Multi-Channel Coherent Detection for Delay-Insensitive Model-Free Adaptive Control

Dimitrios N. Loizos
Electrical and Computer Engineering
The Johns Hopkins University
Baltimore, MD 21218
dloizos@jhu.edu

Paul P. Sotiriadis
Electrical and Computer Engineering
The Johns Hopkins University
Baltimore, MD 21218
pps@jhu.edu

Gert Cauwenberghs
Division of Biological Sciences
University of California, San Diego
San Diego, CA 92093
gert@ucsd.edu

Abstract— A mixed-signal architecture for continuous-time multidimensional model-free optimization is presented. It is based on multi-channel coherent modulation and detection that reliably estimates the objective function's gradient, with respect to the system parameters, in the presence of time delays. The narrow-band nature of the excitation signals reduces the unknown dynamics of the objective function to a single parameter per control channel, the phase delay.

An efficient implementation of the adaptive control architecture is presented; it incorporates parallel control channels with individually selectable 6-level phase delay adjustment. Initial experimental results indicate wide operating range covering almost 7 decades of excitation frequencies.

I. INTRODUCTION

Function optimization and system adaptation using application - specific VLSI circuits has been introduced in many fields. It offers the advantages of hardware minimization and convergence speed augmentation, when compared to equivalent software approaches. Analog implementations are generally preferred in applications demanding low power and high speed. A major challenge in designing these analog systems is to find appropriate optimization techniques immune to fabrication imperfections and mismatches; moreover, in many cases these techniques should consider no prior knowledge of the function or system subject to optimization.

The above problem, generally known as *model-free optimization*, has been broadly studied in Neural Networks, where minimization of the error between the measured and desired response of the system is needed. However, the results can be easily projected to applications requiring minimization/maximization of any convex/concave function. A widely used optimization algorithm is gradient descent; several studies have been published on different variations of it [1]. Most of them use an estimate of the gradient derived by perturbing the input variables of the under optimization function and then correlating the output of the function with the perturbation(s).

According to the nature of the perturbations, the algorithms can be divided into stochastic and deterministic ones. Analog VLSI architectures and/or implementations using stochastic perturbations have been presented for both discrete-time [2], [3] and continuous-time [4], [5]. Algorithms that use deterministic perturbations can be further divided into sequential ones [6], where the dither is applied serially to each of the variables, and parallel ones [7]. To the best of the authors' knowledge, no VLSI circuit has been implemented for either of them.

The main drawback of the sequential approach is that its convergence speed decreases as the number of variables increases.

Stochastic techniques have been much preferred over parallel deterministic ones mainly due to their simpler hardware implementation and scalability with respect to the number of variables. However, they fail in the case of significant and variable propagation delay. On the contrary, narrow-band methods can be made much more robust to delay distortion, and phase/amplitude distortion in general. This problem is addressed in Section II.

In this paper, we present an analog VLSI implementation of a (continuous) gradient descent algorithm for function optimization using narrow-band deterministic (sinusoidal) dithers (perturbations). The specific implementation is meant for applications where adaptation of multivariable systems with significant and time-variant propagation delays is needed. The basic principles of gradient descent optimization and of the delay-insensitive version implemented are given in Section II. Some design aspects are discussed in Section III and initial experimental results are shown in Section IV.

II. THEORETICAL BACKGROUND

Let $\mathbf{u} = (u_1, u_2, \ldots, u_k)^{\mathrm{T}}$ be a $k \times 1$ real vector and $J(\mathbf{u})$ be a scalar cost function. The basic form of the (continuous) gradient descent algorithm is

$$\frac{d\mathbf{u}}{dt} = -\epsilon \nabla J(\mathbf{u}), \tag{1}$$

where ϵ is a positive constant. Assuming that J has no saddle points, $\mathbf{u}$ reaches a (local) minimum (almost always). The convergence rate depends partially on ϵ. A variant of (1) that is easily implementable in mixed-signal circuits is given by (2).

$$\frac{d\mathbf{u}}{dt} = -\epsilon \operatorname{sgn}\big(\nabla J(\mathbf{u})\big). \tag{2}$$

In both cases, maximization can be achieved using the same rules but with $\epsilon < 0$.

The gradient, ∇J, can be extracted by superimposing a small-amplitude perturbation $\boldsymbol{\theta}$ to signal $\mathbf{u}$. For the stochastic

case, $\boldsymbol{\theta}$ is a vector of independent, zero mean, noise sources $n_i(t)$, $i = 1, ..., k$, whereas in the deterministic case, it is a vector of narrow-band signals with known carrier frequency. For simplicity, it is assumed they are sinusoidal signals of frequencies ω_i, $i = 1, ..., k$ [7]. The resulting vector $\tilde{\mathbf{u}}$ is

$$\tilde{\mathbf{u}} = \mathbf{u} + \boldsymbol{\theta} = \begin{cases} \mathbf{u} + \alpha \left[n_1(t), n_2(t), \ldots, n_k(t) \right]^{\mathrm{T}} \\ \text{or} \\ \mathbf{u} + \alpha \left[\cos(\omega_1 t), \cos(\omega_2 t), \ldots, \cos(\omega_k t) \right]^{\mathrm{T}} \end{cases}$$

where α is a small scalar parameter. Evaluating the resulting metric J, using Taylor series, around $\boldsymbol{\theta_0} = 0$, we get

$$J(\tilde{\mathbf{u}}) = J(\mathbf{u} + \boldsymbol{\theta}) = J(\mathbf{u}) + \left(\nabla J(\mathbf{u}) \right)^{\mathrm{T}} \boldsymbol{\theta} + \ldots$$

For the case of stochastic perturbations, the partial derivative of J with respect to variable u_i can be estimated by correlating J with dither $n_i(t)$ over a time-integral period of predefined length T [8]

$$\int_0^T J(\tilde{\mathbf{u}}) \cdot n_i(t) dt \simeq \int_0^T J(\mathbf{u}) \cdot n_i(t) dt + \alpha \int_0^T \left. \frac{\partial J(\mathbf{u})}{\partial u_i} \right|_{\mathbf{u}} n_i^2(t) dt +$$
$$+ \alpha \int_0^T \sum_{j, j \neq i} \left. \frac{\partial J(\mathbf{u})}{\partial u_j} \right|_{\mathbf{u}} n_i(t) n_j(t) dt \quad (3)$$

Assuming that $J(\mathbf{u})$ remains fairly constant within period T, the first term of the R.H.S. (right hand side) is expected to be close to zero since the noise sources (are assumed to be ergodic and) have zero mean. Moreover, the third term of the R.H.S. is also expected to be close to zero since the noise sources are (jointly ergodic and) uncorrelated. Higher order terms can be neglected since their amplitude will be small. Therefore, correlation provides us with an estimate of the partial derivative of J with respect to variable u_i.

Following the same procedure, this time with sinusoidal dithers, it can be proved that correlation of the metric J with each of the dithers will result to

$$\int_0^{T'} J(\tilde{\mathbf{u}}) \cos(\omega_i t) dt \simeq \frac{\alpha}{2} \left. \frac{\partial J}{\partial u_i} \right|_{\mathbf{u}}. \quad (4)$$

It should be noted that T' can be significantly smaller than the integration period T used for correlation in the stochastic case (3).

A difference between the two methods arises when a considerable propagation delay, between the time the perturbation is applied to the variables and the time correlation is performed, should be accounted for. In the case of stochastic signals, due to their random nature, there is no, at least easy, way of compensating for the delay. For a "reasonably" accurate gradient estimate, the integration time T should be no less than $\tau \cdot k \cdot m$, where τ is the average propagation delay, k the number of variables of the metric J and m a parameter proportional to the accuracy of the estimate (e.g. if an error of less than 2% is required, m should be greater than 50) [8]. Obviously, for large delays, large integration times are needed, which reduce the convergence speed. Moreover, the bandwidth of the stochastic perturbations needs to scale inversely proportionally to $m \cdot \tau$ [8].

In the case of sinusoidal perturbations, as shown in [9], a propagation delay scales the estimate of the partial derivative of J with respect to u_i by a factor $\cos(\varphi_i)$, namely

$$\int_0^{T'} J(\tilde{\mathbf{u}})|_{\tau - \text{delay}} \cos(\omega_i t) dt \simeq \frac{\alpha}{2} \left. \frac{\partial J}{\partial u_i} \right|_{\mathbf{u}} \cos(\varphi_i), \quad (5)$$

where $\varphi_i = \omega_i \cdot \tau$. In other words, the propagation delay in the metric reduces to a single parameter, phase φ_i, that can be easily compensated for.

III. IMPLEMENTATION

The design of the implemented VLSI circuit is based on the architecture described in [9]. A brief review is given and further insight on the actual implementation of the main building blocks is provided. Moreover, the aspect of propagation delay compensation is described. The high level architecture of the control channel for each variable of the objective function is given in Fig. 1.

A. General architecture

A 3-phase oscillator generates sinusoidal signals at frequencies ω_i, $i = 1, \ldots, k$, where $\omega_i \neq \omega_j$ for $i \neq j$. For each channel (each channel corresponds to a variable), one of the 3-phases (fixed) is superimposed to variable u_i and the perturbed variable $\tilde{u}_i$ is applied to metric J. The gradient descent algorithm is then applied according to rule (2). Instead of exact correlation, the value of $\left. \frac{\partial J(\mathbf{u})}{\partial u_i} \right|_{\mathbf{u}}$ is estimated by multiplying the metric J with dither u_i and low-pass filtering the product. According to the sign of the gradient and whether we need minimization or maximization, u_i is updated accordingly with a rate specified by the charge pump [10].
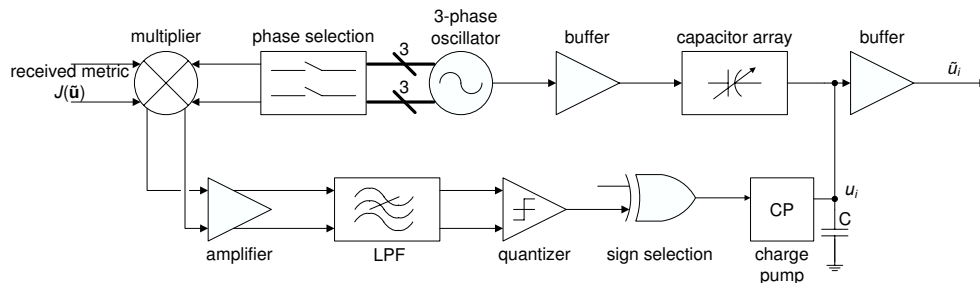

Fig. 1. Block diagram for one channel of the phase controller

B. Three-phase oscillator

The three-phase oscillator is a ring of 3 differential $G_m - C$ filters in shunt with tunable resistors R, as shown in Fig. 2. The gain of each $G_m - R - C$ block is $|H(j\omega)| = G_m R / \sqrt{1 + (\omega R C)^2}$ and the phase introduced is $\tan^{-1}(H(j\omega)) = -\omega R C$.
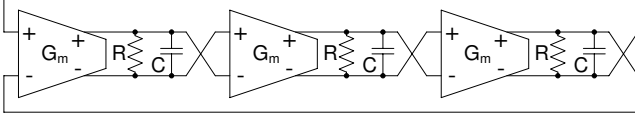

Fig. 2. High-level architecture of the 3-phase oscillator. Transconductors G_m and resistances R are tunable.

The Barkhausen oscillation criterion, namely that the gain around the loop should be 1 and that the sum of the phases should equal 360°, applied on the topology of Fig. 2 gives the following conditions for oscillation

$$R = \frac{2}{G_m} \tag{6}$$

$$\omega = \frac{\sqrt{3} G_m}{2C} \tag{7}$$

According to (7), the oscillation frequency can be controlled by transconductance G_m; the value of G_m is set by a bias current. For oscillations to be sustained, R should also be tunable and scale inversely proportionally to G_m.

C. Low-pass filter

The low-pass filter removes higher frequency components generated by the multiplier. These components are terms both at the perturbation frequencies ω_i as well as at the sums and differences $\omega_i + \omega_j$ and $\omega_i - \omega_j$. In order to ensure satisfactory compression of the unwanted terms, the filter needs to have a steep roll-off and the perturbation frequencies ω_i need to be apart at least twice the bandwidth of the low-pass filter, i.e., $f_{-3dB} < \min_{i,j} |\omega_i - \omega_j| / 2$.

A 5th order Chebyshev filter was implemented using biquads and 1st order $G_m - C$ filters. All trasconductors in the filter have the same topology and their gains (G_m) are linearly controlled by replicas of the same current. In this way, the bandwidth of the filter is controlled using only one current without any distortion of the "shape" of the transfer function.

D. Propagation delay compensation

As shown in Section II, in the case of sinusoidal perturbations, propagation delay introduces a factor $\cos(\varphi_i)$, $i = 1, \ldots, k$ to the gradient. For the purposes of this work, knowledge of phase delays φ_i within some bounds has been assumed.

The oscillator generating the sinusoidal perturbations, provides 3 phases, dividing the phase domain into 3 equal 120° sections. Moreover, the sign selection block offers an extra degree of freedom in phase selection, splitting the phase domain further into 6 equal 60° sections, as shown in Fig. 3.

Knowing approximately the values of the phase delays φ_i, it is possible to use for multiplication, and for each of the channels, the phase of the oscillator that is closest to φ_i. Assuming that the selected oscillator phase for channel i is φ_{oi}, the product of multiplication and low-pass filtering can be proved to be

$$\int_0^{T'} J(\tilde{\mathbf{u}})|_{\tau - \text{delay}} \cos(\omega_i t - \varphi_{oi}) dt \simeq \frac{\alpha}{2} \left. \frac{\partial J}{\partial u_i} \right|_{\mathbf{u}} \cos(\varphi_i - \varphi_{oi}). \tag{8}$$

Since the difference $|\varphi_i - \varphi_{oi}|$ cannot be greater than 60°, the value of the cos factor in (8) will be always positive, thus not affecting the sign of $\left. \frac{\partial J(\mathbf{u})}{\partial u_i} \right|_{\mathbf{u}}$ used for updating the value of variables u_i.
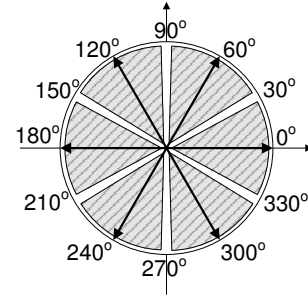

Fig. 3. Phase domain splitted in 60° sections. The arrows represent the selectable phases in the system. Depending on which gray area φ_i lies on, the appropriate phase φ_{oi} is selected, that minimizes the phase error $\varphi_i - \varphi_{oi}$.

IV. EXPERIMENTAL RESULTS

Based on the discussed theoretical model, an 8-channel VLSI implementation was designed, fabricated and tested. Several specifications were set for the design with most important one, the wide tuning range of the dithering frequencies, from 100Hz up to 1GHz. Due to the broad range of operating frequencies, bipolar transistors were preferred and most of the building blocks were designed using translinear topologies. The actual size of the chip is 3mm x 3mm with a minimum size length of 0.5μm.

The update (increase or decrease) rates of the charge pumps are independently programmable, and the amplitude of the dithers superimposed to the outputs are adjustable for each channel, extending the versatility of the implementation. It should be also noted, that the specific design allows cascading of multiple chips in a parallel architecture, expanding the operation of the system to optimization of a virtually n-variable function.

The first experiment to be conducted was the characterization of the 3-phase oscillators. Fig. 4 shows the oscillation frequency with respect to biasing current. Operation from less than 100Hz to almost 1GHz is demonstrated, with the bias current controlling linearly the frequency in a range of almost 7 decades.

In order to verify optimization, a simple 2-variable function was implemented using diodes and resistors. The exact function was $f(V_1, V_2, V_{ref}) = \max(V_1, V_2, V_{ref}) -$
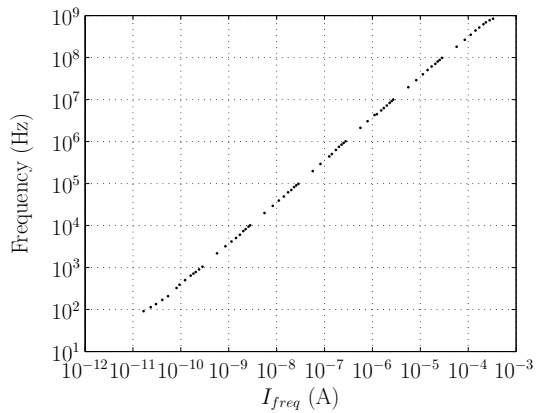
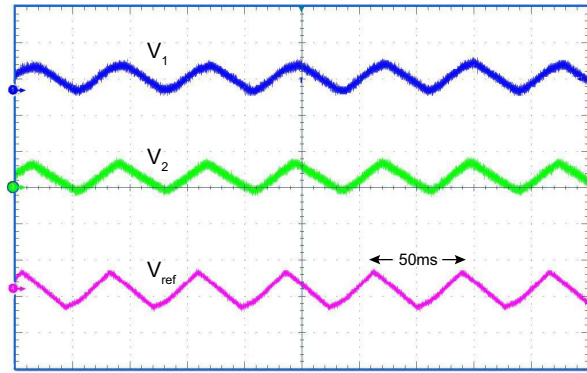Fig. 4.   Linear control of the oscillation frequency with the biasing current

Fig. 5.   Function optimization: The top two waveforms represent V_1 and V_2, while the bottom one V_{ref}

$\min(V_1, V_2, V_{ref}) - 2V_F$, where $V_1 > 0$ and $V_2 > 0$ were the voltage outputs from 2 channels of the system, $V_{ref} > 0$ was a reference voltage provided by a function generator, and V_F was the forward voltage drop of the used diodes. Function f has a global minimum, reached when $V_1 = V_2 = V_{ref}$. For the purposes of the experiment, the dithering frequencies for the two channels used were set to 20MHz and 26MHz, while V_{ref} was a triangular wave of period 50ms. For the specific implementation and the specific perturbation frequencies, the phase delay introduced by the metric was around 180°. This led to a cos factor close to -1, multiplying the information of the gradient; to compensate, the sign of the gradient was flipped by the sign selection block. The response of the system is illustrated in Fig. 5 and shows how variables V_1 and V_2 follow V_{ref}.

V. CONCLUSION

An analog VLSI implementation of a multi-variable function optimization system has been presented. Adaptation is achieved using the gradient descent algorithm, and gradient information is retrieved using sinusoidal perturbations. The deterministic nature of the signals makes propagation delay compensation feasible. The main blocks of the design have been analyzed. Some initial experimental results have been demonstrated showing the broad range of achievable perturbation frequencies with the implemented design and an example of adaptation for a simple 2-variable function.

VI. ACKNOWLEDGEMENT

REFERENCES

[1] P. D. Moerland and E. Fiesler, "Hardware-friendly learning algorithms for neural networks: An overview," in *Proc. MicroNeuro '96*, 1996, pp. 117–124.
[2] G. Cauwenberghs, "A fast stochastic error-descent algorithm for supervised learning and optimization," in *Adv. Neural Information Processing Systems (NIPS '92)*, vol. 5, San Mateo, CA: Morgan Kaufman, 1993, pp. 244–251.
[3] J. Alspector, R. Meir, B. Yuhas, A. Jayakumar, and D. Lippe, "A parallel gradient descent method for learning in analog VLSI neural networks," in *Adv. Neural Information Processing Systems (NIPS '92)*, vol. 5, San Mateo, CA: Morgan Kaufman, 1993, pp. 836–844.
[4] D. B. Kirk, D. Kerns, K. Fleischer, and A. H. Barr, "Analog VLSI implementation of multi-dimensional gradient descent," in *Adv. Neural Information Processing Systems (NIPS '92)*, vol. 5, San Mateo, CA: Morgan Kaufman, 1993, pp. 789–796.
[5] R. T. Edwards, M. Cohen, G. Cauwenberghs, M. A. Vorontsov, and G. W. Carhart, "Analog VLSI parallel stochastic optimization for adaptive optics," in *Learning on Silicon*, G. Cauwenberghs and M. A. Bayoumi, Eds. Boston, MA: Kluwer Academic, 1999, ch. 1, pp. 359–382.
[6] M. Jabri and B. Flower, "Weight perturbation: an optimal architecture and learning technique for analog VLSI feedforward and recurrent multilayer networks," *IEEE Trans. Neural Networks*, vol. 3, no. 1, pp. 154–157, Jan. 1992.
[7] T. R. O'Meara, "The multidither principle in adaptive optics," *J. Opt. Soc. Am.*, vol. 67, no. 3, pp. 306–315, Mar. 1977.
[8] A. Dembo and T. Kailath, "Model-free distributed learning," *IEEE Trans. Neural Networks*, vol. 1, no. 1, pp. 58–70, Mar. 1990.
[9] D. N. Loizos, P. P. Sotiriadis, and G. Cauwenberghs, "A robust continuous-time multi-dithering technique for laser communications using adaptive optics," in *Proc. Int. Symp. Circuits and Systems (ISCAS '06)*, May 2006, pp. 3626–3629.
[10] G. Cauwenberghs, "Analog VLSI stochastic perturbative learning architectures," *Int. J. Analog Integrated Circuits and Signal Processing*, vol. 13, no. 1-2, pp. 195–209, May-June 1997.