On Evaluating Temperature as Observable for CMOS Technology Variability

J. Altet, D. Gómez, C. Dufis, J.L. González, D. Mateo, X. Aragonés, F. Moll and A. Rubio
Universitat Politècnica de Catalunya
Dept. of Electronic Engineering, Barcelona, Spain

Abstract—**The temperature at surface of a silicon die depends on the activity of the circuits placed on it. In this paper, it is analyzed how Process, Voltage and Temperature (PVT) variations affect simultaneously some figures of merit (FoM) of some digital and analog circuits and the power dissipated by such circuits. It is shown that in some cases, a strong correlation exists between the variation of the circuit FoM and the variation of the dissipated power. Since local temperature increase at the silicon surface close to the circuit linearly depends on dissipated power, the results show that temperature can be considered as an observable magnitude for CMOS technology variability monitoring.**

I. Introduction

The impact and complexity of process variations is increasing as technology scales down to nanometer dimensions. Several sources contribute to this increasing importance: photolithography, etching and CMP related variations, line-edge and line-width roughness, layout-dependence stress, random dopant fluctuations, proximity effects, among others. While some of them can be alleviated using special design techniques, other effects are intrinsic to nanoscale devices and therefore unavoidable. In addition, ambient temperature and voltage variations increases the total uncertainty.

Process variations can be classified in terms of scale or nature. Different variation mechanisms may affect different scales, for example leading to Within Die (WID), Die to Die (D2D) or Wafer to Wafer (W2W) variations. On the other hand, the nature of variations is due either to a systematic or to a random cause.

One of the challenges in deep-submicron CMOS design is to minimize the effect of variability in system performance and yield. For both digital and analog circuits, variability of the manufacturing process forces the designer to extend the overdesign margins to maintain production yield, usually at the expenses of an increased power consumption, or a reduced circuit speed.

Several research lines are currently addressed to minimize the effects of variability. One of them is based on the use of monitors. The purpose of monitoring circuits is to track the circuit variability by measuring one circuit parameter and thus detect if compensatory actions need to be activated. Those actions can be either to activate feedback mechanisms by making use of active knobs included in the circuit expressly for this purpose, or to adapt the circuit operating conditions to guarantee a reliable and within specs operation. Examples of monitors can be found in both digital and analog circuits. E.g.: in digital circuits monitors are based on sensing leakage current [1], timing monitors [2] or measuring the period in ring oscillators [3]. In analog circuits, most of the techniques are based on sensing an electrical variable, using, for example, RF power detectors [4]. On the other hand, thermal coupling through the semiconductor substrate generates a rise in temperature in the vicinity of a circuit/device that depends on the circuit/device power dissipation. Therefore, the thermal map at the silicon surface $T(x, y)$ has a direct dependence on the activity of the circuit placed on it.

In this paper the use of a new magnitude is proposed to monitor variability and mismatch in digital and analog circuits: the increase of temperature at selected points of the silicon surface. There are several advantages when considering temperature as an observable for technology variability. First, measuring temperature is a mature subject, either with off chip temperature sensors or with temperature sensors embedded in the same silicon die. The first option can be used in production tests and reliability analysis, whereas the use of built-in temperature sensors would allow self healing strategies. Second, since temperature measurements must be performed at low frequencies (maximum of hundreds of kHz [5]) due to the characteristics of the heat coupling mechanism, the signal processing and design of the monitor circuit are simpler. Third, temperature is a parameter that enhances observability when access to internal electrical variables is not possible: e.g. the detection of hot spots has been classically used in failure analysis to locate defective sections in digital circuits [6]. Fourth, temperature is measured without electrically loading the circuit under monitoring. This is especially important in high frequency analog circuits, where any load alters its performance. In digital circuits, it is well known that either monitoring current consumption or altering loading capacitances reduces system performances.

As a last remark, the use of temperature would allow a unified monitoring strategy for both analog and digital circuits, reducing the complexity of variability monitoring in systems-on-chip with circuits of diverse nature.

The novel analysis presented in this paper underlines the correlation between the effects of variability on both the power dissipated by the devices of a circuit (and therefore the temperature increase that they generate in their vicinity) and the value of some figures of merit of the circuit under analysis: gain and linearity in analog circuits and delay in digital circuits. The presence of these correlations suggests that temperature can be used to track the effects of PVT variability

on circuit performance and opens the possibility of using embedded temperature sensors in self-healing strategies. It is not the goal of this work to present final conclusions regarding the feasibility and applications of temperature measurements for self-healing and variability corrections, but to present a document to start a rich discussion. This is an incipient research topic and preliminary results will be presented here.

In section II some fundamentals of thermal coupling in ICs are reviewed. Being the local temperature increase proportional to the dissipated power, sections III and section IV present some examples of how the power dissipated in digital and analog circuits changes due to PVT variation. Finally, section V presents final remarks and conclusions.

II. Thermal coupling in ICs

A. Fundamental Review

The relation between temperature increase in a point of the silicon surface $\Delta T(x, y, t)$ —x, y refers to surface coordinates and t refers to time— and the power dissipated by the different devices $P(x, y, t)$ is given by the heat transfer equation due to conduction [7]. Power and temperature increase are linked by physical properties of the silicon die: thermal conductivity and thermal capacitance.

Beyond the mathematical formalism of the thermal coupling mechanism, some facts are important from the engineering point of view. The relation between power and temperature increase can be considered linear, behaving as a low pass filter with a typical cut-off frequency of tens of kHz [5], [6]. This fact has two consequences: first, the power dissipated by the devices in the MHz or GHz range will not provoke temperature increases at such frequencies. Second, temperature increase in one point can be obtained by the superposition principle, as the addition of the individual contributions of the different devices that dissipate power. As the distance between the devices that dissipate power and the point where temperature is measured increases, the attenuation of the thermal coupling increases and the cut off frequency of the thermal coupling decreases.

Both analog and digital circuits have two components of dissipated power: static and dynamic. Static power is caused by sub-threshold and leakage current in digital circuits and by biasing currents in linear class A analog circuits. Usually, this component of the dissipated power is around DC, generating DC temperature increases in the silicon surface. Dynamic power dissipation is originated by the processing of information. Although temperature signals have small bandwidth, this fact does not hinder using them to monitor the high frequency electrical behavior of circuits.

In digital circuits dynamic power is mainly caused by charging and discharging capacitances. The average value of the dissipated power will generate a DC temperature increase, which adds to the temperature increase generated by the static dissipation. Similarly, in analog circuits, the DC component of the power dissipated by the devices has a contribution due to the dynamic activity of the circuit.

Fig. 1 represents the two elements that constitute a typical amplifier: a transconductance stage (usually formed by an active device and a cascaded device) and a load. We can write for the AC component of the signal:

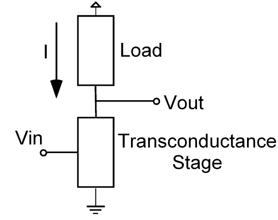

Fig. 1. Simplified schematic of a linear amplifier.

$$V_{out} = G_m R_o V_{in} = A_v V_{in} \tag{1}$$

Where G_m is the global transconductance of the transconductance stage, R_o is the output resistance of the amplifier and A_v is the voltage amplification gain. We can express the power dissipated by the transconductance stage as $P = V_{out} I$, where I is the current flowing through the amplifier. If we now consider only the power dissipated by the dynamic activity provoked by the amplification of a sinusoidal signal of amplitude A and frequency f applied to the input node, the power dissipated by the transconductance stage is:

$$P = \frac{1}{2} G_m A^2 A_v \left(1 + \cos(4\pi f t)\right) \tag{2}$$

From the two spectral components present in the Power expression, the first one adds up to the static power dissipation, being the overall DC power dissipated by the transconductance stage:

$$P_{DC} = P_{static} + \frac{1}{2} A^2 \frac{A_v^2}{R_o} \tag{3}$$

Therefore, in both digital and analog circuits, DC temperature increases provoked by the circuit activity depend on both the static and dynamic performances of the circuit.

B. Embedded temperature sensors to track circuit activity

Measuring temperature in integrated circuits is a mature subject with a large list of related works published in the literature, e.g. [8].

Temperature sensing strategies can be classified as off-chip or embedded.

High sensitivity is achieved with embedded temperature sensors when a differential sensing strategy is used [6]. These sensors are formed with two temperature sensing devices, one located close to the circuit under observation and the other one far away from it. Similarly to differential voltage amplifiers, the electrical output signal is proportional to the difference of temperature between these two sensing devices. Moreover, differential temperature sensors have shown low sensitivity to any offset that may be added to the surface thermal map (which acts as common mode signal), such as ambient temperature changes. These sensors are very sensitive to temperature gradients that are generated in the silicon

surface due to the power dissipated by devices and circuits, being therefore suitable to be used to monitor the power dissipated by circuits. For instance, results published in [6] report a sensitivity of up to 400mV/mW. These results were obtained when the circuit to be monitored was placed at 20 μm distance from the temperature sensor. Larger distances between the circuit and the temperature sensor would reduce the sensitivity to track the power dissipated by the circuit.

III. Digital circuits

In order to study the sensitivity of digital circuit performance with PVT variations, a chain of inverters in two different configurations has been considered: forming a ring oscillator with 5 stages, and in a simple combinational chain with 15 stages. The goal of the first circuit is to emulate the behavior of asynchronous circuits where the activity of the circuit does not depend of an external clock. In the second circuit, a pattern of 1-0-1-0 has been introduced with a period of 1.2 ns, as an example of a typical combinational stage in a synchronous circuit. These circuits were simulated using BSIM4 device models from a CMOS 45 nm technology.

The nature of transistor parameter variations is fundamentally random in nanometer technologies because of atomistic effects. For this reason, each single transistor has unique variations in physical parameters: L, W, t_{ox}, and V_{TH}, and they are essentially uncorrelated even to neighboring transistors.

Furthermore, interconnect parameter variation (modeled here as additional node capacitance C_L) can be thought of as being essentially global, affecting equally all interconnects in a given region of the chip. The same happens with voltage (V_{DD}) and ambient temperature (T). In addition, a global component of V_{TH} variation is added to consider possible global V_{TH} tuning of the whole block.

Given the above considerations, Montecarlo simulations were performed to evaluate the behavior of power and delay to PVT variations. Parameters varied are specified in Table I, indicating whether they are considered as local or global variations.

TABLE I
Parameters variation range and scope in the digital circuit

Parameter	Local or Global	Range
L	Local	10%
W	Local	5%
t_ox	Local	5%
V_{TH}	Local+Global	5%+ 20%
C_L	Global	20%
V_{DD}	Global	20%
Temp	Global	40–120°C

A. Inverters connected in Ring Oscillator (RO) Topology

The performance of this circuit is determined by the characteristics of the technology, bias voltage and temperature. It is, therefore, a circuit very sensitive to PVT variations.

The DC value of the power consumption is independent of the number of stages. In a first order approach, the average value of the dynamic power dissipation is the energy dissipated in a period divided by the period, which after some simplifications is:

$$P = \frac{C_L V_{DD}^2}{2t_p} \quad (4)$$

where C_L is the capacitance of each node, and t_p is the average propagation delay (rise or fall) of an inverter in the chain. Fig. 2 shows the result of 250 Montecarlo simulations. The vertical axis shows the resulting period, $T = 2Nt_p$ (where N is the number of stages, 5 in our example) of each simulation, and the horizontal axis shows the average dissipated power computed over 2 ns (10 periods approximately). Two different analysis are superimposed, the red '+' symbol shows the results when PVT variations are considered, whereas green '×' symbol shows the results considering only process variations. In any case, the relationship expressed by (4) indicates a strong correlation between the variation of delay and static power.

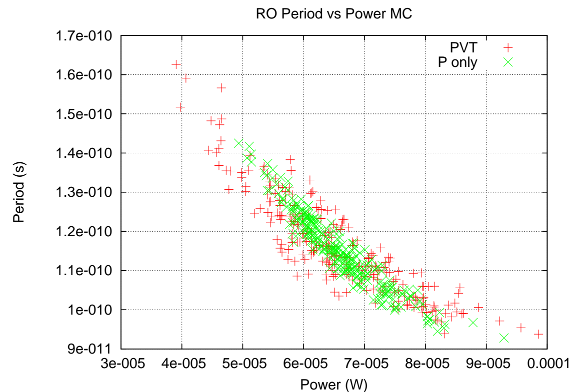

Fig. 2. RO: dependence between period and power.

Another interesting magnitude to consider is MOSFET threshold voltage, V_{TH}. A common strategy to adapt power consumption in digital circuits is the tuning of body bias [9] to modify V_{TH}. Fig. 3 plots the global component of the threshold voltage variation against static dissipated power. It is seen how both magnitudes are correlated, which opens the possibility to use temperature measurement as a way to infer the effective threshold voltage variation obtained by body biasing.

B. Chain of inverters

In this case, the dynamic power is dominated by the frequency of the signal driving the circuit (i.e., clock frequency):

$$P = NfC_L V_{DD}^2 \quad (5)$$

Gate delay does not appear explicitly in the power expression. However, delay and power are correlated by their respective dependence on voltage, threshold voltage, and other parameters. This is shown in Fig. 4, where the propagation

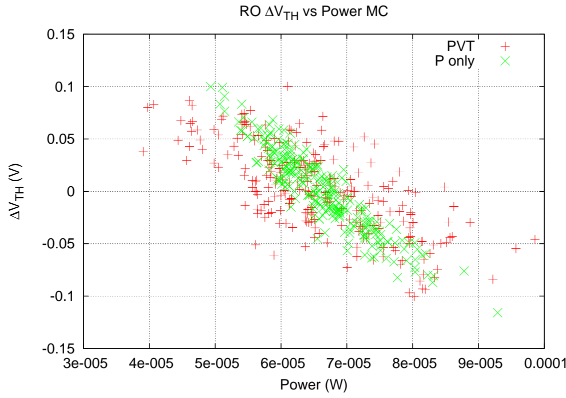

Fig. 3. RO: dependence between global component of V_{TH} deviation and power.

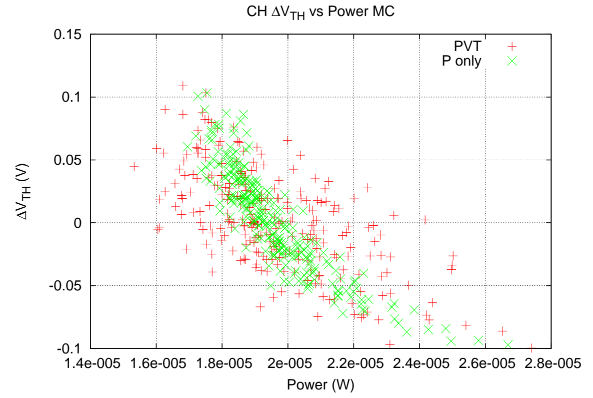

Fig. 5. Chain of inverters: dependence between global component of V_{TH} deviation and power.

time of 14 inverters is plotted against average power. The correlation is not as good as in the case of the RO, but still a correlation is denoted, meaning that power dissipation and eventually local temperature increase can be used to evaluate circuit delay variability.

Similarly to the RO example shown above, Fig. 5 shows the correlation between power and the global component deviation of the threshold voltage, to show the possibility of using temperature as a measure for V_{TH}.

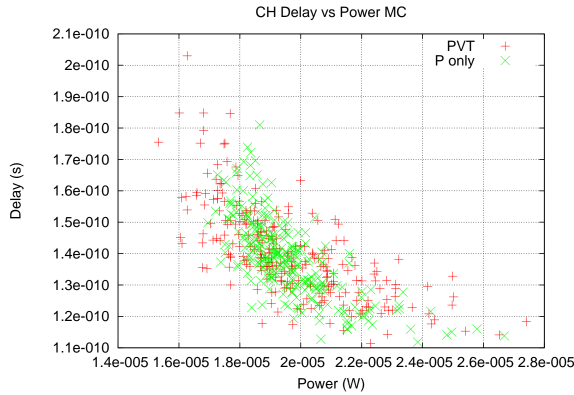

Fig. 4. Chain of inverters: dependence between delay and power.

In summary, for both kind of circuits here analyzed, PVT variations cause a larger dispersion of values than just process (P) variations, as should be expected. However, the point to be noted is that a variation in dissipated power causes a variation in local temperature increase that could be detected with an appropriate sensor as explained in section II.

IV. Analog circuits

Two different circuits have been considered as example of analog circuits. Both are class A amplifiers implemented in a 65 nm CMOS technology, but the first is a Low Noise Amplifier (LNA), whereas the second one is a Power Amplifier (PA).

A. LNA circuit

The LNA schematic is shown in Fig. 6. It consists in a common source cascode configuration (bias not included in the figure), with inductive degeneration and a T matching network at the input. Table II shows the nominal performances of the amplifier.

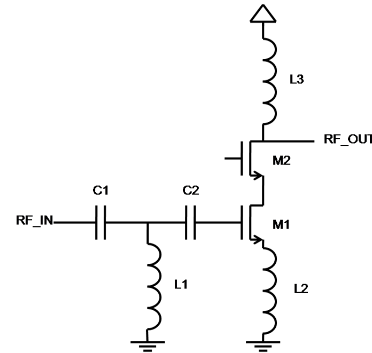

Fig. 6. Common source LNA considered.

TABLE II
Common-source LNA specifications

V_{DD}=1.2 V, T=27°C, f=2.45 GHz	
Voltage Gain	23 dB
IIP3	-2.2 dBm
CP1dB	-12.2 dBm
NF	5 dB
S11	-30 dB
PDC @ 1.2V	413.7 μW

The first experiment performed consists in obtaining the DC component of the power dissipated by the cascode transistor M2 together with the voltage gain, when a 2.45 GHz,

−30 dBm tone is applied at the LNA input. Fig.7 shows the voltage gain squared (see (3)) versus the M2 dissipated power for 100 Montecarlo simulation runs, considering only process variation and mismatch. We observe a high linear correlation between both magnitudes, which demonstrates the possibility of inferring the voltage gain departure from the expected due to process variation by means of measuring the DC temperature increase caused by the power dissipation of some component of the LNA.

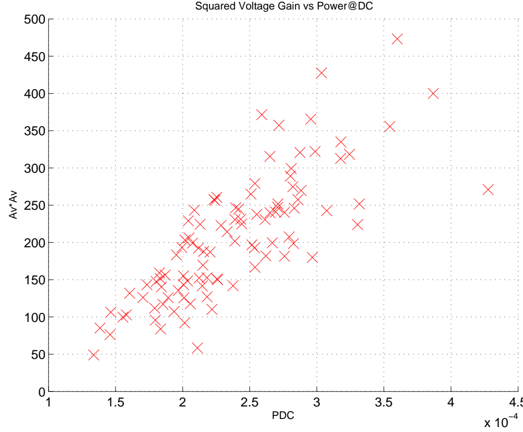

Fig. 7. LNA Voltage gain squared, versus DC component of the power dissipated by M2 transistor.

The expression (3) shows that the power dissipated at DC is contributed by the static and dynamic behavior of the LNA. The work in [10] shows a procedure to obtain a low frequency power dissipation that depends only on the dynamic behavior of the circuit: two tones are applied at the LNA input (f1=2.45 GHz and f2=2.46 GHz) and the amplitude of the spectral component of the dissipated power at f2-f1 (10 MHz) is monitored by temperature sensing. Using this driving strategy, a better correlation is obtained between the voltage gain and the magnitude of the dissipated power, as it is shown in Fig. 8.

B. PA circuit

The topology implemented is a pseudo-differential cascode amplifier working in class-A conditions [11]. The cascode transistor is an NMOS with standard V_{TH} while the input transistor is low-power NMOS with reduced V_{TH}. The PA nominal performance is summarized in Table III and the schematic in Figure 9. The circuit has been simulated using GoldenGate within the Cadence environment. In the simulation, we have included the model of a QFN24 package and external components: baluns, output DC block capacitors and choke inductors. We have studied the effects of PVT variations performing a corner analysis. The device models consider three different process corners, named Slow (SSA), Typical, and Fast (FFA), whereas the corner VDD values are 0.99 V, 1.1 V (nominal) and 1.21 V, and the corner Temperature values are 0°C, 50°C (nominal) and 105°C. For all the possible

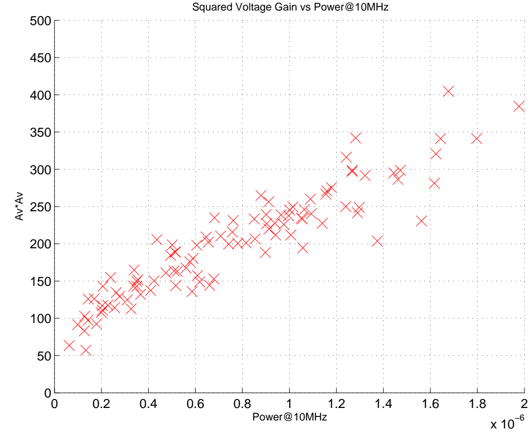

Fig. 8. LNA Voltage gain squared, versus the 10 MHz component of the power dissipated by M2 transistor.

PVT combinations (27 possible cases) we have performed two different experiments. In the first one, we obtained the static and dynamic components of the power dissipated at DC by each of the two transistors of one branch when a −10 dBm signal at 2.25 GHz is applied to its input. Next, conventional one tone and two tones analysis are done to extract the power gain, 1-dB compression point (CP1) and output third-order intermodulation products intercept point (OIP3).

TABLE III
PA CHARACTERISTICS

V_{DD}	1.1 V
Output power (OFDM signal)	0 dBm with 50 Ω load
Power Consumption	95 mW
Frequency range	300 MHz –2.5 GHz
MultiTone Power Ratio	-42 dBm
Power Gain	18 dB
Compression point	9.86 dBm

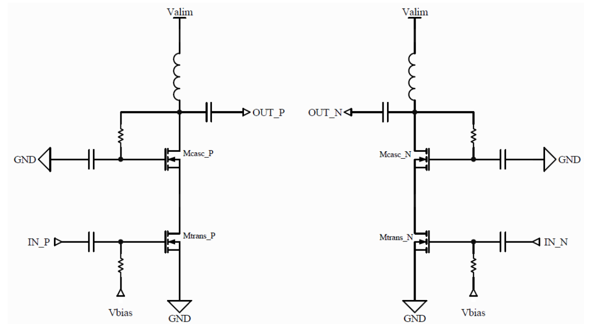

Fig. 9. Schematic of the PA.

Figure 10 compares the OIP3 versus the power dissipated by the gain transistor for each of the PVT cases. The graph shows a good correlation between the two magnitudes. This means the by measuring the temperature increase variation in comparison with the reference circuit cause by the deviation

of the power dissipated by the PA gain transistor the deviation in the OIP3 value from the nominal one can be detected. Similar results are obtained for the power gain, as shown in the same figure, whereas the CP1 shows little correlation with the transistor gain power dissipation.

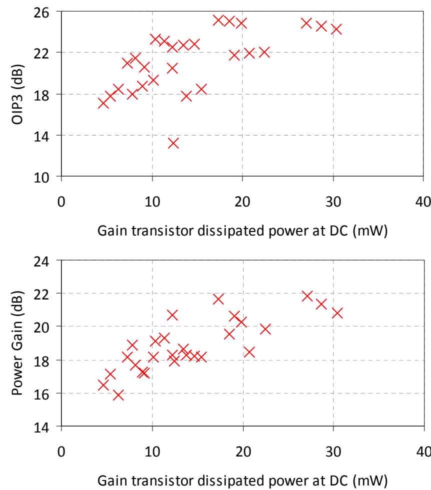

Fig. 10. OIP3 and Power Gain versus dissipated power.

V. Discussion and conclusions

Recent research works are considering temperature, the average temperature of the die or the small temperature fluctuation in specific locations of the circuit, as an efficient observable magnitude. Embedded temperature sensors exhibit a very high sensitivity and readability with low silicon overhead. Temperature has as a unique characteristic that does not cause an extra loading in the measuring circuits because of the natural coupling mechanism with the circuit under test.

In this paper authors have presented results about the existing correlation between parameter variations and power dissipation fluctuations. With the objective to introduce it as a proof of concept it has been shown that this correlation appears in all type of circuits showing as example several representative circuits of both digital and analog type. As a consequence of power fluctuations and following a linear law, temperature fluctuations would allow the observation of a specific component or circuit section.

In the case of digital circuits it has been shown the significant correlation between the propagation delay of an inverters chain, the period of a ring oscillator and in general the fluctuation of the devices threshold voltage with the power dissipation. Consequently temperature measurements may be used to detect deviations of such magnitudes caused by parameter variation. This has a strong interest in the case of failure analysis or yield improvements especially for adapting to or detecting deviations by using built-in temperature sensors and tuning knobs.

The use of temperature as state observable is very attractive because it provides a non-invasive monitoring technique for analog circuits. In this case the technique uses an intrinsic dynamic measurement that based in the injection of one or two tones is able to sense FoM deviation (such as gain and IP3) caused by PVT variations, using DC or dynamic local temperature increase sensors working at frequencies much lower than the corresponding RF circuits.

Consequently temperature is an efficient and non-invasive measuring technique that may be used to read and compensate or adapt key factor fluctuations caused by manufacturing deviations or external factors. Future research progresses will investigate the suitability of this strategy in the test, deviation reduction mechanisms and reliability assurance of future integrated circuits.

Acknowlegments

This work was partially supported by projects ENIAC MODERN (Spanish MICINN PLE2009-0024), MICINN TEC2008-01856, and Catrene CA101 PANAMA (Spanish Avanza I+D project TSI-020400-2009-59).

References

[1] C. Kim, K. Roy, S. Hsu, R. Krishnamurthy, and S. Borkar, "A process variation compensating technique with an on-die leakage current sensor for nanometer scale dynamic circuits," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 14, no. 6, pp. 646–649, 2006.

[2] B. Rebaud, M. Belleville, E. Beigne, M. Robert, P. Maurine, and N. Azemard, "An innovative timing slack monitor for variation tolerant circuits," in *IC Design and Technology, 2009. ICICDT '09. IEEE International Conference on*, 2009, pp. 215–218.

[3] M. Bhushan, A. Gattiker, M. Ketchen, and K. Das, "Ring oscillators for CMOS process tuning and variability control," *Semiconductor Manufacturing, IEEE Transactions on*, vol. 19, no. 1, pp. 10–18, 2006.

[4] A. Valdes-Garcia, R. Venkatasubramanian, J. Silva-Martinez, and E. Sanchez-Sinencio, "A broadband CMOS amplitude detector for on-chip RF measurements," *Instrumentation and Measurement, IEEE Transactions on*, vol. 57, no. 7, pp. 1470–1477, 2008.

[5] N. Nenadovic, S. Mijalkovic, L. Nanver, L. Vandamme, V. d'Alessandro, H. Schellevis, and J. Slotboom, "Extraction and modeling of self-heating and mutual thermal coupling impedance of bipolar transistors," *Solid-State Circuits, IEEE Journal of*, vol. 39, no. 10, pp. 1764–1772, 2004.

[6] J. Altet, A. Rubio, E. Schaub, S. Dilhaire, and W. Claeys, "Thermal coupling in integrated circuits: application to thermal testing," *Solid-State Circuits, IEEE Journal of*, vol. 36, no. 1, pp. 81–91, 2001.

[7] H. S. Carslaw and J. C. Jaegger, *Conduction of Heat in Solids*. Clarendon Press, 1986.

[8] D. Blackburn, "Temperature measurements of semiconductor devices - a review," in *Semiconductor Thermal Measurement and Management Symposium, 2004. Twentieth Annual IEEE*, 2004, pp. 70–80.

[9] N. Azizi and F. Najm, "Compensation for within-die variations in dynamic logic by using body-bias," in *IEEE-NEWCAS Conference, 2005. The 3rd International*, 2005, pp. 167–170.

[10] D. Mateo, J. Altet, E. Aldrete-Vidrio, and J. Gonzalez, "Frequency characterization of a 2.4 GHz CMOS LNA by thermal measurements," in *Radio Frequency Integrated Circuits (RFIC) Symposium, 2006 IEEE*, 2006, pp. 4 pp.–517.

[11] Y. Ding and R. Harjani, "A high-efficiency CMOS +22-dBm linear power amplifier," *Solid-State Circuits, IEEE Journal of*, vol. 40, no. 9, pp. 1895–1900, 2005.