Noise-Shaping Gradient Descent based Online Adaptation Algorithms for Digital Calibration of Analog Circuits

Shantanu Chakrabartty, *Senior Member, IEEE*, Ravi Shaga, Kenji Aono, *Student Member, IEEE*

***Abstract*—Analog circuits that are calibrated using digital-to-analog converters use a DSP based algorithm for real-time adaptation and programming of system parameters. In this paper, we first show that this conventional framework for adaptation yields sub-optimal calibration properties due to artifacts introduced by quantization noise. We then propose a novel online stochastic optimization algorithm called noise-shaping or $\Sigma\Delta$ gradient descent, that can shape the quantization noise out of the frequency regions spanning the parameter adaptation trajectories. As a result, the proposed algorithms demonstrate superior parameter search properties compared to floating-point gradient-methods and better convergence properties than conventional quantized gradient-methods. In the second part of this paper, we apply the $\Sigma\Delta$ gradient descent algorithm to two examples of real-time digital calibration:(a) balancing and tracking of bias currents; and (b) frequency calibration of a band-pass Gm-C biquad filter biased in weak-inversion. For each of these examples, the circuits have been prototyped in a 0.5-μm CMOS process and we demonstrate that the proposed algorithm is able to find the optimal solution even in the presence of spurious local minima which are introduced due to the non-linear and non-monotonic response of calibration DACs.**

***Index Terms*—online learning, sigma-delta learning, noise-shaping, stochastic optimization, digitally assisted analog circuits, analog VLSI, quantization**

I. Introduction

DIGITALLY calibrated analog circuits exploit the precision and reliability of digital signal processing techniques to compensate and calibrate for artifacts (mismatch or non-linearity) encountered in analog components [1]–[3]. Digital calibration (real-time or offline) has been successfully applied to the design of neuromorphic systems [5]–[8], to the design of precision analog-to-digital converters (ADCs) [4] and to the design of current references and memories [9], [10]. Nowhere is the effect of analog artifacts more pronounced than for ultra-low-power CMOS analog circuits where the MOSFET transistors are biased in weak-inversion [5], [11], [13]. As a result, precise and real-time calibration of the circuit requires an exhaustive search and adaptation over the system parameters. In digitally calibrated analog circuits, the system parameters are typically stored on on-chip digital-to-analog converters (DACs) which are programmed using a digital signal processor (DSP). While many adaptation strategies could be implemented on the DSP [1], the most simple and commonly used adaptation algorithm involves the use of online gradients [3]. This method is illustrated in Fig. 1 where the analog circuit to be calibrated is embedded within a digital adaptation loop and the system parameters $\boldsymbol{\theta}$ are incrementally updated based on the gradient of an error or quality metric, $g(\boldsymbol{\theta})$ [1], [12], [14].

S. Chakrabartty and K. Aono are with the Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI, 48824 USA. R. Shaga is currently with the Intel Corporation, OR, USA (e-mails: {shantanu,shagarav,aonokenj}@egr.msu.edu.

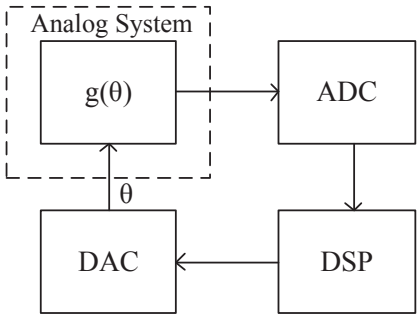


Fig. 1. Calibration of an analog circuit (system) using a digital adaptation loop.

However, when the error gradients are measured directly or indirectly using an analog-to-digital converter (ADC) (see Fig. 1) and when the system parameters are programmed using non-ideal digital-to-analog converters (DAC), quantization noise and non-linearity is injected into the adaptation process. In the first part of this paper, we first analyze the effect of quantization noise on the adaptation process and show that the the conventional calibration approach leads to a sub-optimal trade-off between the precision of the solution and speed of convergence of the algorithm. This effect is illustrated in Fig. 2 which plots a typical system adaptation trajectory in time and spectral domain. Fig. 2(a) plots the system error (computed as the norm of the difference between the instantaneous system parameter $\boldsymbol{\theta}$ and the target/ideal parameter $\boldsymbol{\theta}^*$) with respect to time as the system continuously adapts. For any asymptotically consistent adaptation algorithm, the system error should converge to zero (absolutely or in probability) over time. In the spectral domain, the adaptation trajectory can be visualized to be band-limited within a low-frequency region as shown in Fig. 2(b). When the error gradient is directly quantized, the quantization noise affects the adaptation trajectory (shown in the time plot in Fig. 2(a))

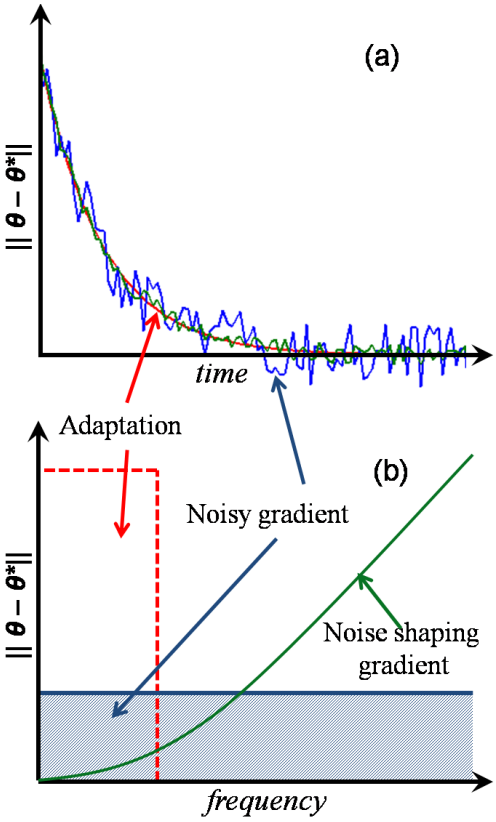


Fig. 2. Effect of quantization noise on adaptation trajectories: in (a) time-domain; and in (b) spectral-domain.

by injecting noise into the adaptation frequency band (shown in spectral plot Fig. 2(b).) The injection of noise affects the precision and hence the speed of the convergence.

But addition of quantization-noise during the adaptation is beneficial because it makes the online algorithm robust to being trapped in local minima. The robustness is important when the calibration DACs used in Fig. 1 are non-ideal. DAC non-ideality arise due to design constraints (like small silicon area and sub-threshold biasing) that ensure that the calibration circuitry does not consume too much power or area. This implies that the system error metric being optimized is a non-linear function of the system parameters. Thus, the magnitude of the quantization noise determines the trade-off between the speed/precision of convergence and the parameter search properties of the gradient-based calibration algorithm.

In this paper, we propose a class of noise-shaping online optimization algorithms called $\Sigma\Delta$ gradient-descent algorithms that achieve a better trade-off between the speed/precision of convergence and the robustness to spurious local minima. The concept is illustrated in Fig. 2(b) where the algorithm shapes the quantization noise out of the adaptation frequency-band. Note that this approach in principle, is similar to the conventional $\Sigma\Delta$ modulation [15], however, its use for online optimization algorithms like gradient-descent has not been reported. In particular, we demonstrate the effectiveness of the $\Sigma\Delta$ gradient-descent algorithm for digital calibration of analog circuits.

The paper is organized as follows: Section II compares the spectral properties of conventional and the proposed noise-shaping optimization algorithms. Section III describes the application of the proposed algorithm for digital calibration of some example analog circuits and Section IV concludes the paper with some discussion on future research directions. Before we describe the mathematical model of the proposed online learning algorithms, we define some of mathematical notations that will be used throughout this paper.

$\mathbf{A}$ (bold capital letters)	A matrix with its elements denoted by $a_{ij}, i = 1, 2, ..; j = 1, 2, ...$
x (normal lowercase letters)	A scalar variable.
$\mathbf{x}$ (bold lowercase letters)	A vector with its elements denoted by $x_i, i = 1, 2, ...$
$\|\|\mathbf{x}\|\|_p$	The L_p norm of a vector and is given by $\|\|\mathbf{x}\|\|_p = (\sum_i \|x_i\|^p)^{1/p}$.
$\mathbf{A}^T$	Transpose of $\mathbf{A}$.
$\mathbf{x}_n$	Discrete-time sequence of vectors where n denotes the time-index.
$\mathcal{R}^D$	denotes a D-dimensional space of real-numbers.

II. Mathematical Modeling of Noise-shaped Online Optimization

We will denote the parameters of the analog system by a vector $\boldsymbol{\theta} \in \mathcal{R}^D$. For example, the system parameters of a band-pass filter-bank could be a vector comprising the center frequencies or its Quality factors. The objective of a calibration or a system optimization procedure is to estimate $\boldsymbol{\theta}^*$ such that

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta} \in \mathcal{C}}{\operatorname{argmin}}\, L(\boldsymbol{\theta}) \tag{1}$$

where $L : \mathcal{R}^D \rightarrow \mathcal{R}$ is a system error function and $\mathcal{C}$ is the constraint space over which the minimization is performed. It is assumed that $L(.) \geq 0$ and is continuous and differentiable w.r.t $\boldsymbol{\theta}$. Also, the calibration algorithm has access to the system gradient function $\mathbf{g}(\boldsymbol{\theta}) \in \mathcal{R}^D$ and is given by

$$\mathbf{g}(\boldsymbol{\theta}) = \frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}. \tag{2}$$

A. Steepest Gradient-descent Algorithm

The simplest form of online calibration uses the steepest gradient-descent (GD) algorithm [16], [17] where the parameters $\boldsymbol{\theta}$ are iteratively updated according to

$$\boldsymbol{\theta_n} = \boldsymbol{\theta_{n-1}} - \epsilon_n \mathbf{g}(\boldsymbol{\theta_{n-1}}). \tag{3}$$

$n = 1, 2, \ldots$ denotes the iteration or time index and ϵ_n denotes a learning-rate parameter. It has been shown that GD algorithm converges to a local minimum when the learning rate asymptotically satisfies $\sum_{n=1}^{\infty} \epsilon_n \rightarrow \infty$ and $\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$ [16]. In literature, different forms of steepest GD algorithms have been proposed based on the choice of the error function $L(.)$ and the trajectory of the learning-rate parameter. A popular variant that is only applicable to convex mean-square error functions are the least-mean-square (LMS) algorithms (e.g. the recursive LMS (RLMS) or time-varying LMS (TV-LMS) [18] algorithms) which are commonly used in the design of adaptive filters. Also, many of the GD algorithms start with a larger value of the learning-rate to speed-up the

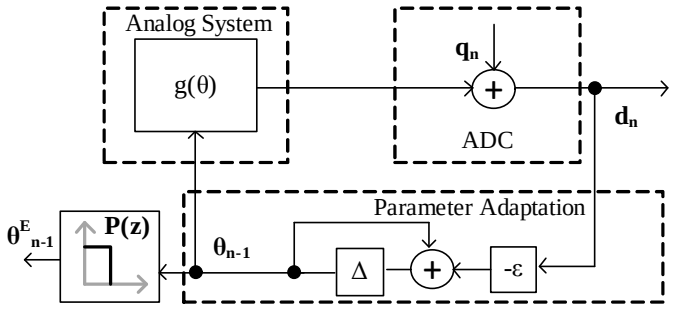


Fig. 3. System architecture showing the signal-flow of a quantized gradient descent (QGD) algorithm.

convergence, but then decrease the magnitude of the learning-rate to precisely converge to the local minima. However, this asymptotic property is not useful when the objective is real-time tracking and calibration of system parameters. Also, when $L(.)$ has multiple local minima, smaller values of ϵ_n increases the likelihood of that iteration in equation (3) of being trapped in the local minimum. In our comparison we have used a fixed value of the learning rate $\epsilon_n = \epsilon$, even though the procedure could potentially be generalized to include time-varying learning-rate parameters.

B. Quantized Gradient-descent Algorithms

In a real-world analog circuit or system, the system-gradient $\mathbf{g}(.)$ is measured using an ADC, the GD adaptation is implemented using a DSP or a micro-controller and the DACs are used for generating the analog parameters. The signal-flow architecture is shown in Fig. 3 where $\mathbf{g}(.)$ is quantized before being processed by the GD based adaptation algorithm. From an analytical point of view, the ADC and the DACs could be combined and the output vector $\mathbf{d}_n$ can be modeled by a quantization noise vector [19] $\mathbf{q}_n$ added to the input of the ADC/DAC module as shown in Fig. 3. The GD algorithm in equation (3) can be accordingly modified as:

$$\boldsymbol{\theta_n} = \boldsymbol{\theta_{n-1}} - \epsilon_n \left[\mathbf{g}(\boldsymbol{\theta_{n-1}}) + \mathbf{q}_n\right]. \tag{4}$$

The algorithm in equation (4) has a similar form as a noisy-GD algorithm [20], [21] and sign-based adaptive filters [22], both of which have been extensively studied in literature. When the quantization noise $\mathbf{q}_n$ admits certain statistical properties [23] (like zero-mean and bounded-variance), it can be shown that $\boldsymbol{\theta_n}$ is bounded and converges *almost surely* to a local minimum. Also, addition of noise to the GD algorithm improves the search properties of the algorithm (to find a more optimal solution) as the stochastic nature of the algorithm would facilitate escaping the neighborhood of the local minima [20], [24], [25]. Note that depending on the choice of learning parameter ϵ_n, the final value of $\boldsymbol{\theta}$ will oscillate randomly about the optimal point $\boldsymbol{\theta}$* and will affect the precision of convergence. In literature, two approaches have been proposed to alleviate this effect of noise and improve the precision of convergence. The first approach uses simulated annealing techniques [26], [27] where larger value of the adaptation parameter ϵ is chosen initially and is successively damped allowing the initial iterations to move $\boldsymbol{\theta}$ out of the local minima neighborhood. As the number of iterations is increased, the effect of the additive noise decreases as the adaptation parameter goes to zero. However, the choice of the noise floor level and annealing rate depends on the loss function being minimized and in most cases the nature of the loss-function and hence the gradient is not known a-priori. Also, if the objective is the real-time parameter estimation and tracking, asymptotic methods like annealing are not very useful because the magnitude of the noise and the rate of annealing has to continuously adjusted. The second approach uses a post-filtering technique on the parameter sequence $\boldsymbol{\theta}_n$. This is shown in Fig. 3 where the post-processing filter $P(z)$ (typically a low-pass filter) filters out the noise and produces smoothed-out estimates the parameter $\boldsymbol{\theta}_n^E$. For real-time tracking and adaptation, the second approach is more practical and hence has been chosen for our analysis and implementation.

We analyze the spectral properties of the quantized GD (QGD) algorithm in the proximity of the local minimum $\boldsymbol{\theta}$* where the gradient function $\mathbf{g}(\boldsymbol{\theta})$ can be approximated using a Taylor series expansion as:

$$\mathbf{g}(\boldsymbol{\theta}) \approx \mathbf{g}(\boldsymbol{\theta}^*) + \mathbf{G}(\boldsymbol{\theta}^*)^T(\boldsymbol{\theta} - \boldsymbol{\theta}^*) \tag{5}$$

$$\mathbf{g}(\boldsymbol{\theta}) \approx \mathbf{G}(\boldsymbol{\theta}^*)^T(\boldsymbol{\theta} - \boldsymbol{\theta}^*). \tag{6}$$

$\mathbf{G}(\boldsymbol{\theta}*) \in \mathcal{R}^P \times \mathcal{R}^P$ is a positive definite Hessian matrix computed at the local minima $\boldsymbol{\theta}$* where $\mathbf{g}(\boldsymbol{\theta}^*) = 0$. Substituting the equation (6) into equation (4)

$$\boldsymbol{\theta_n} = \boldsymbol{\theta_{n-1}} - \epsilon_n \left[\mathbf{G}(\boldsymbol{\theta}^*)^T(\boldsymbol{\theta_{n-1}} - \boldsymbol{\theta}^*) + \mathbf{q}_n\right] \tag{7}$$

which can be rewritten as

$$\boldsymbol{\theta_n} - \boldsymbol{\theta}^* = \boldsymbol{\theta_{n-1}} - \boldsymbol{\theta}^* - \epsilon_n \left[\mathbf{G}(\boldsymbol{\theta}^*)^T(\boldsymbol{\theta_{n-1}} - \boldsymbol{\theta}^*) + \mathbf{q}_n\right] \tag{8}$$

Denoting the error signal as the difference between the instantaneous parameter vector and the sub-optimal parameter vector as $\mathbf{e}_n = \boldsymbol{\theta_n} - \boldsymbol{\theta}^*$, equation (8) leads to

$$\mathbf{e}_n = \mathbf{e}_{n-1} - \epsilon_n \left[\mathbf{G}(\boldsymbol{\theta}^*)^T\mathbf{e}_{n-1} + \mathbf{q}_n\right]. \tag{9}$$

If $\epsilon_n = \epsilon$ is kept constant, then equation (9) can be analyzed in the spectral domain using z-transforms which leads to:

$$\mathbf{E}(z) = -\left[\mathbf{I} + z^{-1}(\epsilon\mathbf{G}(\boldsymbol{\theta}^*) - \mathbf{I})\right]^{-1}\epsilon\mathbf{Q}(z) \tag{10}$$

To understand the spectral properties of equation (10), consider the case when $\theta \in \mathcal{R}$ is a scalar parameter. Equation (10) can be reduced to

$$E(z) = \frac{-\epsilon Q(z)}{1 - z^{-1} + \epsilon G(\theta^*)z^{-1}}. \tag{11}$$

At frequencies $\omega \ll \omega_s$, where ω_s is the frequency of parameter updates $z^{-1} \approx 1 - j\frac{\omega}{\omega_s}$ and hence the power spectral density (PSD) of the error signal $E(\omega)$ can be written as

$$\frac{|E(\omega)|}{|Q(\omega)|} = \frac{1}{G(\theta^*)}\left[1 + \frac{\omega^2}{\omega_p^2}\right]^{-1/2} \tag{12}$$

where

$$\omega_p = \frac{\epsilon\omega_s G(\theta^*)}{1 - \epsilon G(\theta^*)}. \tag{13}$$

If $e_n^E = \theta_n^E - \theta^*$ denotes the error between the estimated value of the system parameter θ_n^E and the actual parameter θ^*, then the mean-square error (MSE) $\mathcal{E}_{QGD}$ is given by

$$\mathcal{E}_{QGD} = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} |e_n^E|^2. \tag{14}$$

Because the adaptation process is much slower than the rate at which the system parameters are updated, we can assume that the adaptation process is bandlimited between $[-\omega_a \ \omega_a]$ Hz, with $\omega_a \ll \omega_s$. Under this assumption, an ideal reconstruction filter $P(z)$ in Fig. 3 would be a brick-wall filter given by

$$P(\omega) = \begin{cases} 1; & |\omega| \leq \omega_a \\ 0; & |\omega| > \omega_a \end{cases} \tag{15}$$

Thus, the spectrum of the error e_n^E is related to $E(\omega)$ according to

$$E^E(\omega) = P(\omega)E(\omega) \tag{16}$$

and equation (14) can be rewritten using Parseval's theorem as

$$\begin{aligned} \mathcal{E}_{QGD} &= \int_{-\infty}^{\infty} |E^E(\omega)|^2 d\omega & (17)\\ &= \int_{-\infty}^{\infty} |P(\omega)|^2 |E(\omega)|^2 d\omega & (18)\\ &= \int_{-\omega_a}^{\omega_a} |E(\omega)|^2 d\omega & (19) \end{aligned}$$

Combining equations (19) and (12) leads to

$$\mathcal{E}_{QGD} = \int_{-\omega_a}^{\omega_a} \frac{|Q(\omega)|^2}{G(\theta^*)} \left[1 + \frac{\omega^2}{\omega_p^2}\right]^{-1} d\omega. \tag{20}$$

To simplify the analysis, we will assume that the quantization noise exhibits a white-noise spectrum (a commonly used assumption [15], [19]) within the frequency band $[-\omega_s \ \omega_s]$. If the combined resolution of the ADC and the DAC in Fig. 1 is assumed to be B bits, then the power spectral density of the quantization noise is given by $|Q(\omega)|^2 = 2^{-2B}/12\omega_s$ [15], [19]. Note that the white-noise assumption is not entirely valid for a single-bit quantizer. In this case, one could employ a "dithering" technique where white-noise is synthetically injected before the quantizer [15]. The total power $\mathcal{E}$ within the adaptation band can be found to be

$$\mathcal{E}_{QGD} = \frac{2^{-2B}\omega_p}{6G(\theta^*)\omega_s} \tan^{-1}\left(\frac{\omega_a}{\omega_p}\right). \tag{21}$$

C. First-order Noise Shaping Gradient Descent Algorithm

One of the approaches for reducing the effect of quantization noise is to use noise-shaping techniques to move the noise power out of the adaptation band. Noise-shaping principles have been used extensively in the design of $\Sigma\Delta$ analog-to-digital converters [15]. However, the method has not been applied to online optimization with the exception of our previous work [28], [29] where $\Sigma\Delta$ modulation was obtained as a consequence of an online learning algorithm. An architecture implementing first-order noise-shaping for the $\Sigma\Delta$ gradient-descent (SGD) algorithm is shown in Fig. 4.

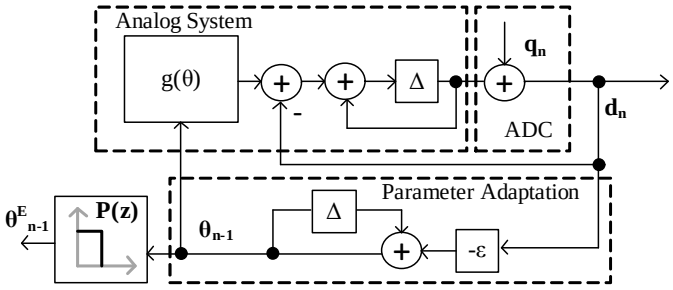


Fig. 4. System architecture showing the signal-flow of a $\Sigma\Delta$ gradient descent algorithm.

When compared to the previous approach as shown in Fig. 3, a $\Sigma\Delta$ modulator is used instead of the conventional quantizer (ADC). Please note that since the system parameter $\boldsymbol{\theta}$ is being continuously adapted in both the cases (using the output from the quantizer), the comparison between the proposed and the conventional approach is reasonable. Also, the sampling-rate or parameter update-rate is the same for SGD and QGD and the traditional argument of $\Sigma\Delta$ oversampling is not directly applicable in the case of online learning. In fact, the novelty of the proposed $\Sigma\Delta$ gradient descent approach arises from the fact that the steepest gradient descent algorithm in (3) naturally performs low-pass filtering, and therefore SGD achieves a better calibration/convergence compared to the conventional QGD method.

The updates for the parameter $\boldsymbol{\theta}$ can be expressed as

$$\boldsymbol{\theta_n} = \boldsymbol{\theta_{n-1}} - \epsilon \left[\mathbf{g}(\boldsymbol{\theta_{n-1}}) + \mathbf{q}_n - \mathbf{q}_{n-1}\right]. \tag{22}$$

Using the Taylor series approximation of $\mathbf{g}(\boldsymbol{\theta_{n-1}})$ given by equation (6) in equation (22) leads to

$$\boldsymbol{\theta_n} = \boldsymbol{\theta_{n-1}} - \epsilon \left[\mathbf{G}(\boldsymbol{\theta^*})^T(\boldsymbol{\theta_{n-1}} - \boldsymbol{\theta^*}) + \mathbf{q}_n - \mathbf{q}_{n-1}\right]. \tag{23}$$

Similar to the analysis for the QGD algorithm, the z-transform of the error $\mathbf{e}_n = \boldsymbol{\theta_n} - \boldsymbol{\theta^*}$ can be obtained as

$$\mathbf{E}(z) = -\left[\mathbf{I} + z^{-1}(\epsilon\mathbf{G}(\boldsymbol{\theta^*}) - \mathbf{I})\right]^{-1} \epsilon(1 - z^{-1})\mathbf{Q}(z), \tag{24}$$

which for a one-dimensional system parameter can be expressed as

$$E(z) = \frac{-\epsilon\left(1 - z^{-1}\right)Q(z)}{1 - z^{-1} + \epsilon G(\theta^*)z^{-1}}. \tag{25}$$

At frequency $\omega \ll \omega_s$, with ω_s being the frequency of parameter updates, $z^{-1} \approx 1 - j\frac{\omega}{\omega_s}$. The error transfer function at frequency ω is given by

$$\frac{|E(\omega)|}{|Q(\omega)|} = \frac{\omega}{G(\theta^*)\omega_s} \left[1 + \frac{\omega^2}{\omega_p^2}\right]^{-1/2}. \tag{26}$$

where ω_p is given by equation (13). For the reconstruction filter $P(z)$ given by equation (15), the MSE $\mathcal{E}_{SGD}$ between the estimated value of the system parameter θ_n^E and the actual parameter θ^* is given by

$$\mathcal{E}_{SGD} = \int_{-\omega_a}^{\omega_a} \frac{\omega^2|Q(\omega)|^2}{G(\theta^*)\omega_s^2} \left[1 + \frac{\omega^2}{\omega_p^2}\right]^{-1} d\omega. \tag{27}$$

The MSE reduces to

$$\mathcal{E}_{SGD} = \frac{2^{-2B}}{6G(\theta^*)} \frac{\omega_p^2}{\omega_s^2} \left[\frac{\omega_a}{\omega_s} - \frac{\omega_p}{\omega_s} \tan^{-1}\left(\frac{\omega_a}{\omega_p}\right) \right] \tag{28}$$

under the assumption that quantization noise exhibits a white-noise spectrum.

The relative reduction in MSE can be expressed as the ratio of equations (28) and (21) according to

$$\frac{\mathcal{E}_{SGD}}{\mathcal{E}_{QGD}} = \frac{\omega_p^2}{\omega_s^2} \frac{\left[\frac{\omega_a}{\omega_p} - \frac{\omega_p}{\omega_s} \tan^{-1}\left(\frac{\omega_a}{\omega_p}\right) \right]}{\frac{\omega_p}{\omega_s} \tan^{-1}\left(\frac{\omega_a}{\omega_p}\right)}. \tag{29}$$

For the case $\omega_a \ll \omega_p$, equation (29) can be approximated by

$$\frac{\mathcal{E}_{SGD}}{\mathcal{E}_{QGD}} = \frac{1}{3}\left(\frac{\omega_a}{\omega_s}\right)^2. \tag{30}$$

In equation (30) we have used the Taylor series approximation $\tan^{-1}(z) \approx z - z^3/3$. Thus, the reduction in MSE within the adaptation band $R(dB) = 10\log_{10}\left(\frac{\mathcal{E}_{SGD}}{\mathcal{E}_{NGD}}\right)$ is given by

$$R(dB) = -4.7 + 20\log_{10}\left(\frac{\omega_a}{\omega_s}\right). \tag{31}$$

and is independent of the learning rate parameter ϵ and is only a function of the ratio of the sampling frequency ω_s and the magnitude of the adaptation bandwidth $2\omega_a$. This ratio can be viewed as an online learning equivalent of "oversampling ratio", a metric commonly used to characterize the resolution of a $\Sigma\Delta$ analog-to-digital converter. Thus, for larger values of ϵ, the relative MSE R is independent of the learning rate and can be significantly reduced by increasing the sampling frequency ω_s.

For the case $\omega_a \approx \omega_p$, equation (29) can be approximated by

$$\frac{\mathcal{E}_{SGD}}{\mathcal{E}_{QGD}} = \frac{4-\pi}{3\pi}\left(\frac{\omega_p}{\omega_s}\right)^2. \tag{32}$$

which can be written in terms of the learning rate parameter ϵ as

$$R(dB) = -1 + 20\log_{10}\left[\frac{1-\epsilon G(\theta^*)}{\epsilon G(\theta^*)}\right]. \tag{33}$$

Thus, for smaller values of ϵ the relative reduction in MSE is independent of the sampling frequency and is a monotonic function of ϵ. Note that the relative improvement in performance is not significant in the regions where the cost function has plateaus or regions around the local minima where $G(\theta^*) \approx 0$. We now verify these results for a specific example.

Example I: Consider a time-dependent cost function $L(\theta; n)$ of a one-dimensional parameter θ, given by

$$L(\theta; n) = \frac{1}{2}\left[\sin(n\omega_0 T_s) - \theta\right]^2 \tag{34}$$

where $\sin(n\omega_0 T_s)$ is a discrete-time sinusoidal signal at frequency ω_0 and sampled at every T_s seconds. Minimizing $L(\theta; n)$ with respect to θ will favor θ which closely tracks the sinusoidal signal. The gradient of L estimated at θ_{n-1} is given by $g(\theta_{n-1}) = \sin(n\omega_0 T_s) - \theta_{n-1}$ and the Hessian is given by $\mathbf{G}(\theta) = 1$. Thus, application of equation (12) leads to

$$\frac{|E(\omega)|}{|Q(\omega)|} = \left[1 + \frac{\omega^2}{\omega_p^2}\right]^{-1/2} \tag{35}$$

with

$$\omega_p = \frac{\epsilon\omega_s}{1-\epsilon}. \tag{36}$$

Fig. 5 compares the trajectory of the variable θ_n generated by the steepest gradient descent (GD), quantized gradient descent (QGD) and the proposed $\Sigma\Delta$ gradient descent (SGD) algorithms. The ideal or the target parameter θ^* is also shown in Fig. 5 for comparison. As expected, both QGD and SGD algorithms generate noisy representation of the target

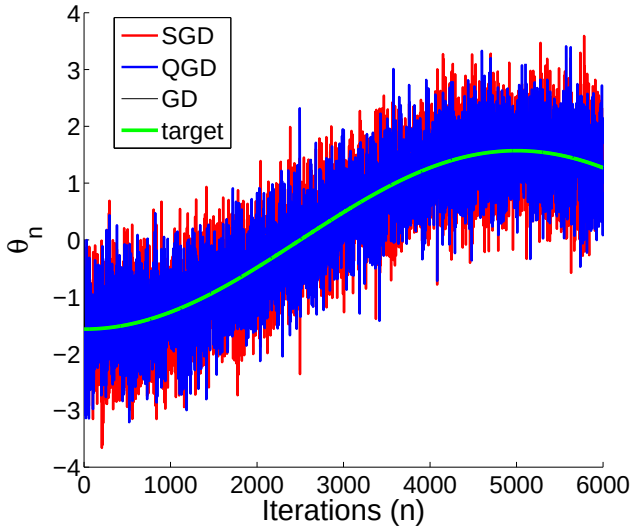


Fig. 5. Plot showing θ_n generated by the GD, QGD and SGD algorithm along with the ideal θ^*. The output of the GD algorithm is overlaid by the ideal trajectory

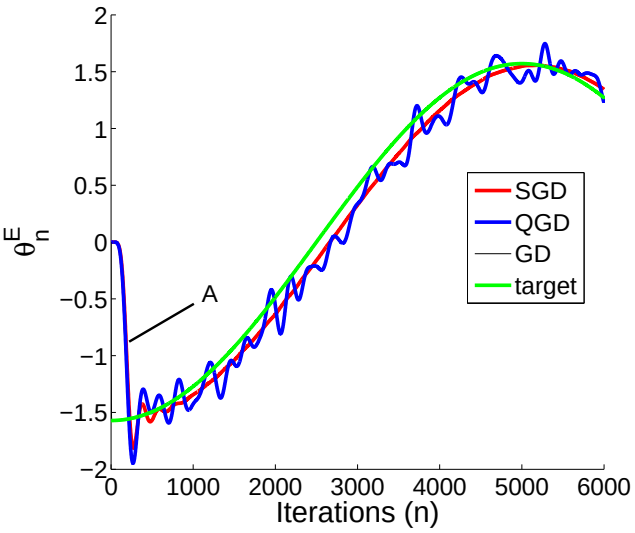


Fig. 6. Plot showing θ_n^E generated by the GD, QGD and SGD algorithm when additive white noise (-6dB) is used as a model for the quantization noise. The output of the GD algorithm is overlaid by the ideal trajectory

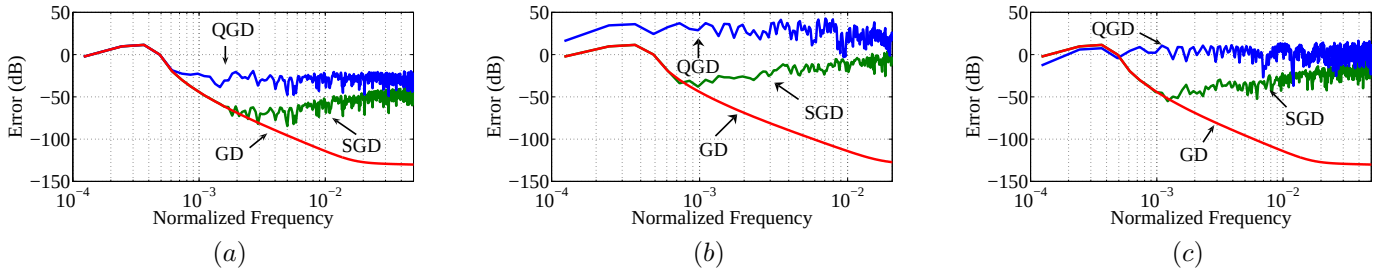


Fig. 7. Magnitude spectrum of the error signal in the case of ideal(GD),quantized(QGD) and $\Sigma\Delta$ gradient descent(SGD) algorithms for: (a) quantization noise level -30dB (relative to the magnitude of the input signal), (b) a 1-bit quantizer with ϵ=0.5; and (c) a 4-bit quantizer with ϵ=0.5.

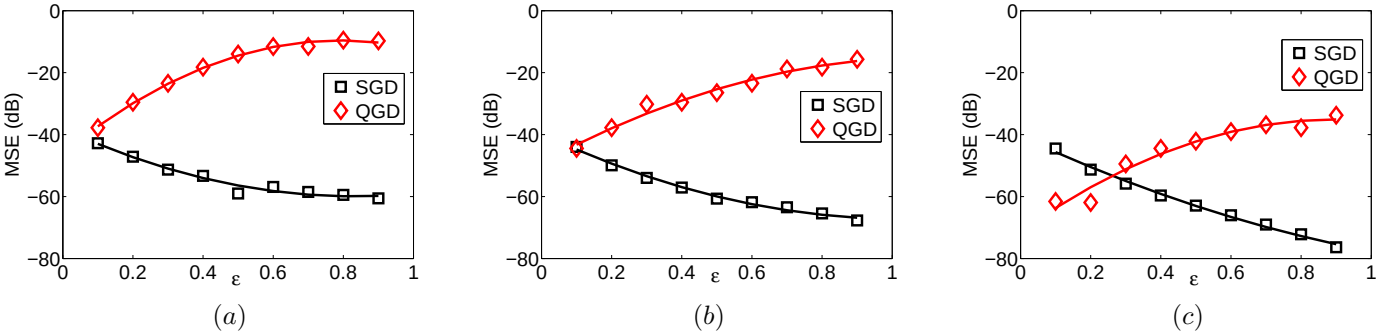


Fig. 8. SNR(in dB) vs adaptation parameter ϵ of gradient descent(GD), quantized gradient descent(QGD) and 1st order $\Sigma\Delta$ gradient descent(SGD) optimization algorithms with a n-bit quantizer for (a) n=1,(b) n=2 and (c) n=4.

parameter, where as the traditional GD algorithm can perfectly track the ideal parameter trajectory. However, when the noisy sequence θ_n is filtered using the reconstruction filter $P(z)$, one can see significant differences in convergence properties exhibited by QGD and SGD algorithms, as shown in Fig. 6. In Fig. 6, an 8^{th}-order Butterworth low-pass filter has been used as the reconstruction filter $P(z)$ to obtain θ_n^E. Note that the Butterworth filter "approximately" satisfies equation (15). Also, for this plot the quantization noise has been modeled using an additive white noise with power -6dB. The convergence plot can be divided into two regions: in region labeled A, the algorithm converges to the target trajectory starting from an initialization condition; where as in the second region the algorithms trajectory tracks the target trajectory. Note that there is a phase (or processing) delay between the parameter sequence produced by all the algorithms and the target sequence. This delay is determined by the choice of the learning rate parameter ϵ. It can be clearly seen from Fig. 6 that the convergence of SGD algorithm is superior (in terms of precision) when compared to the QGD algorithm.

The effect of quantization noise and the benefits of noise-shaping in SGD learning algorithm can be understood by analyzing the spectrum of the error signal $E(\omega)$. Fig. 7(a) shows the case when the quantization noise is modeled by additive white Gaussian noise at magnitude -30dB. All mathematical operations for this simulation have been implemented using floating-point arithmetic. Fig. 7(b) shows the equivalent result when a 1-bit quantizer is used as an ADC and Fig. 7(c) shows the case when 4-bit quantizer is used. The results are shown only for the adaptation frequency band which is $\omega \ll \omega_s$. The results clearly show the noise-shaping effect of the SGD learning algorithm when compared to the QGD algorithm, which results in a lower residual error within the adaptation band. Note that when a 1-bit quantizer is used, the SGD algorithm exhibits "idle-tones" in its spectral response similar to the case for a single-bit $\Sigma\Delta$ modulator [15]. However, the frequency of the "idle tones" are located at frequencies outside the adaptation band and hence can not be observed in 7(b). Also note that even for the floating-point gradient descent (GD) algorithm, the residual error is non-zero, which is due to the phase delay between target and the parameter sequence generated by different algorithms (see Fig. 6). The MSE for each of the algorithms can be computed by integrating the power of the error-signal $E(\Omega)$ within the adaptation band according to equations (19) and (27).

Fig. 8 (a)-(c) compares the MSE for the QGD and SGD algorithms where the MSE has been computed according to $MSE(dB) = 10\log_{10}\mathcal{E}$. The results illustrate several interesting trends which can also be explained based on the mathematical analysis given in Section II:

- As the value of the learning-rate parameter ϵ is increased, the MSE of the QGD algorithm increases. This is expected as near convergence, the magnitude of the oscillation (or noise) is proportional to ϵ. Mathematically, this is explained by equation (21) where ω_p increases as ϵ is increased. However, for the SGD algorithm, the relative MSE reduces with the increase in ϵ as can be seen when the equations (31) and (33) are compared. Conceptually, this phenomenon occurs due to two counter-balancing principles. Increasing the learning rate makes the learning process faster which enables the algorithm to converge faster or track a time-varying system parameter. However,

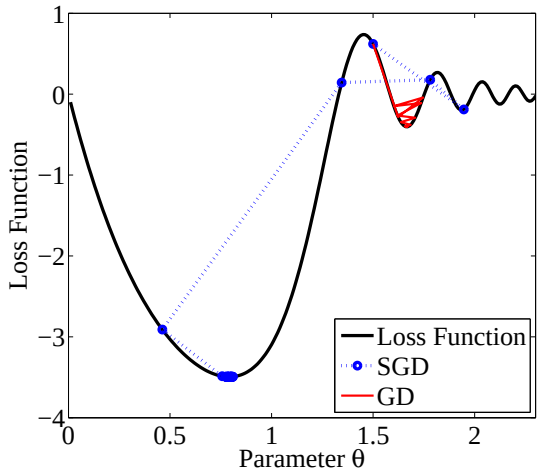


Fig. 9. Convergence of ΣΔ gradient descent algorithm to global minima.

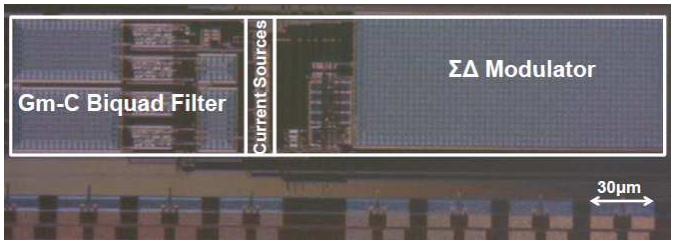


Fig. 10. Micrograph of the prototype integrating different circuit modules used to verify the proposed ΣΔ gradient-descent based calibration.

increase in learning rate also increases the effect of quantization noise on the convergence to the local minima. In the case of the SGD algorithm, noise-shaping moves the quantization noise out of the adaptation band and hence when the learning rate is increased, the deterioration due to increase in the precision error is smaller compared to performance improvement due to the reduction in tracking error (or speed of convergence). Thus, a larger value of ϵ enhances the noise-shaping property of SGD and, as we will see later, also enhances the robustness of the SGD algorithm to spurious local minima.

- At lower values of ϵ, the difference between the MSE achieved by the QGD and SGD algorithm reduces. In fact, at higher resolution of the ADC, the MSE of the QGD algorithm is lower (not significantly) than that of the SGD algorithm. This effect is also explained by comparing the MSE equations (21) and (28) where the contribution due to the inverse tangent function remains bounded irrespective of the value of ϵ. However, note that reducing ϵ also increases the phase-delay between the target parameter and the estimated parameter as is illustrated in Fig. 6.

The consequence of results in Fig. 8 (a)-(c), is that at a larger magnitude of the learning rate ϵ, the SGD algorithm has superior convergence properties compared to QGD. However, larger value of ϵ along with the stochastic nature of the SGD updates makes the algorithm less susceptible to getting trapped in local minima. This is illustrated in Fig. 9, where we have considered an example of a non-linear cost function. The conventional floating-point GD gets trapped to a local minimum whereas the SGD algorithm successfully converges to a more optimal minimum. Note that due to quantization noise, QGD algorithm will also have similar search properties as SGD. Thus, the proposed SGD algorithm achieves a superior trade-off between the precision or speed of convergence and robustness when compared against other online gradient-based techniques. Note that larger magnitude of ϵ would also imply that the SGD algorithm could potentially escape from the global minima. Therefore, the value of ϵ needs to be appropriately chosen based on an a-priori knowledge about the nature of the cost function. Note that the exact mathematical form of the cost-function need not be known a-priori.

III. Applications in Digitally Calibrated Analog Circuits

In this section, we demonstrate the application of SGD algorithm for digital calibration of analog circuits. The circuits have been prototyped in a 0.5μm CMOS process and the micrograph of the prototype is shown in Fig. 10. The circuit modules include a continuous-time first-order ΣΔ modulator which implements the signal-flow shown in Fig. 4 and a current-mode digital-to-analog converter (DAC) with a serial programming interface (SPI) for storage and adaptation of circuit parameters. The output of the modulator is sampled by an off-chip field-programmable gate array (FPGA) which implements the online adaptation according to the Fig. 4. The FPGA then programs the on-chip current DAC. The prototype also integrates current sources and continuous-time Gm-C filters which have been used in the examples discussed in this section.

A. Current tracking and current balancing

Our first setup, shown in Fig. 11(a) implements the tracking example which was described in Section II-C. The tracking signal is a current sink I_{SIG} which is shown in Fig. 11(a) and the system parameter is the current I_{DAC} which is programmed using a DAC as shown in Fig. 11(a). The schematic of the current DAC is shown in Fig. 11(c) which uses a MOS based resistive divider that has been previously reported in literature [30], [31]. For this example, the objective function $L(I_{DAC})$ is given by

$$L(I_{DAC}) = \frac{1}{2}\left(I_{SIG} - I_{DAC}\right)^2 \tag{37}$$

and the gradient with respect to I_{DAC} is given by

$$g(I_{DAC}) = \frac{\partial L}{\partial I_{DAC}} = I_{SIG} - I_{DAC}. \tag{38}$$

As shown in Fig. 11(a), the gradient is the input current to the first-order continuous time ΣΔ modulator whose schematic is shown in Fig. 11(b). The circuit level description of the ΣΔ has been reported elsewhere [32] and has been omitted here for the sake of brevity.

Fig. 12 shows the measured I_{DAC} when the binary input to the DAC is varied. The results are shown for two different levels of reference current I_{REF}. The results show a

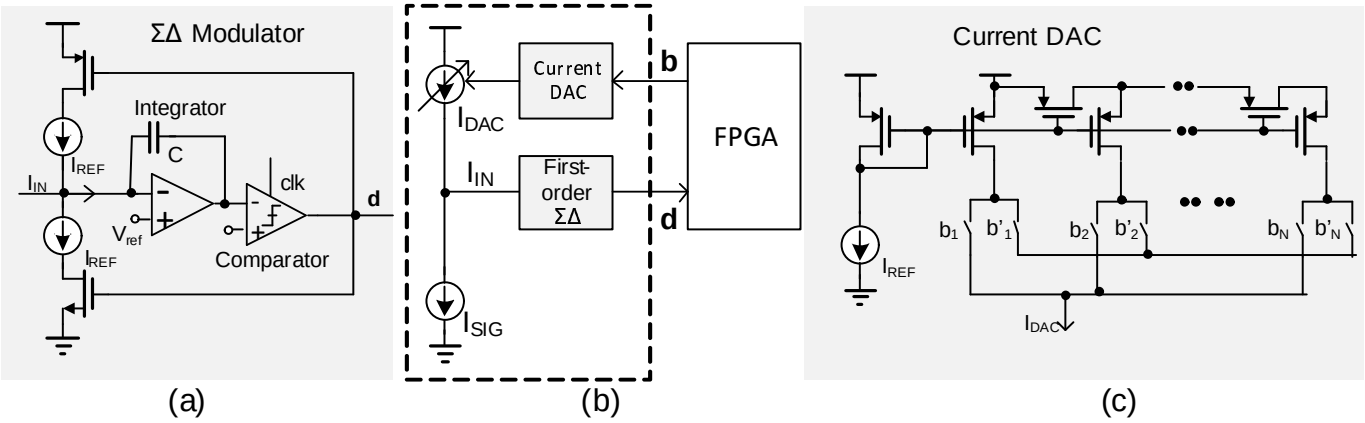


Fig. 11. Circuit illustrating the application of first-order SGD algorithm for current tracking and balancing.

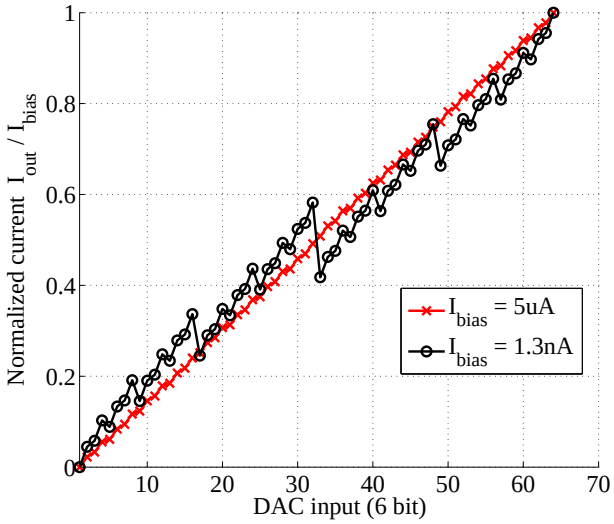


Fig. 12. Non-linear response of the current DAC when biased in weak-inversion.

non-monotonic response and the monotonicity of the DAC deteriorates when the reference current I_{REF} is reduced. This behavior is well documented in literature for the DAC topology [31] and is attributed to the mismatch between the MOS transistors which becomes amplified at low bias currents. The calibration DACs therefore provide 10-bit of programmability of the system parameters but exhibit only 6-bits of linearity. However, these artifacts are ideal to demonstrate the benefits the $\Sigma\Delta$ gradient-descent algorithm. The measured cost-function $L(I_{DAC})$ is shown in Fig. 13(a) which shows the presence of multiple local minima. Fig. 13(b) shows the running average of $\Sigma\Delta$ modulator output when the parameter is adapted according to the SGD algorithm. The result shows that after the initial search, running average of the modulator output or equivalently the average of the system gradient converges to zero. Thus, the proposed architecture converges to a solution where the cost function approaches the global minima ($I_{DAC} \approx I_{SIG}$).

B. Center-frequency Calibration in band-pass Gm-C Filters

The second example applies the SGD algorithm for calibrating the center-frequency of continuous-time band-pass Gm-C filters that have been commonly used for designing silicon cochlea [33]–[35]. Many of these applications require that the filters consume a minimal amount of power and in many instances could have center frequencies as low as 100Hz. This requires biasing the transistors in weak-inversion which, makes the design more susceptible to analog artifacts like mismatch and distortion [11]. One of the possible approaches to calibrate for these analog artifacts is to incorporate programmable circuit elements (e.g. DACs) which can be calibrated to obtain the desired performance metric.

Fig. 14(b) shows the circuit level schematic of a standard Gm-C biquad filter stage which consists of three linearized transconductors whose transconductance parameters are given by g_{m1}, g_{m2} and g_{m3}. The schematic of the transconductor is shown in Fig. 14(c) and was reported and described in [32]. The circuit uses an active source degeneration technique based on the cross-coupled pMOS transistors $M5 - M8$. Note that the input and the mirrors in the schematic are cascoded (even though it is not explicitly shown in the schematic). Assuming that the transconductor are linear, the output of the band-pass biquad filter stage is given by:

$$V_{bp}(s) = \frac{G\frac{\omega_0}{Q}s}{s^2 + \frac{\omega_0}{Q}s + \omega_0^2} V_{in1}(s) \tag{39}$$

where the center frequency ω_0,the filter gain G and the quality factor Q of the filter are given by

$$\omega_0 = \frac{\sqrt{g_{m2}g_{m3}}}{C} \tag{40}$$

$$Q = \sqrt{\frac{g_{m3}}{g_{m1}}} \tag{41}$$

$$G = \frac{g_{m1}}{g_{m2}}. \tag{42}$$

The transconductance parameters $g_{m1} - g_{m3}$ are determined by the bias current of the transconductor as shown in Fig. 14(c), which are programmed using the on-chip 10-bit current DACs (schematic shown in Fig. 12). The magnitude response of the filter is measured by rectifying the filter

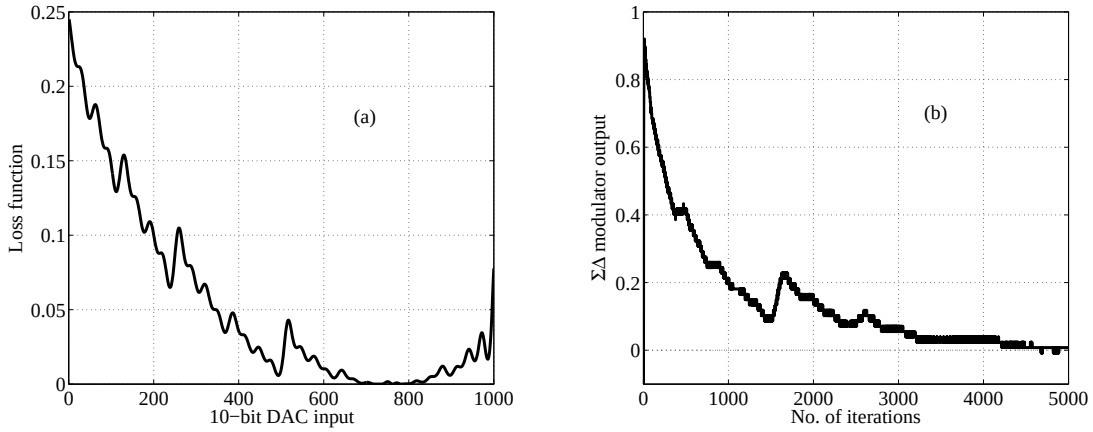


Fig. 13. Figure showing (a) the loss function $\frac{1}{2}\hat{I}_{in}^2$ when the input to the 10-bit current DAC is varied; and (b) the running average of the $\Sigma\Delta$ modulator output. The adaptation rate for this experiment was $\omega_s = 250$kHz.

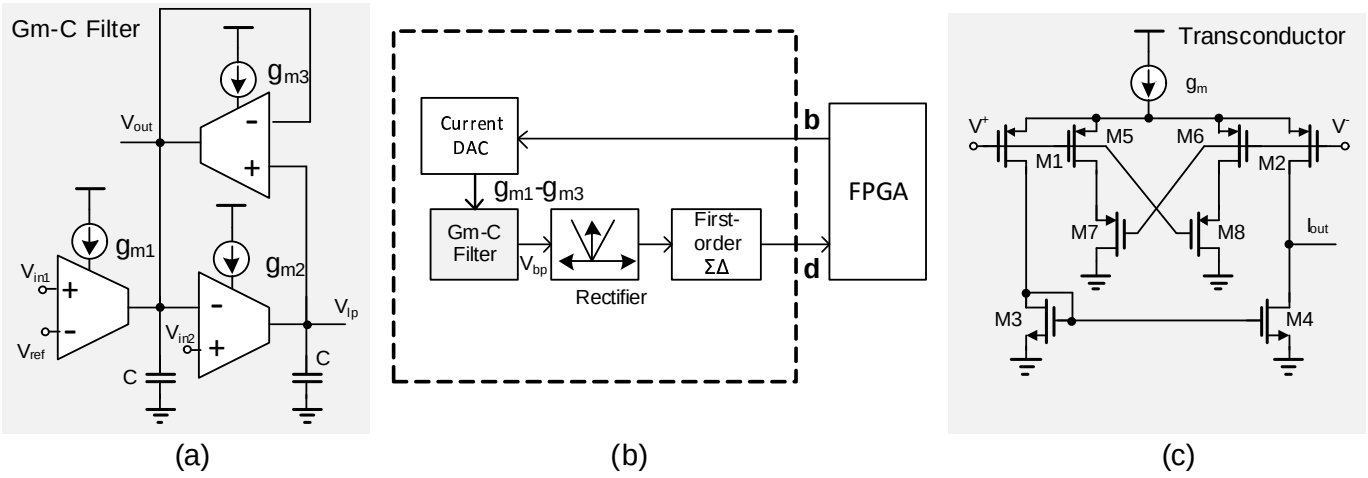


Fig. 14. Circuit illustrating the application of $\Sigma\Delta$ gradient descent algorithm for center frequency calibration of a Gm-C bandpass filter.

output followed by digitization using a first-order continuous-time $\Sigma\Delta$ modulator, as shown in Fig. 11(a). Note that the input stage of the continuous-time $\Sigma\Delta$ modulator automatically performs low-pass filtering of the rectified output, thus eliminating the need for an envelope detector commonly used in the design of silicon cochlea. Fig. 15 shows the measured magnitude response using the set-up in Fig. 11. In this experiment, the gain G and the quality factor Q are kept constant while the center frequency ω_0 is incrementally programmed by adjusting the transconductances $g_{m1} - g_{m3}$. The DAC reference current is set to 1nA which ensures that the MOS transistors are biased in the deep sub-threshold regime. The measured response shows that the center frequencies and the filter gains vary non-monotonically and is primarily attributed to the non-linearity of the programming DACs (see Fig. 12). We now show that, in spite of the non-monotonic response of the filters, the proposed $\Sigma\Delta$ gradient descent algorithm can be used for robust calibration of the filter center frequencies.

To model the calibration procedure as an equivalent minimization problem, we exploit the fact that the biquad filter in Fig.11(a) exhibits a low-pass characteristic with respect to the input V_{in2} of the transconductor g_{m3}. Using superposition, the output V_{out} can be written as

$$V_{out}(s) = H_{bp}(s)V_{in1}(s) + H_{lp}(s)V_{in2}(s) \tag{43}$$

where

$$H_{lp}(s) = \frac{G}{s^2 + \frac{\omega_0}{Q}s + \omega_0^2} \tag{44}$$

and H_{bp} is the transfer function as in equation (39).

Fig. 16(left) shows the frequency response of the band-pass H_{bp} and low-pass H_{lp} filter stages of the biquad filter. If a tone with frequency ω_0 is applied to the inputs V_{in1} and V_{in2}, the magnitude of the output V_{out} reaches its minimum when the center frequency of the biquad matches ω_0. This property can now be exploited to formulate an optimization problem whose solution is a filter whose center-frequency is calibrated to the desired frequency ω_0. The band-pass filter, however, has a zero at DC which contributes to an additional 90^o phase shift in the transfer function (43). Therefore, a 90^o phase difference is synthetically introduced between the inputs V_{in1} and V_{in2}. Fig. 16(center) shows the signals V_{in1} and V_{in2} and the output of the filter according to the equation (43).

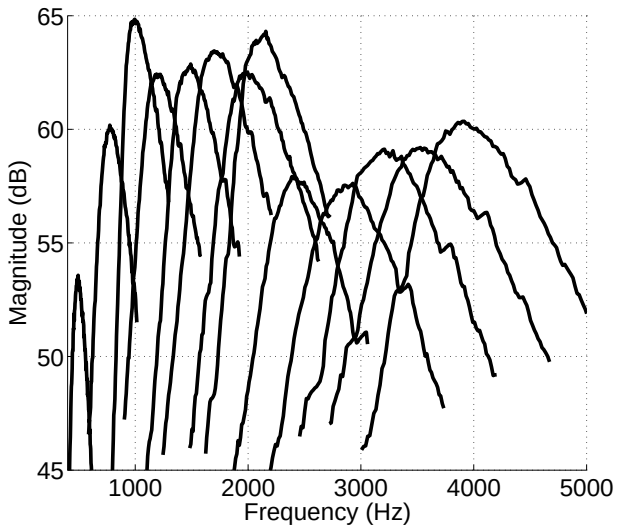


Fig. 15. Non-monotonic response of the biquad filter when the transconductance parameters g_{m1}, g_{m2} and g_{m3} are varied such that $G = 1$ and $Q = 1$.

TABLE I
MEASURED SPECIFICATIONS OF THE GM-C BANDPASS FILTER

Technology	0.5μm CMOS process
Size (including $\Sigma\Delta$ modulator)	1000μm $\times$ 300μm
Programmable Frequency Range	10Hz - 10kHz
Filter Linear Range (at Q = 1)	240mV
Power - Bandpass filter	300pW (100Hz) - 33nW (10kHz)

The continuous-time $\Sigma\Delta$ modulator computes the magnitude of the error signal and adjusts the value of g_{m2} while keeping the gain G and the quality factor Q constant. Note that in this optimization only g_{m2} is a free variable, where as other variables g_{m1} and g_{m3} are determined by the constraints. Fig. 16(right) shows the output of the modulator when the $\Sigma\Delta$ gradient-descent adaptation is applied. Indeed, the output of the modulator reaches a minima at the desired center frequency ω_0, in spite of the DAC and transconductor non-linearities. Note that the value of the minima is non-zero implying that the magnitude of the error signal cannot be reduced to zero. This limitation arises due to finite programming resolution of the current DACs. We have used this procedure to program the center frequencies of a biquad filter according to a logarithmic scale or Mel-scale which is desirable for many auditory frontends [33], [36]. Fig. 17 shows the measured response of the calibrated filter where the center frequencies of the filter have been successfully programmed to the desired frequencies. Table I summarizes the measured specifications of the Gm-C filter and the calibration DAC, where the nanowatt power dissipation is attributed to the sub-threshold biasing of the MOSFET transistors.

IV. CONCLUSIONS AND DISCUSSIONS

All online gradient-descent algorithms use some form of a first-order recursive low-pass filtering of the system gradient to estimate the optimal set of parameters. In this paper, we show that if a $\Sigma\Delta$ modulation step is inserted before the filtering, the quantization noise inserted by the use of finite-resolution ADCs and DACs is shaped out of the frequency bands containing the adaptation trajectories. We show that the resulting class of SGD algorithm not only exhibits superior convergence properties compared to the conventional QGD algorithms but it also exhibits superior parameter search properties compared to the floating-point gradient descent algorithms. This enables the SGD algorithm to avoid spurious local minima and search for more optimal values of the system parameters. This property of the SGD algorithm is beneficial, in particular, for calibration of digitally assisted analog circuits that use DACs for digital storage of analog parameters. Due to non-linearity and non-monotonicity of the DAC, the system error (quality) metric being optimized exhibits numerous spurious local minima. The hardware portion of this paper demonstrated the robustness of the SGD based calibration algorithm towards the DAC non-linearities for two specific applications: (a) current balancing and offset cancellation; and (b) center-frequency calibration of biquad filters biased in weak-inversion. Even though we have used a first-order $\Sigma\Delta$ modulator in the proposed SGD algorithm, the proposed framework can be extended to higher-order noise-shaping as well. However, in this case the first-order gradient updates given by equation (3) will be replaced by a higher-order recursive low-pass filter. Investigation into the effects of higher-order noise-shaping on the trade-off between the convergence and search properties of the online gradient-descent algorithm will be a topic of future research.

A limitation of the SGD approach is that the system gradients need be measured directly, which is not always possible. One possible solution which could be the basis of future research in this area is that the system gradient could be approximated using perturbation methods such as: simultaneous perturbative stochastic approximation (SPSA) [37] or finite different stochastic approximation (FDSA) [38], which only requires direct measurement of the system error (quality) metric (instead of its gradient). Another related area of research could be extending the formulation to higher-dimensional spaces where the proposed SGD framework could lead to noise-shaping not only in the frequency domain but also in the spatial domain [20], [28].

REFERENCES

[1] M. Pastre and M. Kayal, *Methodology for the Digital Calibration of Analog Circuits and Systems: with Case Studies*, Springer, 1st ed., 2006.
[2] B. Murmann, "Digitally assisted analog circuits," *IEEE Micro*, vol. 26, pp. 38–47, Mar. 2006.
[3] G. Cauwenberghs and M. Bayoumi, *Learning on Silicon*, Boston, MA: Kluwer Academic Publishers, 1999.
[4] B. Murmann, B. Boser, "A 12-bit 75-MS/s pipelined ADC using open-loop residue amplification," *IEEE J. Solid-State Circuits*, vol. 38, no. 12, pp. 2040–2050, Dec. 2003.
[5] J. A. Lenero-Bardallo, T. Serrano-Gotarredona, and B. Linares-Barranco, "A Calibration Technique for Very Low Current and Compact Tunable Neuromorphic Cells: Application to 5-bit 20-nA DACs," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 55, no. 6, pp. 522–526, Jun. 2008.
[6] R. J. Kier *et al.*, "Design and implementation of multipattern generators in analog VLSI," *IEEE Trans. Neural Netw.*, vol. 17, no. 4, pp. 1025–1038, Jul. 2006.

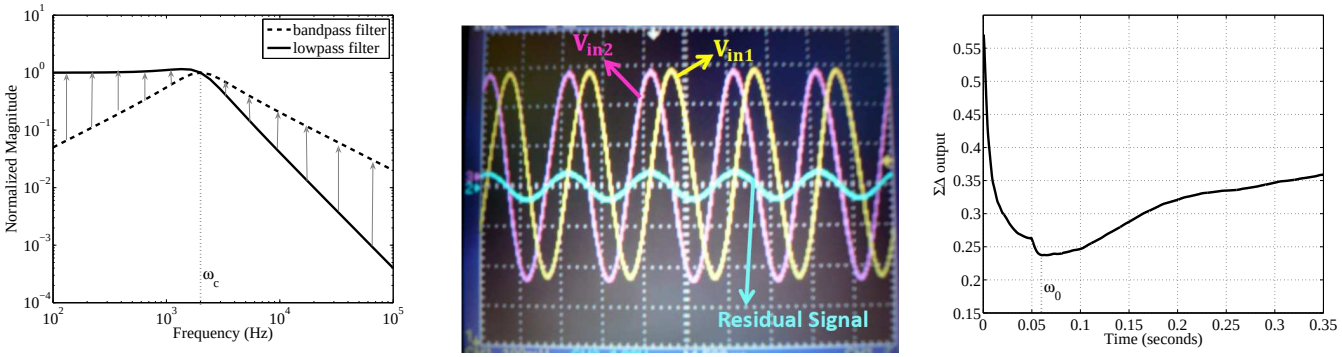


Fig. 16. Frequency calibration of the biquad bandpass filter: (a) figure showing the band-pass and the low-pass filter responses along with the error signal;(b) oscilloscope figure showing the input signals V_{in1} and V_{in2} along with the error output from the biquad filter; and (c) output of the $\Sigma\Delta$ modulator measured from the fabricated prototype, V_{out}.

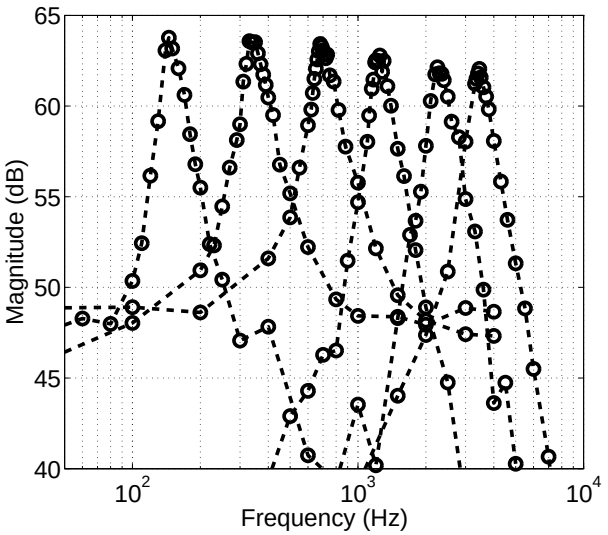


Fig. 17. Measured response of the biquad filter which is calibrated using the $\Sigma\Delta$ gradient descent algorithm to center frequencies chosen according to the Mel-scale [36].

[7] B. Linares-Barranco, T. Serrano-Gotarredona, and R. Serrano-Gotarredona,"Compact low-power calibration mini-DACs for neural massive arrays with programmable weights," *IEEE Trans. Neural Netw.*, vol. 14, no. 5, pp. 1207–1216, Sep. 2003.

[8] H. Juan, A. Murray, W. Dongqing, "Adaptive Visual and Auditory Map Alignment in Barn Owl Superior Colliculus and Its Neuromorphic Implementation", *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no: 9, pp. 1486 - 1497, 2012.

[9] T. Delbruck, R. Berner, P. Lichtsteiner, C. Dualibe, "32-bit Configurable bias current generator with sub-off-current capability", *Proceedings of 2010 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2010.

[10] D. W. Graham, E. Farquhar, B. Degnan, C. Gordon, P. Hasler, "Indirect Programming of Floating-Gate Transistors", *IEEE Transactions on Circuits and Systems I*, vol. 54, no: 5, pp. 951-93, 2007.

[11] P. R. Kinget, "Device mismatch and tradeoffs in the design of analog circuits," *IEEE J. Solid-State Circuits*, vol. 40, no. 6, pp. 1212–1224, Jun. 2005.

[12] G. Cauwenberghs and G. C. Temes, "Adaptive Digital Correction of Analog Errors in MASH ADCs – Part I. Off-Line and Blind On-Line Calibration," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 47, no. 7, pp. 621–628, Jul. 2000.

[13] G. Carvajal, M. Figueroa, D. Sbarbaro, W. Valenzuela, "Analysis and Compensation of the Effects of Analog VLSI Arithmetic on the LMS Algorithm", *IEEE Transactions on Neural Networks*, vol. 38 , no. 7, pp. 1046-1060, 2011.

[14] M. Cohen and G. Cauwenberghs, "Floating-Gate Adaptation for Focal-Plane On-Line Nonuniformity Correction," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 48, no. 1, pp. 83–89, Jan. 2001.

[15] J. C. Candy and G. C. Temes, *Oversampled methods for A/D and D/A conversion in Oversampled Delta-Sigma Data Converters*, Piscataway, NJ: IEEE Press, pp. 1–29, 1992.

[16] D. Bertsekas. *Non-linear Programming*, MA: Athena Scientific, 1995.

[17] S. Haykin, *Neural Networks: A comprehensive foundation*, 2nd ed., Prentice Hall, 1998.

[18] S. S. Haykin, *Least-Mean-Square Adaptive Filters*, B. Widrow Eds., Wiley, 2003.

[19] R. M. Gray,"Spectral analysis of quantization noise," *IEEE Trans. Inf. Theory*, vol. 36, pp. 1220–1244, Nov. 1990.

[20] L. Bottou, "Stochastic Learning," *Advanced Lectures in Machine Learning*, Springer, 2003, pp. 146–168.

[21] H. Robbins and S. Monro, "A Stochastic approximation method," Ann. Math. Statist., vol. 22, no. 3, pp. 400–407, 1951.

[22] H. Kushner, *Introduction to Stochostic Control*, New York: Holt, Rinehart and Winston, 1971.

[23] J. R. Blum, "Approximation methods which converge with probability one," Ann. Math. Statist., vol. 25, pp. 382-386, 1954.

[24] C. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, 1995.

[25] S. Gelfald and S. K. Mitter, "Recursive Stochastic Algorithms for Global Optimization," *J. Control and Optim.*, vol. 29, pp. 999–1018, Sep. 1991.

[26] H. Fang, G. Gng, and M. Qian, "Annealing of Iterative Stochastic Schemes," *J. Control and Optim.*, vol. 35, pp. 1886–1907, 1997.

[27] S. Gelfald and S. K. Mitter, "Simulated annealing with noisy or imprecise energy measurements," *J. Optim. Theor. Appl.*, vol. 62, pp. 49–62, 1989.

[28] A. Gore and S. Chakrabartty, "Min-Max Optimization Framework for Designing $\Sigma\Delta$ Learners: Theory and Hardware," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 57, no. 3, pp. 604–617, 2010.

[29] A. Fazel, A. Gore, and S. Chakrabartty, "Resolution Enhancement in sigma-delta Learners for Super-Resolution Source Separation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1193–1204, 2010.

[30] K. Bult and J. G. M. Geelen,"An inherently linear and compact MOST-only current division technique," IEEE J. Solid-State Circuits, vol 27, pp.1730-1735, Dec.1992.

[31] B. Linares-Barranco and T. Serrano-Gotarredona, "On the design and characterization of femtoampere current-mode circuits, IEEE Journal of Solid-State Circuits, Vol. 38, no:8, pp:1353 - 1363, 2003.

[32] A. Gore, S. Chakrabartty, S. Pal and E. C. Alocilja,"A Multichannel Femtoampere-Sensitivity Potentiostat Array for Biosensing Applications," in IEEE Transactions of Circuits and Systems I:Regular Papers,Vol. 53,Issue 11, Nov. 2006.

[33] L. Watts *et al.*, "Improved implementation of the silicon cochlea," *IEEE J. Solid-State Circuits*, vol. 27, no. 5, pp. 692–700, May 1992.

[34] V. Chan, S. C. Liu, and A. van Schaik, "AER EAR: A matched silicon cochlea pair with address event representation interface," *IEEE Trans.*

Circuits Syst. I: Special Issue on Smart Sensors, vol. 54, no. 1, pp. 48–59, 2007.

[35] R. Sarpeshkar, *et al.*, "An Ultra-Low-Power Programmable Analog Bionic Ear Processor," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 4, pp. 711–727, April 2005.

[36] A. Fazel, S. Chakrabartty, "Statistical Pattern Recognition Techniques for Speaker Verification," *IEEE Circuits Syst. Mag.*, vol. 11, no. 2, pp. 62–81, 2011

[37] J. C. Spall, "Multivariate stochastic approximation using a simultaneous perturbation gradient approximation," *IEEE Trans. Automat. Control*, vol. 37, pp. 332–341, 1992.

[38] J. Kiefer and J. Wolfowitz, "Stochastic estimate of a regression function," *Ann. Math. Statistics.*, vol. 23, pp. 462–466, 1952.

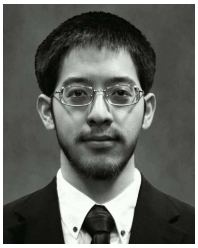
Shantanu Chakrabartty (SM'99-M'04-S'09) received his B.Tech degree from Indian Institute of Technology, Delhi in 1996, M.S. and Ph.D in Electrical Engineering from Johns Hopkins University, Baltimore, MD in 2002 and 2005 respectively.

He is currently an associate professor in the department of electrical and computer engineering at Michigan State University (MSU). From 1996-1999 he was with Qualcomm Incorporated, San Diego and during 2002 he was a visiting researcher at The University of Tokyo.

Dr. Chakrabartty's work covers different aspects of analog computing, in particular non-volatile circuits, and his current research interests include energy harvesting sensors and neuromorphic and hybrid circuits and systems. Dr. Chakrabartty was a Catalyst foundation fellow from 1999-2004 and is a recipient of National Science Foundation's CAREER award, University Teacher-Scholar Award from MSU and the 2012 Technology of the Year Award from MSU Technologies. Dr. Chakrabartty is a senior member of the IEEE and is currently serving as the associate editor for *IEEE Transactions of Biomedical Circuits and Systems*, associate editor for the *Advances in Artificial Neural Systems* journal and a review editor for *Frontiers of Neuromorphic Engineering* journal.

Ravi K. Shaga received his B.Tech degree from Indian Institute of Technology, Kharagpur in 2007 and a M.S. degree in Electrical and Computer Engineering from Michigan State University, East Lansing, MI in 2011.

He is currently with Intel Corp., Hillsboro, Oregon, and his research interests include analog integrated circuits and high-speed digital circuits.



Kenji Aono (S'09) received both a B.S. in electrical engineering and a B.S. in computer engineering, with honors, from Michigan State University, East Lansing, MI, U.S.A. in 2011.

He was employed as a Technical Aide in the Adaptive Integrated Microsystems Laboratory after finishing his undergraduate work in early 2011. He is currently a Research Assistant at the same facility while enrolled at Michigan State University to complete work towards a master's degree in electrical engineering.

Mr. Aono is an active officer for the Michigan Alpha chapter of Tau Beta Pi and the Gamma Zeta chapter of IEEE-Eta Kappa Nu. He is a recipient of the graduate fellowship from the Michigan Space Grant Consortium for the 2012-2013 year.