11.

The question of performance with a faster update rate is partially addressed in this article. On one hand, a faster update rate has the potential to allow significantly faster convergence. On the other hand, as update rates increase, integration time over the focal-plane array decreases since update rates and integration times are inversely proportional. So while faster update rates offer the promise of better convergence in wind, there will be a definite trade-off between the faster update rates required by the gradient descent algorithm and the resulting shorter integration times and resultant loss of SNR at the input to the gradient descent algorithm. This is a topic for future research, and no such trade-off analysis has been carried out here. However, the potential of gradient descent algorithms

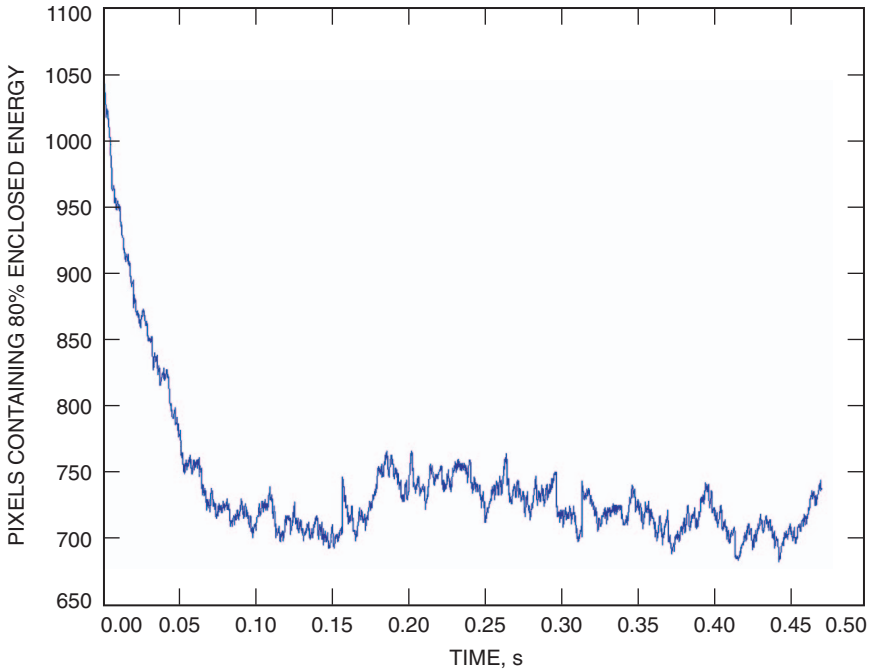

Fig. 7. Convergence of gradient descent algorithm for 100 coefficients.

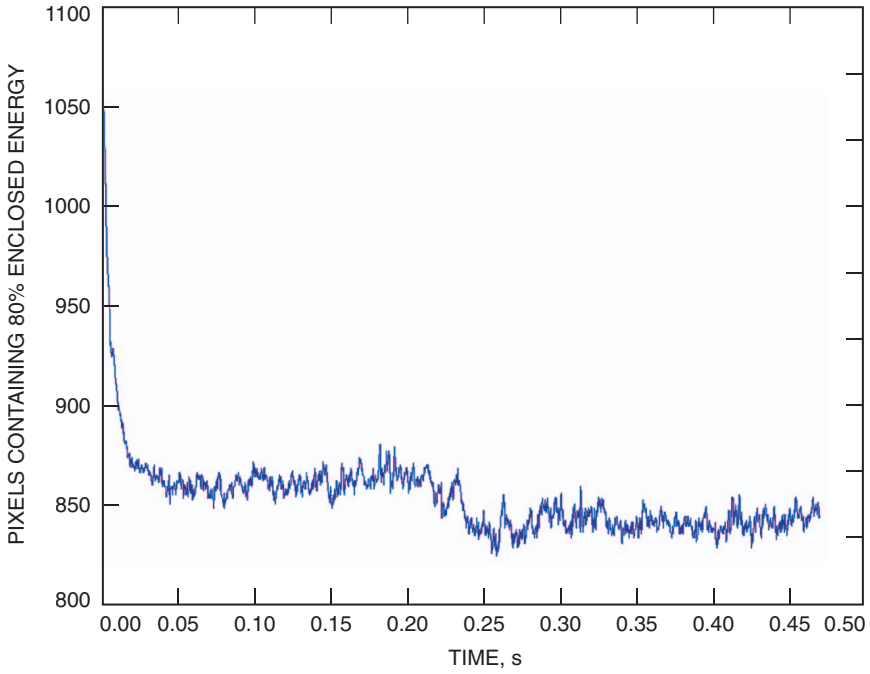

Fig. 8. Convergence of the gradient descent algorithm for 20 coefficients.

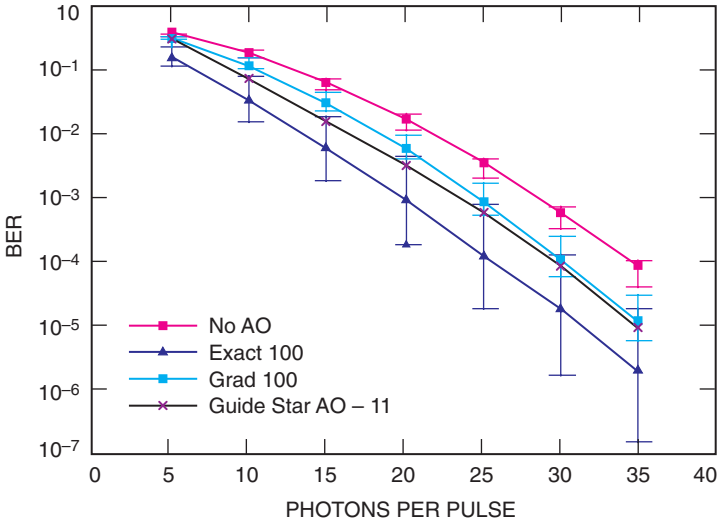

Fig. 9. Bit-error rate performance using the optimal subset method for gradient correction of 100 Zernike coefficients with 1-m/s wind.

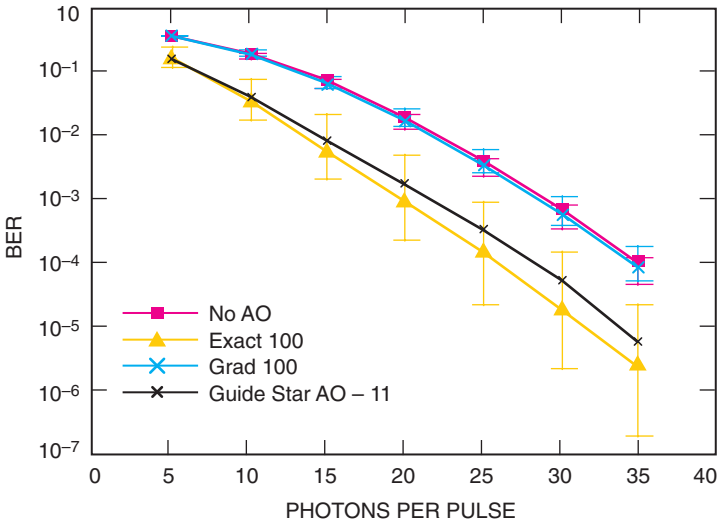

Fig. 10. Bit-error rate performance for gradient correction of 100 coefficients in 5-m/s wind with the optimal subset method.

is illustrated by the following. Suppose that the assumption of very high SNR at the input to the gradient descent algorithm continues to hold even at a 2-MHz update rate, and assume that the wind speed is now a far more realistic 30 m/s (for upper atmosphere winds) [8]. Under such assumptions, gradient descent converges, as shown in Figs. 12 and 13.

It is seen that while gradient descent with a 2-MHz update rate yields convergence performance similar to the static atmosphere case even in 30-m/s wind under the high SNR assumption, gradient descent completely fails to converge with a 10-kHz update rate. The effects of finite integration time and the resulting finite SNR must be taken into consideration in future work, however, and there will be a trade-off between update rates and integration times that has not been addressed here.

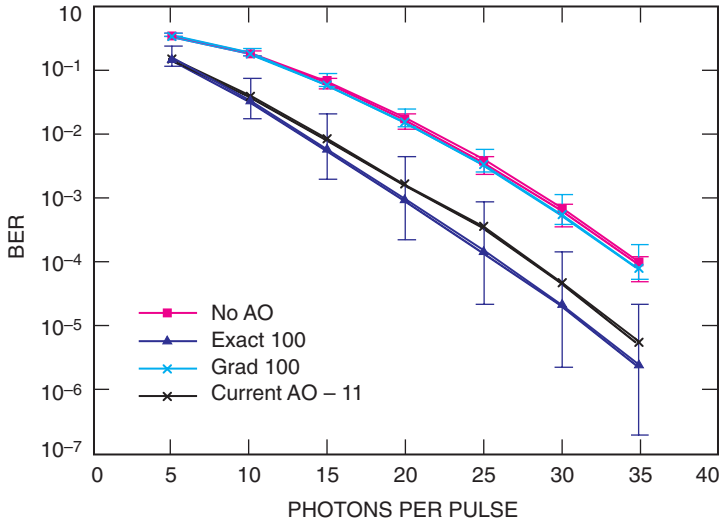

Fig. 11. Bit-error rate performance for gradient correction of 25 coefficients in 5-m/s wind with the optimal subset method.

The present algorithm assumes that a known set of coefficients is applied at each iteration, the phase element (the DM) is permitted to reach that state, and then a measurement of fitness is taken. However, if random perturbations were to be applied to the mirror at a rate much less than 2 MHz, and if sensors were to be used to sample the mirror's position at high accuracy at a 2-MHz update rate, then assuming that the mirror moves rapidly enough to produce significant changes from sample to sample, there could still be sufficient information available to enable faster convergence of the gradient descent algorithm. This is a topic for future research, but the ability to converge with 2-MHz updates under high SNR conditions indicates the need for a careful trade-off study, where the main trade-off occurs between an increased update rate and the need for sufficient integration time to allow reliable performance of the gradient descent algorithm.

V. Discussion

Both genetic algorithms and gradient descent methods can, in principle, extract information from a focal-plane intensity distribution and correct that distribution. The primary difficulty lies in the trade-off between updates rates and integration times. A fast phase-correction element operating in the megahertz range could enable tracking of a dynamic atmosphere, although increases in update rates and associated shorter integration times will result in less reliable gradient estimates. Work also needs to be done to determine whether sampling a continuously moving deformable mirror at a high rate may indeed provide adequate information to enable gradient descent to converge even without faster mirror responses, but the trade-off involving shorter integration times remains an open issue.

Another issue to be considered is that of local minima in the search space. The genetic algorithm results from Section III.D suggest, but do not prove, that the search space is characterized by large numbers of local minima. There are two approaches to this problem. In one approach, a different cost function may be chosen for minimization, and there is a possibility that a better cost function will yield fewer local minima. This does, however, seem unlikely given the highly non-linear nature of this optimization problem. In a second approach, there exist methods for improving upon basic gradient descent to keep it from getting stuck in local minima and to get it out of local minima. A large number of these methods, including the use of "momentum" and "fuzzy logic" controlled adaptation are discussed in the neural network literature, and some of these methods have been applied to back-propagation learning in neural networks [13]. Since the back-propagation learning problem is a non-linear optimization problem that

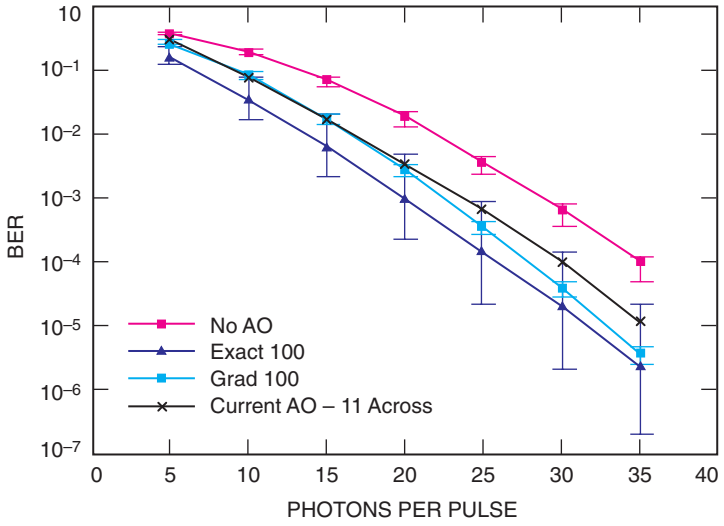

Fig. 12. Performance of gradient descent on 100 coefficients assuming 30-m/s wind and a 2-MHz update rate.

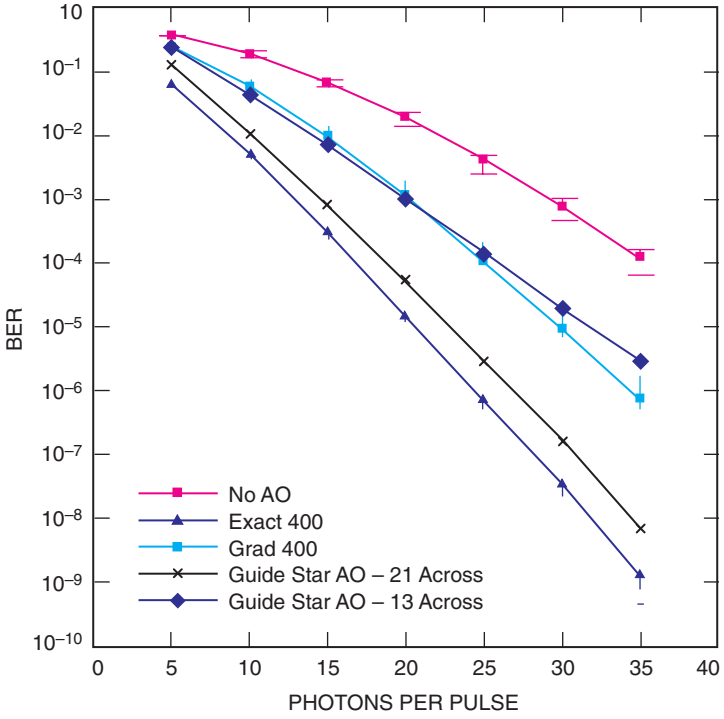

Fig. 13. Performance of gradient descent on 400 coefficients assuming 30-m/s wind and a 2-MHz update rate.

has some features in common with gradient-descent-based AO [3], it is likely that methods for escaping local minima from the neural network literature will prove useful in this case as well.

The intensity distribution is rich, although incomplete, in information on the phase of the optical signal in the pupil plane. Future algorithms for focal-plane-array-based AO may need to exploit this information in order to achieve improved performance. The development of such algorithms is a potential direction for future research. Such algorithms may augment genetic algorithms and gradient descent

methods by providing a better fitness function, or such algorithms may replace genetic algorithms and gradient descent altogether.

In summary, gradient descent algorithms have the potential to yield improvements in the performance of optical communication systems while avoiding the complexity of guide-star-based AO. More research is necessary to determine this, and several topics for future work have been identified.

References

[1] G. Bartlett, "Genie: A First GA," *Practical Handbook of Genetic Algorithms: Applications*, L. Chambers, ed., vol. I, Boca Raton, Florida: CRC Press, 1995.

[2] M. Born and E. Wolf, *Principles of Optics*, Cambridge, United Kingdom: Cambridge University Press, 1999.

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

[4] G. W. Carhart, G. J. Simer, and M. A. Vorontsov, "Adaptive Compensation of the Effects of Non-Stationary Thermal Blooming Based on the Stochastic Parallel Gradient Descent Optimization Method," *SPIE Proceedings*, San Diego, California, December 2003.

[5] G. W. Carhart, M. A. Vorontsov, M. Cohen, G. Cauwenberghs, and R. T. Edwards, "Adaptive Wavefront Correction Using a VLSI Implementation of the Parallel Gradient Descent Algorithm," *SPIE Proceedings*, Denver, Colorado, November 1999.

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

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

[8] J. Everaerts, N. Lewyckyj, and D. Fransaer, "Pegasus: Design of a Stratospheric Long Endurance UAV System for Remote Sensing," *Proceedings of the XXth ISPRS Congress*, Istanbul, Turkey, commission II, vol. XXXV, part B2, pp. 29–33, July 12–23, 2004.

[9] A. Givéon, N. J. Kasdin, R. J. Vanderbei, D. N. Spergel, M. G. Littman, and P. Gurfil, "Feasible Optimal Deformable Mirror Shaping Algorithm for High Contrast Imaging," *SPIE Proceedings*, San Diego, California, pp. 288–297, December 2003.

[10] A. Givéon, N. J. Kasdin, R. J. Vanderbei, D. N. Spergel, M. G. Littman, and P. Gurfil, "Stochastic Optimal Phase Retrieval Algorithm for High Contrast Imaging," *SPIE Proceedings*, San Diego, California, pp. 276–287, December 2003.

[11] J. W. Goodman, *Introduction to Fourier Optics*, 2nd. ed., Boston, Massachusetts: McGraw-Hill, 1996.

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

[13] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Upper Saddle River, New Jersey: Prentice-Hall, 1998.

[14] M. Srinivasan, V. Vilnrotter, M. Troy, and K. Wilson, "Adaptive Optics Communications Performance Analysis," *The Interplanetary Network Progress Report*, vol. 42-158, Jet Propulsion Laboratory, Pasadena, California, pp. 1–14, August 15, 2004. http://ipnpr.jpl.nasa.gov/progress_report/42-158/158B.pdf

[15] V. Vilnrotter and M. Srinivasan, "Adaptive Detector Arrays for Optical Communications Receivers," *The Telecommunications and Mission Operations Progress Report 42-141, January–March 2000*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–22, May 15, 2000.
http://tmo.jpl.nasa.gov/progress_report/42-141/141H.pdf

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

[17] M. A. Vorontsov, G. W. Carhart, and J. W. Gowens, "Target-in-the-Loop Adaptive Optics: Wavefront Control in Strong Speckle-Modulation Conditions," *SPIE Proceedings*, Seattle, Washington, November 2002.

[18] T. Weyrauch, M. A. Vorontsov, T. G. Bifano, and M. K. Giles, "Adaptive Optics System with Micromachined Mirror Array and Stochastic Gradient Descent Controller," *SPIE Proceedings*, San Diego, California, November 2000.

[19] T. Weyrauch, M. A. Vorontsov, T. G. Bifano, A. Tuanranont, V. M. Bright, J. R. Karpinsky, and J. A. Hammer, "Performance Evaluation of Micromachined Mirror Arrays for Adaptive Optics," *SPIE Proceedings*, San Diego, California, November 2000.

[20] T. Weyrauch, M. A. Vorontsov, and J. W. Gowens, "Adaptive Compensation of Atmospheric Effects with a High-Resolution Micro-Machined Deformable Mirror," *SPIE Proceedings*, Seattle, Washington, November 2002.