Adaptive wavefront correction using a VLSI implementation of the parallel gradient descent algorithm

G. W. Carhart[a], M. A. Vorontsov[a], M. Cohen[b], G. Cauwenberghs[b] and R. T. Edwards[b]

[a]Army Research Laboratory, Adelphi, MD 20783
[b]Johns Hopkins University, Baltimore, MD

ABSTRACT

An adaptive laser beam focusing system using a 127 channel liquid crystal phase modulator is presented. The controller for the system is a circuit built with prototype VLSI chips that implement the stochastic parallel gradient descent algorithm. The controller is driven by a scalar laser beam quality metric and can run at the rate of 150 kHz. The system performance is characterized and a secondary control loop manipulating one of the algorithm parameters is experimentally investigated. The performance of the system is reported and performance improvements obtained by using the recent history of the beam quality metric to control an algorithm parameter is demonstrated.

Keywords: adaptive optics, adaptive focusing, model-free optimization, parallel gradient descent, analog control circuitry.

1. INTRODUCTION

As high resolution phase modulators become more widely available for use in adaptive optics, the problem of how to control these devices becomes a major concern. Although phase conjugation techniques based on rapid wavefront measurement have been found useful in astronomical applications [1], this approach to the control problem is more difficult to use in situations where wavefront distortions can not be measured directly, for example in anisoplanatic imaging conditions. When a performance metric can be defined, model-free optimization techniques provide an alternative approach to the control problem that does not require the use of any *a priori* knowledge of a system model. A common strategy used by model-free optimization techniques is to consider the performance metric as a function of the control parameters and then follow the gradient of the function in the direction of improved performance until an *extrenum* is found. Unfortunately, the function typically has many *extrenema* and the one found may not be the global optimum. Furthermore, estimating the gradient requires sequentially perturbing each control parameter and then measuring the performance criterion to compute all of gradient components. Since the time that is required for each step is proportional to the number of control parameters, this procedure does not scale well with increase of the number of control parameters. A more efficient implementation of the gradient optimization procedure, referred to as the stochastic parallel gradient descent algorithm, is first discussed in the literature as an unsupervised learning algorithm employed for the evaluation of the weighting coefficients in neural networks [2]. The stochastic parallel gradient descent approach has been applied recently to adaptive optics applications [3, 4]. In the stochastic parallel gradient descent technique perturbations are applied to all the control parameters simultaneously. The performance metric is measured only once to obtain the information required for the gradient estimation used at each step. This provides a significant advantage over the sequential perturbation technique since the time that is required per step for this method is independent of the number of control parameters. The parallel nature of the perturbation and update of the control parameters lends itself to the design of parallel control circuitry based on this algorithm which can provide for fast control of a multi-element wavefront corrector that can be used for many adaptive optical applications. The Army Research Laboratory and Johns Hopkins University have jointly developed a VLSI circuit that implements much of the stochastic parallel gradient descent algorithm electronically [5]. In this paper we report the use of this circuit to control a simple adaptive focusing system, characterize the system performance and present experimental results demonstrating performance improvements obtained using a modified control algorithm based on adaptive change of the update rule gain coefficient.

2. STOCHASTIC PARALLEL GRADIENT DESCENT CIRCUIT

To use the parallel gradient descent algorithm to control an adaptive optical device, such as a liquid crystal phase modulator or a deformable mirror, the spatial light modulator should be a part of an optical system the performance of which can be measured. Let J be a scalar function of system performance and $u_1,\dots,u_j,\dots,u_N$ represent the control voltages applied to a wavefront corrector: $J(u_1,\dots,u_j,\dots,u_N)$. At each iteration of the algorithm the first step is to construct a vector of

Part of the SPIE Conference on High-Resolution Wavefront Control: Methods, Devices, and Applications • Denver, Colorado • July 1999
SPIE Vol. 3760 • 0277-786X/99/$10.00

perturbations, δu_j, where the magnitude of the δu's are the same but their signs are random. Next apply the perturbation to each of control voltages and measure the perturbed metric value:

$$J_+^{(n)} = J(u_1^{(n)} + \delta u_1, ..., u_j^{(n)} + \delta u_j, ..., u_N^{(n)} + \delta u_N). \quad (1)$$

Then apply the perturbation in the other direction and measure the oppositely perturbed metric value:

$$J_-^{(n)} = J(u_1^{(n)} - \delta u_1, ..., u_j^{(n)} - \delta u_j, ..., u_N^{(n)} - \delta u_N). \quad (2)$$

Now obtain the metric change, $\delta J^{(n)}$,

$$\delta J^{(n)} = J_+^{(n)} - J_-^{(n)}. \quad (3)$$

In accordance with the stochastic gradient descent procedure the control voltage update can be implemented as follows [3,4]:

$$u_j^{(n+1)} = u_j^{(n)} + \gamma \delta J^{(n)} \delta u_j, \quad (4)$$

where γ is a gain coefficient which scales the size of the control parameter corrections.

A VLSI circuit that implements this algorithm is described in [5]. In this VLSI implementation, γ and the magnitude of the δu's are parameters that can be controlled outside the chip. Each chip is designed to control 19 channels and can be easily connected to others to build controllers for devices with more channels. An adaptive optical system controlled by circuitry made with these chips will be able to operate at a rate limited by the spatial light modulator used. The fastest available devices still run at rates less than 2 kHz

3. ADAPTIVE LASER BEAM FOCUSING SYSTEM

The spatial light modulator the chip was designed to control is a Meadowlark HEX127 liquid crystal phase modulator. This device, shown in Fig. 1, has 127 simultaneously addressable independent channels arranged in a hexagon array like a honeycomb. Each channel is driven by a 2 kHz bipolar square wave in the range 0 to ±10V that produces no modulation at 0V and maximum modulation of about 2.5π radians at ±10V. The response curve is shown in Fig. 1 (bottom right). Unfortunately, the relaxation time for the liquid crystals used in this device is 150 ms prohibiting drive rates higher than about 10 Hz. The VLSI circuit produces 0V to ±5V output signals and cannot exploit the full range of the SLM without amplification. As can be seen from the response curve, however, most of the device response occurs in the range 0V to ±5V. The response is also more linear in this range and acceptable performance can be obtained without further amplification. Another concern associated with this device is the need to keep the average phase centered in the dynamic range. The motive for maintaining this condition is that it keeps most of the

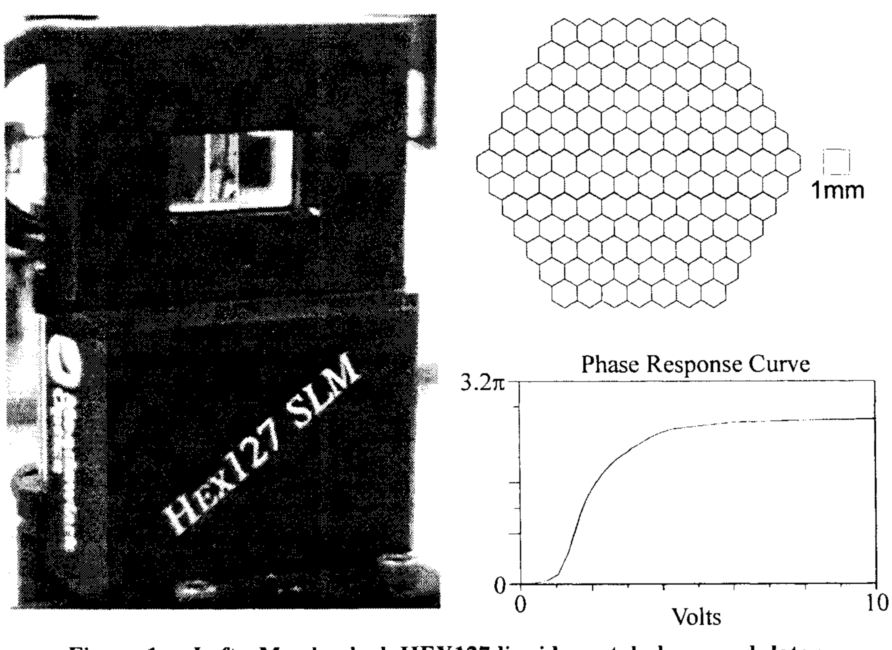

Figure 1: Left - Meadowlark HEX127 liquid crystal phase modulator. Top right – The geometry of the SLM's active area. Bottom right – The SLM's phase response curve.

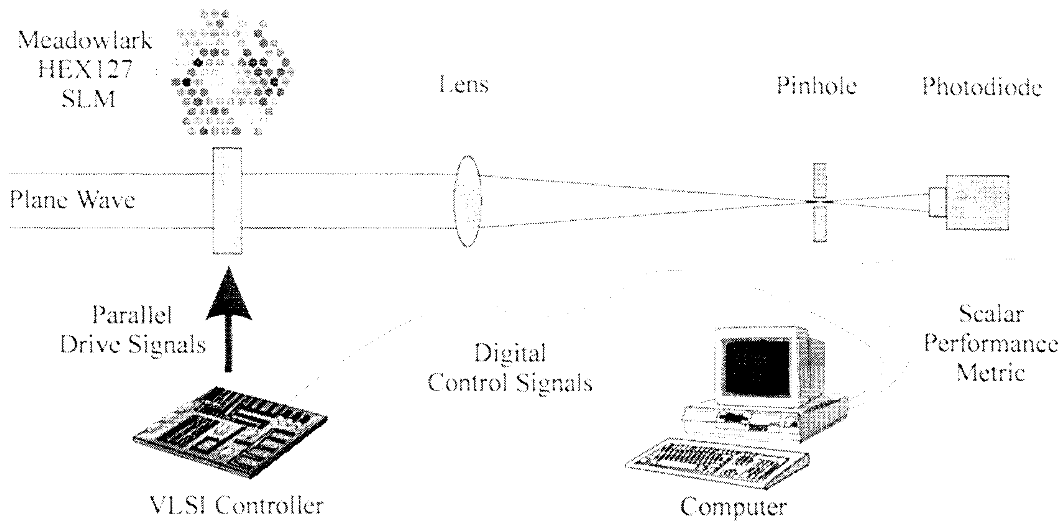

Figure 2: Optical setup and SLM controller scheme.

elements in the driving range where their response is linear. To accomplish this, the chips have an output that is the sum of all the drive voltages and provides a means to monitor their average. Periodically, a constant shift is applied to all channels in the proper direction to keep the average from drifting away from the desired midrange value. This implies the following modification to update equation (4),

$$u_j^{(n+1)} = u_j^{(n)} + \gamma \delta J^{(n)} \delta u_j + \beta\left(u_0 - \frac{1}{N}\sum_{i=1}^{N} u_i \right), \tag{5}$$

where u_0 is the desired mean and β is a coefficient which scales the mean correction rate [3]. It has been shown that adding another term to the update equation that provides global coupling of the voltages can improve the convergence rate [3,6], but this version of the chip does not implement this idea.

To test the chip and characterize its performance, a simple adaptive focusing system was constructed which uses the energy emergent from a pinhole placed in the focal plane of the system as a performance criterion. The light passing the pinhole was collected with a photodiode and the resulting voltage digitized by the computer. To control the SLM device, a control circuit

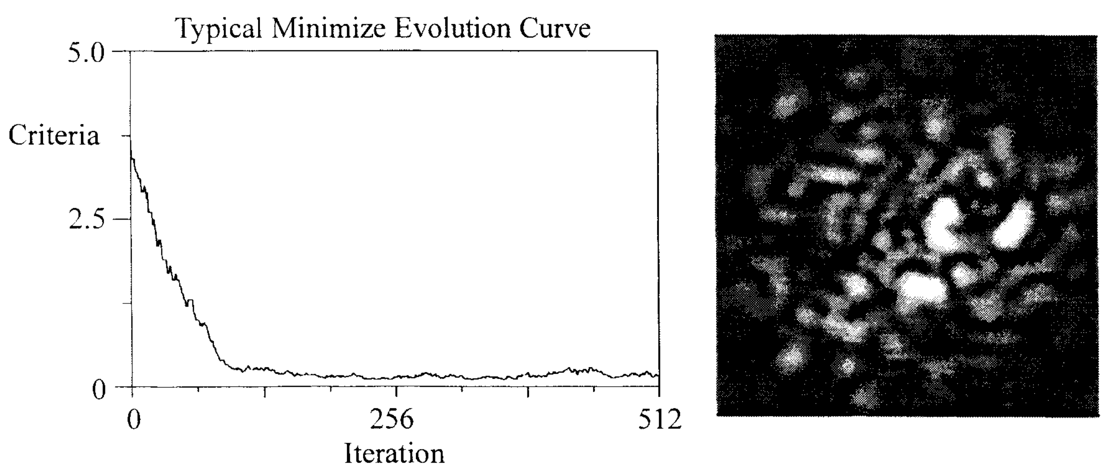

Figure 3. Left: A typical minimization evolution curve. Right: The intensity distribution in the focal plane at the end of the minimization cycle.

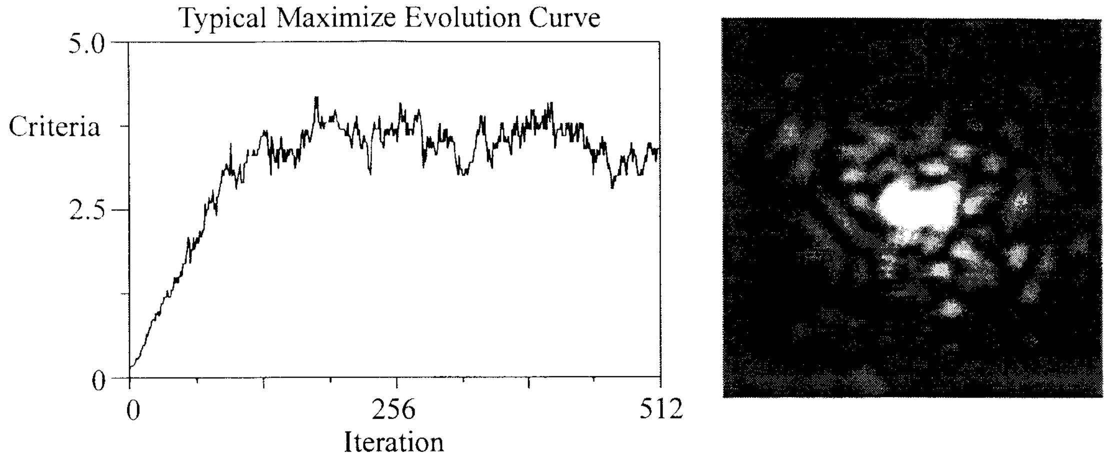

Figure 4. Left: A typical maximization evolution curve. Right: The intensity distribution in the focal plane at the end of the maximization cycle.

was built using 7 chips to drive all 127 channels of the device and operated in the 0V to ±5V range. A 350 MHz Dell Dimension computer with a Computer Boards CIO-DAS 1602/12 multifunction I/O card was used to obtain δJ (by analog-to-digital conversion) and drive the VLSI chips (which require external digital control signals). A schematic of the system is shown in Fig. 2. The system's mode of operation, i.e., whether attempting to maximize or minimize the criterion, was controlled with the sign of the γ coefficient. Typical evolution curves for the two operating modes and their final intensity distributions in the focal plane are shown in Figs. 3 and 4. Optimum values for γ and the perturbation size were found by examining the average performance of 50 such pairs of evolution curves for several different values of each. Selecting an optimum value for γ was found to involve a trade-off between convergence rate and stationary state variance. As can be expected, a large γ produced faster convergence but also a larger variance once an *extrenum* was located. Conversely, a small γ reduced the variance at the *extrenum* but also yielded slower convergence. Fig. 5 shows minimize and maximize evolution curve mean behaviors and standard deviations obtained from 100 typical curves produced with the optimum parameter choices. The results show the system performs about 95% of either transition in less than 256 iterations and a small further improvement occurs afterward.

4. ADAPTIVE GAIN PARAMETER

When searching for the optimum algorithm parameters to use for the results presented in Figs. 3 and 4, we noticed that there

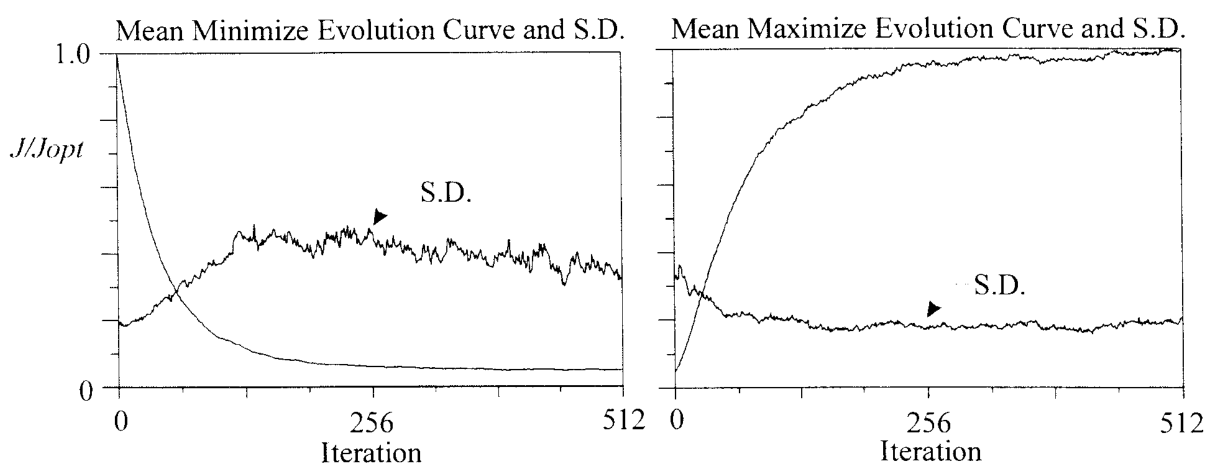

Figure 5. Mean minimize and maximize evolution curves and their standard deviations obtained from 100 typical evolution curves.

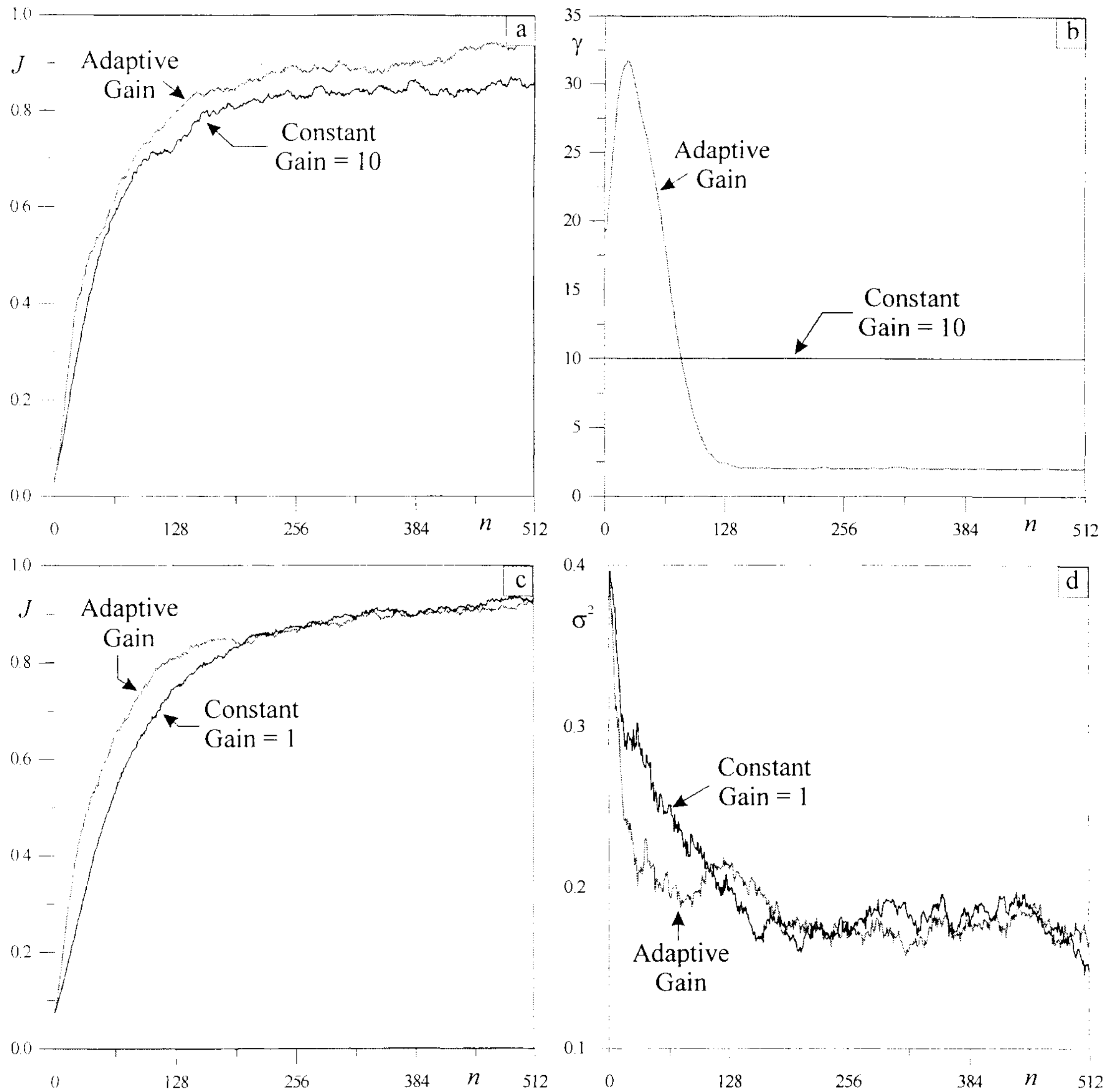

Figure 6: a - Average evolution curves: with adaptive gain and constant gain of γ =10.
b - Behavior of γ during the adaptation for case (a).
c -Average evolution curves: with adaptive gain and constant gain of γ =1.
d- metrics variances for the case (c).

could be an advantage to controlling γ while the system operates. As previously noted, a large γ optimized the convergence rate and a small γ optimized the variance once an *extrenum* was located. Clearly the system should adjust γ to be optimal for its current operating mode. To do this, the system must detect whether it is in transition or at a stationary state. This information can be obtained from the recent history of the criterion behavior.

To use this idea we must measure the unperturbed criterion, $J(u_1,\ldots,u_j,\ldots,u_N)$, at each iteration as well as the two perturbed ones. Using these values, define the positive perturbation metric change as

$$\delta J_+^{(n)} = J_+^{(n)} - J(u_1^{(n)},\ldots,u_j^{(n)},\ldots,u_N^{(n)}) \qquad (6)$$

and the negative perturbation metric change as

$$\delta J_{-}^{(n)} = J_{-}^{(n)} - J(u_1^{(n)},...,u_j^{(n)},...,u_N^{(n)}). \tag{7}$$

Assuming the system is attempting to maximize the metric, then if the system is stationary at a maximum both the positive and negative perturbation metric changes will be negative and when the system is in transition these values will have opposite signs. Now form the weighted sum of the signs of these values over the last k iterations:

$$H = \sum_{i=0}^{k-1}\left(\frac{k-i}{k}\right)\left(\frac{\delta J_{+}^{(n-i)}}{\left|\delta J_{+}^{(n-i)}\right|} + \frac{\delta J_{-}^{(n-i)}}{\left|\delta J_{-}^{(n-i)}\right|}\right). \tag{8}$$

The linear weighting factor is included to place more significance on the most recent history. We use only the signs of the perturbed metric changes because this was observed to worked better than using the actual values. With this sum we can control γ in the desired fashion as follows:

$$\gamma^{(n+1)} = \gamma^{(n)} + \kappa(H + C) \tag{9}$$

where κ is a scaling factor controlling the correction size and C is a constant.

The constant C acts as a lag before γ begins decreasing when the system becomes stationary and a maximum increase of γ per iteration when the system is in transition. In practice, γ must remain within an allowable range of values so the calculated value must be clipped to the allowable maximum or minimum if it has exceeded the acceptable range. If the system is trying to minimize the metric the definitions of the positive and negative perturbation metric changes must be negated and then equations (8) and (9) will still work correctly.

Fig. 6 shows mean evolution curves obtained from 100 typical evolution curves for both operation modes that compare the behavior of the system operating with a constant or an adaptive gain coefficient. It also shows the variances for the case of the mean minimization curves and shows the average behavior of the adaptive gain coefficient for the mean maximization curves. As can be seen, the system operating with an adaptive gain provides both a faster convergence rate and a smaller variance at the *extrenum*.

5. REFERENCES

1. R. K. Tyson, *Principles of Adaptive Optics*, (Academic, Boston, 1991).
2. J.C. Spall, "A Stochastic Approximation Technique for Generating Maximum Likelihood Parameter Estimates," Proc. of the 1987 Am. Control Conf., Minneapolis, MN, 1987.
3. 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).
4. M.A. Vorontsov and V.P. Sivokon, "Stochastic parallel gradient descent technique for high-resolution wavefront phase distortion correction," *J. Opt. Soc. Am.* A, **15**, pp. 2745-2758, 1998.
5. R.T. Edwards, M. Cohen, G. Cauwenberghs, M.A.Vorantsov and G.W. Carhart, "Analog VLSI parallel stochastic optimization for adaptive optics" in *Learning on silicon*, ed. G. Cauwenberghs (Kluwer Academic, Boston, 1999).
6. G.W. Carhart, J.C. Ricklin, V.P. Sivokon and M.A. Vorontsov, "Parallel perturbation gradient descent algorithm for adaptive wavefront correction," SPIE Proc. 3126, 221-227 (1997).