Stochastic parallel-gradient-descent technique for high-resolution wave-front phase-distortion correction

M. A. Vorontsov

U.S. Army Research Laboratory, Adelphi, Maryland 20783

V. P. Sivokon

New Mexico State University, Las Cruces, New Mexico 88003

Received January 20, 1998; revised manuscript received June 4, 1998; accepted June 11, 1998

A new optimization technique, stochastic parallel-gradient descent, is applied for high-resolution adaptive wave-front correction. A performance criterion for parallel-perturbation-based algorithms is introduced and applied to optimize adaptive system architecture. We present numerical simulation results for an adaptive imaging system based on the stochastic parallel-perturbation technique, along with experimental results obtained for a white-light adaptive imaging system with 37 control channels. An adaptive system with a self-organized (adaptive) control channel hierarchy is introduced and analyzed. © 1998 Optical Society of America [S0740-3232(98)01310-6]

OCIS codes: 010.0010, 010.1080, 110.0110.

1. INTRODUCTION

The recent rapid development of various tools for high-resolution wave-front phase-distortion correction (liquid-crystal phase modulators,[1,2] phase liquid-crystal television panels,[3] and microelectromechanical systems[4–6]) has introduced a new problem for adaptive optics: how to organize and control a system that has a large number N of control channels ($N \sim 10^3$–10^6). In addition to the basic scientific interest in solving this problem, there are several important applications that are stimulating the transition to high-resolution adaptive optics: airborne and ground-based adaptive laser systems, object identification and free-space communication, wave-front sensing, and adaptive imaging.

In this paper we address this challenging problem through the analysis of a new optimization technique known as stochastic parallel-perturbation gradient descent, applied here to high-resolution adaptive wave-front control. This technique was originally developed for model-free supervised learning in deterministic artificial neural networks[7–11] and only recently has been used for adaptive phase-distortion correction in white-light and active (laser-based) imaging systems.[12,13]

In Section 2 we present a brief synopsis of various optimization methods that can be applied for adaptive wave-front correction, with an emphasis on stochastic parallel-gradient-descent (parallel-perturbation) optimization algorithms. Application of this technique for adaptive wave-front control is discussed in Section 3. We introduce a performance criterion for parallel-perturbation-control algorithms by accounting for the statistical properties of wave-front aberrations. This criterion is applied in Section 4 to optimize an adaptive system architecture that is influenced by wave-front distortions corresponding to the Kolmogorov turbulence model. A new control algorithm—parallel perturbation with global coupling—is introduced and analyzed. Numerical simulation results for a high-resolution adaptive imaging system based on a stochastic parallel-perturbation technique, along with experimental results obtained for an adaptive imaging system that has a deformable mirror with 37 control channels, are presented in Section 5.

Convergence speed, the main concern for gradient-descent-based optimization techniques, can be increased by using an adaptive system with a self-organized (adaptive) control system architecture. In Section 6 we discuss an adaptive optical system that has a self-organized control channel hierarchy ("adaptive" adaptive imaging system). In this system, changes in the quality metric are reflected by corresponding changes to the control algorithm structure. Numerical simulations of an adaptive imaging system with a self-organized control structure demonstrate a property that is unusual for model-free optimization methods: the algorithm convergence speed did not change—at times it even increased—when the number of control channels N increased.

2. MODEL-FREE OPTIMIZATION TECHNIQUES: POTENTIAL FOR HIGH-RESOLUTION WAVE-FRONT CONTROL

A. Adaptive Systems Based on Direct System Performance Metric Optimization

The transition to high-resolution ($N \sim 10^3$–10^6) wave-front control is perhaps the most challenging problem for adaptive optics. Traditional approaches commonly used in low-resolution ($N \sim 10^2$) adaptive systems cannot automatically be extended to high-resolution adaptive op-

0740-3232/98/102745-14$15.00

tics. In adaptive systems based on wave-front information, this would require the same order-of-magnitude resolution in phase measurements—a difficult task even with modern phase analysis techniques.[14–17] Although for specific applications this problem can be overcome by using nonlinear phase conjugation[18,19] or control methods operating solely with intensity information,[20,21] it is still unclear how these techniques can be applied for adaptive systems based on direct system performance metric optimization, the most general type of adaptive optics system.

In an adaptive system based on direct system performance metric optimization, the control algorithm is independent of the system model ("model-free" or "blind" optimization techniques[7]). A measured quality metric $J = J(\mathbf{u})$ is a function of the control parameters $\mathbf{u} = \{u_1,\ldots, u_j,\ldots, u_N\}$, typically voltages applied to wave-front-corrector electrodes. To optimize J, various gradient-descent methods can be applied.[22] A gradient-descent algorithm essentially performs an incremental adjustment of control parameters u_j by means of a real-time estimation for the gradient components $J_j' = \partial J/\partial u_j$. The values J_j' may be estimated by applying small perturbations δu_j one at a time and measuring the corresponding changes δJ_j in the quality metric J.[23] In the control algorithm the ratios $\delta J_j/\delta u_j$ are used directly instead of the true gradient components J_j'. The critical issue for this sequential perturbation technique is adaptation speed. When the number of control channels N is increased, the optimization (adaptation) process convergence time t_c increases dramatically, at least as N^2.

Time losses caused by sequential perturbations can be reduced by use of the multidither technique, which can be regarded as a parallel analog implementation of the conventional gradient-descent method.[24] In this method, perturbations in the form of harmonic signals that have small amplitudes and different dithering frequencies are simultaneously applied to all wave-front-corrector electrodes. Estimates of the gradient components can be obtained by performing synchronous signal detection in each control channel. With this technique the increase in N requires a corresponding increase in the system frequency bandwidth, typically limited by system hardware (primarily by the frequency bandwidth of the wave-front corrector).[25] Among additional problems are the low signal-to-noise ratio in the gradient component measurements and the enormous increase in the system's electronic complexity for large N. Accordingly, the number of control channels for the multidither technique commonly does not exceed 10^2 (Ref. 25).

Other model-free optimization methods that we can refer to include various realizations of random descent, stimulated annealing, and genetic algorithms.[26–29] However, the main drawback of these algorithms that limits their potential application to high-resolution dynamic wave-front phase-distortion correction is their extremely slow divergence.

B. Stochastic Parallel-Gradient-Descent Optimization Technique

Among model-free optimization techniques that have recently appeared, stochastic parallel-gradient-descent optimization is perhaps the most promising for high-resolution adaptive optics. This algorithm performs gradient descent that on average has a maximum attainable average convergence time that is a factor of $N^{1/2}$ faster than for conventional gradient descent based on sequential perturbations.[9] The greatest potential benefit for adaptive optics applications may come from the fact that the parallel-perturbation technique is well suited for analog very-large-scale-integration implementation.[8–11]

The method is based on the rather simple and elegant idea described below.[7,9,10] Assume that small perturbations $\{\delta u_j\}$ $(j = 1,\ldots, N)$ are applied simultaneously (in parallel) to the system control parameters $\{u_j\}$. The resulting change in system performance quality metric $\delta J \equiv J(u_1 + \delta u_1,\ldots u_j + \delta u_j,\ldots,u_N + \delta u_N) - J(u_1,\ldots u_j,\ldots u_N)$ can be represented as the Taylor series expansion:

$$\delta J = \sum_{j=1}^{N} \frac{\partial J}{\partial u_j}\,\delta u_j + \frac{1}{2}\sum_{j,l}^{N} \frac{\partial^2 J}{\partial u_j \partial u_i}\,\delta u_j \delta u_i + \ldots. \tag{1}$$

Consider the products $\{\delta J\,\delta u_l\}$ calculated by use of the known perturbations $\{\delta u_l\}$ and the measured value δJ. From Eq. (1) we obtain

$$\delta J\,\delta u_l = \frac{\partial J}{\partial u_l}(\delta u_l)^2 + \psi_l, \tag{2}$$

where

$$\psi_l = \sum_{j\neq l}^{N} \frac{\partial J}{\partial u_j}\,\delta u_j \delta u_l + \frac{1}{2}\sum_{j,i}^{N} \frac{\partial^2 J}{\partial u_j \partial u_i}\,\delta u_j \delta u_i \delta u_l + \ldots.$$

The product $\delta J\,\delta u_l$ in Eq. (2) consists of two terms: the true gradient component J_l' and the noise term ψ_l. Assume that the perturbations $\{\delta u_l\}$ are random variables. For the expectations $\langle \delta J\,\delta u_l\rangle$ and $\langle \psi_l\rangle$ we have

$$\langle \delta J\,\delta u_l\rangle = \frac{\partial J}{\partial u_l}\langle(\delta u_l)^2\rangle + \langle \psi_l\rangle, \tag{3}$$

where

$$\langle \psi_l\rangle = \sum_{j\neq l}^{N} \frac{\partial J}{\partial u_j}\langle \delta u_j \delta u_l\rangle + \frac{1}{2}\sum_{j,i}^{N} \frac{\partial^2 J}{\partial u_j \partial u_i}\langle \delta u_j \delta u_i \delta u_l\rangle + \ldots. \tag{4}$$

In the parallel-perturbation gradient-descent method, the perturbations $\{\delta u_j\}$ are typically chosen as statistically independent variables having zero mean and equal variances: $\langle \delta u_j \delta u_i\rangle = \sigma^2 \delta_{ji}$ and $\langle \delta u_j\rangle = 0$, where δ_{ji} is the Kronecker symbol ($\delta_{ji} = 1$ for $j = i$ and 0 otherwise). The probability densities $\{p(\delta u_i)\}$ are assumed symmetric about their mean values. For statistically independent random variables, $\langle \delta u_j \delta u_i \delta u_l\rangle = 0$ for all j, i, and l. Note that $\langle(\delta u_l)^3\rangle = 0$ owing to the fact that it is an odd central moment of random variables that have a symmetric probability density.[30] As a result, instead of Eq. (3) we obtain $\langle \psi_l\rangle = O(\sigma^4)$, and hence expression (4) may be represented as $(1/\sigma^2)\langle \delta J\,\delta u_l\rangle = (\partial J/\partial u_l) + O(\sigma^2)$.

The random values $\{\delta J \delta u_l/\sigma^2\}$ (stochastic gradient components) obtained from a single perturbation, along with subsequent measurement of the quality-metric variation δJ, provide an estimation on average of the true gradient components $\{J_l'\}$ to an accuracy of $O(\sigma^2)$. This property of the stochastic gradient is the major motivation for replacing the true gradient components $\{J_l'\}$ in conventional gradient-descent algorithms with the products $\{\delta J \delta u_l\}$.

Consider a standard gradient-descent iterative procedure for incremental update of the control parameters $\{u_l\}$ $(l = 0, 1,..., N)$: $u_l^{(m+1)} = u_l^{(m)} - \mu(\partial J/\partial u_l)$, $m = 0, 1,\ldots$, where m is the iteration number and μ is a small constant (positive for the case of quality-metric minimization). A similar iterative procedure based on the stochastic gradient components can be written as

$$u_l^{(m+1)} = u_l^{(m)} - \mu \delta J^{(m)} \delta u_l^{(m)}, \qquad l = 0, 1,..., N. \tag{5}$$

Now consider a variation of the quality metric $\Delta J = J(\mathbf{u}^{m+1}) - J(\mathbf{u}^m)$ resulting from iterative procedure (5). From Eqs. (3), (4), and (5) we have, respectively,

$$\begin{aligned}\Delta J &\cong \sum_{i=1}^{N} \frac{\partial J}{\partial u_i}(-\mu \delta J \delta u_i) \\ &= -\mu \sum_{i=1}^{N} \frac{\partial J}{\partial u_i}\left[\frac{\partial J}{\partial u_i}(\delta u_i)^2 + \psi_i\right] \\ &\cong -\mu \sum_{i=1}^{N}\left(\frac{\partial J}{\partial u_i}\delta u_i\right)^2 - \mu \sum_{i,j\neq i}^{N} \frac{\partial J}{\partial u_i}\frac{\partial J}{\partial u_j}\delta u_i \delta u_j .\end{aligned} \tag{6}$$

For statistically independent perturbations, averaging, expression (6) gives

$$\langle \Delta J \rangle = -\mu\sigma^2 \sum_{i=1}^{N}\left(\frac{\partial J}{\partial u_i}\right)^2 + O(\mu\sigma^4). \tag{7}$$

For small $\mu\sigma$ the expectation $\langle \Delta J \rangle$ in Eq. (7) is negative. This means that algorithm (5) provides a decrease in the quality metric on average. Note that the last term in expression (6) consists of a sum of statistically independent random summands and for large N approaches zero for even a single perturbation (without averaging over an ensemble of perturbations). For this reason even a single iterative step typically leads to a decrease in the quality metric: $\Delta J < 0$. The perturbations $\{\delta u_j\}$ in algorithm (5) may not be truly statistically independent. This statistical dependence may result in a decrease in the number of effective (actually operating) control channels. As shown in Section 4, this feature can be used to speed up algorithm convergence.

3. PHASE-DISTORTION CORRECTION BASED ON PARALLEL-GRADIENT-DESCENT OPTIMIZATION

A. Problem Statement

Consider an adaptive optical system for phase-distortion compensation based on parallel-gradient-descent optimization of a quality metric $J = J[\phi(\mathbf{r})]$. The residual phase $\phi(\mathbf{r}) = \varphi(\mathbf{r}) + u(\mathbf{r})$, wave-front distortion $\varphi(\mathbf{r})$, and phase $u(\mathbf{r})$ introduced by the wave-front corrector are continuous functions ($\mathbf{r} = \{x, y\}$ is a vector in the plane orthogonal to the optical axes). Assume that J approaches an absolute minimum value $J = J_0$ at the point $\phi = 0$.

In contrast with neural network applications of the parallel-perturbation technique, here the quality metric J is a functional that is dependent on the continuously distributed function $\phi(\mathbf{r})$. Hence in adaptive optics we deal with a rather different problem statement: model-free optimization for a system with spatially distributed control. This problem may be transformed into a conventional parameter optimization problem by accounting for the fact that the phase $u(\mathbf{r})$ depends on a limited number of parameters $\{u_j\}$ $(j = 1,..., N)$-actuator drive signals, or some combinations thereof. In this case $u(\mathbf{r})$ can be represented as the expansion

$$u(\mathbf{r}) = \sum_{j=1}^{N} u_j S_j(\mathbf{r}) \tag{8}$$

over the wave-front-corrector influence functions $\{S_j(\mathbf{r})\}$. Consequently, the quality metric is a function of N variables: $J = J(u_1,\ldots u_j,\ldots u_N)$.

In spite of this standard statement for the optimization problem, several important features distinguish parameter optimization in adaptive optics from that in neural network applications. Typically, the number of control parameters N (number of neurons) is a built-in characteristic of a particular neural network. In adaptive optics the number of control channels N is determined by the control system architecture and can be changed by combining wave-front-corrector electrodes or by introducing an additional wave-front corrector. The number of control channels N is strongly dependent on the choice of the influence functions $\{S_j(\mathbf{r})\}$—an additional tool for improving the performance of the control algorithm.

B. Perturbation Statistics

We now consider problems related to choice of the control system architecture, using as an example the quality metric

$$J = \frac{1}{s}\int \phi^2(\mathbf{r}) d^2\mathbf{r}, \tag{9}$$

where s is the aperture size of the wave-front corrector.

The quality (phase-error) metric (9) was selected primarily because of the simplicity and convenience in dealing directly with the residual phase. In this case we may derive analytical expressions for metric perturbation and analyze the efficiency of the stochastic gradient-descent algorithm. The phase-error metric (9) cannot be observed directly unless we measure the wave front; however, if wave-front aberrations are small, most quality metrics used in adaptive optics can be expressed in terms of the phase-error metric (9).[15,16,25]

We now introduce a small phase perturbation into the control parameters $\{\delta u_j\}$, which results in the wave-front change $\delta u(\mathbf{r}) = \Sigma_{j=1}^{N} \delta u_j S_j(\mathbf{r})$. The perturbation $\delta u(\mathbf{r})$

results in the quality-metric perturbation $\delta J = J[\phi + \delta u] - J[\phi]$ needed for calculation of the stochastic gradient components $\{\delta J \delta u_j\}$.

Convergence speed in conventional gradient-descent algorithms depends on the amplitudes of the gradient component $\{J_j'\}$. Typically, the larger the absolute value of $\{J_j'\}$, the steeper is the metric descent. A similar conclusion can be drawn for parallel gradient descent. In this case, instead of $\{J_j'\}$ we must consider the stochastic gradient components $\{\delta J \delta u_j\}$. Thus we may speed up stochastic parallel-gradient-descent convergence by maximizing the absolute value of the quality-metric perturbation expectation $\langle \delta J \rangle$. Here averaging is defined over an ensemble of random phase distortions $\varphi(\mathbf{r})$ and perturbations $\delta u(\mathbf{r})$. To exclude the dependence of $\langle \delta J \rangle$ on perturbation amplitude, we restrict the aperture-averaged perturbation value by

$$\frac{1}{s} \int \langle \delta u^2(\mathbf{r}) \rangle \mathrm{d}^2\mathbf{r} \leqslant P^2 \tag{10}$$

where $P \ll 1$ is a constant. For the expectation $\langle \delta J \rangle$ from Eq. (9) we have

$$\langle \delta J \rangle = \frac{1}{s} \left[\int \langle \delta u^2(\mathbf{r}) \rangle \mathrm{d}^2\mathbf{r} + 2 \int \langle \delta u(\mathbf{r}) \phi(\mathbf{r}) \rangle \mathrm{d}^2\mathbf{r} \right] \tag{11}$$

and, correspondingly,

$$\langle \delta J \rangle \leqslant \langle \delta J \rangle_{\max} = P^2 + \frac{2}{s} \max_{\delta(\mathbf{r})} \left| \int \langle \delta u(\mathbf{r}) \phi(\mathbf{r}) \rangle \mathrm{d}^2\mathbf{r} \right|. \tag{12}$$

As seen from relation (12), the maximum attainable expectation value $\langle \delta J \rangle_{\max}$ depends on the aperture-averaged cross-correlation coefficient $\eta = 1/s \int \langle \delta u(\mathbf{r}) \phi(\mathbf{r}) \rangle \, \mathrm{d}^2\mathbf{r}$. This formally confirms the following conclusions, which are somewhat obvious from a physical point of view:

(a) The applied perturbation realizations $\delta(\mathbf{r})$ should match the statistical properties of the residual phase distortion, with respect to obtaining maximum cross correlation between $\delta(\mathbf{r})$ and $\phi(\mathbf{r})$.

(b) To achieve the maximum convergence rate, statistical properties of the random perturbations $\delta(\mathbf{r})$ should be dependent on the adaptation process and follow the changes in the statistics of the residual phase distortions.

To estimate the maximum expectation value $\langle \delta J \rangle_{\max}$, we assume that the applied perturbations $\delta u(\mathbf{r})$ and the residual phase-distortion realizations $\phi(\mathbf{r})$ are statistically dependent random functions:

$$\delta u(\mathbf{r}) = \alpha \phi(\mathbf{r}),$$
$$\langle \delta u(\mathbf{r}) \phi(\mathbf{r}) \rangle = \alpha \langle \phi^2(\mathbf{r}) \rangle, \tag{13}$$

where $\alpha \ll 1$ is a small constant. In this case inequality (10) may be transformed into $(\alpha^2/s) \int \langle \phi^2(\mathbf{r}) \rangle \mathrm{d}^2\mathbf{r} \leqslant P^2$ or $\alpha \leqslant P/(\langle J \rangle)^{1/2}$, where $\langle J \rangle$ is the quality-metric expectation. For $\alpha = P/(\langle J \rangle)^{1/2}$ and $P \ll (\langle J \rangle)^{1/2}$, instead of relations (11) and (12) we obtain

$$\langle \delta J \rangle = \langle \delta J \rangle_{\max} \approx 2P(\langle J \rangle)^{1/2}. \tag{14}$$

The maximum attainable expectation value $\langle \delta J \rangle_{\max}$ is dependent on the current state of the adaptation process (quality metric $\langle J \rangle$) and the aperture-averaged perturbation amplitude P.

In practice, the expectation value $\langle \delta J \rangle_{\max}$ in Eq. (14) cannot be achieved. The primary problem is the requirement for complete statistical dependence [Eqs. (13)] between $\delta u(\mathbf{r})$ and $\phi(\mathbf{r})$, which cannot be realized without knowledge of the residual phase $\phi(\mathbf{r})$. In addition, wavefront correctors have a limited spatial-frequency bandwidth and cannot provide perturbations in the spectral domain beyond their operational bandwidth. Nevertheless, an estimate of the expectation value $\langle \delta J \rangle_{\max}$ can be used for comparative analysis and for optimization of control system architecture.

Consider the perturbations $\delta u(\mathbf{r})$ formed by a wavefront corrector with a set of N orthogonal influence functions $\{S_j(\mathbf{r})\}$:

$$\frac{1}{s} \int S_j(\mathbf{r}) S_i(\mathbf{r}) \mathrm{d}^2\mathbf{r} = \delta_{ji}, \qquad \delta u(\mathbf{r}) = \sum_{j=1}^{N} \delta u_j S_j(\mathbf{r}). \tag{15}$$

Using conditions (15), we can represent the cross-correlation term in relations (11) and (12) in the form

$$\eta_N \equiv \eta = \frac{1}{s} \int \langle \delta u(\mathbf{r}) \phi(\mathbf{r}) \rangle \mathrm{d}^2\mathbf{r} = \sum_{j=1}^{N} \langle b_j \delta u_j \rangle, \tag{16}$$

where $\{b_j\}$ are coefficients of the residual phase $\phi(\mathbf{r})$ decomposition in terms of $\{S_j(\mathbf{r})\}$:

$$\phi(\mathbf{r}) = \sum_{j=1}^{\infty} b_j S_j(\mathbf{r}), \qquad b_j = \frac{1}{s} \int \phi(\mathbf{r}) S_j(\mathbf{r}) \mathrm{d}^2\mathbf{r}.$$

The maximum value of η_N corresponds to random perturbations $\{\delta u_j\}$ correlated with the coefficients $\{b_j\}$: $\langle \delta u_j b_j \rangle = \alpha \langle b_j^2 \rangle$. In this case, from Eq. (16) we obtain

$$\eta_N = \alpha \sum_{j=1}^{N} \langle b_j^2 \rangle. \tag{17}$$

To fulfill the limitation on perturbation amplitude (10), we set $\alpha = \alpha_0$, where

$$\alpha_0 = P \left(\sum_{j=1}^{N} \langle b_j^2 \rangle \right)^{-1/2}. \tag{18}$$

In this case instead of Eqs. (17) and (11) we have, respectively,

$$\eta_N = P \left(\sum_{j=1}^{N} \langle b_j^2 \rangle \right)^{1/2},$$
$$\langle \delta J \rangle \approx 2\alpha_0 \sum_{j=1}^{N} \langle b_j^2 \rangle = 2P \left(\sum_{j=1}^{N} \langle b_j^2 \rangle \right)^{1/2}. \tag{19}$$

The goal of the optimization of control system architecture is to decrease the number of control channels N while inflicting little or no damage to adaptive system performance. From this point of view, the set of orthogonal functions $\{S_j(\mathbf{r})\}$ is optimal (in comparison with any other orthogonal set of functions) if for a fixed N the cross-

correlation coefficient η_N and hence the quality-metric perturbation $\langle \delta J \rangle$ in relations (19) attain the maximum.

4. PERTURBATION STATISTICS FOR THE KOLMOGOROV TURBULENCE MODEL

A. Modal Control: Uncompensated Phase Aberrations

With phase distortions corresponding to the Kolmogorov turbulence model, in the inertial subrange the power spectrum of phase fluctuations $\varphi(\mathbf{r})$ is expressed as

$$G(q) = (0.023/r_0^{5/3})q^{-11/3}, \tag{20}$$

where q is spatial frequency and r_0 is the Fried parameter.[31] We assume first that the residual phase $\phi(\mathbf{r})$ coincides with $\varphi(\mathbf{r})$, corresponding to the beginning of the adaptation process. On the basis of results given in Refs. 15 and 32, it can be shown that for fixed N the maximum value of the sum in relations (19) corresponds to the Karhunen–Loève expansion. Thus for Kolmogorov turbulence, the maximum average convergence speed of the parallel-gradient-descent algorithm can be achieved if Karhunen–Loève basis functions are chosen as the set of the influence functions. Statistics of the perturbations $\{\delta u_j\}$ are determined by the statistical properties of the expansion coefficients $\{b_j\}$. Thus the perturbations $\{\delta u_j\}$ should be Gaussian statistically independent random variables such as $\langle \delta u_j \rangle = 0$ and $\langle \delta u_j \delta u_i \rangle = \alpha_0^2 \sigma_j^2 \delta_{ji}$, where $\sigma_j^2 = \langle b_j^2 \rangle$ are the Karhunen–Loève expansion coefficient variances [parameter α_0 is defined in Eq. (18)]. The coefficients σ_j^2 are calculated in Ref. 32.

Zernike polynomials $\{Z_j(\mathbf{r})\}$ are more convenient for analysis, and are near optimum for relatively low-order aberrations.[33] Using Zernike polynomials, instead of relations (19) we have

$$\langle \delta J \rangle \equiv \langle \delta J \rangle_Z \approx 2P\left(\sum_{j=2}^{N} \delta J_j \right)^{1/2} = 2P\theta_N (D/r_0)^{5/6}, \tag{21}$$

where $\delta J_j = \langle b_j^2 \rangle = \beta_j^2 (D/r_0)^{5/3}$ and D is the wave-front-corrector diameter. Note that the coefficient b_1 corresponding to the aperture-averaged phase is omitted, because the averaged phase is not significant for most adaptive optics applications. The values $\{\delta J_j\}$ characterize average sensitivities of the quality metric with respect to the selected perturbation statistics and influence functions. The coefficient θ_N accounts for overall sensitivity of the quality metric to the introduced perturbations. The first nine coefficients β_j^2 and θ_N are given in Table 1. When $N > 10$, the following approximation for θ_N can be used: $\theta_N^2 \approx 1.0299 - 0.2944 N^{-\sqrt{3}/2}$ (Ref. 33).

Estimation (21) for the quality-metric perturbation $\langle \delta J \rangle$ is valid only for the control-parameter perturbations $\{\delta u_j\}$ determined by statistics of the random coefficients $\{b_j\}$. These optimal perturbations $\{\delta u_j\}$ are Gaussian statistically independent random variables with zero mean and variances proportional to $\{\beta_j^2\}$. Here we neglect the cross correlation of the coefficients $\{b_j\}$. Since low-order aberrations (tilts, defocus, astigmatism, etc.) have the largest coefficients β_j^2, the corresponding wave-front-perturbation components $\{\delta u_j Z_j(\mathbf{r})\}$ have the most significant impact on quality-metric perturbation. The result is that gradient-descent optimization occurs primarily through this relatively small number of active control channels. Owing to the introduced control-channel hierarchy, gradient descent is effectively dependent on a relatively small number ($M \ll N$) of active control channels, which results a potential increase in algorithm convergence speed.

B. Perturbation Statistics during the Adaptation Process

Assume that during the adaptation process the first M aberrations are compensated. Similarly to relations (19) and (21) for the expectation $\langle \delta J \rangle$, we obtain

$$\langle \delta J \rangle \approx 2 \sum_{j=M+1}^{N} \langle b_j \delta u_j \rangle = 2\alpha_0 \sum_{j=M+1}^{N} \langle b_j^2 \rangle = 2\alpha_0 \sum_{j=M+1}^{N} \beta_j^2 (D/r_0)^{5/3}. \tag{22}$$

Since perturbation statistics are assumed unchanged, the coefficient α_0 is still determined by Eq. (18). Substituting Eq. (18) into relation (22) yields

$$\langle \delta J \rangle \equiv \langle \delta J \rangle_M \approx 2P \sum_{j=M+1}^{N} \beta_j^2 \left(\sum_{j=2}^{N} \beta_j^2 \right)^{-1/2} (D/r_0)^{5/6} = 2P\nu_{N,M}(D/r_0)^{5/6},$$

where $\nu_{N,M} = \theta_{N,M}^2/\theta_N$. For $M > 10$ the approximation $\theta_{N,M}^2 \approx 0.2944(M^{-\sqrt{3}/2} - N^{-\sqrt{3}/2})$ can be used. The beginning of the adaptation process corresponds to $M = 1$. Coefficient $\nu_{N,M}$ as a function of the number of compensated aberrations M is shown in Fig. 1(a). The coefficient $\nu_{N,M}$ and hence the corresponding expectation value $\langle \delta J \rangle_M$ drop rapidly as M is increased: $\nu_{N,M} = \theta_N \approx 1$ for $M = 1$, and $\nu_{N,M} < 0.04$ for $M > 10$. This leads to a potential decrease in the amplitudes of the stochastic gradient components and slows the adaptation process.

We can improve on this situation by changing the statistics of the perturbations $\{\delta u_j\}$ during the adaptation process. Since the quality-metric expectation $\langle \delta J \rangle$ in relation (22) does not depend on the first M random variables δu_j ($j = 1,..., M$), we may exclude these perturbations. We introduce the following control-parameter perturbations: $\delta u_j = 0$ for $j = 1,..., M$ and all other per-

Table 1. Coefficients β_j^2 and θ_N for Kolmogorov Turbulence Model

	j/N								
Coefficient	2	3	4	5	6	7	8	9	10
β_j^2	0.448	0.448	0.0232	0.0232	0.0232	0.0062	0.0062	0.0062	0.0062
θ_N	0.669	0.946	0.958	0.970	0.982	0.985	0.9853	0.991	0.995

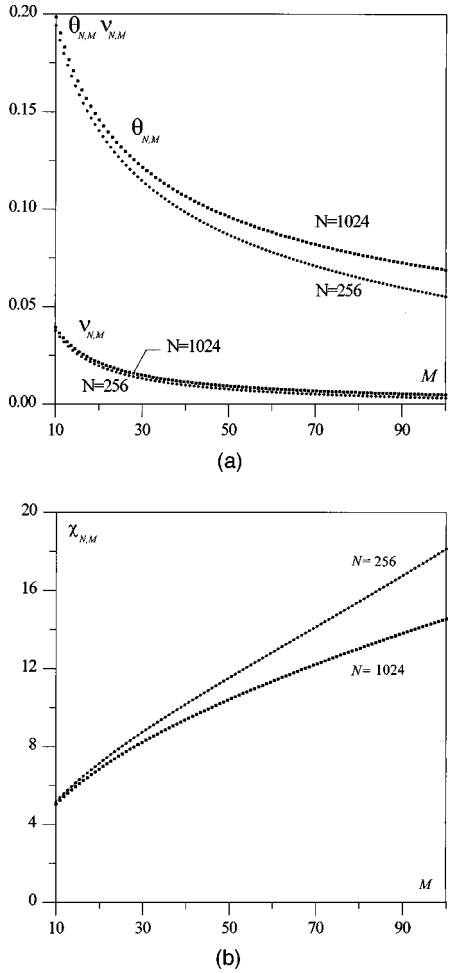

Fig. 1. (a) Coefficients $\theta_{N,M}$ and $\nu_{N,M}$ and (b) the ratio χ_M versus number of compensated aberrations M for different values N of the number of control channels.

turbations δu_j ($j = M + 1,..., N$) are statistically independent Gaussian random variables, such as

$$\langle \delta u_j \rangle = 0, \qquad \langle \delta u_j \delta u_i \rangle = \alpha_M^2 \langle b_j^2 \rangle \delta_{ji},$$

$$\text{where } \alpha_M = P\left(\sum_{j=M+1}^{N} \langle b_j^2 \rangle \right)^{-1/2}. \tag{23}$$

The expected value of the quality-metric perturbation then becomes

$$\langle \delta J \rangle \equiv \langle \delta J \rangle_0 \approx 2P\left(\sum_{j=M+1}^{N} \beta_j^2 \right)^{1/2} (D/r_0)^{5/6}$$

$$= 2P\theta_{N,M}(D/r_0)^{5/6}.$$

The coefficient $\theta_{N,M}$ as a function of M is shown in Fig. 1(a). The ratio $\chi_M = \langle \delta J \rangle_0 / \langle \delta J \rangle_M$ characterizes the potential benefit that can be achieved by using perturbations that depend only on residual aberrations. As Fig. 1(b) shows, ratio χ_M increases when the number of aberrations removed increases. Changes required in the perturbations are rather obvious from a physical point of view: with application of the perturbations in Eq. (23), information is injected primarily through channels corresponding to aberrations that have the maximum impact on the quality metric. In addition, the perturbation amplitudes grow ($\sim \alpha_M$) following a decrease in residual aberrations.

C. Pixelated Wave-Front Corrector: Piston-Type Wave-Front Control

To complete our analysis, we consider a case that is important for practical applications: that of a pixelated wave-front corrector that has a two-dimensional array of $N = n \times n$ elements of size a with controlled individual phase shifts. The influence functions $\{S_j(\mathbf{r})\}$ are stepwise functions centered at points $\mathbf{r}_j : S_j(\mathbf{r}) = S(\mathbf{r} - \mathbf{r}_j)$, where $S(r) = 1$ for $|x| \leq a/2$, $|y| \leq a/2$, and $S(r) = 0$ otherwise. Assume that the perturbations $\{\delta u_j\}$ are Gaussian statistically independent variables having zero mean and equal variances: $\langle \delta u_j^2 \rangle = \alpha_0^2 \sigma^2$, where σ^2 is the mean phase-distortion variance for a single element, independent of element location. The value of σ^2 was calculated in Ref. 34. For phase distortions with aperture-averaged mean phase over D and tilts removed, $\sigma^2 = 0.14(d/r_0)^{5/3}$ where d is the diameter of a small single aperture. Using this result for the pixelated wave-front corrector with $a = d$, we obtain the following estimate for the expectation $\langle \delta J \rangle$:

$$\langle \delta J \rangle \equiv \langle \delta J \rangle_P \approx 2P\sqrt{0.141}N^{-5/12}(D/r_0)^{5/6}.$$

When N increases, the expected value $\langle \delta J \rangle_P$ and hence the average stochastic gradient components decrease almost as $N^{-1/2}$. This may result in a potential $\sim N^{-1/2}$ decrease in the adaptation process convergence speed if we apply the standard parallel-gradient-descent procedure to drive the pixelated wave-front corrector. Note that a similar convergence rate was observed in neural network applications of the stochastic parallel-gradient-descent optimization technique.[9]

D. Parallel-Perturbation-Gradient-Descent Technique with Global Coupling

We consider two possible ways that the standard parallel-perturbation-gradient-descent technique as applied to pixelated wave-front-corrector control can be made faster. First, an additional low-resolution modal wave-front corrector can be used to separately compensate a relatively small number ($M \ll N$) of large-scale aberrations. Phase correction can then be represented in the form

$$u(\mathbf{r}) = \sum_{j=1}^{N} v_j S_j(\mathbf{r}) + \sum_{k=2}^{M} a_k W_k(\mathbf{r}),$$

where $\{W_k(\mathbf{r})\}$ and $\{a_k\}$ are the influence functions and the control signals of the modal wave-front corrector, respectively. The parallel-perturbation technique can be applied simultaneously to drive both wave-front correctors. In this case, one can expect two characteristic phases in the adaptation process: a relatively fast convergence corresponding to large-scale aberration correction with a convergence rate $\sim M^{1/2}$ and a slow process ($\sim N^{-1/2}$) related to residual small-scale aberration correction. From our preceding analysis, the maximum convergence speed can be achieved if the Zernike polynomials chosen are a set of the influence functions $\{W_k(\mathbf{r})\}$.

The second possibility for increasing parallel-gradient-descent convergence speed is based on an approximation of Zernike polynomials with use of stepwise functions of the pixelated corrector $S_j(\mathbf{r}) = S(\mathbf{r} - \mathbf{r}_j)$. An additional modal wave-front corrector is not required. We represent the phase introduced by the pixelated corrector as a sum of two components:

$$u(\mathbf{r}) = \sum_{j=1}^{N} v_j S_j(\mathbf{r}) + \sum_{k=2}^{M} a_k \hat{Z}_k(\mathbf{r}). \tag{24}$$

Here $\{\hat{Z}_k(\mathbf{r})\}$ are an approximation of the Zernike polynomials $\{Z_k(\mathbf{r})\}$:

$$\hat{Z}_k(\mathbf{r}) = \sum_{j=1}^{N} c_{k,j} S_j(\mathbf{r}),$$

$$c_{k,j} = \frac{1}{a^2} \int Z_k(\mathbf{r}') S_j(\mathbf{r}') \mathrm{d}^2\mathbf{r}',$$

$$k = 2,..., M. \tag{25}$$

The polynomial $Z_1(\mathbf{r})$ corresponding to the average phase is omitted. From Eqs. (24) and (25) we obtain

$$u(\mathbf{r}) = \sum_{j=1}^{N} u_j S_j(\mathbf{r}), \qquad u_j = v_j + \Sigma_{k=2}^{M} c_{kj} a_k, \tag{26}$$

which are actual drive signals applied to the pixelated wave-front-corrector electrodes. The variables $\{v_j\}$ and $\{a_k\}$ in Eq. (24) can be regarded as new control parameters on which the quality metric depends: $J = J(v_1,..., v_N, a_2,..., a_M)$. The small random perturbations $\{\delta v_j\}$ and $\{\delta a_k\}$ result in a corresponding change in the quality metric δJ. We can write the iterative algorithm as

$$v_j^{(m+1)} = v_j^{(m)} - \mu \delta J \delta v_j, \qquad j = 1,..., N,$$

$$a_k^{(m+1)} = a_k^{(m)} - \mu \delta J \delta a_k, \qquad k = 2,..., M. \tag{27}$$

The actual number of control elements is still N, but now each element participates in both local (through the variables $\{v_j\}$) and global (through the variables $\{a_k\}$) wave-front-shape control. We use the term "control algorithm with global coupling" to distinguish iterative procedure (27) from the conventional parallel-gradient-descent algorithm.

Consider statistically independent Gaussian random perturbations $\{\delta v_j\}$ and $\{\delta a_k\}$ such as

$$\langle \delta v_j \delta a_l \rangle = 0, \qquad \langle \delta v_j \delta v_l \rangle = \sigma^2 \delta_{jl},$$

$$\langle \delta a_j \delta a_i \rangle = \alpha_0^2 \langle a_j^2 \rangle \delta_{ji}, \tag{28}$$

where $\langle a_j^2 \rangle = \beta_j^2 (D/r_0)^{5/3}$ are the Zernike expansion coefficient variances [parameter α_0 is defined in Eq. (18)]. The actual driving signal's perturbations $\{\delta u_k\}$ corresponding to these perturbations are no longer statistical independent. From Eqs. (28) and (26) we have

$$\langle \delta u_j \delta u_l \rangle = \sigma^2 \delta_{jl} + \sum_{k=2}^{M} c_{kj} c_{kl} \langle a_k^2 \rangle$$

$$= \sigma^2 \delta_{jl} + \alpha_0^2 \left(\sum_{k=2}^{M} c_{kj} c_{kl} \beta_k^2 \right) (D/r_0)^{5/3}. \tag{29}$$

In principle, statistical dependence of perturbations is not forbidden in the parallel-gradient-descent optimization technique as long as the expectations $\langle \delta J \delta u_j \rangle$ in Eq. (3) can provide a correct estimation of the true gradient components $\{J_j'\}$. From Eqs. (3), (4), and (29),

$$\langle \delta J \delta u_j \rangle \cong \frac{\partial J}{\partial u_j} \left[\sigma^2 + \kappa \sum_{k=2}^{M} (c_{kj}^2 \langle \beta_k^2 \rangle) \right]$$

$$+ \kappa \sum_{l \neq j}^{N} \frac{\partial J}{\partial u_l} \sum_{k=2}^{M} (c_{kj} c_{kl} \langle \beta_k^2 \rangle), \tag{30}$$

where $\kappa = \alpha_0^2 (D/r_0)^{5/3}$. Calculations show that the first term in relation (30) is significantly larger than the noise term containing the sum of the coefficients with different signs. Accordingly, the parallel-perturbation gradient-descent technique can still be applied despite the fact that the true perturbations $\{\delta u_j\}$ are no longer statistically independent.

To implement the parallel-gradient-descent algorithm with global coupling, we can apply two approaches. First, iterative procedure (27) can be implemented directly by using the variables $\{v_j\}$ and $\{a_k\}$ and the statistically independent perturbations in Eqs. (28). In this case the driving signals $\{u_j\}$ should be determined separately from Eqs. (26). The second method is based on direct usage of the true variables $\{u_j\}$ (driving signals) along with standard iterative procedure (5). To provide global coupling, random perturbations $\{\delta u_k\}$ should be chosen as random statistically dependent variables with covariance matrix elements $\langle \delta u_j \delta u_l \rangle$ determined by expressions (29). Thus global coupling corresponding to the Zernike polynomial wave-front expansion can be realized with parallel-gradient-descent optimization only through a change in perturbation statistics.

5. ADAPTIVE IMAGING BASED ON THE PARALLEL-PERTURBATION TECHNIQUE

Our analysis of the stochastic parallel-gradient-descent technique is based on phase-error metric (9). Quality metric choice in actual adaptive systems is dependent on the type of adaptive system used. How robust are the results obtained in the preceding sections with respect to the choice of quality metric? To address this problem, in Sections 5 and 6 we consider an adaptive imaging system operating with a quality metric that is different from the phase-error metric examined here. The results demonstrate that optimal perturbation statistics and adaptive system architecture appear to be dependent not primarily on the choice of the quality metric used but on the phase-aberration statistics and the nature of the stochastic gradient-descent technique.

A. Numerical Model

The efficiency and performance of the parallel-gradient-descent optimization technique were analyzed through numerical simulation of an incoherent adaptive imaging system with a random aberration $\varphi(\mathbf{r})$ located in the exit pupil. The general optical scheme of the system is shown in Fig. 2. This system includes a camera, an image-quality analyzer that can either measure or calculate an image quality metric J, electronic components for hardware implementation of parallel-gradient-descent optimization, and a high-resolution wave-front corrector. In our calculations we used both a pixelated wave-front-corrector model (square array of $N = n \times n$ elements having piston-only-type influence functions as described in Subsection 4.C) and a wave-front-corrector model with Zernike-polynomial-type influence functions.

Calculations were performed on a 128×128 uniform square grid. To model the incoherent imaging system numerically, we applied a standard approach based on the optical transfer function for a system with aberrations.[15] The following image-quality metric was used for adaptive wave-front control[35]:

$$J = \int |F\{\exp[i\gamma I(\mathbf{r})]\}|^4 \mathrm{d}^2\mathbf{q}. \tag{31}$$

Here $I(\mathbf{r})$ is the normalized image-plane intensity distribution, γ is a parameter, and $F\{\cdot\}$ is the Fourier transform operator. The global minimum of J corresponds to an undistorted image.[35] Image-quality metric (31) can be measured in real time by means of an analog coherent optoelectronic processor.[36] To characterize the input $\varphi(\mathbf{r})$ and residual $\phi(\mathbf{r})$ phase aberrations, we used standard deviations σ_φ and σ_ϕ, and the corresponding values of the Strehl ratio St.

B. Parallel Perturbations with Global Coupling: Kolmogorov Turbulence Model

Numerical simulations were performed for an adaptive imaging system with phase distortions that had a Kolmogorov spatial power spectrum [Eq. (20)]. Turbulence strength was characterized by the ratio D/r_0. Calculations were for a phase-distortion realization corresponding to $D/r_0 = 10$ with aperture-averaged phase and global tilts removed. Figures 3(a) and 3(b) show gray-scale representations of the phase-distortion realization and the corresponding distorted image. The following stochastic parallel-gradient-descent procedures were analyzed:

(I) Conventional algorithm (5) with a pixelated wave-front corrector. A pixelated corrector model having $N = 16 \times 16$ elements was used in the calculations. For the perturbations $\{\delta u_j\}$ we used a standard model for artificial neural network applications of the parallel-gradient-descent technique: $\{\delta u_j\}$ are random statistically independent variables with fixed amplitudes ($|\delta u_j| = \sigma$) and random signs with equal probabilities for $\delta u_j = \sigma$ and $\delta u_j = -\sigma$ (binary perturbations).[9,10]

(II) Conventional algorithm (5) with modal wave-front control and Gaussian perturbation statistics. Calculations were performed for a modal wave-front corrector with Zernike polynomials $\{Z_j(r)\}$ ($j = 4,\ldots, N$) as the set of influence functions ($N = 36$). In wave-front representation (8) the aperture-averaged phase and tilts were removed ($u_1 = u_2 = u_3 = 0$); these aberrations have no impact on the quality of short-exposure images. Control parameter perturbations $\{\delta u_j\}$ were Gaussian statistically independent random variables with zero mean and variances $\langle \delta u_j^2 \rangle = \sigma_j^2$. Perturbation models with both equal variances ($\sigma_j^2 = \sigma^2$) and variances proportional to $\{\beta_j^2\}$ were analyzed.

(III) Parallel-perturbation algorithm with global coupling (27). As with algorithm (I), a pixelated corrector model with $N = 16 \times 16$ elements was used. In wave-front representation (24), Zernike polynomials $\{Z_k(\mathbf{r})\}$, $k = 4,\ldots, M$ ($M = 36$) were expanded over $N = 256$ piston-type influence functions of the pixelated corrector. Per-

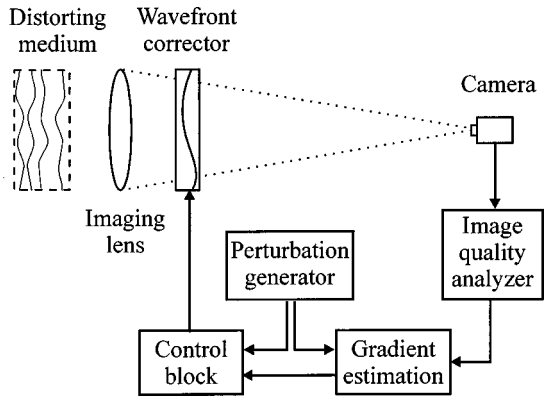

Fig. 2. General schematic for an adaptive imaging system based on image-quality-metric optimization with use of the parallel-perturbation technique.

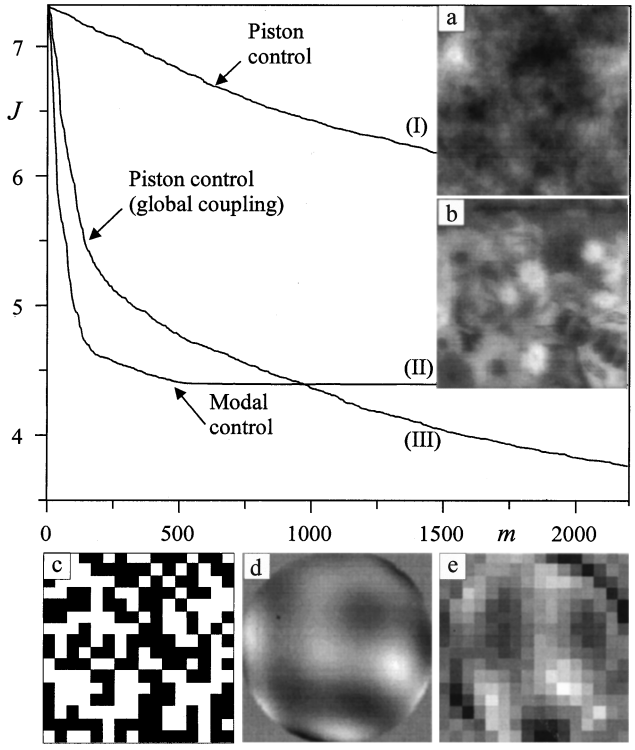

Fig. 3. Evolution curves (J versus iteration number m) for algorithms (I)–(III): (a) phase-distortion realization ($\sigma_\varphi = 1.8$ rad, St = 0.11), (b) distorted image, (c)–(e) random perturbation realizations corresponding to algorithms (I)–(III).

turbations $\{\delta v_j\}$ and $\{\delta a_k\}$ in Eqs. (27) were Gaussian statistically independent random variables with zero mean and variances $\langle \delta v_j^2 \rangle = \sigma_v^2$ and $\langle \delta a_k^2 \rangle = \sigma_a^2 (\sigma_a^2/\sigma_v^2 = 0.5)$.

A single iteration for the wave-front-correction process includes the following steps:

1. Measurement (calculation) of the image-quality metric corresponding to the current control-parameter values.
2. Wave-front perturbation with use of the generated control-parameter perturbations.
3. Measurement (calculation) of the image-quality metric corresponding to the perturbed wave front.
4. Calculation of the image-quality metric perturbation δJ and the corresponding stochastic gradient components.
5. Wave-front control-parameter update.

Examples of wave-front-perturbation realizations corresponding to algorithms (I)–(III), as well as adaptation curves for the algorithms under discussion, are shown in Fig. 3. The fastest convergence rate was observed with modal control [curve (II)] corresponding to the adaptive system with $N = 34$ control channels and perturbation variances $\langle \delta u_j^2 \rangle$ proportional to β_j^2. Using perturbations with equal variances $\langle \delta u_j^2 \rangle = \sigma^2$ slightly decreased the convergence speed. In both cases the stationary-state image-quality metric value ($J_0 \cong 4.3$) achieved was significantly large than the optimal value ($J_{\text{opt}} \cong 3.3$). This indicates that the low-resolution adaptive system was not able to compensate small-scale phase-aberration components. Transitioning to high-resolution wave-front control (pixelated wave-front corrector with $N = 256$ elements) slowed the convergence rate significantly, and an even larger value of $J_0 \cong 6.0$ was achieved after 2000 iterations [curve (I)]. Curve (III) in Fig. 3 corresponds to the parallel-perturbation algorithm with global coupling. Although the number of control channels was the same ($N = 256$), convergence of the adaptation process was significantly faster. As clearly seen in curve III of Fig. 3, there are two characteristic phases in the adaptation process: initial rapid convergence followed by a decrease in the convergence rate. This adaptation behavior was anticipated in Subsection 4.D.

Corrected images obtained after 200 and 2000 iterations are presented in Fig. 4. Relatively good-quality images were achieved after 200 iterations with use of modal control [Fig. 4(b)] and the algorithm with global coupling [Fig. 4(c)]. Succeeding iterations improved image quality only slightly for both the conventional algorithm [Fig. 4(d)] and the algorithm with global coupling [Fig. 4(f)]. The results presented in Figs. 3 and 4 are typical for moderate turbulence strength. When D/r_0 was increased ($D/r_0 \sim 20$), small-scale aberration components significantly degraded image quality and higher wave-front-corrector resolution was required ($N \sim 10^3$–10^4). This case is considered in Section 6.

C. Adaptive Imaging Based on the Parallel-Perturbation Technique: Experimental Results

The conventional parallel-gradient-descent algorithm with binary perturbations was recently applied to wave-front control in an adaptive imaging system with a pixelated 127-element liquid-crystal phase modulator.[12] Here we analyze the parallel-perturbation technique in an adaptive imaging system with a modal wave-front corrector (deformable mirror).

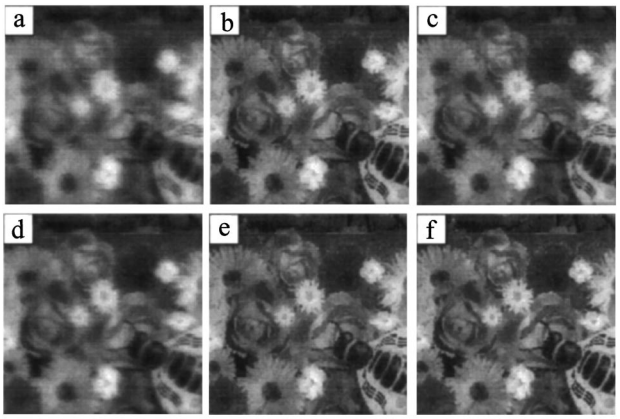

Fig. 4. Images obtained after (a)–(c) 200 and (d)–(f) 2000 iterations for algorithms (I)–(III): (a), (d) conventional stochastic parallel-gradient-descent algorithm (I); (b), (e) algorithm with modal control (II); (c), (f) algorithm with global coupling (III).

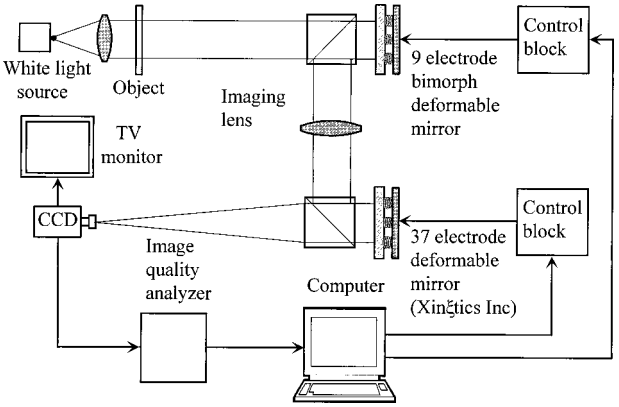

Fig. 5. Experimental setup for an adaptive imaging system based on parallel-gradient-descent optimization.

A schematic diagram of the experimental system is shown in Fig. 5. An object (gray-scale slice) was illuminated by a white-light source. Two different types of deformable mirrors were placed in the front and rear planes of the imaging lens (focal length 150 cm). A nine-electrode bimorph deformable mirror (diameter 40 mm) was used to generate phase distortions. The second deformable mirror (diameter 60 mm) from Xinetics Inc. had 37 control electrodes and Gaussian-type influence functions. This mirror was used to compensate the introduced phase aberrations. The object's image was registered by a CCD camera, with the corresponding video signal sent to an image-quality analyzer implemented as an analog coherent optical processor.[36] Thus the image-quality metric $J(t)$ was measured in real time without digital processing of the registered image-plane intensity distribution.

Parallel-perturbation algorithm (5) was implemented by use of a personal computer connected to both deformable mirrors. The computer provided for the following

signal processing: digitizing the input signal $J(t)$ proportional to image-quality metric (31), generating Gaussian random perturbations with zero mean and equal variances applied to the deformable-mirror electrodes, calculating the corresponding image-quality perturbation, and updating control voltages. In addition, the computer introduced phase aberrations by generating random signals that were applied to the bimorph mirror electrodes.

Results of system operation are shown in Fig. 6. At the beginning of the adaptation process ($n < n_0$), the system operated without phase distortions. At $n = n_0$, random voltages were applied to all nine electrodes of the bimorph mirror, causing phase aberrations and increasing the image-quality-metric value. The adaptive system reduced the image-quality metric to near its optimal value in approximately 100 iterations and significantly improved image quality (~100 ms were required for one iteration). At $n = n_1$ and $n = n_2$, new random voltages were applied to create different phase-distortion realizations. Examples of the distorted and the corrected images are shown in Figs. 6(a) and 6(b), respectively.

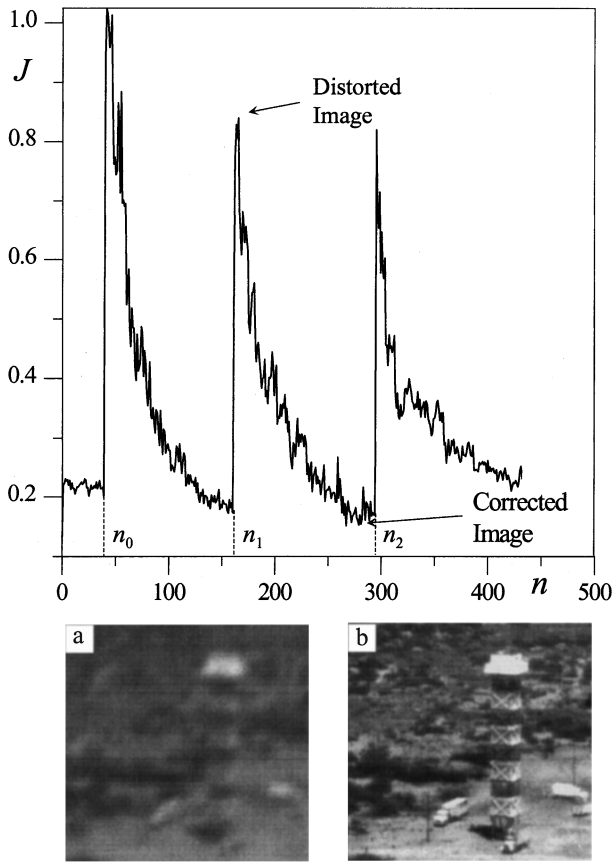

Fig. 6. Evolution curve of the adaptation process for a white-light adaptive imaging system. Points n_0, n_1, and n_2 indicate the moment at which phase distortions were introduced: (a) distorted image ($\sigma_\varphi \cong 2.2\pi$ rad) for $n = n_1$, (b) corrected image.

6. HIGH-RESOLUTION (SPATIALLY DISTRIBUTED) CONTROL AND "ADAPTIVE" ADAPTIVE OPTICS

A. Generalized Parallel-Perturbation Technique

In the preceding analysis of phase-distortion correction, we considered relatively weak and moderate turbulence ($D/r_0 \sim 10$). In this case, compensation of only large-scale phase-distortion components can be sufficient to provide substantial image-quality improvement. With strong turbulence, wave-front resolution on the order of $N \sim 10^4$–10^6 elements is required. For example, when $D/r_0 = 20$, more than $N = 400$ Zernike polynomials are needed to achieve the Strehl ratio $\mathrm{St} \cong 0.8$ (Refs. 16 and 32), corresponding to a pixelated wave-front corrector with $N \sim 10^4$ elements. Consider a generalization of the parallel-perturbation technique for an adaptive system with spatially distributed wave-front control—a limiting case of an adaptive system having a high-resolution pixelated corrector. Assume that $u^{(m)}(\mathbf{r})$ and $\delta u^{(m)}(\mathbf{r})$ are, correspondingly, the introduced phase and random perturbations at the mth step of the iterative procedure:

$$u^{(m+1)}(\mathbf{r}) = u^{(m)}(\mathbf{r}) - \mu \delta J^{(m)} \delta u^{(m)}(\mathbf{r}), \qquad m = 0, 1, \ldots, \quad (32)$$

where $\delta J^{(m)}$ is the quality-metric perturbation. Algorithm (32) is a generalization of parallel-perturbation algorithm (5) for a system with spatially distributed control. In accordance with the analysis in Subsection 3.B, the efficiency of the parallel-perturbation technique depends on how well spatial statistical properties of the applied perturbations $\delta u(\mathbf{r})$ and the residual phase distortions $\phi(\mathbf{r})$ are matched: aperture-averaged cross correlation between $\delta u(\mathbf{r})$ and $\phi(\mathbf{r})$ in relation (12) should approach its maximum value. Thus the generation of random phase screens [perturbations $\delta u(\mathbf{r})$] with controlled spatial statistical properties becomes an important issue for high-resolution adaptive optics based on the parallel-perturbation technique.

B. Perturbation Generation: Computational and Optical Approaches

For matching of the statistical properties of the phase distortions $\varphi(\mathbf{r})$ corresponding to the Kolmogorov turbulence model, the applied perturbations $\delta u(\mathbf{r})$ should represent random field realizations with the power spectrum given by expression (20). Random field realizations that have the required properties can be generated on the basis of known numerical algorithms and can be saved in a perturbation database. To introduce these perturbations into the actual adaptive system, one can use a high-resolution phase modulator connected to the database.[37] The phase modulator can be either separate or the same as that used for wave-front control.

The other (optical) approach can be based on the speckle technique described in Ref. 37. Consider the optical system shown in Fig. 7. A coherent wave with complex amplitude $A(\mathbf{r}) = I_0^{1/2}(\mathbf{r})\exp[i\psi(\mathbf{r})]$ is focused on a rough surface (frosted glass in Fig. 7), thus creating the speckle field realization $I_p(\mathbf{r})$ at the observation plane $z = L$ (plane of camera). The power spectrum $G_S(\mathbf{p})$ for the speckle-field intensity fluctuation is given by (see

Refs. 35 and 38) $G_S(\mathbf{p}) = \int I_S(\mathbf{r}_F) I_S(\mathbf{r}_F - L\mathbf{q}/k) \mathrm{d}^2\mathbf{r}_F$, where $k = 2\pi/\lambda$ and $I_S(\mathbf{r}_F)$ is the intensity distribution on the rough surface. Thus, to create a speckle-field realization $I_p(\mathbf{r})$ with the predetermined intensity fluctuation power spectrum, $G_S(\mathbf{p})$, we should obtain the corresponding intensity distribution $I_S(\mathbf{r}_F)$ on the rough surface. To accomplish this we can use a spatial light modulator (SLM) located in the back focal plane of the lens, as shown in Fig. 7. The SLM forms the input beam intensity $I_0(\mathbf{r})$ and/or phase $\psi(\mathbf{r})$ distributions to produce the required intensity distribution $I_S(\mathbf{r}_F)$ in the focal plane of the lens.

A change in the speckle field realization is brought about by moving the rough surface perpendicular to the direction of light propagation. This same result can be achieved by rotating the frosted glass. As shown in Fig. 7, the wave-front perturbations $\delta u(\mathbf{r})$ are obtained by transforming the speckle-field intensity pattern into its corresponding phase modulation with a high-resolution phase modulator (wave-front corrector) connected to a camera. The most important benefit of this speckle-based perturbation-generation technique is that perturbation statistics can be easily controlled by changing the intensity distribution $I_S(\mathbf{r}_F)$ during the adaptation process.

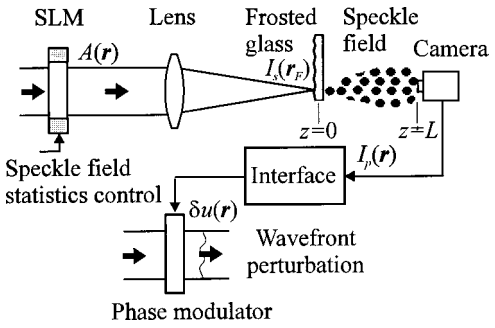

Fig. 7. Optical scheme for generation of wave-front perturbation.

C. "Adaptive" Adaptive Imaging with High-Resolution Wave-Front Control

A schematic illustration of our concept of a high-resolution adaptive imaging system with self-organized perturbation statistics is shown in Fig. 8. Wave-front control is based on the continuous stochastic parallel-gradient-descent algorithm model (32). Two identical high-resolution wave-front correctors are placed behind the imaging lens: one (C_u) to correct the incoming wave front and the second ($C_{\delta u}$) to introduce small wave-front perturbations $\delta u(\mathbf{r})$. The image-plane intensity $I(\mathbf{r})$ is registered by a camera connected to both the image-quality analyzer and the optical-perturbation generator's phase SLM. The phase SLM transforms the intensity modulation $I(\mathbf{r})$ into the phase modulation $\psi(\mathbf{r}) = \alpha_s I(\mathbf{r})$ of the optical-perturbation generator's input wave. Coefficient α_s characterizes the depth of the phase modulation.

After passing through the SLM, the coherent input wave with uniform intensity ($I_0 = \text{const.}$) and phase is transformed into a phase-modulated wave with amplitude $A(\mathbf{r}) = I_0^{1/2} \exp[i\alpha_s I(\mathbf{r})]$ (phase image). This wave is focused by the lens onto the rough surface, producing the intensity distribution $I_S(\mathbf{r}_F)$. The intensity distribution $I_S(\mathbf{r}_F)$ and hence the statistical properties of the speckle-field realizations $I_p(\mathbf{r})$ are dependent on the image-plane intensity $I(\mathbf{r})$, thus creating a feedback between image-quality and speckle-field realization statistics. The phase perturbation $\delta u(\mathbf{r})$ formed by the wave-front corrector $C_{\delta u}$ is proportional to the varying component of $I_p(\mathbf{r})$: $\delta u(\mathbf{r}) = \chi(I_p(\mathbf{r}) - \langle I_p \rangle)$, where χ is a coefficient and $\langle I_p \rangle$ is the aperture-averaged speckle-field realization intensity.

The physical basis for this perturbation-statistics control is rather clear. The better the quality of the registered image $I(\mathbf{r})$, the larger is the characteristic width b_F of the intensity distribution $I_S(\mathbf{r}_F)$ on the rough surface. Recall that $I_S(\mathbf{r}_F)$ is proportional to the phase image spatial spectrum. Correspondingly, at the beginning of the

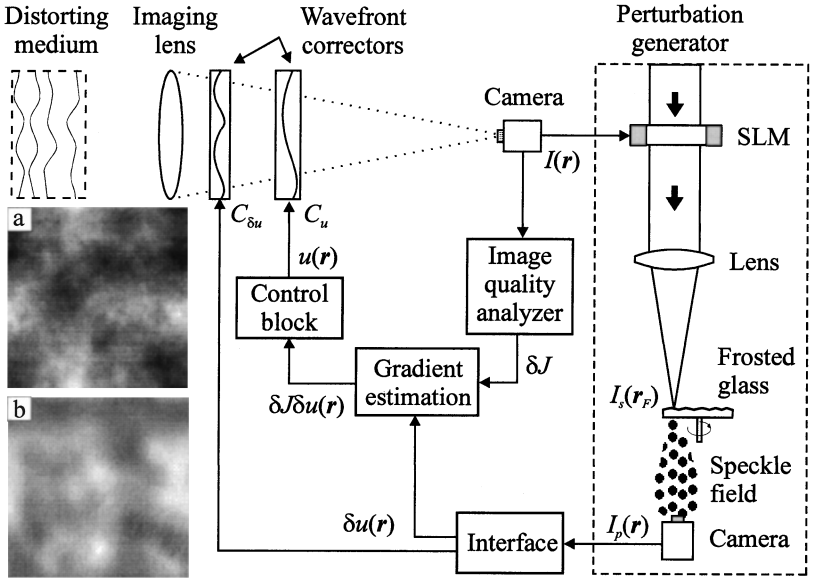

Fig. 8. Schematic for high-resolution adaptive imaging system with self-organized perturbation statistics: (a) phase-distortion realization for $D/r_0 = 20$ ($\sigma_\varphi \cong 3.5$ rad, St = 0.002), (b) distorted image.

adaptation process when an image is highly distorted, the characteristic speckle size $a_S \sim \lambda L/b_F$ is large, which results in large-scale perturbations. When image quality improves, the intensity distribution $I_S(\mathbf{r}_F)$ becomes wider and correspondingly the characteristic speckle and perturbation size becomes smaller.

D. Adaptive System with Self-Organized Perturbation Statistics: Numerical Analysis

In numerical simulations of the previously described high-resolution adaptive imaging system with self-organized perturbation statistics, we used phase aberrations $\varphi(\mathbf{r})$ corresponding to the Kolmogorov turbulence power spectrum model with $D/r_0 = 20$ and overall tilts removed. A typical phase-distortion realization and corresponding distorted image are shown in Figs. 8(a) and 8(b). Calculations were performed for the following control algorithms:

(I) Conventional algorithm (5) with a pixelated corrector having $N = 32 \times 32$ elements (binary perturbations).

(II) Parallel-perturbation algorithm with global coupling (27) with a pixelated corrector having $N = 32 \times 32$ elements and Zernike polynomial expansion $\{Z_k(\mathbf{r})\}$, $k = 4,\ldots, M$ ($M = 36$). Note that algorithms (I) and (II) failed (image quality did not improve noticeably during the first 1000 iterations) when a pixelated corrector having $N \geqslant 64 \times 64$ elements was used.

(III) Parallel-perturbation algorithm (32) for a system with high-resolution (spatially distributed) control and perturbations $\delta u(\mathbf{r})$ statistically matched with phase-aberration realizations $\varphi(\mathbf{r})$. The Kolmogorov turbulence model was used to create both $\delta u(\mathbf{r})$ and $\varphi(\mathbf{r})$ random fields. The adaptive imaging system had a pixelated wave-front corrector with $N = 128 \times 128$ pixels.

(IV) Parallel-perturbation algorithm (32) for a system with high-resolution control ($N = 128 \times 128$ control channels) and self-organized speckle-based perturbations as described in Subsection 6.C.

Evolution of the average values $\langle J \rangle$ of the image-quality metric for these algorithms is shown in Fig. 9. Average (long-exposure) images corresponding to $m = 500$ iterations and $m = 1000$ iterations are presented in Fig. 10. The average image without correction is shown in Fig. 9(a). Over 200 realizations of the random phase screens $\varphi(\mathbf{r})$ were used for statistical averaging.

The fastest convergence rate occurred with stochastic parallel-gradient descent with use of self-organized speckle-based perturbations [curve (IV)], corresponding to an adaptive system with $N = 128 \times 128$ control channels. Examples of perturbation-realization change during the adaptation process are shown in Fig. 11. The perturbation realization $\delta u(\mathbf{r})$ characteristic size d_s was estimated by using the effective width b_u of the spatial spectrum $G_p(\mathbf{q}) = |F\{\delta u(\mathbf{r})\}|$: $d_s = 2\pi/b_u$, where $b_u = \int |\mathbf{q}| G_p(\mathbf{q}) \mathrm{d}^2\mathbf{q} / \int G_p(\mathbf{q}) \mathrm{d}^2\mathbf{q}$. The value d_s is dependent on the phase modulation $\psi(\mathbf{r}) = \alpha_s I(\mathbf{r})$ of the optical perturbation generator's input wave and hence on image quality.

Evolution of d_s during the adaptation process for different coefficients α_s is shown in Fig. 11. The highest convergence speed was observed for $\alpha_s \cong 2\pi$ (curve I), which

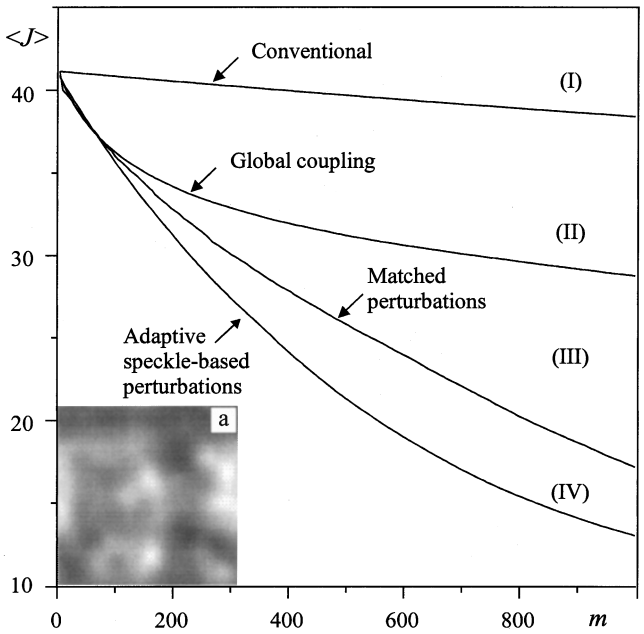

Fig. 9. Evolution curves of the adaptation process for image-quality-metric average values $\langle J \rangle$: (I) conventional stochastic parallel-gradient descent, algorithm (II) algorithm with global coupling, (III) algorithm with perturbations having Kolmogorov statistics ($D/r_0 = 20$), (IV) algorithm with self-organized speckle-based perturbations. (a) Average distorted image.

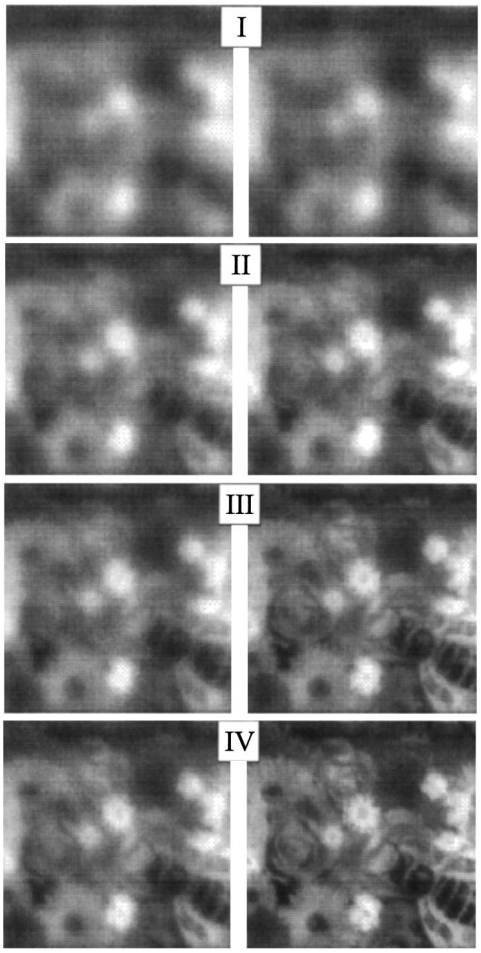

Fig. 10. Average images obtained after 500 (left row) and 2000 iterations (right row) for algorithms (I)–(IV).

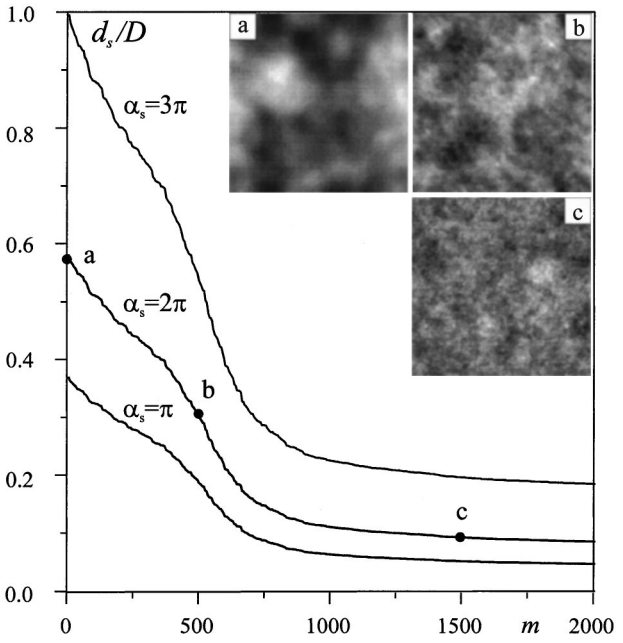

Fig. 11. Evolution of perturbation characteristic size d_s during the adaptation process for different coefficients α_s. Gray-scale images of the perturbation realizations corresponding to $\alpha_s = 2\pi$ for (a) $m = 0$, (b) $m = 500$, and (c) $m = 1500$.

corresponds to a phase modulation with maximum value $\max|\psi(\mathbf{r})| \cong 2\pi$. At the beginning of the adaptation process when large-scale aberrations were dominant, the perturbation-realization spatial scale d_s was approximately $0.6D$. During the adaptation process, large-scale aberration components were compensated, resulting in a significant decrease in the perturbation spatial scale ($d_s \cong 0.025D$ for $m = 1000$). Despite the fact that for the speckle-based technique the perturbations $\delta u(\mathbf{r})$ and residual phase distortions $\phi(\mathbf{r})$ were not statistically matched, we obtained a higher convergence rate for the algorithm with self-organized perturbation statistics than for the algorithm with fixed matched perturbation statistics (compare curves IV and III in Fig. 9). Comparison of curves IV and III shows a property that is extraordinary for the model-free optimization technique: the high-resolution "adaptive" adaptive system with $N = 128 \times 128$ control channels demonstrated a significantly higher convergence rate than the adaptive system in which 16 times fewer control channels were used with the conventional algorithm.

Numerical simulation results for adaptive phase-distortion compensation were obtained only for unchanging (frozen) aberrations corresponding to relatively strong turbulence ($D/r_0 \sim 20$). With high-resolution (spatially distributed) wave-front correction, approximately 300–500 iterations were required for significant image-quality improvement. For achievement of adaptation process convergence during a time interval less than the characteristic time of atmospheric phase-distortion change, the entire adaptive system should operate with a speed of the order of 30–50 kHz. This suggests that even with state-of-the-art phase modulators, this proposed technique may not be effective in the real atmosphere. However, the situation is not so unfavorable for this wave-front control technique as may appear from this estimation. First, there is still room for improvement of algorithm convergence. Second, atmospheric phase distortions are not statistically independent phase screens. Owing to the temporal correlation of wave-front aberrations, even partial compensation will decrease the residual phase amplitude and hence speed up algorithm convergence. Third, rapid progress in the development of fast optical microelectromechanical systems and quantum-well-based phase modulators, together with the coming merger of optical and electronic components on a microchip, keeps alive the hope that the required adaptive system resolution and operational speed might be achieved in a few years. And finally, atmospheric distortions are not the only distortions that adaptive optics are able to compensate. There are a number of applications in which phase distortions are either time independent or at least not so fast as the atmospheric ones.

7. CONCLUSION

Fast, inexpensive, and high-resolution wave-front correction—the ultimate adaptive optics goal—cannot be achieved unless rather complicated control problems are addressed. For adaptive systems with a large number N of control parameters, these control problems constitute the main obstacle to applying model-free optimization techniques to adaptive optics applications. With the recent appearance of the stochastic parallel-gradient-descent algorithms, the speed of optimization process convergence, the main drawback of model-free optimization techniques, can be significantly increased. Still, even with very favorable estimation, convergence time increases, at least as a factor of $N^{1/2}$.

The results presented here demonstrate that by accounting for the statistical properties of wave-front-aberration, one can significantly increase the convergence speed of the conventional parallel-perturbation technique. The convergence time of the adaptation process can even be independent of the number of control channels, creating the potential for practical implementation of high-resolution (spatially distributed) adaptive optical systems based on direct system performance metric optimization. The key factor here is the ability to control and adapt the perturbations to changing statistical properties of wave-front aberrations. This creates the potential for adaptive optical systems that are in fact optical systems with automatic control—becoming truly adaptive in the sense of how this term was defined in control theory long before the advent of adaptive optics.

ACKNOWLEDGMENTS

We thank G. Cauwenberghs and P. R. Barbier for helpful discussions and J. C. Ricklin for editorial and technical comments. The experimental results presented in Subsection 5.C were obtained with the participation of R. Dou and G. W. Carhart.

Address correspondence to Mikhail A. Vorontsov, Army Research Laboratory, 2800 Powder Mill Road, Adelphi,

Maryland 20783. Phone, 301-394-0214; fax, 301-394-0225; email, vorontsov@iol.arl.mil.

REFERENCES

1. G. B. Love, J. S. Fender, and S. R. Restaino, "Adaptive wavefront shaping with liquid crystals," Opt. Photonics News **6**, 16–21 (October 1995).
2. P. R. Barbier and G. Moddel, "Spatial light modulators: processing light in real time," Opt. Photonics News **8**, 16–21 (March 1997).
3. U. Efron, ed., *Spatial Light Modulator Technology: Materials, Devices, and Applications* (Marcel Dekker, New York, 1995).
4. M. C. Roggeman, V. M. Bright, B. M. Welsh, S. R. Hick, P. C. Roberts, W. D. Cowan, and J. H. Comtois, "Use of micro-electro-mechanical deformable mirrors to control aberrations in optical systems: theoretical and experimental results," Opt. Eng. (Bellingham) **36**, 1326–1338 (1997).
5. J. M. Younse, "Mirrors on a Chip," IEEE Spectr. **30**, 27–31 (November 1993).
6. G. V. Vdovin and P. M. Sarro, "Flexible mirror micromachined in silicon," Appl. Opt. **34**, 2968–2972 (1995).
7. A. Dembo and T. Kailath, "Model-free distributed learning," IEEE Trans. Neural Netw. **1**, 58–70 (1990).
8. G. Cauwenberghs, "Analog VLSI stochastic perturbative learning architectures," Int. J. Analog Integr. Circ. Signal Process. **13**, 195–209 (1997).
9. G. Cauwenberghs, "A fast stochastic error-descent algorithm for supervised learning and optimization," in *Advances in Neural Information Processing Systems*, S. J. Hanson, ed. (Morgan Kaufman, San Mateo, Calif., 1993), Vol. 5, pp. 244–251.
10. J. Alspector, R. Meir, B. Yuhas, and A. Jayakumar, "A parallel gradient descent method for learning in analog VLSI neural networks," in *Advances in Neural Information Processing Systems*, J. D. Cowan, ed. (Morgan Kaufman, San Mateo, Calif., 1993), Vol. 5, pp. 836–844.
11. D. B. Kirk, D. Kerns, K. Fleischer, and A. H. Barr, "Analog VLSI implementation of multidimensional gradient descent," in *Advances in Neural Information Processing Systems*, C. L. Giles, ed. (Morgan Kaufman, San Mateo, Calif., 1993), Vol. 5, pp. 789–796.
12. M. A. Vorontsov, G. W. Carhart, and J. C. Ricklin, "Adaptive phase-distortion correction based on parallel gradient-descent optimization," Opt. Lett. **22**, 907–909 (1997); G. W. Carhart, J. C. Ricklin, V. P. Sivokon, and M. A. Vorontsov, "Parallel perturbation gradient descent algorithm for adaptive wavefront correction," in *Adaptive Optics and Applications*, R. Tyson and R. Fugate, eds., Proc. SPIE **3126**, 221–227 (1997).
13. V. I. Polejaev and M. A. Vorontsov, "Adaptive active imaging system based on radiation focusing for extended targets," in *Adaptive Optics and Applications*, R. Tyson and R. Fugate, eds., Proc. SPIE **3126**, 216–220 (1997).
14. Special issues on Atmospheric-Compensation Technology, J. Opt. Soc. Am. A **11(1) and (2)**, 255–451, 779–945 (1994).
15. M. C. Roggemann and B. M. Welsh, *Imaging Through Turbulence* (CRC Press, Boca Raton, Fla., 1996).
16. J. M. Beckers, "Adaptive optics for astronomy: principles, performance, and applications," Annu. Rev. Astron. Astrophys. **31**, 13–62 (1993).
17. R. G. Paxman, T. J. Schulz, and J. R. Fienup, "Joint estimation of object and aberration using phase diversity," J. Opt. Soc. Am. A **9**, 1072–1085 (1992).
18. B. Ya. Zel'dovich, N. F. Pilipetsky, and V. V. Shkunov, *Principles of Phase Conjugation* (Springer-Verlag, Berlin, 1985).
19. T. G. Alley, M. A. Kramer, D. R. Martinez, and L. P. Schelonka, "Single-pass imaging through a thick dynamic distorter using four-wave mixing," Opt. Lett. **15**, 81–83 (1990).
20. V. P. Sivokon and M. A. Vorontsov, "High-resolution adaptive phase distortion suppression based solely on intensity information," J. Opt. Soc. Am. A **15**, 234–247 (1998).
21. D. M. Pepper, C. J. Gaeta, and P. V. Mitchell, "Real-time holography, innovative adaptive optics, and compensated optical processors using spatial light modulators," in *Spatial Light Modulator Technology: Materials, Devices, and Applications*, U. Efron, ed. (Marcel Dekker, New York, 1995), pp. 585–654.
22. M. Avriel, *Nonlinear Programming: Analysis and Methods* (Prentice-Hall, Englewood Cliffs, N.J., 1976).
23. M. A. Vorontsov, G. W. Carhart, D. V. Pruidze, J. C. Ricklin, and D. G. Voelz, "Adaptive imaging system for phase-distorted extended source/multiple distance objects," Appl. Opt. **36**, 3319–3328 (1997).
24. T. R. O'Meara, "The multidither principle in adaptive optics," J. Opt. Soc. Am. **67**, 315–325 (1977).
25. R. K. Tyson, *Principles of Adaptive Optics* (Academic, Boston, 1991).
26. H. J. Kushner and D. S. Clark, *Stochastic Approximation Methods for Constrained and Unconstrained Systems* (Springer-Verlag, New York, 1987).
27. D. Lawrence, *Genetic Algorithm and Stimulated Annealing* (Morgan Kaufman, Los Altos, Calif., 1987).
28. P. J. M. van Luarhoven and E. H. L. Aarts, *Stimulated Annealing: Theory and Applications* (Reidel, Dordrecht, The Netherlands, 1987).
29. U. Mahlab and J. Shamir, "Iterative optimization algorithms for filter generation in optical correlators: a comparison," Appl. Opt. **31**, 1117–1125 (1992).
30. L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics* (Cambridge U. Press, Cambridge, UK, 1995).
31. J. W. Goodman, *Statistical Optics* (Wiley, New York, 1985).
32. D. L. Fried, "Statistics of a geometric representation of wavefront distortion," J. Opt. Soc. Am. **55**, 1427–1435 (1965).
33. R. J. Noll, "Zernike polynomials and atmospheric turbulence," J. Opt. Soc. Am. **66**, 207–211 (1976).
34. D. P. Greenwood and D. L. Fried, "Power spectra requirements for wave-front-compensative systems," J. Opt. Soc. Am. **66**, 193–206 (1976).
35. M. A. Vorontsov, G. W. Carhart, D. V. Pruidze, J. C. Ricklin, and D. G. Voelz, "Image quality criteria for an adaptive imaging system based on statistical analysis of the speckle field," J. Opt. Soc. Am. A **13**, 1456–1466 (1996).
36. R. Dou, D. V. Pruidze, J. C. Ricklin, V. P. Sivokon, and M. A. Vorontsov, "Coherent optical processor for image-quality metric measurement," in *Propagation and Imaging through the Atmosphere*, L. Bissonnette and C. Dainty, eds., Proc. SPIE **3125**, 339–343 (1997).
37. M. A. Vorontsov, J. C. Ricklin, and G. W. Carhart, "Optical simulation of phase distorted imaging systems: nonlinear and adaptive approach," Opt. Eng. (Bellingham) **34**, 3229–3238 (1995).
38. L. I. Goldfischer, "Autocorrelation function and power spectral density of laser-produced speckle patterns," J. Opt. Soc. Am. **55**, 247–252 (1965).