An Ultra-small On-Chip Sensor for Temperature and Thermal Gradient Measurements

Yen-Ting Wang[1,2], Chen Zhao[1], Degang Chen[1], Shu-Chuan Huang[2], and Randall Geiger[1]
Iowa State University[1] and Tatung University[2]
e-mail: {yentingw,zhaochen,djchen,rlgeiger}@iastate.edu, schuang@ttu.edu.tw

Abstract- **An ultra-small, low-power CMOS temperature sensor designed for multi-site temperature and temperature gradient monitoring to support power/thermal management in integrated circuits where electromigration limits system reliability is proposed. The temperature sensor utilizes the temperature characteristics of the threshold voltage of MOS transistors to sense temperature and is highly linear throughout the temperature range required for power/thermal management and has good accuracy throughout this range with calibration at a single temperature.**

Prototype temperature sensors have been designed in a TSMC 0.18µm process with a 1.8V supply voltage. The sensing device has a very small die area of only 2.5µm×6.2µm and a low power consumption of about 0.95µW at a 1% duty cycle making it suitable for multi-site thermal monitoring.

Key Words: Temperature Sensor, Thermal Gradient, Thermomigration, Threshold Extraction

I. INTRODUCTION

As feature size and frequency in emerging processes scale faster than supply voltage, power density is increasing as is the presence of large time-dependent thermal gradients. Systems operated under stress caused by high current density, high temperature, or large temperature gradients are plagued by a decrease in lifetime and reliability. Thermal management has become a key consideration in the design of integrated circuits (ICs) and is necessary to maintain lifetime and reliability at acceptable levels in systems operating under high thermal/electrical stress. Most power/thermal management approaches are based upon monitoring the temperature at one or more locations on a chip and using this information to change or adapt the usage or performance to guarantee that the temperature does not exceed a predetermined trigger temperature. When the sensed temperature exceeds the trigger temperature, the clock frequency is often throttled or the supply voltage is reduced to keep the system operating at a temperature [1] that will provide an acceptable reliability metric. Several techniques have been developed to reduce the average temperature and/or reduce the peak temperature in order to manage cooling costs [2]-[3].

Interconnect lifetime is one of the dominant contributors to reliability degradation under electrical/thermal stress in VLSI circuits and is of increasing concern with scaling of size and packing density. Interconnect degradation and ultimately failures occur in the presence of electrical and thermal stress associated with high current levels, high temperature levels, fast temperature cycling, and large thermal gradients. This degradation is often attributed to electromigration and has been a major research topic in the reliability field within the IC industry over the past two decades [4]-[6].

Though thermomigration also causes degradation in interconnects, these effects are often insignificant compared to the reliability reduction caused by electromigration. But some studies have shown that interconnect lifetime under electrical stress degrades significantly in the presence of static thermal gradients too suggesting that thermomigration (or temperature-gradient-enhanced electromigration) is also of major concern [7]-[10] and these thermal gradients will likely increase in emerging processes with projected increases in both the total number of processing cores and the total die power dissipation. Experimental results have shown that negative or positive thermal gradients can either improve or reduce mean time to failure (MTF) [7] in the presence of electrical stress. Assessments of the severity of temperature gradients on electromigraiton and reliability are discussed in [8-10]. The authors of [9] make a reliability comparison between a simple AlSiCu interconnect under electrical stress operating at a constant temperature (T=202˚C) with the same interconnect operating with the same electrical stress but with a static thermal gradient (T=152˚C to T=202˚C) that averages about 0.19^{o}C/µm. A reduction in the MTF of 50% was reported in the presence of an unfavorably-directed static thermal gradient. There is limited material in the literature focusing on the magnitude of the static thermal gradients that will cause significant reliability reduction but repeated concerns about managing thermal gradients to circumvent reliability problems bear out the importance of the problem. This suggests that power/thermal management algorithms that depend upon using a single trigger temperature, even if monitored at many sites on a die, are not adequate to accurately maintain reliability in electrically-stressed circuits where large thermal gradients are present.

These results suggest that significant improvements in reliability can be obtained if both local die temperature and local thermal gradients are used as a part of the power/thermal management algorithm. Since many large integrated circuits have a large number of locations on the die where high temperatures can occur or where high thermal gradients can occur, and since these regions can be quite small, it is critical that sensors that measure temperature and temperature gradients be extremely small.

The temperature-dependence of the I-V characteristics of the PN junction diode [11] has been widely used for sensing temperature but PN diodes are invariably large and dissipate considerable power. Aside from the size and power concerns, diode-based temperature sensors are still not practical in fine feature CMOS processes for several reasons. One reason is

because in most CMOS processes, a good junction diode is not available but rather a parasitic vertical BJT is diode-connected to behave as a diode with a P diffusion in an N-well over the P-type substrate (PNP). This substrate PNP has a temperature-dependent β, a large temperature-dependent base-spreading resistance, and a large temperature-dependent collector resistance. These factors all degrade performance. In addition, the temperature-dependent models of the parasitic bipolar devices are not well-developed in emerging CMOS processes making it difficult to accurately predict performance.

In this work, a threshold-voltage-based supply-insensitive temperature sensor [12], [13] is adapted to provide both thermal gradient and average temperature readings. The temperature-sensing part of this circuit is very small making it suitable for inclusion in critical parts of a circuit where hot spots are most problematic as well as making it useful for measuring localized thermal gradients. The overall area is also small and the total supply current is very small making it applicable for sensing temperature and temperature gradients at multiple locations on a die. In Section II, the proposed temperature sensor is analytically characterized. Simulation results are presented in Section III. Section IV describes the circuit layout strategy for both average temperature and thermal gradient measurements. Section V concludes this work.

II. TEMPERATURE SENEOR CIRCUIT DESCRIPTION

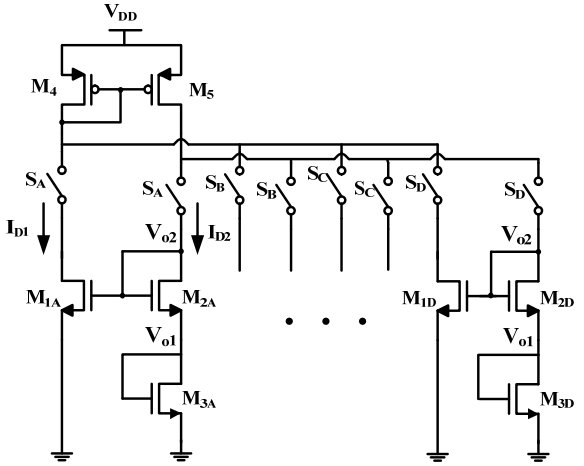

Fig.1 Proposed temperature sensor (start-up not shown)

The proposed temperature sensor shown in Fig. 1 is derived from the inverse-Widler self-stabilized temperature sensor [14]. Sensors B,C, … are not shown in the schematic. Neglecting output conductance effects and assuming that the transistors are operating in saturation and are characterized by the ideal square-law model, it follows that:

$$I_{D1} = \frac{\mu_{n1} C_{ox}}{2}\left(\frac{W_1}{L_1}\right)(V_{O2} - V_{tn1})^2 \tag{1}$$

$$I_{D2} = \frac{\mu_{n2} C_{ox}}{2}\cdot\left(\frac{W_2}{L_2}\right)(V_{O2} - V_{O1} - V_{tn2})^2 \tag{2}$$

$$I_{D2} = \frac{\mu_{n3} C_{ox}}{2}\left(\frac{W_3}{L_3}\right)(V_{O1} - V_{tn3})^2 \tag{3}$$

$$I_{D2} = M \cdot I_{D1} \tag{4}$$

where M is the gain of the M_4:M_5 current mirror and where$W_{1\text{-}3}$ and $L_{1\text{-}3}$ are the equivalent width and length of the $M_{1\text{-}3}$ transistors respectively as determined by the number of switches that are closed.

Assuming that M_4 and M_5 are operating at the same temperature with same size and matched mobilities, it follows from (1)-(4) that V_{O1} and V_{O2} can be expressed as:

$$V_{O1} = \frac{(V_{tn2} - V_{tn1})\cdot\sqrt{\frac{\mu_{n2}(W/L)_2}{\mu_{n3}(W/L)_3}} + V_{tn3}\cdot\left(\sqrt{\frac{\mu_{n2}(W/L)_2}{\mu_{n1}(W/L)_1}} - 1\right)}{\sqrt{\frac{\mu_{n2}(W/L)_2}{\mu_{n1}(W/L)_1}} - \sqrt{\frac{\mu_{n2}(W/L)_2}{\mu_{n3}(W/L)_3}} - 1} \tag{5}$$

$$V_{O2} = \frac{V_{tn2}\cdot\sqrt{\frac{\mu_{n2}(W/L)_2}{\mu_{n1}(W/L)_1}} + V_{tn3}\cdot\sqrt{\frac{\mu_{n2}(W/L)_2}{\mu_{n1}(W/L)_1}} - V_{tn1}\cdot\left(1 + \sqrt{\frac{\mu_{n2}(W/L)_2}{\mu_{n3}(W/L)_3}}\right)}{\sqrt{\frac{\mu_{n2}(W/L)_2}{\mu_{n1}(W/L)_1}} - \sqrt{\frac{\mu_{n2}(W/L)_2}{\mu_{n3}(W/L)_3}} - 1} \tag{6}$$

The expressions show that the output voltages V_{O1} and V_{O2} are independent of the supply voltage V_{DD} and have a linear relationship with the threshold voltage. Since the threshold voltage is nearly linear with temperature and since temperature μ_{n2}/μ_{n3} and μ_{n2}/μ_{n1} will cancel if M_1, M_2, and M_3 are operating at the same temperature, it follows that V_{O1} and V_{O2} are nearly linear w.r.t. temperature.

The temperature dependence of the threshold voltage in the BSIM model [15], given in (7), shows explicitly the temperature dependence of the threshold voltage used in most simulators.

$$V_{tn}(T) = V_{tn0} + (KT1 + \frac{KT1L}{L_{eff}} + KT2\cdot V_{bseff})\cdot(\frac{T}{T_{NOM}} - 1) \tag{7}$$

In this model, *KT*1, *KT*1*L*, and *KT*2 are process dependent constants, T_{NOM} is equal to 300K, L_{eff} is the effective length, and V_{bseff} is the effective bulk to source voltage. If V_{bseff} is not zero, a weak nonlinearity of threshold voltage w.r.t. temperature is introduced but overall the threshold voltage is nearly linear w.r.t. temperature. For notational convenience, we will express $V_{tn}(T)$ as

$$V_{tn}(T) = V_{tn00} + \gamma T \tag{8}$$

To maintain operation in the saturation region for the circuit of Fig. 1, the relationship in (5) must be satisfied:

$$\sqrt{\frac{\mu_{n2}(W/L)_2}{\mu_{n1}(W/L)_1}} > 1 + \sqrt{\frac{\mu_{n2}(W/L)_2}{\mu_{n3}(W/L)_3}} \tag{9}$$

If we neglect the bulk effect on the threshold voltage of M_2 and assume M_1, M_2 and M_3 are operating at the same temperature and assume that the threshold voltages V_{tn1} and V_{tn2} are matched, (5) simplifies to:

$$V_{O1} = V_{tn3}\cdot\left(\frac{\sqrt{\frac{(W/L)_2}{(W/L)_1}} - 1}{\sqrt{\frac{(W/L)_2}{(W/L)_1}} - \sqrt{\frac{(W/L)_2}{(W/L)_3}} - 1}\right) \tag{10}$$

From (10), only the temperature dependence of the threshold voltage of M_3 contributes to the temperature dependence of V_{O1} and from (7) and (10), the temperature is given by

$$T = \frac{\theta V_{O1} - V_{tn300}}{\gamma} \tag{11}$$

where $\theta = \frac{\sqrt{\frac{(W/L)_2}{(W/L)_1}} - \sqrt{\frac{(W/L)_2}{(W/L)_3}} - 1}{\sqrt{\frac{(W/L)_2}{(W/L)_1}} - 1}$, and where the temperature coefficient is given by γ/θ. The value for V_{tn300}, which is not temperature dependent, can be obtained from a single-point calibration and the parameter γ from a single batch calibration. So, the strategy for designing an ultra-small on-chip temperature sensor is to use three diode-connected transistors M_1-M_3 (or parallel/fingered devices as determined by S_A and S_D) as the temperature sensing device with M_1-M_3 being placed at temperature-critical locations on a die.

To measure a one-dimensional temperature gradient, consider two n-channel transistor cells, $\{M_{1A}, M_{2A}, M_{3A}\}$ and $\{M_{1B}, M_{2B}, M_{3B}\}$ with the distance between the geometric centroids of the cells being d_{AB}. Designate the output with The S_A switches closed as V_{O1A} and the output with the S_B switches closed as V_{O1B}. It T_A and T_B are the temperatures of the "A" cell and the "B" cell respectively, it follows from (11) that

$$T_A - T_B = \frac{\theta(V_{O1A} - V_{O1B}) + V_{tn3B00} - V_{tn3A00}}{\gamma} \quad (12)$$

and the thermal gradient G_{AB} in the direction from A to B is given by

$$G_{AB} = \frac{T_A - T_B}{d_{AB}} = \frac{\theta(V_{O1A} - V_{O1B}) + V_{tn3B00} - V_{tn3A00}}{\gamma d_{AB}} \quad (13)$$

If T_A and T_B are measured using (11), the temperature at the geometric centroid of the "A" and "B" cells, denoted at T_{AB}, can be obtained by taking the average of T_A and T_B which is expressed as

$$T_{AB} = \frac{\theta(V_{O1A} + V_{O2A}) + V_{tn3B00} + V_{tn3A00}}{2\gamma} \quad (14)$$

The parameters V_{Tn3A00} and V_{Tn3B00} can be obtained with a single point calibration.

Multiple M_1-M_3 cells can be appropriately spaced to monitor two-dimensional gradients as discussed below. The location of the devices M_4, and M_5 is not critical and appropriate interdigitated layout methods can be used to assure that these devices are operating at the same temperature.

The key distinction between this temperature sensor and other inverse-Widlar structures discussed in the literature is that it is not necessary to use the whole five-transistor structure as a temperature sensor. Here, the sensor output is temperature insensitive to the temperature of the upper p-channel devices provided they operate at the same temperature so the position and the local temperature of those two transistors are not critical. Since the sensor is low current, the voltage drop along the lines connecting the core to individual M_1-M_3 cells is negligible small.

III. SIMULATION RESULTS

The temperature sensor can operate over a very wide range from a low temperature of about -70˚C to very high temperatures. However, the simulation results in this paper focus on a higher and narrow temperature range, 60°C to 90°C, where temperature sensors are a key component in power/thermal management. In this range, electrical/thermal stress affects the reliability of an integrated circuit and in contrast to existing approaches where only maximum temperature is used to throttle operation, and the proposed temperature sensors will provide both average temperature and temperature gradient information that can be used to better manage system reliability.

Table 1

Device	W/L	multiplier
M_1	0.3μ/0.9 μ	1
M_2	5μ/0.3 μ	1
M_3	5μ/0.3 μ	1
M_4	0.56μ/0.3 μ	2
M_5	0.56μ/0.3 μ	2

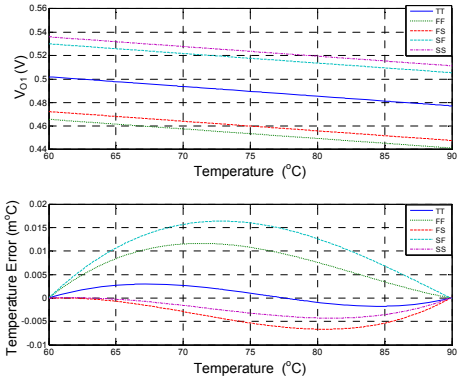

Fig. 2 Simulation results with the original model

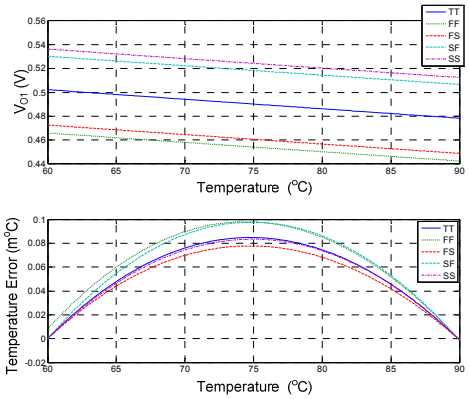

Fig. 3 Simulation results with 2nd order

The temperature sensor has been designed in a TSMC 0.18μm process with a 1.8V supply voltage. Device sizes are given in Table 1. This design has a temperature coefficient, γ/θ of 0.88mV/°C. The simulation results in Fig. 2 for a single output with all devices operating at the same temperature show very good linearity over process corners, typically designated as TT, FF, FS, SF, and SS. Simulation results show the nonlinearity error is bounded by ±0.02 ˚C over a 30 ˚C range. The voltage level changes due to the device mismatch and process variations corresponding to the shifts in the curves can be corrected with a one-point calibration.

The temperature-dependent threshold voltage model of (7) does not account for higher-order temperature nonlinearities

which are known to exist in real devices. This nonlinearity will introduce some nonlinear errors in the proposed temperature sensor. Measurements of the temperature dependence of the threshold voltage of MOS devices from a single run in this process correlate much better with simulation results if a second-order temperature coefficient is added to the BSIM model. This can be implemented by replacing the coefficient *KT*1 with

$$KT1 = -9.2 \cdot 10^{-3}(\frac{T}{T_{NOM}} - 1) - KT1_O \quad (15)$$

where $KT1_O$ is the constant value of the original model. Simulation results with the modified 2nd order temperature coefficient model are shown in Fig.3. The nominal voltage level is about the same as with the original model but the nonlinear errors increase to about ±0.15°C. These more realistic errors are still very small and of little concern when this temperature sensor is used for temperature or temperature gradient measurements for power/thermal management applications.

IV. Layout and Placement for Thermal Gradient Measurement

Different layout and placements of the sensors can provide various temperature gradient information. Four different placements of the M_1-M_3 transistors are shown in Fig. 4. The regions labeled with the letters A, B, C, and D correspond to the gate regions of the corresponding M_1-M_3 transistors. The spacing between these regions is shown as being close in the figure but the distance between the regions can be arbitrary. The layout of Fig. 4a is useful for obtaining linear gradients in the horizontal direction and two segments, $M_{1\text{-}3A}$ and $M_{1\text{-}3B}$ are used to measure the gradient along with two corresponding output voltages V_{O1A} and V_{O1B} obtained by closing first switch S_A and then switch S_B. Vertical and horizontal gradients can be obtained by using four gate regions as shown in Fig. 4b and 4c. To measure the vertical gradient, switches S_A and S_B are closed and output voltage V_{O1AB} is obtained. Then switches S_C and S_D are closed and V_{O1CD} is obtained. The vertical gradient is than calculated from (14) using the voltages V_{01AB} and V_{01CD}. To measure the horizontal gradient, switches S_A and S_C are closed and V_{O1AC} is obtained. Then switches S_B and S_D are both closed and V_{01BD} is obtained. The horizontal gradient is then calculated from (14). The layout of Fig. 4d can be used to measure gradients in the -45° direction.

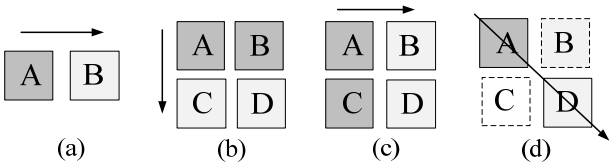

Fig. 4 Layout for temperature gradient sensing

V. Conclusion

An ultra-small temperature sensor has been designed in a TSMC 0.18 μm process and the active area for the temperature sensing element is only 15.5μm^2. This temperature sensor can measure temperature at arbitrary locations on a die as well as thermal gradients in any direction. It provides good accuracy with a single point temperature calibration for use in power/thermal management applications where electrical and thermal stress and stress gradients limit the reliability of the circuit.

References

[1] Duarte, D.E.; Taylor, G.; Wong, K.L.; Mughal, U.; Geannopoulos, G.; , "Advanced thermal sensing circuit and test techniques used in a high performance 65nm processor," *Low Power Electronics and Design (ISLPED), 2007 ACM/IEEE International Symposium on* , vol., no., pp.304-309, 27-29 Aug. 2007

[2] Chaparro, P.; Gonzalez, J.; Gonzalez, A.; , "Thermal-aware clustered microarchitectures," *Computer Design: VLSI in Computers and Processors, 2004. ICCD 2004. Proceedings. IEEE International Conference on* , vol., no., pp. 48- 53, 11-13 Oct. 2004

[3] Oleg Semenov, Arman Vassighi, and Manoj Sachdev. "Impact of selfheating effect on long-term reliability and performance degradation in CMOS circuits". Device and Materials Reliability, IEEE Transactions, 6(1), 2006.

[4] Nguyen, H.V.; Salm, C.; Krabbenborg, B.H.; Bisschop, J.; Mouthaan, A.J.T.; Kuper, F.G.; , "Fast thermal cycling-enhanced electromigration in power metallization," *Device and Materials Reliability, IEEE Transactions on* , vol.4, no.2, pp. 246- 255, June 2004

[5] Matsunaga, T.; Uegai, Y.; , "Thermal Fatigue Life Evaluation of Aluminum Wire Bonds," *Electronics Systemintegration Technology Conference, 2006. 1st* , vol.2, no., pp.726-731, 5-7 Sept. 2006

[6] Smorodin, T.; Wilde, J.; Alpern, P.; Stecher, M.; , "A Temperature-Gradient-Induced Failure Mechanism in Metallization Under Fast Thermal Cycling," *Device and Materials Reliability, IEEE Transactions on* , vol.8, no.3, pp.590-599, Sept. 2008

[7] Guo, W.; Li, Z.; Zhu, H.; Zhang, W.; Ji, Y.; Sun, Y.; Shen, G.; , "Temperature gradient impact on electromigration failure in VLSI metallization," *Semiconductor Thermal Measurement and Management Symposium, 1998. SEMI-THERM Proceedings 1998., Fourteenth Annual IEEE* , vol., no., pp.122-127, 10-12 Mar 1998

[8] Kim Jonggook; Tyree, V.C.; Crowell, C.R.; , "Temperature gradient effects in electromigration using an extended transition probability model and temperature gradient free tests. I. Transition probability model," *Integrated Reliability Workshop Final Report, 1999. IEEE International* , vol., no., pp.24-40, 1999

[9] Nguyen, H.V., Salm, C.; Krabbenborg, B.;Weide-Zaage, K.; Bisschop, J.; Mouthaan, A.J.; Kuper, F.G.; , "Effect of thermal gradients on the electromigration life-time in power electronics," *Reliability Physics Symposium Proceedings, 2004.42nd Annual. 2004 IEEE International* , vol., no., pp. 619- 620, 25-29 April 2004

[10] Zhijian Lu; Wei Huang; Stan, M.R.; Skadron, K.; Lach, J.; , "Interconnect Lifetime Prediction for Reliability-Aware Systems," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on* , vol.15, no.2, pp.159-172, Feb. 2007

[11] Bakker, A.; Huijsing, J.H.; , "Micropower CMOS temperature sensor with digital output," *Solid-State Circuits, IEEE Journal of* , vol.31, no.7, pp.933-937, Jul 1996

[12] Jun He; Chen Zhao; Sheng-Huang Lee; Peterson, K.; Geiger, R.; Degang Chen; , "Highly linear very compact untrimmed on-chip temperature sensor with second and third order temperature compensation," *Circuits and Systems (MWSCAS), 2010 53rd IEEE International Midwest Symposium on* , vol., no., pp.288-291, 1-4 Aug. 2010

[13] Sheng-Huang Lee; Chen Zhao; Yen-Ting Wang; Degang Chen; Geiger, R.; , "Multi-threshold transistors cell for Low Voltage temperature sensing applications," *Circuits and Systems (MWSCAS), 2011 IEEE 54th International Midwest Symposium on* , vol., no., pp.1-4, 7-10 Aug. 2011

[14] R.J. Widlar**, "**New developments in IC voltage regulators,"IEEE J. Solid-State Circuits, vol.SC-6, no.1, pp.2–7, Feb**.** 1971**.**

[15] X. Xi, M. Dunga, J. He, W.Liu and C. Hu,"BSIM4.3.0 MOSFET Model Users' Manual." Berkeley, CA: Univ. California, 2001, pp. 12-1.