Low-Power Die-Level Process Variation and Temperature Monitors for Yield Analysis and Optimization in Deep-Submicron CMOS

Amir Zjajo, *Member, IEEE*, Manuel J. Barragan, and José Pineda de Gyvez, *Fellow, IEEE*

Abstract—**This paper reports design, efficiency, and measurement results of the process variation and temperature monitors for yield analysis and enhancement in deep-submicron CMOS circuits. Additionally, to guide the verification process with the information obtained through monitoring, two efficient algorithms based on an expectation–maximization method and adjusted support vector machine classifier are proposed. The monitors and algorithms are evaluated on a prototype 12-bit analog-to-digital converter fabricated in standard single poly six-metal 90-nm CMOS.**

Index Terms—**Analog test, process variation monitoring, temperature monitors, yield enhancement.**

I. Introduction

CMOS technologies move steadily toward finer geometries, which provide higher digital capacity, lower dynamic power consumption, and smaller area resulting in the integration of whole systems, or large parts of systems, on the same chip. However, due to technology scaling, ICs are becoming more susceptible to variations in process parameters and noise effects like power supply noise, crosstalk reduced supply voltage and threshold voltage operation severely impacting the yield [1]. Since parameter variations depend on unforeseen operational conditions, chips may fail despite passing standard test procedures. Similarly, the magnitude of thermal gradients and associated thermomechanical stress increase further as CMOS designs move into nanometer processes and multigigahertz frequencies [1]. Higher temperature increases the risk of damaging the devices and interconnects since major back-end and front-end reliability issues, including electromigration, time-dependent dielectric breakdown, and negative-bias temperature instability, have strong dependence on temperature. Consequently, continuous observation of process variation and thermal monitoring becomes a necessity. Such observation is enhanced with dedicated monitors embedded within the functional cores [2]. To maximize the coverage, the process variation and thermal sensing devices are scattered across the entire chip to meet the control requirements. The monitors are networked by an underlying infrastructure, which provides the bias currents to the sensing devices, collects measurements, and performs analog to digital signal conversion. Therefore, the supporting infrastructure is an on-chip element at a global scale, growing in complexity with each emerging design.

The process variation and temperature monitors for signal integrity measurement systems of very large scale integration (VLSI) circuits should meet several requirements, including compatibility with the target process with no additional fabrication steps, high accuracy, a small silicon area, and low power consumption. In a ring-oscillator-based technique [3], isolation of individual parameters for variability study is challenging due to mixture of the variation of large number of transistors into a single parameter (i.e., the frequency of ring operation). On the other hand, the transistor array based structures [4] enable collection of transistor I–V curves with digital I/O, enabling measurement of I–V characteristics of a larger number of devices than is typically sustained by common DC probing measurement schemes. Such structures use row and column decoders to select an individual transistor in the transistor array and employ different schemes to address the current–resistance drop imposed by the transmission gates on a transistor's selection path. The temperature monitor based on a time-to-digital-converter [5] is constrained by the large area and power overhead at the required sampling rate. Temperature monitor operating in the subthreshold region [6] is prone to dynamic variations as thermal sensitivity increases by an order of magnitude when operating in subthreshold [7]. Consequently, the majority of CMOS temperature monitors are based on the temperature characteristics of parasitic bipolar transistors [8].

In this paper, we present compact, low-area, low-power process variation and temperature monitors with high accuracy and a wide temperature range that does not need to operate with special requirements on technology, design, layout, testing, or operation. The monitors operate at the local power supply and are designed to maximize the sensitivity of the circuit to the target parameter to be measured. The monitors are small, stand alone, and easily scalable and can be fully switched off. All the peripheral circuits, such as decoders and latches, are implemented with thick gate oxide and long channel devices and are, hence, less sensitive to the process variation. To

Manuscript received October 6, 2011; revised November 16, 2011; accepted December 22, 2011. Date of publication February 6, 2012; date of current version July 13, 2012. The Associate Editor coordinating the review process for this paper was Dr. Wendy Van Moer.

A. Zjajo is with the Delft University of Technology, 2628 CD Delft, The Netherlands.

M. J. Barragan is with the National Center of Microelectronics, Microelectronics Institute of Seville (IMSE-CNM), University of Seville, 41092 Seville, Spain.

J. P. de Gyvez is with the Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands, and also with NXP Semiconductors, 5656 AE Eindhoven, The Netherlands.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIM.2012.2184195

0018-9456/$31.00 © 2012 IEEE

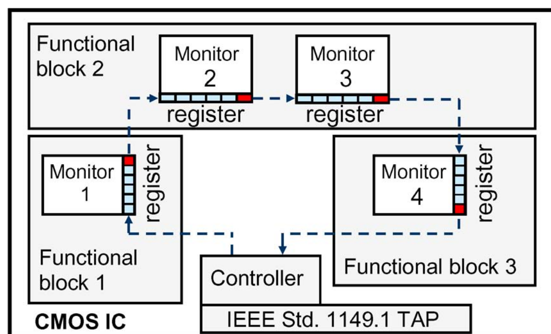

Fig. 1. Architecture of the measurement system.

characterize current process variability conditions and enable test guidance based on the data obtained from the monitors, we utilize the expectation–maximization (EM) algorithm [9] and the adjusted support vector machine (ASVM) classifier [10], respectively.

This paper is organized as follows. Section II focuses on the observation strategy and design of process variation and temperature monitors. Section III discusses the algorithms for verification process and test-limit guidance and update. In Section IV, the proposed monitors and algorithms are evaluated on an application example, namely, dual-residue multistep A/D converter. Finally, Section V provides a summary and the main conclusion.

II. Die-Level Process Variation and Temperature Monitors

A. Observation Strategy

Yield loss can be caused by several factors, e.g., wafer defects and contamination, IC manufacturing process defects and contamination, process variations, packaging problems, and design errors or inconsiderate design implementations or methods. Constant testing in various stages is of utmost importance for minimizing costs and improving quality. Fig. 1 depicts the proposed observation strategy block diagram for dice wafer probing. A family of built-in process variation and temperature sensing circuits is embedded within the functional blocks. The monitors in a core are connected through a bus to the controller [2]. The monitors operate at the local power supply and are designed to maximize the sensitivity of the circuit to the target parameter to be measured.

The monitors are small, stand alone, and easily scalable and can be fully switched off. The analog sensing is converted locally into pass/fail (digital) signals through the data decision circuit. The output of a monitor is a digital signal, which is transferred to the monitoring processor. The interface circuitry allows the external controllability of the test and also feeds out the decision of the detector to a scan chain. This register chain provides a serial connection between the various monitors in the different cores at minimum costs in terms of data communication and wiring. The test control block (TCB) in scan-chain selects through a test multiplexer (TMX) the individual die-level process monitor circuit measurement. Select, reference, and timing window signals are offered to the detector through this interface circuitry. All (critical) signal paths and clock lines have been extensively shielded. All the peripheral circuits, such as decoders and latches, are implemented by I/O devices (thick gate oxide and long channel devices) and, thus, are less sensitive to the process variation. The monitors have a 1-bit output; the accuracy of the measurement is achieved by logarithmically stepping through the range (successive approximation). The scan-chain is implemented through the IEEE Std 1149.4 analog test bus extension to 1149.1. The serial shift register is a user register controlled by an IEEE Std 1149.1 TAP controller [11], which allows access to the serial register, while the device is in functional mode. Furthermore, such controller creates no additional pin counts since it is already available in the system on chip. Another mode of operation allows self-test: the controller continuously interrogates the monitors for their measurements and will react to preset conditions (e.g., too high a temperature in a block). The architecture can also be operated in slave mode: an external controller (e.g., a tester workstation or a PC with 1149.1 control software) will program the monitor settings and evaluate the measured values. The monitors are designed in standard cell format so that they can be automatically located anywhere within each standard-cell block.

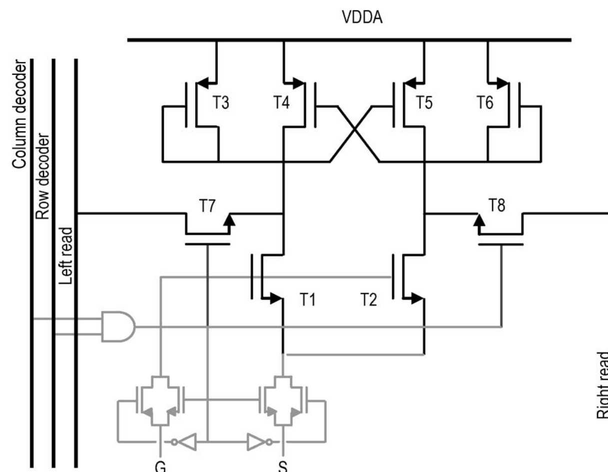

Fig. 2. Schematic of a one-cell of gain-based DLPVM.

B. Die-Level Process Variation Monitors (DLPVMs)

The DLPVM measurements are directly related to asymmetries between the branches composing the circuit, giving an estimation of the offset when both DLPVM inputs are grounded or set at predefined common-mode voltage. In this paper, we propose three distinctive DLPVMs, namely, gain-, decision-, and reference-based monitors, each covering characteristic analog structures. As shown in Fig. 2, the gain-based monitor consists of a differential input pair (transistors T_1 and T_2) with active loading (T_3 and T_4) and some additional gain (transistors T_5 and T_6) to increase the monitor's resolution and transistors T_7 and T_8 to connect to read lines (lines leading to a programmable data decision circuit).

The drain voltage of the different transistors in each die-level process monitor are accessed sequentially though a switch

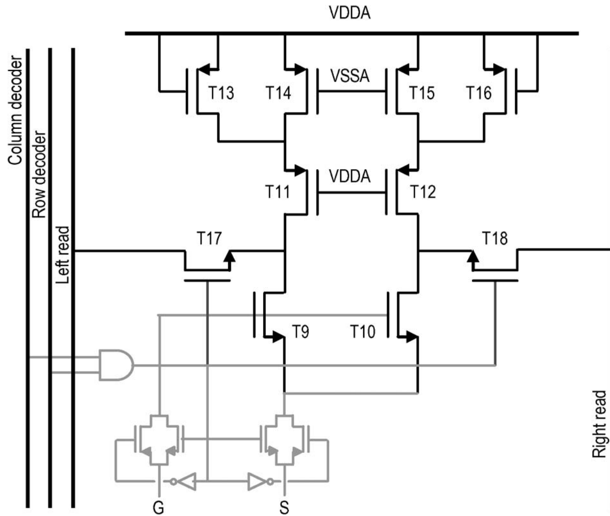

Fig. 3. Schematic of a one-cell decision-based DLPVM.

matrix that connects the drain of the transistor pairs under test to the detector; the drains of the other transistors are left open. The switch matrix connects the gate of the transistor pairs under test to the gate voltage source and connects the gates of the other rows to ground. The different device arrangements in the matrix include device orientation and the nested device environment. The matrix is placed several times on the chip to obtain information from different chip locations and distance behavior. As shown in Fig. 3, in the decision-based monitor, the common dynamic latch (transistors T_{11} to T_{16}) has been broken to allow a DC current flow through the device needed for the intended set of measurements. In addition to these two, internal reference voltages monitoring circuits, as shown in Fig. 4, sense the mismatch between two of the unit resistors. The current that flows through the resistors is fixed using a current mirror. Since the current is fixed, the voltage drop between the nodes labeled V_1 and V_2 is a measurement of the mismatch between the resistors. The feedback amplifier is realized by the common-source amplifier consisting of T_5 and its current source I_5. The amplifier keeps the drain–source voltage across T_3 as stable as possible, irrespective of the output voltage. The circuit consisting of T_7, T_9, T_{11}, I_1, and I_2 operates almost identically to a diode-connected transistor; however, it is employed instead to guarantee that all transistor bias voltages are accurately matched to those of the output circuitry consisting of T_1, T_3, T_5, and I_5. Consequently, I_{R1} will very accurately match I_1 [12]. As transistors T_3 and T_9 are biased to have drain–source voltages larger than the minimum required, V_{eff3}, this can pose a limitation in very low power supply technologies. To prevent this, we add diode-connected transistors, which act as level shifters in front of the common-source enhancement amplifier [13]. At the output side, the level shifter is the diode-connected transistor T_7, biased with current I_2. The circuitry at the input acts as diode-connected transistor while ensuring that all bias voltages are matched to the output circuitry. Although the power dissipation of the circuit is almost doubled over that of a classical cascode current mirror, by biasing the enhancement circuitry at lower densities, sufficient power dissipation savings are made.

C. Detector and Interface Circuit

The complete interface circuit including DLPVMs, a detector, the switch matrix to select the reference levels for a decision window, the interface to the external world, control blocks to sequence events during test, the scan chain to transport the pass/fail decisions, and the external tester is illustrated in Fig. 5. For clarity, only eight DLPVMs are shown. The analog decision is converted into pass/fail (digital) signals through the data decision circuit (transistors T_{1-24}). The TCB selects through a TMX the individual die-level process monitor circuit measurement. Select, reference, and calibration signals are offered to the detector through this circuitry. The data detector compares the output of the die level process monitor against a comparison reference window. The reference voltages defining the decision windows are related to the performance figures under study. The robustness against process variations is provided by an auto-zeroing scheme [14]. The data decision circuit operates on a two phase nonoverlapping clock. The comparison references needed to define the monitor decision windows are controlled through the DC signals labeled $refp$ and $refn$. The differencing network samples reference voltage during phase clk onto capacitor C, while the input is shorted giving differential zero. During phase $clkn$, the input signal is applied at the inputs of both capacitors, causing an input differential voltage to appear at the input of the comparator preamp. At the end of $clkn$, the regenerative flip-flop is latched to make the comparison and produce digital levels at the output. In the test mode, two main phases can be distinguished according to the state of signal ϕ. If ϕ is high, the inputs of the detector are shorted to the analog ground to perform a test of the detector itself, e.g., the circuit is in the auto-zeroing mode, whereas if ϕ is low the particular die-level process monitor circuit is connected to the detector and tested.

D. Temperature Monitor

To convert temperature to a digital value, a well-defined temperature-dependent signal and a temperature-independent reference signal are required. For constant collector current, base–emitter voltage V_{be} of the bipolar transistors has negative temperature dependence around room temperature.

This negative temperature dependence is cancelled by a proportional-to-absolute temperature (PTAT) dependence of the amplified difference of two base–emitter junctions. These junctions are biased at fixed but at unequal current densities resulting in the relation directly proportional to the absolute temperature. This proportionality is, however, rather small (0.1–0.25 $\mathrm{mV}/^{\circ}\mathrm{C}$) and needs to be amplified to allow further signal processing. The proposed temperature monitor is illustrated in Fig. 6. The right part of this circuit, comprising a voltage comparator (transistors T_{13-21}), creates the output signal of the temperature sensor. The rest of this circuit consists of the temperature sensing-circuit, amplifier and start-up. To enable a certain temperature detection, the voltage comparator requires the following two signals with different temperature dependence: 1) an increasing PTAT voltage V_{int} across the resistor network $N_T R$ and 2) a decreasing PTAT voltage V_{inr} at the comparator positive input. Adjustable resistors $N_R R$

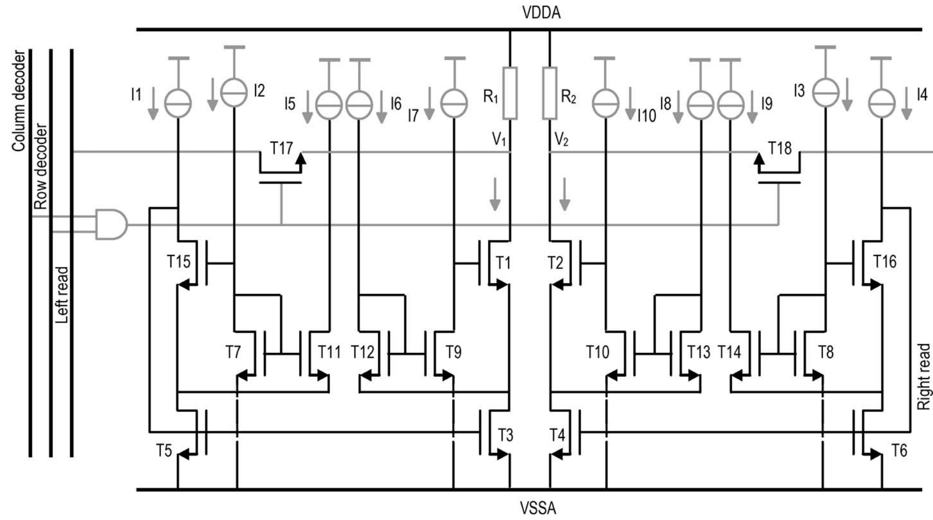

Fig. 4. One cell of reference-based DLPVM with a modified wide-swing current mirror.

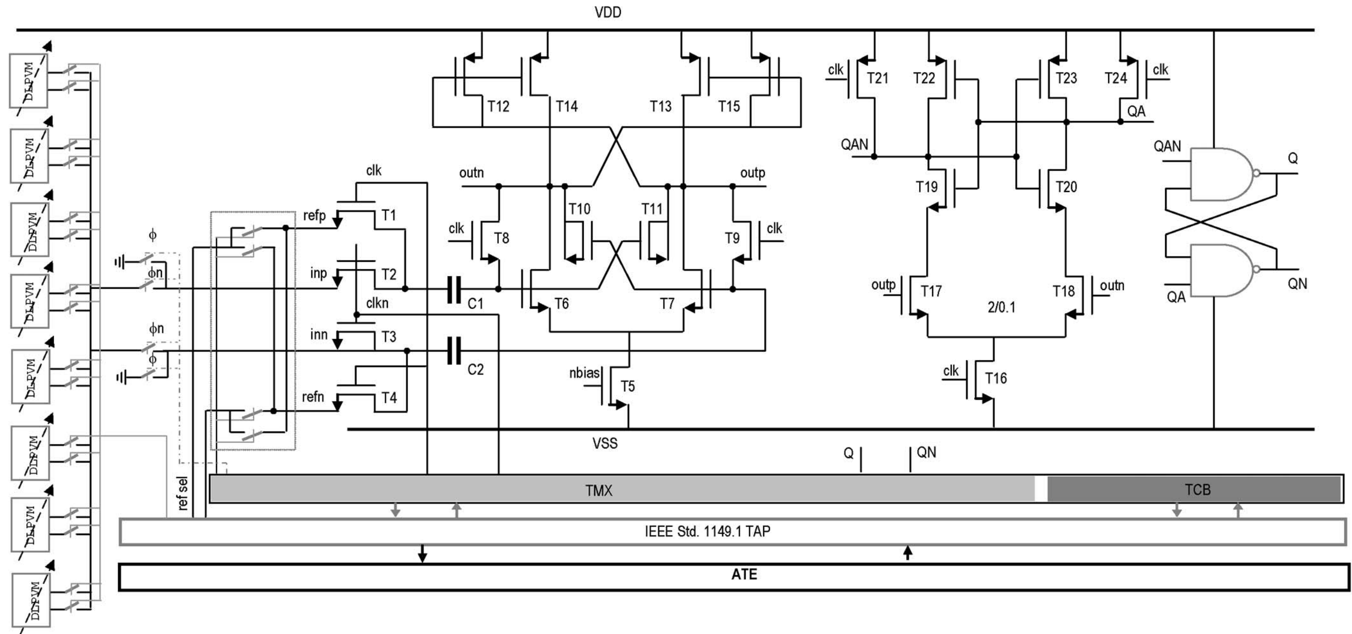

Fig. 5. Detector and interface circuit.

are employed for V_{be} (of transistors Q_{1-2}) curvature compensation [15]. The amplifier (transistors T_{1-6}) consists of a non-cascoded operational transconductance amplifier (OTA) with positive feedback to increase the loop-gain. Due to the asymmetries, the inaccuracy of the circuit is mainly determined by the offset and flicker noise of the amplifier. Several dynamic compensation techniques such as auto-zeroing, chopping, or dynamic element matching [16] might be employed to decrease offset and flicker noise. However, inherently, such techniques require very fast amplifier, whose noise is typically several orders of magnitude larger and consumes considerably more power. In addition, chopping ads switching noise due to, e.g., charge dump and clock interference. Such characteristics make these techniques unsuitable for thermal monitoring of VLSI circuits. In this design, to lower the effect of offset, the systematic offset is minimized by adjusting transistor dimensions and bias current in the ratio, while the random offset is reduced by a symmetrical and compact layout. Additionally, the collector currents of bipolar transistors Q_1 and Q_2 are rationed by a predefined factor, e.g., transistors are multiple parallel connections of unit devices. A start-up circuit consisting of transistors T_{7-9} drives the circuit out of the degenerate bias point when the supply is tuned on. The scan chain delivers a four-bit thermometer code for the selection of the resistor value $N_T R$. The nodes in between each resistor have different voltages depending on their proximity to V_{int}. By using thermometer decoding on the digital signal one specific node can be selected as the correct analog voltage. The resistor-ladder network is inherently monotonic as long as the switching elements are designed correctly. Similarly, since no high-speed operation is required, parasitic capacitors at a tap point will not create significant voltage glitch.

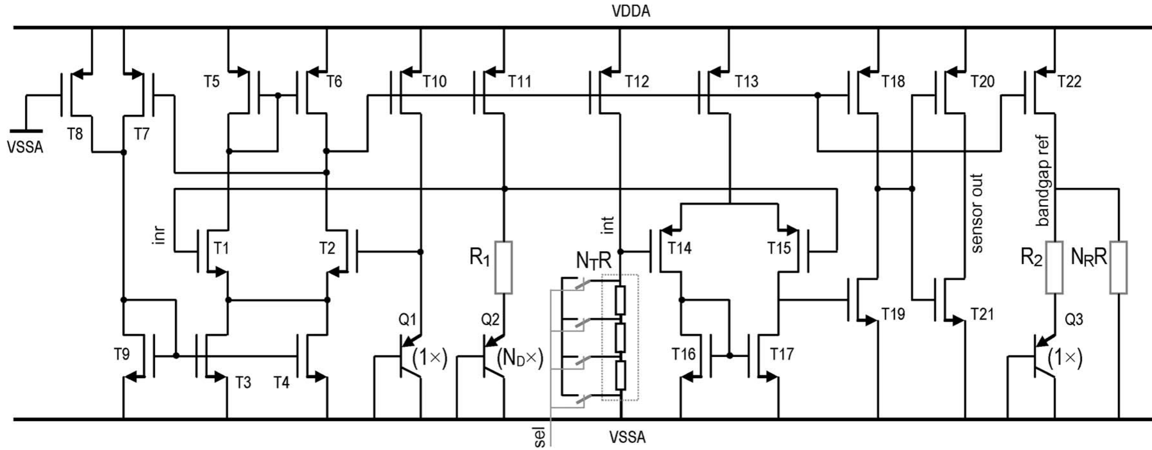

Fig. 6. Temperature monitor.

III. Characterization of Process Variability Conditions and Test-Limit Updates and Guidance

The complexity of yield estimation, coupled with the iterative nature of the design process, makes yield maximization computationally prohibitive. Worst case analysis is very efficient in terms of designer effort and, thus, has become the most widely practiced technique for statistical verification. However, the worst case performance values obtained are extremely pessimistic and as a result lead to unnecessarily large and power hungry designs to reach the desired specifications. In this paper, statistical data extracted through the monitor measurements allow us possibilities not only to enhance observation of important design and technology parameters, but also to characterize current process variability conditions of certain parameters of interest, enabling optimized design environment as well.

A. Characterization of Process Variability Conditions

A maximum likelihood (ML) estimation involves estimation of parameter vector (threshold voltage variation, resistor width variation, etc., obtained through monitor's observation) $\theta \in \Theta$, where Θ is a parameter space for which the observed data is the most likely, e.g., marginal probability $p_{X|\Theta}(x|\theta)$ is a maximum, given the vector of the DLPVM's observations $x_i \in X$, where X is a measurement space, at temperature T. The $p_{X|\Theta}(x|\theta)$ is the Gaussian mixture model given by the weighted sum of the Gaussian distributions. The logarithm of the probability $p(T_X|\theta)$ is referred to as the log-likelihood $L(\theta|T_X)$ of θ with respect to T_X.

The input set T_X is given by $T_X = \{(x_1, \ldots, x_l)\}$, which contains only vectors of DLPVM's observations x_i. The log-likelihood can be factorized as

$$L(\theta|T_X) = \log p(T_X|\theta) = \sum_{i=1}^{l} \sum_{y \in Y} p_{X|Y,\Theta}(x_i|y_i, \theta) p_{Y|\Theta}(y_i|\theta) \tag{1}$$

for the missing data vector $y_i \in Y$, where Y is the incomplete data set, which are independent and identically distributed according to the probability $p_{XY|\Theta}(x, y|\theta)$. The problem of ML estimation from the set of DLPVM observations T_x can be defined as

$$\theta^* = \max_{\theta \in \Theta} L(\theta|T_X) = \max_{\theta \in \Theta} \sum_{i=1}^{l} \sum_{y \in Y} p_{X|Y,\Theta}(x_i|y_i, \theta) p_{Y|\Theta}(y_i|\theta). \tag{2}$$

Obtaining optimum estimates through the ML method involves the following two steps: 1) computing the likelihood function and 2) maximizing over the set of all admissible sequences. Evaluating the contribution of the random parameter θ requires computing an expectation over the joint statistics of the random parameter vector, a task that is analytically intractable. Even if the likelihood function L can be obtained analytically, it is invariably a nonlinear function of θ, which makes the maximization step (which must be performed in real time) computationally unfeasible. In such cases, EM algorithm [9] allows obtaining the ML estimates of the unknown parameters by a computational procedure that iterates, until convergence, between two steps.

Instead of using the traditional incomplete-data density in the estimation process, the EM algorithm uses the properties of the complete-data density. In doing so, it can often make the estimation problem more tractable and also yield good estimates of the parameters for small sample sizes [17]. Thus, with regard to implementation, the EM algorithm holds a significant advantage over traditional steepest descent methods acting on the incomplete-data likelihood equation. Moreover, the EM algorithm provides the values of the log-likelihood function corresponding to the ML estimates based uniquely on the observed data.

The EM algorithm builds a sequence of parameter estimates $\theta^{(0)}, \theta^{(1)}, \ldots, \theta^{(t)}$, such that the log-likelihood $L(\theta^{(t)}|T_X)$ monotonically increases, i.e., $L(\theta^{(0)}|T_X) < L(\theta^{(1)}|T_X) < \ldots < L(\theta^{(t)}|T_X)$ until a stationary point $L(\theta^{(t-1)}|T_X) = L(\theta^{(t)}|T_X)$ is achieved. Using Bayes rule, the log likelihood of x_i can be written as

$$\log p(T_X|\theta) = \log p(X, Y|\theta), \theta(t) - \log p_{X,Y|X}(X, Y|\theta), \theta(t). \tag{3}$$

Taking expectations on both sides of the above equation given x_i and θ, where $\theta^{(t)}$ is an available estimate of θ, we have

$$\begin{aligned}\log p(T_X|\theta) &= E_{\theta^{(t)}}\{\log p(X,Y|\theta)|X,\theta(t)\}\\ &\quad - E_{\theta^{(t)}}\left\{\log p_{X,Y|X}(X,Y|X)|X,\theta(t)\right\}\\ &= Q_n\left(\theta|\theta^{(t)}\right) - P\left(\theta|\theta^{(t)}\right).\end{aligned} \tag{4}$$

By Jensen's inequality, the relation holds that

$$P\left(\theta|\theta^{(t)}\right) \leq P\left(\theta^{(t)}|\theta^{(t)}\right). \tag{5}$$

Therefore, a new estimate θ in the next iteration step that makes $Q(\theta^{(t)}|\theta^{(t)}) \geq Q(\theta|\theta^{(t)})$ leads to

$$\log p(T_X|\theta) \geq \log p\left(T_X|\theta^{(t)}\right). \tag{6}$$

In each iteration, two steps, called E-step and M-step, are involved. In the E-step, the EM algorithm forms the auxiliary function $Q(\theta|\theta^{(t)})$, ($\theta^{(0)}$,$\theta^{(1)}$,...,$\theta^{(t)}$ is a sequence of parameter estimates), which calculates the expected value of the log-likelihood function with respect to the conditional distribution Y of the functional test, given the vector of the DLPVM's observations X under the current estimate of the parameters $\theta(t)$, i.e.,

$$Q\left(\theta|\theta^{(t)}\right) = E\left(\log p(X,Y|\theta)|X,\theta(t)\right). \tag{7}$$

In the M-step, the algorithm determines a new parameter maximizing Q, i.e.,

$$\theta^{(t+1)} = \arg\max_{\theta} Q\left(\theta|\theta^{(t)}\right). \tag{8}$$

At each step of the EM iteration, the likelihood function can be shown to be nondecreasing [17]; if it is also bounded (which is mostly the case in practice), then the algorithm converges. An iterative maximization of $Q(\theta|\theta^{(t)})$ will lead to a ML estimation of θ [17].

EM Algorithm

Initialization
– Initialize the data set $T_{XY} = \{(x_1, y_1), \ldots, (x_l, y_l)\}$.
– Initialize the parameter $\theta^{(0)}$.
Data collection
– Collect N samples from the DLPVMs and temperature monitors
Update parameter estimate
1) Calculate $Q(\theta|\theta^{(n)}) = E(logp(X,Y|\theta)X, \theta^{(n)})$—E step.
2) Reestimate θ by maximizing the θ-function $\theta^{(n+1)} = \arg\max_\theta Q(\theta|\theta^{(n)})$, estimate mean, and variance—M step.
3) Increase the iteration index n.
4) Stop when a stationary point $L(\theta^{(n-1)}|T_{XY}) = L(\theta^{(n)}|T_{XY})$ is found.

B. Algorithm for Test-Limit Updates and Guidance

When an optimum estimate of the parameter distribution is obtained as described in the previous section, the next step is to update the test limit values utilizing an ASVM classifier [10]. In comparison with established classifiers (such as quadratic, boosting, neural networks, and Bayesian networks), the ASVM classifier is particularly resourceful, since it simultaneously minimizes the empirical classification error and maximizes the geometric margin. Assuming that the input vectors (e.g., values defining test limits) belong to *a priori* (nominal values) and *a posteriori* (values estimated with the EM algorithm) classes, the goal is to set test limits that reflect observed on-chip variations. Each new measurement is viewed as an r-dimensional vector and the ASVM classifier separates the input vectors into an $r-1$-D hyperplane in feature space Z. Let $D = \{x_i, c_i)|x_i \in R^r, c_i \in -1, 1\}_{i=1}^n$ be the input vectors belonging to *a priori* and *a posteriori* classes, where the c_i is either 1 or -1, indicating the class to which data x_i from the input vector belong. To maximize the margin, w and b are chosen such that they minimize the nearest integer $\|w\|$ subject to the optimization problem described by

$$c_i(w \cdot x_i + b) \geq 1 \tag{9}$$

for all $1 \leq i \leq n$, where the vector w is a normal vector, which is perpendicular to the hyperplane (e.g., defined as $w \cdot x + b = 0$) the parameter $b/\|w\|$ determine the offset of the hyperplane from the origin along the normal vector w.

In this paper, we solve this optimization problem with a quadratic programming [18]. The equation is altered by substituting $\|w\|$ with $1/2\|w\|^2$ without changing the solution (the minimum of the original and the modified equation have the same w and b). The quadratic programming problem is solved incrementally, covering all the subsets of classes constructing the optimal separating hyperplane for the full data set. Writing the classification rule in its unconstrained dual form reveals that the maximum margin hyperplane and, therefore, the classification task is now only a function of the support vectors, e.g., the training data that lie on the margin

$$\max \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j c_i c_j x_i^T x_j \tag{10}$$

subject to $\alpha_i \geq 0$ and $\sum_{i=1}^n \alpha_i c_i = 0$,

$$w = \sum_i \alpha_i c_i x_i \tag{11}$$

where the α terms constitute the weight vector in terms of the training set. To allow for mislabeled examples a modified maximum margin technique [18] is employed. If there exists no hyperplane that can divide the *a priori* and *a posteriori* classes, the modified maximum margin technique finds a hyperplane that separates the training set with a minimal number of errors. The method introduces nonnegative variables ξ_i, which measure the degree of misclassification of the data x_i, i.e.,

$$c_i(w \cdot x_i + b) \geq 1 - \xi_i \tag{12}$$

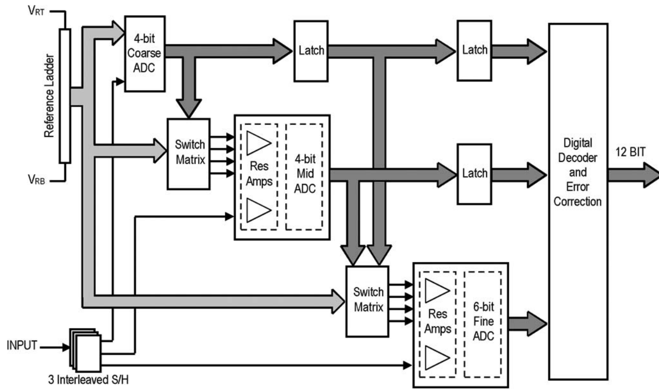

Fig. 7. Block diagram of the 12-bit multistep A/D converter [19].

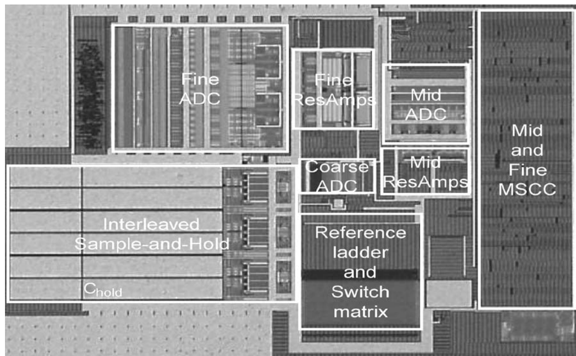

Fig. 8. Chip micrograph of the A/D converter and embedded monitors.

for all $1 \leq i \leq n$. The objective function is then increased by a function that penalizes nonzero ξ_i, and the optimization becomes a tradeoff between a large margin and a small error penalty. For a linear penalty function, the optimization problem now transforms to

$$\min \frac{1}{2}\|w\|^2 + C\sum_i \xi_i^\sigma \tag{13}$$

such that (9) holds for all $1 \leq i \leq n$. For sufficiently large constant C and sufficiently small σ, the vector w and constant b that minimize the functional (13) under constraints in (9) determine the hyperplane that minimizes the number of errors on the training set and separate the rest of the elements with maximal margin. This constraint in (9) along with the objective of minimizing $\|w\|$ is solved using Lagrange multipliers. The key advantage of a linear penalty function is that the variables ξ_i vanish from the dual problem, with the constant C appearing only as an additional constraint on the Lagrange multipliers.

IV. Experimental Results

The proposed monitors and algorithms are evaluated on a 12-bit A/D converter described in [19] (Fig. 7) and fabricated in a standard single poly six-metal 90-nm CMOS (Fig. 8). The stand-alone A/D converter consist of three stages, namely, coarse-, mid-, and fine-stage, occupies an area of 0.75 mm^2, operates at 1.2 V supply voltage, and dissipates 55 mW (without output buffers).

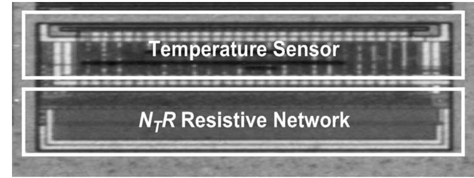

Fig. 9. Chip micrograph of the temperature monitor (zoomed in)

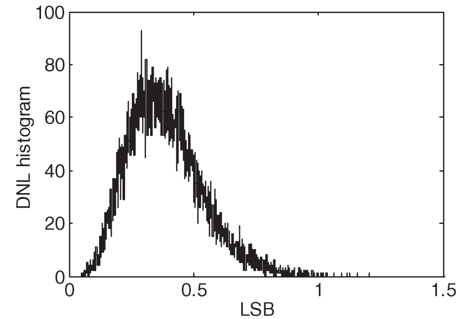

Fig. 10. A/D converter DNL histogram.

Dedicated embedded DLPVMs (12 per stage subdivided into three specific groups and placed in and around the partitioned multistep A/D converter) and the complete design-for-test circuit are restricted to less than 5% of the overall area and consume 8 mW when in active mode. The multistage circuit calibration algorithm [20] requires about 1.5k logic gates as calibration overhead, occupies an area of 0.14 mm^2, and consumes 11 mW of power. A temperature monitor (Fig. 9) is located between coarse A/D converter and fine residue amplifiers. The stand-alone temperature monitor occupies an area of 0.05 mm^2, operates within 1.0–1.8 V, and dissipates 11 μW. In the test silicon, four bits for 16 selection levels are chosen for the temperature settings, resulting in a temperature range from 0 °C to 160 °C in steps of 9 °C, which is sufficient for thermal monitoring of VLSI circuits. If more steps are required, a selection $N_T R$ can be easily extended with a higher resolution resistive network. For the robustness, the circuit is completely balanced and matched both in the layout and in the bias conditions of devices, canceling all disturbances and nonidealities to the first order.

The overall converter employs around 6500 transistors within an analog core and consists primarily of noncritical low-power components, such as low-resolution quantizers, switches, and open-loop amplifiers. The total acquisition time required at wafer-level manufacturing test is in 0.5–1 ms range per functional block. This pales in comparison with $\sim$1 s needed to perform histogram-based static [21] or $\sim$1 s for FFT-based dynamic A/D converter test. Note that the time required to perform these functional tests depends on the speed of the converter and available postprocessing power. The algorithms for the test window generation/update, namely, the EM and ASVM algorithms, are performed off-line and are implemented in Matlab. Fig. 10 illustrates A/D converter differential nonlinearity (DNL) histogram.

Figs. 11–13 illustrate the histogram estimated from 3780 samples extracted from 108 specific DLPVMs and measured across 35 prototype devices. The extracted DLPVM and DNL measurements of each stage of the multistep A/D converter are

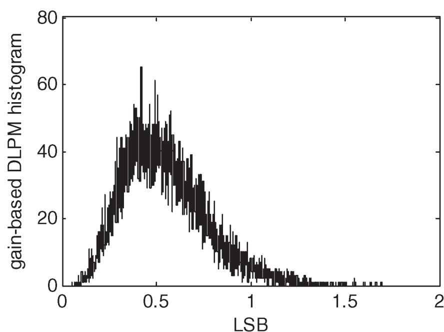

Fig. 11. Gain-based DLPVM histogram.

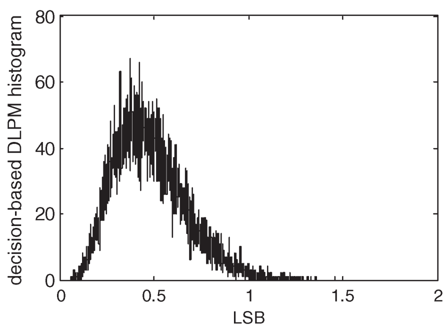

Fig. 12. Decision-based DLPVM histogram.

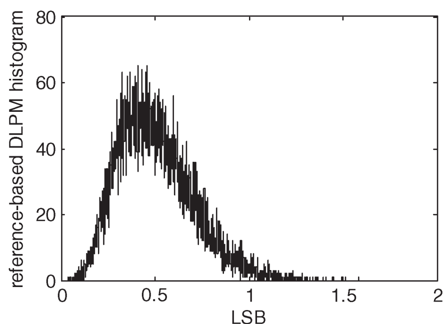

Fig. 13. Reference-based DLPVM histogram.

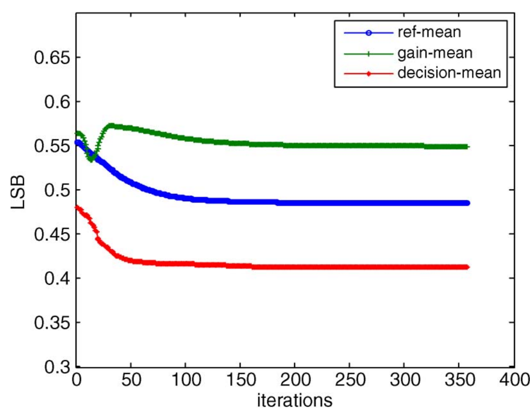

Fig. 14. Estimating mean μ values of gain-, decision-, and reference-based DLPVMs with respect to the number of iterations of the EM at temperature T.

correlated with the EM algorithm. To make the problem manageable, the process parameter variation model is assumed to follow a Gaussian distribution. The mean μ and the variance σ of gain-, decision-, and reference-based DLPVMs are estimated based on the EM algorithm (Figs. 14 and 15). This observed

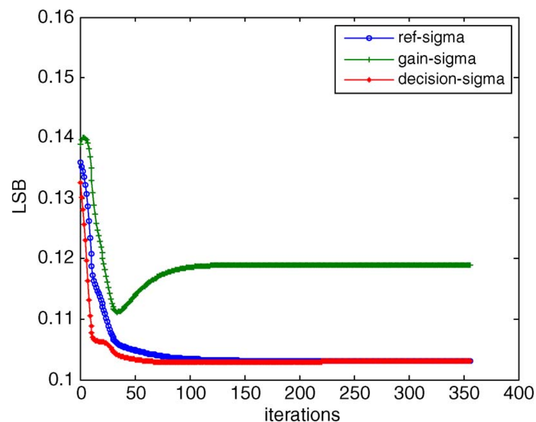

Fig. 15. Estimating variance σ values of gain-, decision-, and reference-based DLPVMs with respect to the number of iterations of the EM at temperature T.

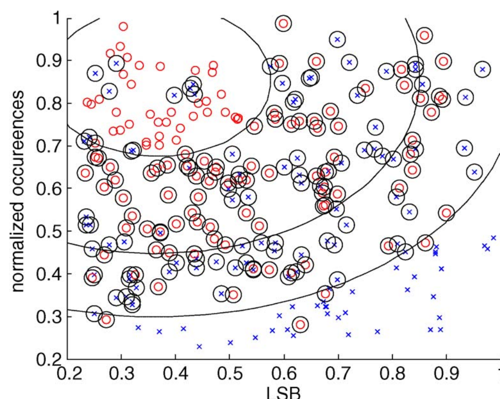

Fig. 16. Fitting *a posteriori* probability to the SVM output. The support vectors, marked with larger circles, define the margin of separation between the classes of multiple runs of DLPVMs (crosses) and DNL measurements (smaller circles).

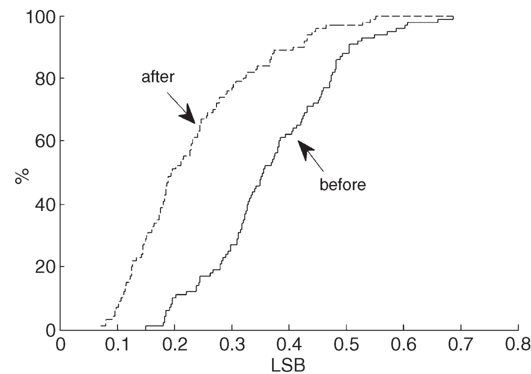

Fig. 17. Yield enhancement; DNL cumulative histograms of 100 000 devices before and after adjusting the tolerance limits.

process-related information allows design recentering, e.g., test limit setting with the ASVM classifier. As illustrated in Fig. 16, the high limit value is updated in the corresponding functional test specs of the stage-under-test with 0.35 least significant bit. This on-the-fly test limit setting leads to an increased yield, as illustrated in Fig. 17. The cumulative DNL is obtained across a projected 100 000 devices showing similar characteristics as a measured prototype.

TABLE I
SUMMARY OF THE TEMPERATURE SENSOR PERFORMANCE AND COMPARISION WITH PRIOR ART

	[5]	[6]	[22]	[23]	[24]	[25]	**[This work]**
CMOS Technology	*0.35 μm*	*1.0 μm*	*0.7 μm*	*0.18 μm*	*0.13 μm*	*65 nm*	*90 nm*
Range (ºC)	*0~100*	*10~100*	*-55~125*	*temp switch*	*0~100*	*0~100*	*0~160*
Supply voltage (V)	*3.0~3.8*	*5*	*2.5~5.5*	*1.0~1.8*	*1.2*	*1.0*	*1.0~1.8*
Inaccuracy (ºC)	*-0.7~+0.9*	*± 1*	*±0.1*	*±1.1*	*-1.8~+2.3*	*±10.0*	*± 0.9*
Sensor type	*temp-to-pulse*	*analog current*	ΔV_{be}	ΔV_{be}	*duall-DLL*	*temp-to-pulse*	ΔV_{be}
Calibration	*two-points*	-	*one-point*	-	*one-point*	*autocalibration*	-
Power (μW)	*490*	*300*	*247*	*13*	*12000*	*55*	*11*
Area (mm²)	*0.175*	*0.023*	*0.16*	*0.03*	*0.16*	*0.01*	*0.05*

For the circuit measurement, a single-frequency sinusoidal input signal is generated by an arbitrary waveform generator and applied at the first to a narrow bandpass filter to remove any harmonic distortion and extraneous noise, and then to the test board. The signal is connected via 50 Ω coaxial cables to minimize external interference.

On the test circuit board, the single-ended signal is converted to a balanced differential signal using a transformer. Potentiometers are used to adjust the reference voltages and the common-mode voltage. The common-mode voltage of the test signal going into the A/D converter is set through matching resistors connected to a voltage reference. The digital output of the A/D converter is buffered with an output buffer to the drive large parasitic capacitance of the lines on the board and probes from the logic analyzer. A clock signal is also provided to the logic analyzer to synchronize with the A/D converter. Repetitive single die-level process monitor measurements are performed to minimize noise errors. Special attention is paid in the layout to obtain a very low resistance in the gate path to eliminate systematic errors during the measurements; very wide source metal connections are used. Since different transistors are measured sequentially the DC repeatability of the DC gate voltage source must be larger than the smallest gate-voltage offset to be measured. The repeatability of the source in the measurement setup was better than six digits. All chips are functional in a temperature range between 0 °C and 160 °C. The measured behavior of the temperature monitor shows the typical bandgap curve, which reaches a maximum at 810 mV close to the target of 800 mV without trimming.

We observe that the improvement of DNL coincident with the fact that the mismatch increases when decreasing the temperature. Therefore, as the worst case mismatch and temperature condition, the lower end (0 °C) of the used temperature scale (0 °C to 90 °C) is observed. The linearity measurements show bathtub-like features since at the higher temperature end mobility degradation deteriorates the circuit performance. The DLPM measurements show that at optimal temperature (30 °C), the standard deviation $Stdev(\Delta V_{T\mathrm{sat}})$ decreases by 0.16 mV. This compares reasonably well with the measured improvement in $I_{D\mathrm{sat}}$ matching of 0.032%. The threshold voltage matching coefficient A_{VT}, the standard deviation of percent ΔI_D, and the current matching coefficient A_{ID} improve by 0.3 mVμm, 0.032% (0.036 μA), and 0.06% μm, respectively. The average error of the temperature monitor at room temperature is around 0.5 °C, with a standard deviation of less than 0.4 °C, which matches the expected error of 0.4 °C within a batch. Nonlinearity is approximately 0.4 °C from 0 °C to 160 °C. The intrinsic base–emitter voltage nonlinearity in the bandgap reference is limited by the compensation circuit. The measured noise level is lower than 0.05 °C. A summary of the temperature monitor performance and comparison with recently published works is shown in Table I. In all-digital temperature sensors [5], [25], the two-temperature-point calibration is required in every sensor; thus, calibration cost is very large in on-chip thermal sensing applications. A current-output temperature sensor [6] does not have a linear temperature reading and is sensitive to process variation, which requires more effort and cost for after-process calibration. Although the dual-DLL-based temperature sensor [24] only needs one-temperature-point calibration, it occupies a large chip area with a high level of power consumption at a microwatt level. The sensors based on the temperature characteristics of parasitic bipolar transistors [22], [23] offer high accuracy and small chip area. However, the high power consumption in [22] and the small temperature range in [23] make these realizations unsuitable for on-chip thermal monitoring.

V. CONCLUSION

The feasibility of the proposed method has been verified by experimental measurements from the silicon prototype fabricated in standard single poly six-metal 90-nm CMOS. The monitors allow the readout of local (within the core) performance parameters as well as the global distribution of these parameters, significantly increasing the obtained yield. The monitors are small, stand alone, and easily scalable and can be fully switched off. The flexibility of the concept allows the system to be easily extended with a variety of other performance monitors. The implemented EM algorithm and ASVM classifier allow us to guide the verification process with the information obtained through monitoring process variations. Fast identification of excessive process parameter and temperature variation effects is facilitated at the cost of at most 5% area overhead and 8 mW of power consumption when in the active mode.

ACKNOWLEDGMENT

The authors would like to thank H. van der Ploeg, B. Butter, G. van der Weide, S. Krishnan, P. Pavithran, V. Zieren of NXP Semiconductors and M. Berkelaar of Delft University of Technology for their contributions.

References

[1] *International Technology Roadmap for Semiconductors (ITRS)*, 2009.

[2] V. Petrescu, M. Pelgrom, H. Veendrick, P. Pavithran, and J. Wieling, "Monitors for a signal integrity measurement system," in *Proc. IEEE Eur. Solid-State Circuit Conf.*, 2006, pp. 122–125.

[3] M. Bhushan, M. B. Ketchen, S. Polonksy, and A. Gattiker, "Ring oscillator based technique for measuring variability statistics," in *Proc. IEEE Int. Conf. Microelectron. Test Structures*, 2006, pp. 87–92.

[4] N. Izumi, H. Ozaki, Y. Nakagawa, N. Kasai, and T. Arikado, "Evaluation of transistor property variations within chips on 300-mm wafers using a new MOSFET array test structure," *IEEE Trans. Semicond. Manuf.*, vol. 17, no. 3, pp. 248–254, Aug. 2004.

[5] P. Chen, C. Chen, C. Tsai, and W. Lu, "A time-to-digital-converter based CMOS smart temperature sensor," *IEEE J. Solid-State Circuits*, vol. 40, no. 8, pp. 1642–1648, Aug. 2005.

[6] V. Szekely, C. Marta, Z. Kohari, and M. Rencz, "CMOS sensors for on-line thermal monitoring of VLSI circuits," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 5, no. 3, pp. 270–276, Sep. 1997.

[7] B. Datta and W. Burleson, "Temperature effects on energy optimization in sub-threshold circuit design," in *Proc. IEEE Int. Symp. Quality Electron. Des.*, 2009, pp. 680–685.

[8] G. C. M. Meijer, G. Wang, and F. Fruett, "Temperature sensors and voltage references implemented in CMOS technology," *IEEE Sensors J.*, vol. 1, no. 3, pp. 225–234, Oct. 2001.

[9] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*. New York: Wiley-Interscience, 1997.

[10] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.

[11] *IEEE Standard Test Access Port and Boundary-Scan Architecture*, IEEE Std. 1149.1-2001, 2001.

[12] E. Sackinger and W. Guggenuhl, "A high-swing, high-impedance MOS cascode circuit," *IEEE J. Solid-State Circuits*, vol. 25, no. 1, pp. 289–298, Feb. 1990.

[13] A. Coban and P. Allen, "A 1.75-V rail-to-rail CMOS op amp," in *Proc. IEEE Int. Symp. Circuits Syst.*, 1994, vol. 5, pp. 497–500.

[14] T. Kumamoto, M. Nakaya, H. Honda, S. Asai, Y. Akasaka, and Y. Horiba, "An 8-bit high-speed CMOS A/D converter," *IEEE J. Solid-State Circuits*, vol. SSC-21, no. 6, pp. 976–982, Dec. 1986.

[15] M. R. Valer, S. Celma, B. Calvo, and N. Medrano, "CMOS voltage-to-frequency converter with temperature drift compensation," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 9, pp. 3232–3234, Sep. 2011.

[16] A. Bakker and J. H. Huijsing, "A low-cost high-accuracy CMOS smart temperature sensor," in *Proc. IEEE Eur. Solid-State Circuit Conf.*, 1999, pp. 302–305.

[17] R. A. Redner and H. F. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *Surveys Math. Ind.*, vol. 26, no. 2, pp. 195–239, 1984.

[18] V. Franc and V. Hlavac, "Multi-class support vector machine," in *Proc. IEEE Int. Conf. Pattern Recognit.*, 2002, vol. 2, pp. 236–239.

[19] A. Zjajo and J. Pineda de Gyvez, "A 1.2 V 55 mW 12 bits self-calibrated dual-residue analog to digital converter in 90 nm CMOS," in *Proc. IEEE Int. Symp. Low Power Electron. Des.*, 2011, pp. 187–192.

[20] A. Zjajo and J. P. de Gyvez, "An adaptive digital calibration of multi-step A/D converters," in *Proc. IEEE Int. Conf. Signal Process.*, 2010, pp. 2456–2459.

[21] H.-W. Ting, B.-D. Liu, and S. J. Chang, "A histogram-based testing method for estimating A/D converter performance," *IEEE Trans. Instrum. Meas.*, vol. 57, no. 2, pp. 420–427, Feb. 2008.

[22] M. A. P. Pertijs, K. A. A. Makinwa, and J. H. Huijsing, "A CMOS smart temperature sensor with a 3σ inaccuracy of ± 0.1 °C from -55 °C to 125 °C," *IEEE J. Solid-State Circuits*, vol. 40, no. 12, pp. 2805–2815, Dec. 2005.

[23] D. Schinkel, R. P. de Boer, A. J. Annema, and A. J. M. van Tuijl, "A 1-V 15 μW high-precision temperature switch," in *Proc. IEEE Eur. Solid-State Circuit Conf.*, 2001, pp. 77–80.

[24] K. Woo, S. Meninger, T. Xanthopoulos, E. Crain, D. Ha, and D. Ham, "Dual-DLL-based CMOS all-digital temperature sensor for microprocessor thermal monitoring," in *Proc. IEEE Int. Solid-State Circuit Conf.*, 2009, pp. 68–69.

[25] C.-C. Chung and C.-R. Yang, "An all-digital smart temperature sensor with auto-calibration in 65 nm CMOS technology," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2010, pp. 4089–4092.

Amir Zjajo (M'02) received the M.Sc. and DIC degrees from the Imperial College London, London, U.K., in 2000 and the Ph.D. degree from the Eindhoven University of Technology, Eindhoven, The Netherlands, in 2010, all in electrical engineering.

In 2000, he joined Philips Research Laboratories as a member of the research staff in the Mixed-Signal Circuits and Systems Group. From 2006 to 2009, he was with Corporate Research, NXP Semiconductors, as a Senior Research Scientist. In 2009, he joined the Delft University of Technology, Delft, The Netherlands, as a Faculty Member in the Circuit and Systems Group. His research interests include mixed-signal circuit design, signal integrity and timing, and yield optimization of VLSI. He is the author of more than 40 papers published in referenced journals and conference proceedings. He is the author of the book *Low-Voltage High-Resolution A/D Converters: Design and Calibration* (Springer, 2010). He is the holder of more then ten U.S. patents or patents pending.

Dr. Zjajo serves as a Member of the Technical Program Committee of the IEEE Design, Automation and Test in Europe Conference, the IEEE International Symposium on Circuits and Systems, and the IEEE International Mixed-Signal Circuits, Sensors and Systems Workshop.

Manuel J. Barragan received the M.S. degree in physics and the Ph.D. degree in microelectronics from the University of Seville, Seville, Spain, in 2003 and 2009, respectively.

He has been with the National Center of Microelectronics, Microelectronics Institute of Seville (IMSE-CNM), University of Seville, Seville, Spain, since 2003, where he currently holds a CSIC JAE-Doc postdoctoral research position financed by FSE. His research interests include test and design for testability of analog, mixed-signal, and RF systems.

Dr. Barragan received a Silver Leaf Award in the Fifth IEEE International Conference on Ph.D. Research in Microelectronics and Electronics in 2009. In 2011, his work was selected for inclusion in the 20th Anniversary Compendium of Most Influential Papers from Asian Test Symposium.

José Pineda de Gyvez (F'09) received the Ph.D. degree from the Eindhoven University of Technology, Eindhoven, The Netherlands, in 1991.

From 1991 to 1999, he was a Faculty Member in the Department of Electrical Engineering at Texas A&M University, College Station. He is a Fellow at NXP Semiconductors, Eindhoven, where he leads the research program on variability-tolerant designs. Since 2006, he has held the professorship "Deep Submicron Integration" with the Department of Electrical Engineering, Eindhoven University of Technology. He is a member of the Editorial Board of the *Journal of Low Power Electronics*. He has more than 100 combined publications in the fields of testing, nonlinear circuits, and low-power design. He is the author or a coauthor of three books and is the holder of a number of granted patents.

Dr. de Gyvez has been an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS PART I AND PART II and the TECHNOLOGY IN IEEE TRANSACTIONS ON SEMICONDUCTOR MANUFACTURING. His work has been acknowledged in academic environments as well as in patent portfolios of many companies. His research has been funded by the Dutch Ministry of Science, U.S. Office of Naval Research, U.S. National Science Foundation, among others.