Wide-Supply-Range All-Digital Leakage Variation Sensor for On-Chip Process and Temperature Monitoring

A. K. M. Mahfuzul Islam, *Member, IEEE*, Jun Shiomi, *Student Member, IEEE*, Tohru Ishihara, *Member, IEEE*, and Hidetoshi Onodera, *Senior Member, IEEE*

***Abstract*—Variation in process, voltage and temperature is a major obstacle in achieving energy-efficient operation of LSI. This paper proposes an all-digital on-chip circuit to monitor leakage current variations of both of the nMOSFET and pMOSFET independently. As leakage current is highly sensitive to threshold voltage and temperature, the circuit is suitable for tracking process and temperature variation. The circuit uses reconfigurable inhomogeneity to obtain statistical properties from a single monitor instance. A compact reconfigurable inverter topology is proposed to implement the monitor circuit. The compact and digital nature of the inverter enables cell-based design, which will reduce design costs. Measurement results from a 65 nm test chip show the validity of the proposed circuit. For a 124 sample size for both of the nMOSFET and pMOSFET, the monitor area is 4500 μm² and active power consumption is 76 nW at 0.8 V operation. The proposed technique enables area-efficient and low-cost implementation thus can be used in product chips for applications such as dynamic energy and thermal management, testing and post-silicon tuning.**

***Index Terms*—Leakage current, MOSFET, on-chip sensor, process variation, reconfigurable, ring oscillator, temperature.**

I. Introduction

VARIATION in PVT (Process, Voltage and Temperature) has become a major performance limiting factor in scaled CMOS process [1]. For high performance applications such as processors, peak performance is often limited by the power/thermal budget [2]. Many chips operate in environmental conditions with high temperature. High temperature accelerates transistor aging causing severe reliability problems. High temperature increases leakage current causing serious energy loss. Especially for low-power and energy-aware applications, leakage power becomes a limiting factor in terms of battery life. Energy-efficient operation of LSI can be related to effective management of leakage energy. Thus, dynamic management of energy and temperature of chips has become a necessity. Various dynamic energy-management techniques such as dynamic voltage and frequency scaling (DVFS), back-gate biasing, power gating, etc are now employed in the product chips [2]–[4]. Under these scenarios, on-chip monitoring of PVT variation is a must.

Besides the power consumption, costs for post-silicon timing validation and testing are increasing with process scaling. Timing failure can be caused by either hardware defects or parametric variations. Parametric faults are caused by Die-to-Die (D2D) and Within-die (WID) process variation. Especially, WID random variation of MOSFET $V_{\rm th}$ has become a major problem. WID variation also results in large leakage consumption [5], thus process variation control is a must for any successful fabrication. Low-cost characterization and modeling of process variation are necessary for process monitoring.

From the above discussions, there are strong needs for on-chip leakage, process and temperature sensors. On-chip sensors can provide designers and users of the chip useful hardware information, and help tune and debug their circuits accordingly. The cost and area over-head are two big factors limiting the wide use of on-chip sensors. Normally, different sensors are required monitoring MOSFET leakage current variation, process variation, temperature variation, and so on. The sensor design becomes more complex if we want to monitor nMOSFET and pMOSFET separately. An area-efficient on-chip process sensing methodology providing multiple parameter monitoring can play an important role to energy-efficient operation.

In this paper, we propose an area-efficient all-digital leakage current variation sensor, which enables monitoring of process, leakage and temperature variation. The proposed technique converts leakage current to time and measures transistor-by-transistor variation utilizing reconfigurable inhomogeneous topology architecture. No external voltage reference or bias generator circuit is required to monitor leakage variation with our circuit. Thus, the circuit can be operated at any supply voltage without any tuning. Leakage current variation does not depend on supply voltage, thus a voltage independent measurement is obtained. This paper assumes that sub-threshold leakage current is the dominant leakage factor in the scaled CMOS process. As sub-threshold leakage current is exponentially related to threshold voltage and temperature, leakage current variation can thus be used for on-chip process and temperature monitoring. A proof-of-concept on-chip sensor circuit is designed and fabricated in a 65 nm

Manuscript received February 20, 2015; revised April 29, 2015, June 18, 2015; accepted July 14, 2015. This paper was approved by Guest Editor Stefan Rusu. This work was supported in part by JSPS KAKENHI Grant Numbers 25280014 and 25·6432.

A. K. M. M. Islam is with the Institute of Industrial Science, University of Tokyo, Tokyo 153-8505, Japan (e-mail: mahfuzul@iis.u-tokyo.ac.jp).

J. Shiomi, T. Ishihara, and H. Onodera are with the Department of Communications and Computer Engineering, Kyoto University, Kyoto 606-8501, Japan.

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/JSSC.2015.2461598

technology for demonstration. The proposed sensor advances the state-of-the-art in the following way.

a) Area-efficient. Only a single instance is used to capture statistical properties of both nMOSFET and pMOSFET leakage currents,
b) Single sensor to monitor process, leakage and temperature variation,
c) Digital in nature both in design and measurement, and
d) Wide-supply-range operation down to 0.5 V.

The digital nature of the sensor provides design automation of placement and routing. Thus, low-cost flexible implementation of the circuit is realized.

This paper is an extension of the paper [6]. We give detailed description of the proposed methodology. We discuss on the various design options and their effects on sensing accuracy. We present measurement results from multiple chips and show the correlation between nMOSFET and pMOSFET leakage currents. We find no significant correlation for WID leakage variation. With detailed description and measurement results, we validate our proposed leakage variation sensing technique. We measure the effect of temperature on the leakage current and show that our circuit captures temperature dependency of leakage current correctly. Temperature dependency of 124 nMOSFETs and 124 pMOSFETs are obtained with our circuit.

The remainder of the paper is organized as follows. In Section II, an overview on the state-of-the-art of on-chip process and temperature sensors is provided. In Section III, the concept of process and temperature monitoring based on leakage current sensing is described in details. Effects of various uncertainties resulting from process variation are also explained here. In Section IV, a topology-reconfigurable leakage monitor cell is proposed which operates digitally. Various design parameters are explained and summarized here. In Section V, an area-efficient reconfigurable leakage variation sensor architecture is proposed. In Section VI, a design example of a demonstration circuit in a 65 nm process is explained. Simulation results are demonstrated here to show the validity of our proposal. Measurement results on leakage current variation are demonstrated in Section VII. Process variation extraction is performed and temperature dependency on leakage current is examined in detail. Finally, Section VIII concludes the paper.

II. Prior Work

Various on-chip circuits are proposed for monitoring PVT variation. For example, an all-digital all-chip process control module is proposed in [7]. An on-chip digital measurement method is proposed in [8]. [9] proposes an on-chip leakage current sensor for measuring D2D process variation. A leakage current sensor is used for compensating process variation [10], [11]. Band-gap reference based temperature sensors are proposed for high accuracy [12]. MOSFET leakage current based temperature sensors are proposed [13], [14]. All-digital time-domain temperature sensor is proposed in [15]. The state-of-the-art sensing technique can be characterized as follows: a) mostly analog, b) design is tuned to monitor either of leakage, or temperature variation, c) cannot operate at a wide supply range, d) separate sensors for transistors with different sizes.

Leakage current is a strong function of MOSFET $V_{\rm th}$ and temperature. Accurate measurements of leakage current variation of transistors thus provide the information of process variation and temperature. An area-efficient all-digital leakage variation sensor circuit can exploit the leakage variation to efficiently monitor both of the process variation and temperature.

MOSFET leakage current sensor circuit design is tricky as leakage current is in the order of pA. High precision I-V measurement equipment is thus required which are not suitable for on-chip implementation. Analog circuitry such as analog-to-digital converters (ADC) can aid the measurement of leakage current. With the help of ADC, leakage variation of large MOSFET array can be measured [16]. However, implementing analog circuitry such as an ADC consumes large power and area. An on-chip 6-channel leakage current sensor is proposed in [10], where bias circuits, and an analog comparator are used. However, WID leakage variation monitoring is not achieved in their approach. Analog approaches require fine tuning of various components which prevents wide supply range operation. This paper for the first time proposes an all-digital leakage current sensor circuit, which operates at wide supply range and can measure leakage variation of large transistor samples from just a single sensor instance. The leakage current variation can then be used for monitoring process and temperature variations. The proposed circuit is digital, thus provides implementation and operation flexibility. In addition, it provides large measurement samples which can be used effectively for improving the accuracy of monitoring. In this paper, we discuss the theoretical aspects of our proposal and prove the proposed concept with measurement results from a 65 nm demonstration circuit. Utilizing the capabilities of the circuit, circuit designers can implement various monitoring schemes according to their needs.

III. Time-Domain Leakage Current Monitoring

A. Leakage Current in an MOS Transistor

With the decrease of $V_{\rm th}$ to cope with supply voltage scaling in deep sub-micron process, sub-threshold leakage current has been increasing exponentially. Besides the sub-threshold leakage, various other leakage paths has been a concern as a result of short channel effect (SCE) and gate oxide scaling [17], [18]. Leakage current can thus occur in the form of gate tunneling, punch-through current between source-drain, reverse pn junction current, etc. Various transistor channel and doping engineering techniques are required to keep these effects within a tolerable range. Since the introduction of high-k metal-gate technology, gate leakage has been kept under control [19]. However, sub-threshold leakage current still remains a concern and consumes a significant portion of the overall power consumption. Sub-threshold leakage current has exponential relationship with transistor threshold voltage and temperature. Thus sub-threshold leakage current suffers the most from process and temperature variations. Conversely, by monitoring sub-threshold leakage variation, underlying process and temperature variations can be extracted. This paper

concentrates on process and temperature variations, and their effects on sub-threshold leakage current variation. Considering sub-threshold leakage current is the dominant factor, this paper proposes a digital circuit technique to monitor leakage current variation on-chip for process and temperature monitoring.

B. Sub-Threshold Leakage Current Model

Sub-threshold leakage current $I_{\rm leak}$ of a MOSFET is a strong function of its threshold voltage $V_{\rm th}$ and temperature T, and can be modeled as follows [20].

$$I_{\rm leak} = I_0 \mathrm{e}^{\frac{-V_{\rm th}+\lambda V_{\rm ds}-\gamma V_{\rm sb}}{n v_T}} \left(1 - \mathrm{e}^{\frac{-V_{\rm ds}}{v_T}}\right) \tag{1}$$

where $I_0 = \mu C_{\rm ox}(W/L){v_T}^2\mathrm{e}^{1.8}$, and $v_T = k_{\rm B}T/q$. μ is mobility, $C_{\rm ox}$ is gate capacitance, W is gate width, L is gate length, v_T is thermal voltage, n is sub-threshold swing coefficient, λ is DIBL coefficient, γ is body effect coefficient, $k_{\rm B}$ is Boltzmann's constant, T is absolute temperature, and q is electron charge. $V_{\rm ds}$ is voltage between drain and source, and $V_{\rm sb}$ is body bias.

C. Leakage to Time Conversion

Leakage current needs to be converted to time-domain parameter such as delay to realize a digital leakage current sensor. A technique to use MOSFET leakage current to pull-down or pull-up a capacitance node is used to develop a digital temperature sensor [13]. Fig. 1 shows this concept by comparing operations of two inverter structures. In Fig. 1(a), a conventional inverter structure is shown where pull-up and pull-down MOSFETs are turned ON and OFF alternately depending on the input. When the input is "L", pull-down nMOSFET turns OFF and pull-up pMOSFET turns ON, and vice versa. Thus in this case, the output node is charged and discharged by the ON currents. However, in the case of the topology in Fig. 1(b), only the pull-up pMOSFET's gate is connected to the input and the pull-down nMOSFET's gate is tied down to "L". When the input is "L", the output node will be charged by the pMOSFET ON current. But, when the input becomes "H", both the MOSFETs turn OFF and a collision between pull-down and pull-up leakage currents occur. Now, if the nMOSFET leakage current is much larger than the pMOSFET leakage current, then the output node will be discharged by the nMOSFET leakage current gradually. The time required for the discharge of the output node thus can be used for sensing leakage current. By altering the topology of Fig. 1(b), pMOSFET leakage current can be converted to delay as well. In order to monitor both of the nMOSFET and pMOSFET $I_{\rm leak}$ and their variations, large number of monitor units are required. We show in this paper that by using a reconfigurable architecture, a compact area-efficient monitoring scheme can be developed.

D. Leakage Driven Delay Model

The delay characteristic is explained next by taking an example of discharging a node by nMOSFET leakage current as shown in Fig. 1(b). Here, when the input is "L", the capacitance $C_{\rm L}$ will be driven by nMOSFET and pMOSFET leakage currents. So, we get the following differential equation.

$$C_{\rm L}\frac{\mathrm{d}V}{\mathrm{d}t} = -(I_{\rm leak,n} - I_{\rm leak,p}). \tag{2}$$

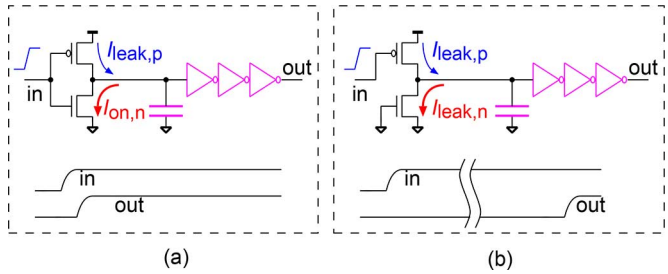


Fig. 1. Converting leakage current to delay. Pull-down nMOSFET is turned off constantly to drive the load with nMOSFET leakage current given that pull-down leakage is much larger than pull-up leakage. (a) Conventional operation. (b) Leakage driven operation.

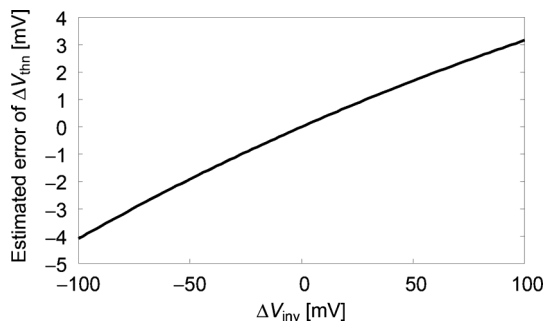


Fig. 2. Monitoring error of leakage current in terms of threshold voltage due to logic threshold variation.

Here, $I_{\rm leak,n}$ is the leakage current of the pull-down path, and $I_{\rm leak,p}$ is the leakage current of the pull-up path. These two currents have opposite effect on the load capacitance. If we design the inverter such that $I_{\rm leak,n} \gg I_{\rm leak,p}$, then we can ignore the pull-up leakage component. We can then solve (2) to obtain the fall delay $t_{\rm fall}$.

$$C_{\rm L}\int_{V_{\rm dd}}^{V_{\rm inv}} \mathrm{e}^{\frac{-\lambda_{\rm n}V}{n_{\rm n}v_T}}\mathrm{d}V = -I_0\mathrm{e}^{\frac{-V_{\rm thn}}{n_{\rm n}v_T}}\int_0^{t_{\rm fall}}\mathrm{d}t, \tag{3}$$

$$t_{\rm fall} = \frac{n_{\rm n}C_{\rm L}}{\lambda_{\rm n}\beta_{\rm n}v_T\mathrm{e}^{1.8}}\mathrm{e}^{\frac{V_{\rm thn}}{n_{\rm n}v_T}} \times \left(\mathrm{e}^{\frac{-\lambda_{\rm n}V_{\rm inv}}{n_{\rm n}v_T}} - \mathrm{e}^{\frac{-\lambda_{\rm n}V_{\rm dd}}{n_{\rm n}v_T}}\right). \tag{4}$$

Here, $V_{\rm dd}$ is supply voltage, $V_{\rm inv}$ is logic threshold of the next inverter, $n_{\rm n}$ is sub-threshold swing coefficient of nMOSFET, $\beta_n = \mu_{\rm n}C_{\rm ox}(W/L)$, $\mu_{\rm n}$ is mobility, and λ_n is DIBL coefficient of nMOSFET. $V_{\rm sb}$ to considered zero here. By taking logarithm for both sides, we obtain

$$\ln(t_{\rm fall}) = K_{\rm n} + \ln\left(\frac{n_{\rm n}}{v_T}\right) + \frac{V_{\rm thn}}{n_{\rm n}v_T} + \ln\left(\mathrm{e}^{\frac{-\lambda_{\rm n}V_{\rm inv}}{n_{\rm n}v_T}} - \mathrm{e}^{\frac{-\lambda_{\rm n}V_{\rm dd}}{n_{\rm n}v_T}}\right), \tag{5}$$

$$K_{\rm n} = \ln(C_{\rm L}) - 1.8 - \ln(\lambda_{\rm n}\beta_{\rm n}). \tag{6}$$

From the equation, fall delay has exponential relationship to $V_{\rm thn}$ and thermal voltage. Fall delay variation thus gives us information on $V_{\rm thn}$ and temperature variations. Similarly,

pMOSFET leakage current driven rise delay can be modeled as (7), and used for monitoring of $V_{\rm thp}$ and T variations.

$$\ln(t_{\rm rise}) = K_{\rm p} + \ln\left(\frac{n_{\rm p}}{v_T}\right) + \frac{|V_{\rm thp}|}{n_{\rm p}v_T} + \ln\left(\mathrm{e}^{\frac{-\lambda_{\rm p}(V_{\rm dd}-V_{\rm inv})}{n_{\rm p}v_T}} - \mathrm{e}^{\frac{-\lambda_{\rm p}V_{\rm dd}}{n_{\rm p}v_T}}\right), \quad (7)$$

$$K_{\rm p} = \ln(C_{\rm L}) - 1.8 - \ln(\lambda_{\rm p}\beta_{\rm p}). \quad (8)$$

E. Sensing Accuracy

The sensing accuracy of leakage current in time-domain can be affected by variations in the MOSFETs other than the DUTs. In (5) and (7), the delay depends on the logic threshold $V_{\rm inv}$ which is affected by process variation. When nMOSFET and pMOSFET vary in the same direction, that is both of the nMOSFET and pMOSFET becomes faster or slower, logic threshold $V_{\rm inv}$ remains constant. However, when nMOSFET and pMOSFET vary in the opposite direction, $V_{\rm inv}$ varies largely and thus affects the leakage driven delay in Fig. 1. The impact of $V_{\rm inv}$ variation is explained next using a simplified model for $t_{\rm fall}$ and $t_{\rm rise}$ as follows by ignoring the DIBL effect.

$$t_{\rm fall} = \frac{C_{\rm L}\cdot V_{\rm inv}}{I_{\rm leak,n}}, \quad (9)$$

$$t_{\rm rise} = \frac{C_{\rm L}\cdot (V_{\rm dd}-V_{\rm inv})}{I_{\rm leak,p}}. \quad (10)$$

$t_{\rm fall}$ and $t_{\rm rise}$ have exponential relationship to $V_{\rm th}$ and temperature variations, but linear relationship to $V_{\rm inv}$ variation. One key observation is that when $V_{\rm inv}$ is lowered, $t_{\rm rise}$ decreases but $t_{\rm fall}$ increases. As leakage current is a strong function of transistor $V_{\rm th}$, leakage current monitoring error can be expressed as $V_{\rm th}$ shift. The effect of $V_{\rm inv}$ on leakage current sensing can be calculated from (5) and (7), and then expressed by $\Delta V_{\rm th,e}$ which is the amount of $V_{\rm th}$ shift. Fig. 2 shows nMOSFET $\Delta V_{\rm th,e}$ against $V_{\rm inv}$ variation, assuming sub-threshold swing coefficient $n_{\rm n}$ value of 1.5. When $V_{\rm inv}$ varies by -100 mV which refers to a fast nMOSFET and slow pMOSFET condition, $\Delta V_{\rm th,e}$ is only -4 mV. When $V_{\rm inv}$ varies by 100 mV which refers to a slow nMOSFET and fast pMOSFET condition, $\Delta V_{\rm th,e}$ is only 3 mV. For accurate WID random variation monitoring, $V_{\rm inv}$ variation can be reduced by enlarging gate area of switching transistors. Thus, the effect of $V_{\rm inv}$ variation has small impact on accurate monitoring of process variation.

F. Process Variation Effect

At constant temperature, the thermal voltage component in (4) becomes constant. At room temperature, the thermal voltage v_T has a value of 26 mV. By measuring multiple instances of the leakage current driven inverter delay, we can perform statistical operation. As $V_{\rm th}$ variation is dominant in scaled CMOS process, the variance of the logarithm of fall delay, $\ln(t_{\rm fall})$, in (5) can be approximated as follows:

$$\sigma_{\ln(t_{\rm fall})}{}^2 = \frac{1}{(n_{\rm n}v_T)^2}\sigma_{V_{\rm thn}}{}^2 + \sigma_{\ln\left(\mathrm{e}^{\frac{-\lambda_{\rm n}V_{\rm inv}}{n_{\rm n}v_T}} - \mathrm{e}^{\frac{-\lambda_{\rm n}V_{\rm dd}}{n_{\rm n}v_T}}\right)}{}^2. \quad (11)$$

The second term in the right hand side of the equation corresponds to the logic threshold variation of the following inverter. We will show in Section VI that logic threshold variation can be made negligible. Thus, the variation of delay logarithm becomes as follows:

$$\sigma_{V_{\rm thn}} = n_{\rm n}\times v_T \times \sigma_{\ln(t_{\rm fall})}. \quad (12)$$

As $n_{\rm n}$ and v_T are constant at a fixed temperature, $V_{\rm th}$ variation is calculated from delay variation with (12). Similarly, pMOSFET $V_{\rm th}$ variation can be estimated using the following relationship.

$$\sigma_{V_{\rm thp}} = n_{\rm p}\times v_T \times \sigma_{\ln(t_{\rm rise})}. \quad (13)$$

Here, $t_{\rm rise}$ is the time to charge the output node of an inverter whose pull-up path is turned completely OFF. Thus the charge is driven by pMOSFET $I_{\rm leak,p}$.

G. Temperature Effect

Transistor characteristics such as mobility μ, threshold voltage $V_{\rm th}$ and sub-threshold swing coefficient n are influenced by temperature. μ and $V_{\rm th}$ dependency on temperature can be approximated by the following equations [21]:

$$\mu(T) = \mu(T_{\rm r})\left(\frac{T}{T_{\rm r}}\right)^{-k_\mu}, \quad (14)$$

$$V_{\rm th}(T) = V_{\rm th}(T_{\rm r}) - k_{\rm vth}(T-T_{\rm r}). \quad (15)$$

Here, T is the absolute temperature, $T_{\rm r}$ is room temperature. k_μ and $k_{\rm vth}$ are fitting parameters with typical values of 1.5 and 1–2 mV/K respectively. Leakage current has exponential relationship with $V_{\rm th}$ and T, and linear relationship with μ. Thus we can ignore the effect of μ variation. Leakage current driven delay $t_{\rm leak}$ can be related to temperature influence parameters as follows:

$$\ln(t_{\rm leak}) \propto -\frac{k_{\rm vth}(T-T_{\rm r})}{nv_T}. \quad (16)$$

IV. Topology-Reconfigurable Leakage Monitor Cell

A. Structure

Fig. 3 shows the reconfigurable leakage monitor cell. The inverter in Fig. 3 has two configuration bits to turn ON or OFF the header/footer transistors. Several nMOS transistors are stacked in the pull-down path. Similarly, several pMOS transistors are stacked in the pull-up path. The use of stack transistors has several effects on the leakage current reduction, which can be exploited to increase sensing accuracy. The header/footer transistors are the transistors under test whose leakage current we want to monitor. In order to characterize transistors as used in the digital circuits, the header/footer transistors have the same gate length and width as those used in the standard cells. Table I summarizes the different configuration modes of the monitor cell.

B. Operation

We explain the operation of our leakage monitor cell taking nMOSFET leakage monitoring as an example. The "C0" nMOSFET will be turned OFF and "C1" pMOSFET will be turned ON in this case. When the input is "L", the pull-up

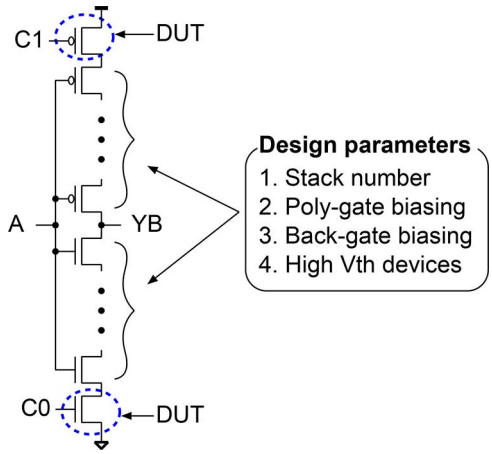


Fig. 3. Reconfigurable leakage monitor cell.

TABLE I
CONFIGURATION OF LEAKAGE MONITOR CELL

Configuration ["C1","C0"]	Function
01	Standard inverter delay
00	nMOSFET leakage monitor
11	pMOSFET leakage monitor
10	Not used

path turns ON and charges the output node. When the input is "H", both of the pull-up and pull-down paths turn OFF. However, all the stack pMOSFETs in the pull-up path turns OFF, whereas all except the DUT nMOSFET turns ON in the pull-down path. Leakage current reduces exponentially with the increase of stack number [22]. By tuning stack number, poly-gate and back-gate biasing, and using high $V_{\rm th}$ transistors, the leakage current ratio can be tuned so that the pull-down current becomes much larger than the pull-up leakage current. The output node is discharged gradually by the nMOSFET leakage current. As nMOSFET DUT is OFF and other stack nMOSFETs are ON, the DUT leakage current is the dominant component of the overall pull-down leakage current. The ratio between pull-down leakage and pull-up leakage can be tuned by the following design parameters: 1) transistor stacking, 2) poly-gate biasing, 3) back-gate biasing, and 4) use of high $V_{\rm th}$ transistors.

C. Design Parameters

1) Transistor Stacking: Transistor stacking affects both the leakage current and its variation. The intermediate nodes in a stacked transistor series have intermediate values between $V_{\rm dd}$ and GND. Fig. 4 shows a single nMOSFET path and a two-stacked nMOSFET path. The gate and the body of the transistors are tied down to GND. In the case of a two-stacked path, the intermediate node "X" rises and results in an increase in the source voltage of the "M1" transistor of Fig. 4(b). The increase of source voltage affects leakage current in the following three ways [22]. First, the gate-source voltage $V_{\rm gs}$ becomes negative. Second, the source-body voltage $V_{\rm sb}$ becomes positive resulting in an increase of channel $V_{\rm th}$. Third, drain-source voltage $V_{\rm ds}$ voltage reduces drastically. These three phenomena affect the $I_{\rm leak}$ exponentially. Leakage current reduction of two-stacked

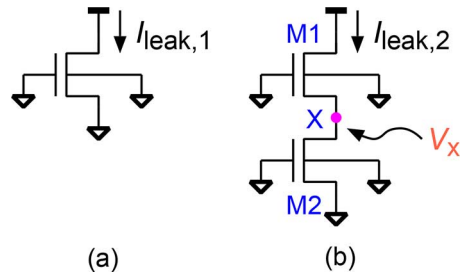


Fig. 4. Effect of transistor stacking on leakage current. (a) Single MOSFET. (b) Two stacked MOSFETs.

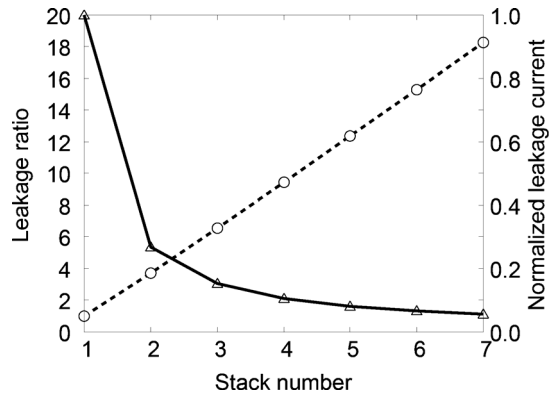


Fig. 5. Leakage current reduction effect versus stack number.

transistor compared with a single transistor can be expressed by the following equation.

$$\frac{I_{\rm leak,2}}{I_{\rm leak,1}} = {\rm e}^{\frac{-V_{\rm X}-\gamma V_{\rm X}-\lambda V_{\rm X}}{nV_T}} \cdot \frac{\left(1-{\rm e}^{\frac{-V_{\rm dd}+V_{\rm X}}{v_T}}\right)}{\left(1-{\rm e}^{\frac{-V_{\rm dd}}{v_T}}\right)}. \qquad (17)$$

Here, $V_{\rm X}$ is the source voltage, $I_{\rm leak,1}$ is the leakage current of a single transistor, and $I_{\rm leak,2}$ is the leakage current of a two-stacked transistor.

Fig. 5 shows the leakage current reduction rate against the transistor stack number. Solid line shows the leakage current and dotted line shows the increase of leakage current ratio. Leakage current ratio increases linearly against the stack number. Although, transistor stacking affects the desired leakage ratio linearly only, stacking is effective in reducing $I_{\rm leak}$ variation. As $(1-{\rm e}^{(-V_{\rm ds})/(v_T)})$ term in (1) becomes dominant, the effect of $V_{\rm th}$ variation is reduced as will be confirmed by Monte Carlo simulation in Section VI.

2) Poly-Gate Biasing: For short-channel transistors, increasing the gate length increases MOSFET $V_{\rm th}$. Thus, gate length of the stacked transistors can be increased to achieve larger leakage current ratio between the pull-up and pull-down paths. Many processes allow poly-gate biasing with the same transistor layout which can reduce the design complexity significantly. One drawback of tuning the gate length is that the effect becomes more unpredictable as accurate models for longer gate lengths are not always provided. Another drawback can be that the poly density will differ by increasing the gate length which may cause different variation profile.

3) Back-Gate Biasing: Back-gate can be tuned to control the $V_{\rm th}$ of the transistors. So, for example, when monitoring nMOSFET leakage variation, pMOSFET can be reverse biased to obtain enhanced leakage ratio. Implementing on-chip bias generators can be costly in terms of area. However, this option can be used effectively during the testing of the chip given the back-gate can be biased externally.

4) High $V_{\rm th}$ Transistors: Another good option is to use high $V_{\rm th}$ transistors for the stacked transistors. As leakage current decreases exponentially with the increase of $V_{\rm th}$, using high $V_{\rm th}$ transistors along with stacking can reduce the area significantly.

5) Parameter Tuning Methodology: Leakage sensing accuracy depends on the leakage ratio between the pull-up and pull-down paths in a stage. Ignoring the DIBL effect, (2) is solved as follows:

$$t_{\rm fall} = \frac{C_L V_{\rm inv}}{I_{\rm leak,n} - I_{\rm leak,p}} = \frac{C_L V_{\rm inv}}{I_{\rm leak,n}(1 - 1/r_{\rm leak})}. \quad (18)$$

Here, $r_{\rm leak} = I_{\rm leak,n}/I_{\rm leak,p}$ is the ratio between pull-down and pull-up leakage. The effect of $r_{\rm leak}$ on leakage monitoring error can be expressed by $V_{\rm thn}$ shift as follows:

$$r_{\rm leak} = \frac{1}{{\rm e}^{\frac{\Delta V_{\rm thn,e}}{n_{\rm n} v_T}} - 1}. \quad (19)$$

Here, $V_{\rm thn,e}$ is the estimation error of nMOSFET $V_{\rm th}$ variation. Fig. 6 shows the relationship between leakage current sensing accuracy and leakage current ratio, $r_{\rm leak}$. Leakage current sensing accuracy is presented by error in $\Delta V_{\rm th}$ here. In order to achieve sensing inaccuracy as low as 1 mV, $r_{\rm leak}$ of 38.5 is required. However, if the sensing inaccuracy is relaxed to 10 mV, the required ratio is only 3.4.

After the desired sensing accuracy is set, the design parameters are tuned to achieve the required ratio. Table II summarizes the effects of the design parameters. Our proposal is to first select a suitable stack number to increase leakage ratio by multiple times and reduce local variation effect. Then, select high $V_{\rm th}$ transistors as the stack transistors other than the DUTs. This will provide leakage ratio improvement in the order of several tens of times. Together with the transistor stacking, leakage ratio over 100 can be achieved which should be sufficient for most cases. However, in order to ensure robust operation for extreme conditions such as slow nMOSFET and fast pMOSFET, or vice versa, transistor gate length can be increased to enhance the ratio further.

V. All-Digital Leakage Variation Sensor

In the previous section, we have shown that inverter delay can be used as a parameter to monitor leakage current. We need an area-efficient architecture to monitor large number of samples for both of the nMOSFET and pMOSFET. In this section, based on a topology-reconfigurable inverter, we propose an area-efficient ring oscillator (RO) based sensor circuit.

A. Proposed Sensor Architecture

Previously, we have proposed a topology-reconfigurable monitor circuit structure to monitor process variation from a single monitor instance [23]. In order to measure $I_{\rm leak}$ variations of nMOSFET and pMOSFET independently, we use a similar topology-reconfigurable architecture. Fig. 7 shows a block diagram of our proposed all-digital leakage variation sensor circuit. The core part of the sensor is an RO which consists of a delay chain from our developed reconfigurable leakage monitor cells. The leakage monitor cell can be configured to three different operation modes as shown in Table I. Each stage in the delay chain can be configured independently. The leakage monitor cell can be configured to drive its output by $I_{\rm leak}$ of either nMOSFET or pMOSFET, or by the ON currents of nMOSFET and pMOSFET. The configuration signals are set by the controller. The controller consists of a state machine and a signal decoder. A frequency counter measures the RO frequency and feeds to a statistical processing unit. The controller then reconfigures the RO to a new configuration and corresponding frequency is measured and fed to the statistical processing unit. By reconfiguring and re-measuring the frequency, the statistical properties are obtained which gives us the $I_{\rm leak}$ variation of nMOSFET and pMOSFET. The whole system works digitally and no additional supply voltage is required.

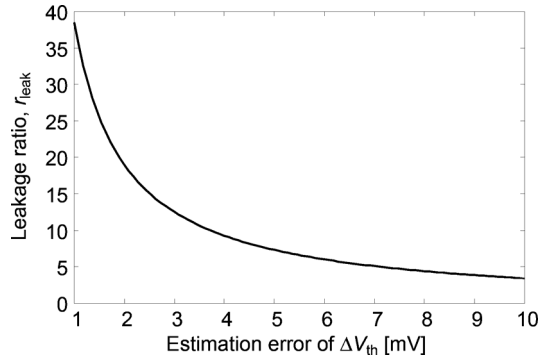


Fig. 6. Required leakage current ratio $r_{\rm leak}$ against estimation error of $\Delta V_{\rm th}$.

TABLE II
Summary of Various Design Parameters to Tune the Leakage Ratio

Parameter	Effect	Design cost	Area overhead	Variability reduction
Gate width	Linear	Low	High	Small
Gate length	Exponential	Low	Low	Small
Stacking	Linear	Medium	High	Large
Back-gate biasing	Exponential	High	Low	Small
High $V_{\rm th}$	Exponential	Low	None	Small

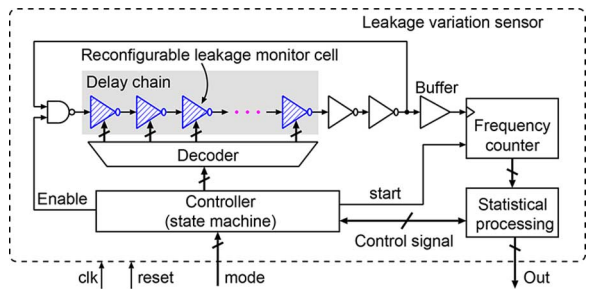


Fig. 7. Concept of our proposed leakage variation sensor architecture.

B. Leakage Variation Monitoring

The proposed sensor can monitor $I_{\rm leak}$ variation as well as average $I_{\rm leak}$ of nMOSFET and pMOSFET. $I_{\rm leak}$ variation is

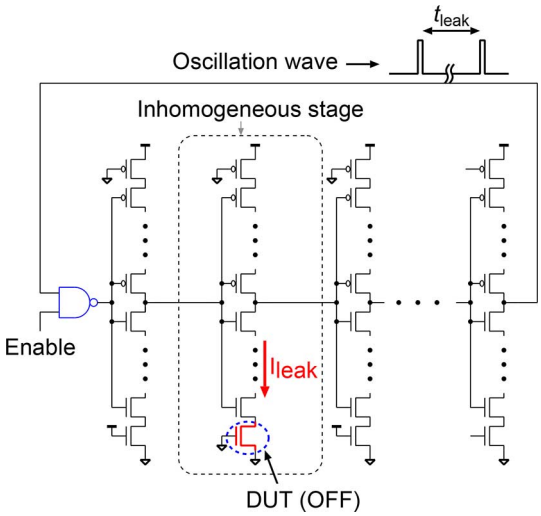


Fig. 8. Inhomogeneous structure to monitor single nMOSFET leakage current.

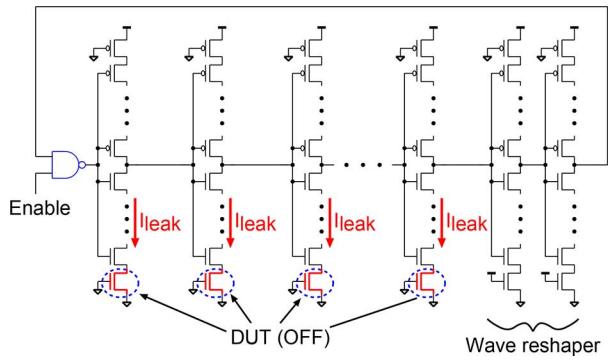


Fig. 9. Homogeneous structure to monitor average nMOSFET leakage current.

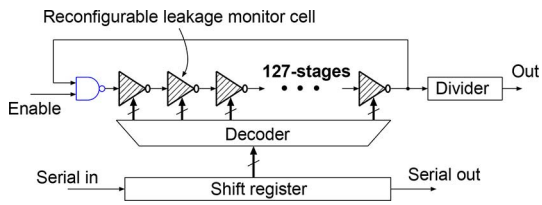


Fig. 10. Schematic of the reconfigurable leakage variation sensor circuit.

measured by configuring the RO as inhomogeneous and then measuring the oscillation periods by swapping the inhomogeneous stage across the inverter chain [24]. Fig. 8 shows an inhomogeneous configuration to monitor $I_{\rm leak}$ of an nMOSFET in an inverter stage. The footer DUT nMOSFET in the inhomogeneous stage is turned OFF and footer nMOSFETs of other stages are turned ON. Thus, the inhomogeneous stage output node will be driven by the footer nMOSFET $I_{\rm leak}$. The corresponding oscillation wave is also shown in the figure. Small width pulses are observed with long intervals. The interval time corresponds to the time $t_{\rm leak}$ needed to drive the inhomogeneous stage output node by $I_{\rm leak}$. The pulse width time corresponds to one cycle delay that is driven by the ON currents of all the stages. $t_{\rm leak}$ is several magnitudes larger than the time interval. Similarly, pMOSFET leakage variation is measured by configuring the inverter as a pMOSFET leakage sensor.

Average $I_{\rm leak}$ for nMOSFET and pMOSFET is measured by configuring the RO as homogeneous. For nMOSFET average $I_{\rm leak}$ monitoring, the inverter stages are configured as nMOSFET $I_{\rm leak}$ monitor cell. Fig. 9 shows the homogeneous configuration for nMOSFET average $I_{\rm leak}$ monitoring. The footer nMOSFETs of all the stages are turned OFF. Fall delay of each stage is driven by nMOSFET $I_{\rm leak}$, thus the oscillation period gives us a sum of each stage fall delay. Similarly, by turning all the header pMOSFETs OFF and turning all the footer nMOSFETs ON, we can monitor the average pMOSFET $I_{\rm leak}$.

C. Oscillation Period Model

The oscillation period of an inhomogeneous configuration can be expressed by the sum of the propagation delays of all the inverter stages.

$$T_{\rm osc} = T_0 + K{\rm e}^{\frac{V_{\rm th}}{nv_T}}. \tag{20}$$

Here, T_0 is the delay component resulting from rise/fall delays driven by ON currents of the MOSFETs other than the DUT. K is a constant here. T_0 is a constant too from the fact that larger N averages out random variation effect and that contribution of ON current delay is negligible compared with the leakage current delay. Thus, $T_{\rm osc}$ can be treated as a shifted log-normal distribution with an offset T_0. $V_{\rm th}$ variation is then expressed with the following equation.

$$\sigma_{V_{\rm th}} = n \times v_T \times \sigma_{\ln(T_{\rm osc}-T_0)}. \tag{21}$$

VI. Demonstration Circuit in a 65 nm Process

A. Chip Design

A proof-of-concept sensor circuit is fabricated in a 65 nm process to demonstrate our concept. We implemented a simplified version of the leakage variation sensor shown in Fig. 7. Fig. 10 shows the block diagram of the demonstration circuit. The circuit consists of a ring oscillator (RO), a decoder, a shift register and a divider. A 127-stage RO is designed where the first stage is an NAND gate to control the oscillation, and the last two stages are used as buffers. Frequency counting and statistical processing is performed outside the chip. Inverter structure with four stacked transistors and the DUT header/footer transistors are designed. Thus, the total number of stack transistors in the pull-up and pull-down path is 5. Minimum gate length is used for all the transistors. pMOSFET and nMOSFET gate widths are the same as those in the standard cell library. The pMOSFET gate width is 1.5 times larger than the nMOSFET gate width. Complete cell-based design flow is adopted for design automation. The monitor cells are placed adjacent to each other using placement constraint. Regular placement ensures that the load capacitance variation between the stages does not affect the leakage variation monitoring. A serial interface is designed to control the configuration of each inverter stage. Fig. 11 shows the chip micro-graph and the layout of the sensor circuit. Total area of the circuit including the shift register and the divider is 4500 μm^2.

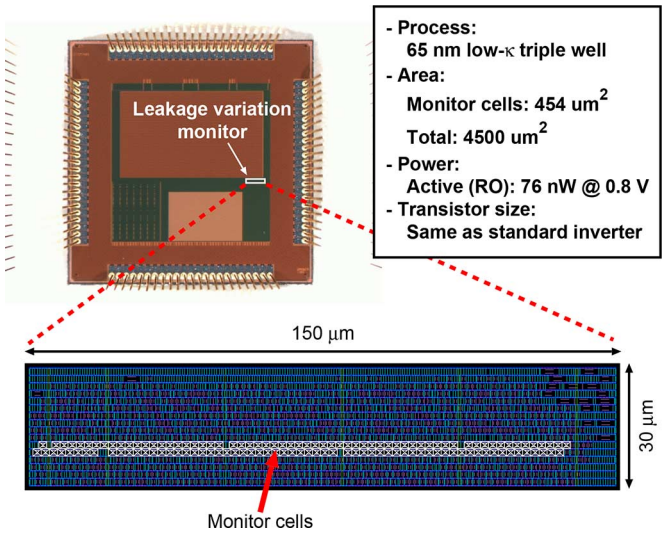


Fig. 11. Chip micro-graph and layout of the leakage variation sensor.

B. Simulation Results

1) Delay Characteristics: First we show the operation of leakage driven charging of a capacitance and the corresponding rise time. A delay chain shown in Fig. 12 is used for calculating the propagation delay of each inverter stage. Supply voltage is set to 0.8 V. Fig. 13 shows the output waves of the first three stages over time after a rising edge is applied to the input of the chain. The first stage fall time is driven by nMOSFET $I_{\rm leak}$. The output voltage of the first inverter falls slowly over time as this node is driven completely by nMOSFET $I_{\rm leak}$. The output voltage of the second inverter starts rising when the first inverter output approaches the logic threshold. Finally, when the second stage output reaches the logic threshold, the third stage output starts falling. Below that point, the output voltages of the following inverters change sharply compare to the rising delay of the first inverter. Thus, the overall delay $t_{\rm total}$ of an N-stage inverter chain will be dominated by $I_{\rm leak}$ driven delay, and thus can be approximated as follows:

$$t_{\rm total} = t_{\rm leak} + \sum_{i=2}^{N} t_i \approx t_{\rm leak}. \qquad (22)$$

Here, $t_{\rm leak}$ is the propagation delay of the first stage, and t_i is the propagation delay of the ith stage. The validity of this approximation is shown in Fig. 14. Fig. 14 shows the propagation delay of each inverter in log-scale. The leakage driven delay is 500 times larger than the following stage delay. With the increase in the stage index, the propagation delay tends to become smaller. This is the effect of large slew caused by the slow rising edge of the first stage output. So, by having the number of stage N in the order of 100, approximation of (22) is valid.

2) Transistor Stacking Effect: Transistor stacking reduces $I_{\rm leak}$ and its variation. Fig. 15 shows a Q-Q plot of DC leakage current distribution flowing through a series of stacked nMOS transistors. X-axis is shown in log-scale. The circuit shown in Fig. 4(b) is used in the simulation. The current distribution is obtained by performing Monte Carlo simulation under $\sigma_{V_{\rm th}} = 25$ mV. An nMOSFET which is not stacked, a 2-stacked nMOSFET, a 5-stacked nMOSFET and a 9-stacked nMOSFET is used for comparison. With the increase of stack number, the mean leakage reduces drastically. At the same time, the variation also reduces. In the case of pMOSFET $I_{\rm leak}$ monitoring with the Fig. 3 structure, the nMOSFETs in the pull-down path act as a stacked series. Whereas the pull-up path header pMOSFET acts as a single transistor because other pMOSFETs

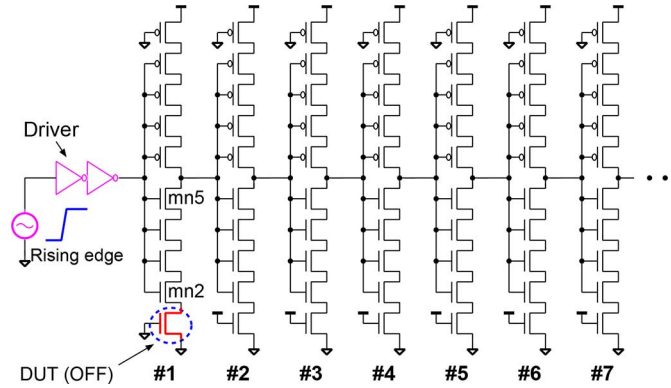


Fig. 12. Delay chain consisting of a topology-reconfigurable leakage sensor inverter structure. The first stage is configured so that fall delay is driven by nMOSFET $I_{\rm leak}$.

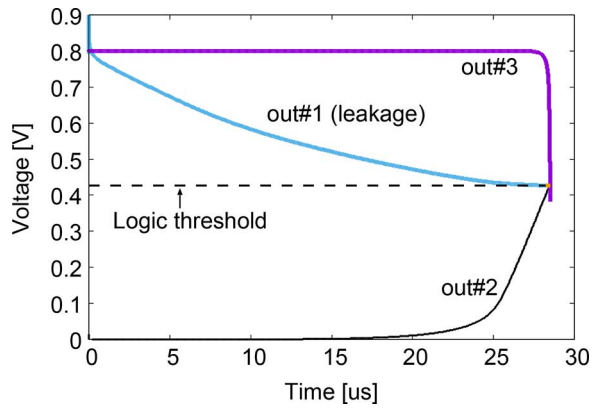


Fig. 13. Wave of inverter outputs of a delay chain when an rise edge is applied to the input. The fall delay of the first stage is driven by nMOSFET $I_{\rm leak}$.

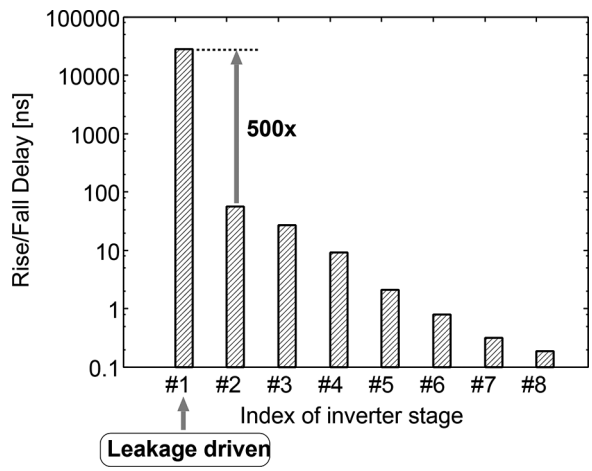


Fig. 14. Propagation delay of each stage in a delay chain. The first stage fall time is driven by nMOSFET $I_{\rm leak}$.

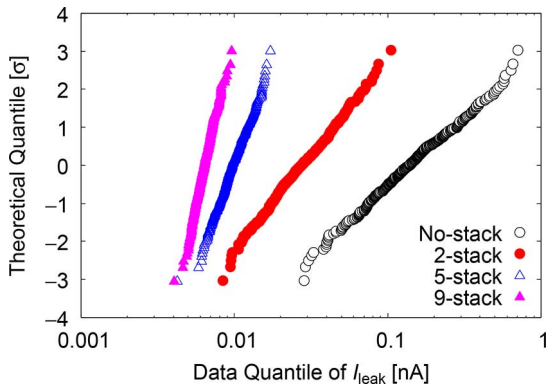


Fig. 15. Effect of transistor stacking on leakage current and its variation. X-axis values are shown in log-scale.

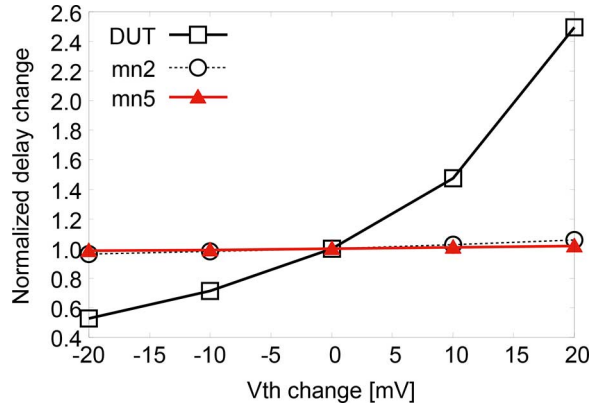


Fig. 16. Delay change according to $V_{\rm th}$ changes in the DUT and the stack transistors.

TABLE III
CONTRIBUTION OF LOGIC THRESHOLD VARIATION TO LEAKAGE DRIVEN LOGARITHMIC DELAY VARIATION

Stack number	$\frac{\sigma_{V_{\rm thp}}^2}{(nv_T)^2}$	$\sigma^2_{\ln\left(e^{\frac{-\lambda V_{\rm inv}}{nv_T}}-e^{\frac{-\lambda V_{\rm dd}}{nv_T}}\right)}$	% Contribution
2	0.231	0.0035	0.98%
5	0.231	0.0010	0.28%
9	0.231	0.00054	0.15%

are turned ON. Thus, transistor stacking can be used effectively to increase the sensing accuracy of our propose monitor cell.

3) Variation Effect on Sensing Accuracy: There are two possible factors that can affect the sensing accuracy. One is the logic threshold variation and the other is the drain-source voltage variation of the DUT.

Theoretical aspects on the impact of logic threshold variation is discussed in Section III. The effect of logic threshold variation is verified by Monte Carlo simulation here. Table III shows the values of the two terms in the right hand side of (11). The table compares the contribution of logic threshold variation for monitor cell structures of different stack numbers. $\sigma_{V_{\rm thp}} = 20$ mV and $\sigma_{V_{\rm thn}} = 25$ mV is assumed in the simulation. The minimum stack number for our leakage monitor cell is 2. Even with the minimum stack number, logic threshold variation contributes to only 0.98%. With total stack number of 5, which is the same as in our implementation, the amount of contribution reduces to 0.15%. Thus, we conclude that logic threshold variation has no impact on leakage driven delay variation measurement.

In (4), the drain-source voltage $V_{\rm ds}$ of the DUT is assumed to be equal to $V_{\rm dd}$. However, because of the stacked transistor topology, drain-source voltage of the DUTs are also affected by process variation. As the stack transistors are turned ON, they have low sensitivity to delay change compare to DUT $V_{\rm th}$ variation sensitivity. Sensitivity analyses based on simulation are therefore performed. Fig. 16 shows the delay change according to $V_{\rm th}$ change of MOSFETs in the stack. $V_{\rm th}$ variation sensitivities of three transistors are shown in the figure. One is the DUT whose leakage current variation we want to monitor. The second is the transistor next to the DUT in the series. This MOSFET is denoted by "mn2" in Fig. 12. The third is the transistor connected to the output load of the inverter, which is denoted by "mn5" in Fig. 12. Fig. 16 confirms that the delay is highly sensitive to DUT $V_{\rm th}$ change, but not sensitive to $V_{\rm th}$ changes of the other stack transistors. As variation is inversely proportional to gate area, the variation in the stack transistors can be reduced further by enlarging the gate area. Thus, accurate process variation measurement can be achieved with our proposed circuit topology.

C. Power Consumption

As the oscillation period is dominated by $I_{\rm leak}$ driven delay, the oscillation frequency is in the order of kHz. Thus, the dynamic power of the RO is low and insignificant. However, when the RO is configured as inhomogenous to monitor $I_{\rm leak}$ of particular transistor, the output node voltage of the inhomogenous stage remains in an intermediate value for a long period of time. As a result, the pull-up and pull-down paths of the next inverter stage become partially ON, and through current flows from power to ground. Even with through current flowing in one stage, the overall power consumption for the RO is simulated to be 76 nW at 0.8 V supply which we consider to be small enough to be used in any application.

VII. MEASUREMENT RESULTS

A. RO Oscillation

Fig. 17 shows the image capture of an oscilloscope screen for the oscillation wave. Supply voltage is set to 0.8 V in this measurement. The wave here corresponds to a pMOSFET leakage dominant oscillation. The pulse width is extremely small compare to the pulse interval. The pulse width represents the delay component driven by ON currents, and the pulse interval represents the delay component driven by a single MOSFET leakage. Thus from silicon measurement, it is confirmed that delay component from the ON currents can be neglected and the oscillation period can be approximated with the delay driven by a single MOSFET $I_{\rm leak}$.

The capability of independent monitoring of nMOSFET and pMOSFET leakage current is validated by applying body bias. Body biasing tunes the channel threshold voltage. When the sensor is monitoring nMOSFET leakage, the output should response only when nMOSFET $V_{\rm th}$ is tuned, and vice versa. In

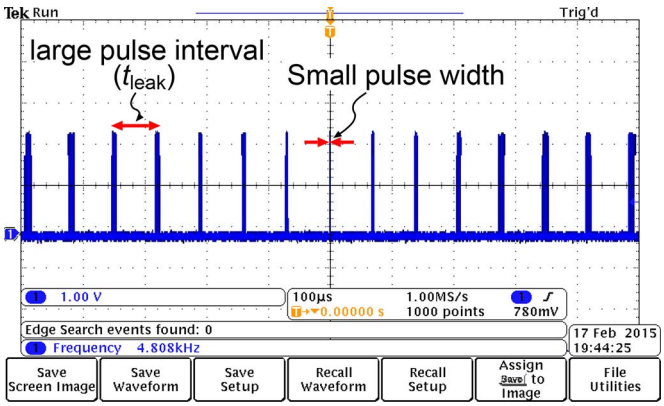


Fig. 17. Capture of oscilloscope waveform of pMOSFET driven inhomogeneous RO oscillation.

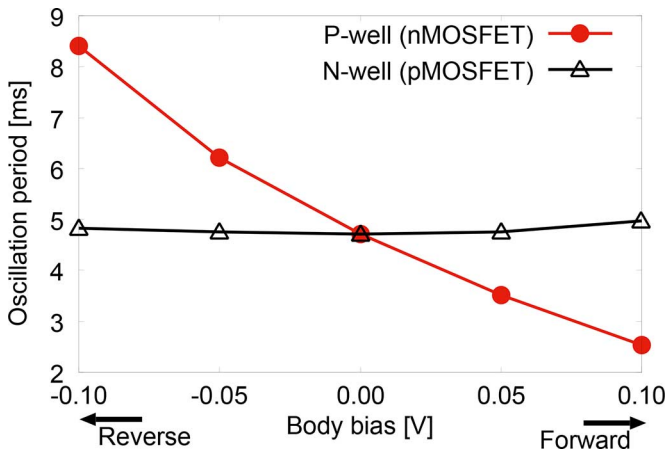


Fig. 18. Change of oscillation period against well bias for nMOSFET leakage sensor at 0.8 V operation.

order to validate the sensor capability, oscillation period is measured by varying independent body bias to either nMOSFET or pMOSFET. Homogeneous configuration is used for monitoring average leakage current monitoring in the measurement. Fig. 18 shows the oscillation period versus N-well and P-well bias values for nMOSFET leakage monitoring. Oscillation period changes exponentially when nMOSFET body is biased, whereas the period remains almost constant when pMOSFET is biased. Similarly, Fig. 19 shows the period change against body bias for pMOSFET leakage monitoring. The sensor is sensitive to pMOSFET threshold voltage variation, but insensitive to nMOSFET threshold voltage variation. The results thus confirm that the proposed sensor monitors leakage variation correctly.

B. Leakage Current Variation

Leakage variation is measured for both of the nMOSFET and pMOSFET by varying supply voltages to demonstrate the wide-supply-range operation. Fig. 20 shows the measured distribution of oscillation period for nMOSFET $I_{\rm leak}$. Fig. 21 shows the measured distribution of oscillation period for pMOSFET $I_{\rm leak}$. Supply voltage is 0.8 V in both the cases. Both the distributions have long tails suggesting they are not normal distributions.

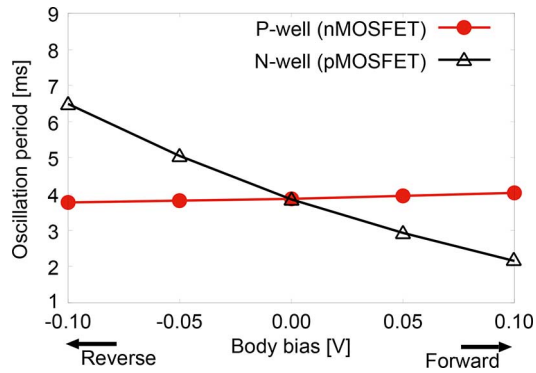


Fig. 19. Change of oscillation period against well bias for pMOSFET leakage sensor at 0.8 V operation.

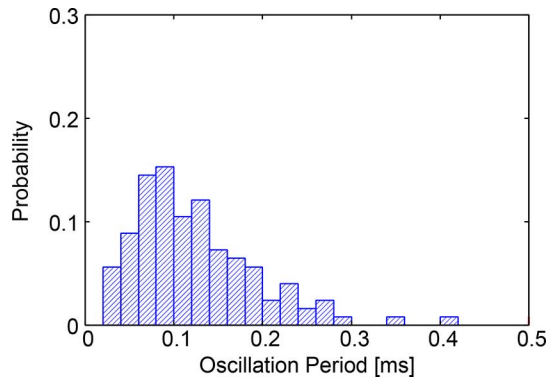


Fig. 20. Histogram of delay distribution for WID nMOSFET leakage variation.

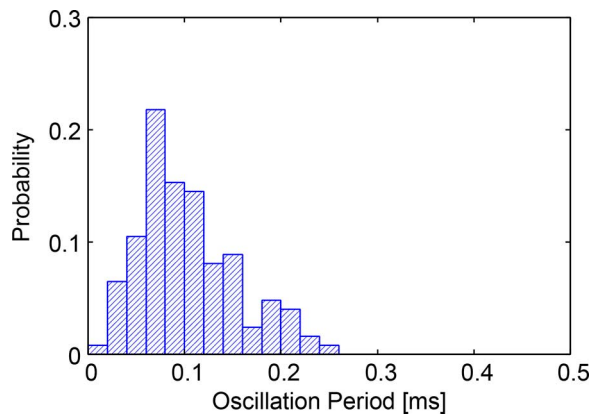


Fig. 21. Histogram of delay distribution for WID pMOSFET leakage variation.

Next we confirm the effect of logic threshold $V_{\rm inv}$ variation on leakage current variation measurement. As explained in Sections IV and VI, increase in $V_{\rm inv}$ results in an increase in the nMOSFET $I_{\rm leak}$ driven fall time and a decrease in pMOSFET $I_{\rm leak}$ driven rise delay. Thus, if pMOSFET $I_{\rm leak}$ delay is plotted nMOSFET against $I_{\rm leak}$ delay, strong negative correlation will be observed if the logic threshold variation effect and leakage driven delay variation are comparable. Fig. 24 shows the correlation between pMOSFET and nMOSFET leakage currents. Each point in the figure corresponds to the leakage currents of nMOSFET and pMOSFET of the same inverter stage. Logarithm is taken for the leakage currents. No significant correlation is observed meaning we do not have any effect

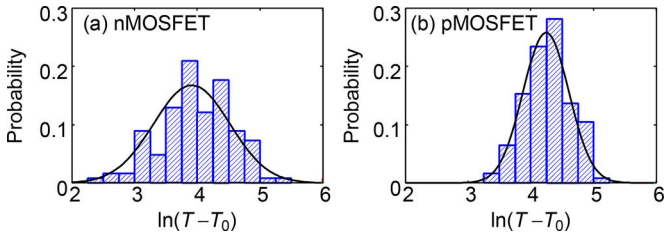


Fig. 22. Log-normal fitting of WID leakage-current driven delay variation. (a) nMOSFET, (b) pMOSFET.

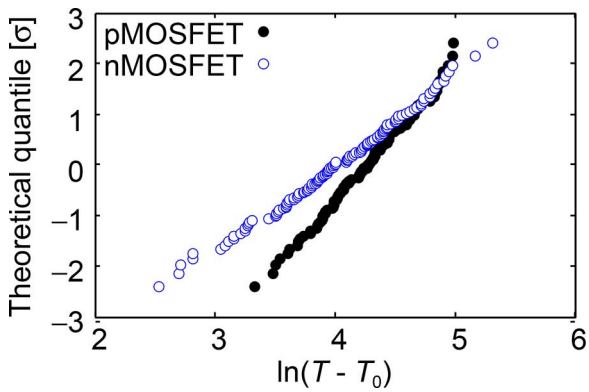


Fig. 23. Q-Q plot of WID leakage variation for nMOSFET and pMOSFET. The distribution fits log-normal distribution.

coming from the load and the logic threshold variation. Also, the stack effect is working well as we do not see any correlation between the two opposite configurations for $I_{\rm leak}$ monitoring.

Leakage current has exponential relationship to $V_{\rm th}$ variation. As $V_{\rm th}$ variation is reported to follow normal distribution, monitored delay distribution is expected to follow log-normal distribution. Fig. 22 shows the delay distributions at 0.8 V supply for nMOSFET and pMOSFET after fitting the measured data to a log-normal distribution. The delay distributions follow log-normal distribution. This validates that the proposed monitoring technique is able to detect leakage variation. Q-Q plots of the measured delay variations are shown in Fig. 23. nMOSFET has larger distribution than pMOSFET. Fig. 25 shows the measured delay variation for different supply voltages. As (12) suggests, leakage variation does not depend on supply voltage. Measured variation is constant across wide range of supply voltages which validates our concept.

When the monitor circuit is configured as a homogeneous structure, the total delay due to the N DUT MOSFETs' leakage current can be obtained. Fig. 26 shows the oscillation periods for nMOSFET and pMOSFET measured at different supply voltage operation. Here, smaller the delay is, larger the leakage current is. In this particular chip, pMOSFET leakage is monitored larger than nMOSFET leakage. Next, average $I_{\rm leak}$ for nMOSFET and pMOSFET of several chips are measured. Fig. 27 plots pMOSFET $I_{\rm leak}$ against nMOSFET $I_{\rm leak}$ for nine chips. High correlation of 0.83 is observed which suggests that the D2D variation mechanism is the same for nMOSFET and pMOSFET.

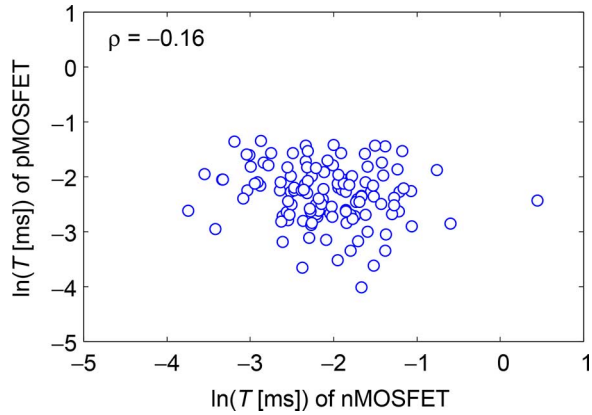


Fig. 24. Correlation between nMOSFET and pMOSFET leakage of the same inverter. Inhomogeneous configuration is used to capture WID leakage-current driven delay variation.

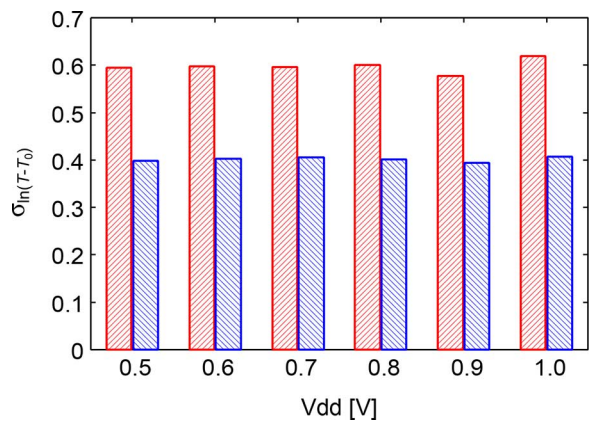


Fig. 25. Measured leakage variation at several supply voltage.

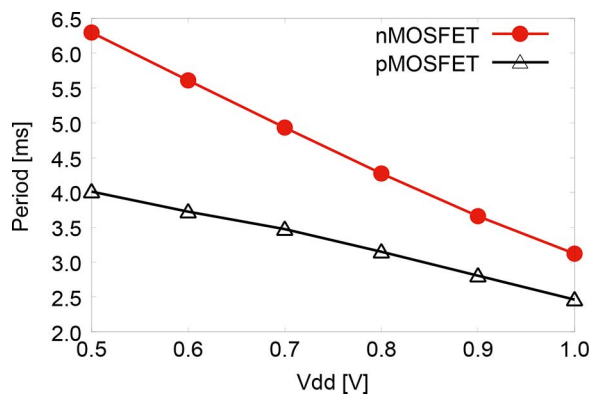


Fig. 26. Measured oscillation period due to leakage current for 124 nMOSFETs and 124 pMOSFETs at different supply voltages.

C. Process Variation Monitoring

An one-to-one mapping between $V_{\rm th}$ and monitored delay is performed using (5) and (7) to monitor $V_{\rm th}$ variation of nMOSFET and pMOSFET. Thus, the distribution shape of $V_{\rm th}$ variation is the same as the leakage current distribution as shown in Fig. 23 only that the X-axis is transformed to $V_{\rm th}$ variation. $V_{\rm th}$ variation of nMOSFET and pMOSFET is then estimated using (12) and (13). Because of non-disclosure agreement, we are unable to show the absolute values of the

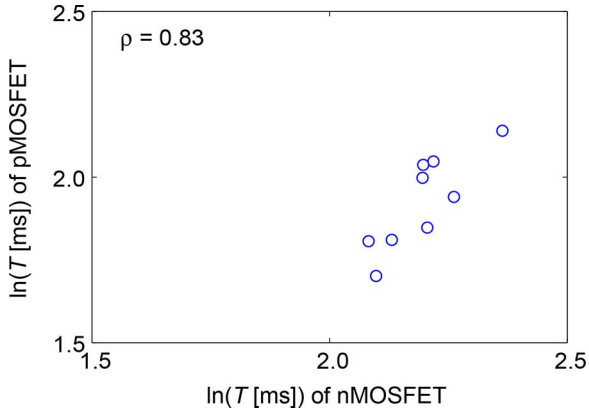


Fig. 27. Correlation between inter-die nMOSFET and pMOSFET leakage. Homogeneous configuration is used to capture sum of 124 leakage-current driven delay.

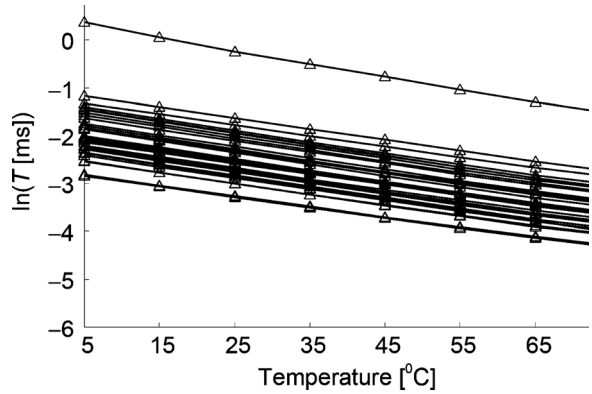


Fig. 28. nMOSFET leakage current driven delay versus temperature at 0.8 V operation. Y-axis is the logarithm of oscillation period.

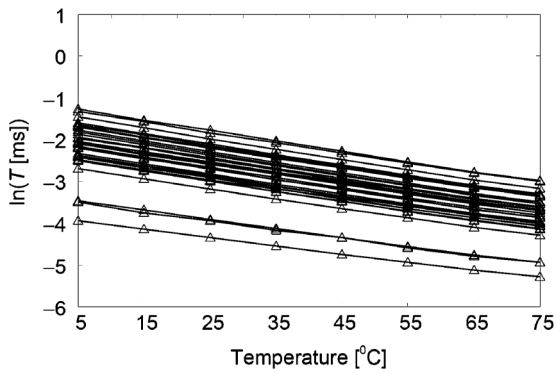


Fig. 29. pMOSFET leakage current driven delay versus temperature at 0.8 V operation. Y-axis is the logarithm of oscillation period.

estimated $V_{\rm th}$ variations. nMOSFET $V_{\rm th}$ variation is estimated 1.6 times larger than pMOSFET $V_{\rm th}$ variation which agrees with literature results considering the MOSFET sizes used in our circuit [25], [26]. Thus, the proposed sensor can be used for monitoring $V_{\rm th}$ variation.

D. Temperature Monitoring

As leakage current is sensitive to temperature variation, our proposed leakage sensor can be used for on-chip temperature monitoring. Oscillation period for each of the transistors is measured by varying the temperature of the chip. Fig. 28 shows the change of measured oscillation period for 30 nMOSFET transistors within a chip against the temperature change of the sensor circuit. The 30 nMOSFETs are taken from 30 inverters in a series from the same monitor circuit here. The logarithm of oscillation period is taken and plotted against temperature. The logarithm of delay changes linearly with temperature which agrees with the model (16). We can observe large variation in the offset values of delay between the 30 nMOSFET transistors. This variation corresponds to $V_{\rm th}$ variation between the transistors. However, we also observe some variation in the gradients of these curves. The logarithm of delay thus can be expressed by the following equation:

$$\ln(T_{osc}) = a_T + b_T \cdot T \tag{23}$$

where T is the absolute temperature, a_T and b_T are temperature coefficients. The variation in the gradient implies that the effect of temperature on MOSFET $V_{\rm th}$ is not uniform. Fig. 29 shows the measured oscillation period change against temperature for 30 pMOSFET transistors within the same chip. Similar trends in the offset and gradient are observed for pMOSFETs as well.

As both of the coefficients in (23) vary from transistor to transistor, single MOSFET $I_{\rm leak}$ based temperature sensor will require a two-point calibration. Figs. 30 and 31 show the temperature monitoring error when a single transistor is chosen to monitor temperature after a two-point calibration at 15°C and 65°C. Fig. 30 shows the temperature monitoring result using nMOSFET leakage current. Whereas, Fig. 31 shows the temperature monitoring result using pMOSFET leakage current. An error range of -2.1°C to 1.2°C is observed when nMOSFET leakage current is used. Although a two-point calibration is required, the proposed circuit gives the flexibility in design and measurement. In addition, the proposed monitor provides measurements of multiple transistors, which can be exploited for accuracy improvement. For example, the average leakage between multiple transistors can easily be measured by configuring the circuit as homogeneous. Figs. 32 and 33 show the logarithm of measured delay for three different dies when the monitor circuit is configured as homogeneous. Fig. 32 thus represents the delay for nMOSFET average $I_{\rm leak}$ and Fig. 33 represents the delay for pMOSFET average $I_{\rm leak}$. The temperature dependency of three dies have similar slope, thus an one-point calibration can be performed. Fig. 34 shows the monitoring error when nMOSFET average $I_{\rm leak}$ is used. Calibration is performed at 15°C. An error range of -1.3°C to 1.4°C is observed. Similarly, monitoring error using pMOSFET average $I_{\rm leak}$ is shown in Fig. 35. An error range of -1.1°C to 1.6°C is observed. The error curves in both the figures show systematic nature which are caused by non-linearity of the sensor output. In order to increase the accuracy, the effect of non-linearity therefore needs to be cancelled out. Non-linearity can occur from a delay offset that is caused by the inverter stages other than the one with the DUT. Another possible reason for the non-linearity is the degradation of leakage ratio, $r_{\rm leak}$, because of process variation which is explained in Section IV. The offset and non-linearity issues can be

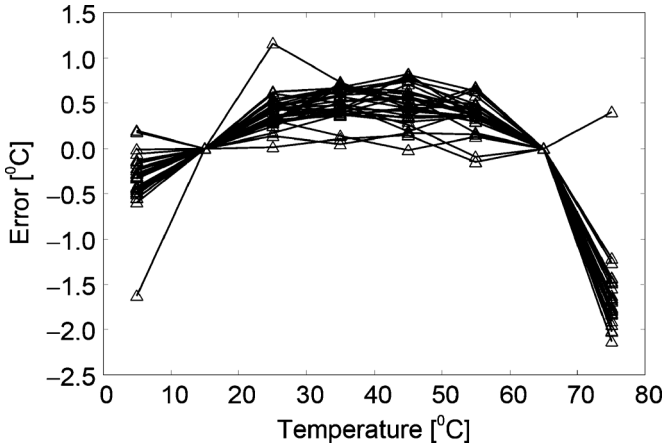


Fig. 30. Sensing error versus temperature after two-point calibration for 30 individual nMOS transistors within the same die.

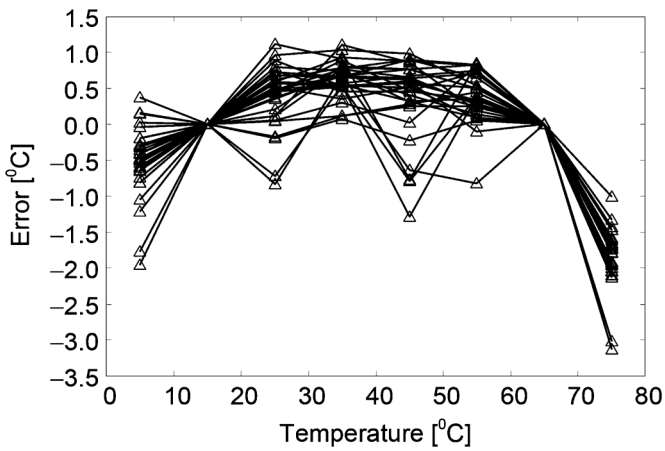


Fig. 31. Sensing error versus temperature after two-point calibration for 30 individual pMOS transistors within the same die.

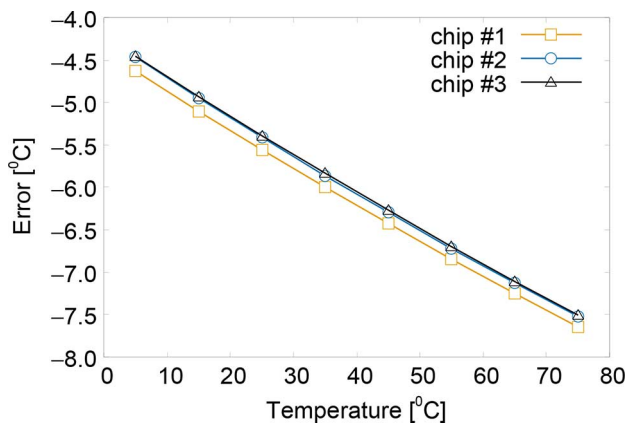


Fig. 32. Logarithm of delay period of homogeneous configuration for average nMOSFET I_{leak} monitoring against temperature.

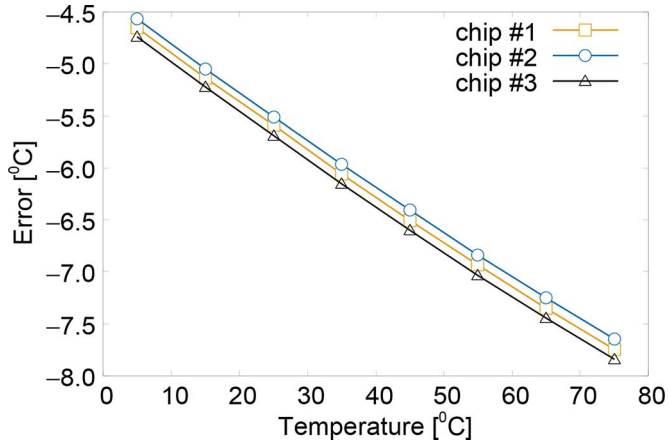


Fig. 33. Logarithm of delay period of homogeneous configuration for average nMOSFET I_{leak} monitoring against temperature.

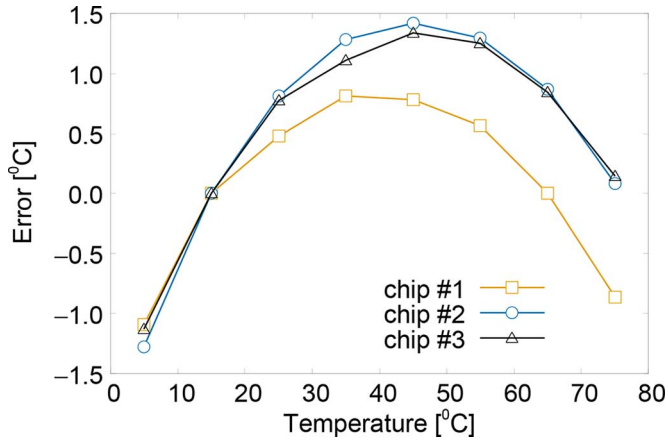


Fig. 34. Sensing error versus temperature for three different dies using nMOSFET average I_{leak} monitoring homogeneous configuration after one-point calibration.

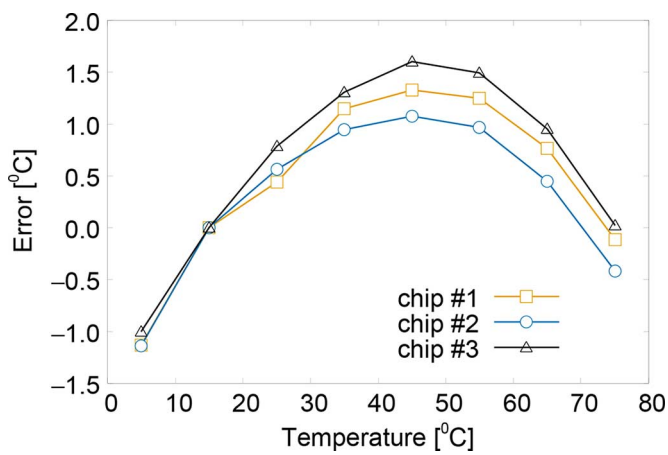


Fig. 35. Sensing error versus temperature for three different dies using pMOSFET average I_{leak} monitoring homogeneous configuration after one-point calibration.

minimized by optimizing the design parameters. Finally, trimming and offset removing algorithms can be applied to increase the accuracy.

Next, in order to achieve wide-voltage-range operation for the temperature monitor without the need for temperature calibration at each supply voltage, the voltage dependency needs to be calibrated and compensated. Although (4) has an implication that monitoring delay has strong dependency on temperature but weak dependency on the supply voltage, some offset will be introduced in the sensing error when supply voltage is varied. Several algorithms can be put forward to encounter this problem. One example can be to take the delay ratio of two transistors to

TABLE V
COMPARISON WITH STATE-OF-THE-ART ON-CHIP TEMPERATURE SENSORS

Reference	[12]	[15]	[28]	[29]	**This work**
Sensor type	BJT	MOSFET	MOSFET	MOSFET	MOSFET
Technology	0.16 μm	0.18/0.22 μm	32 nm	65 nm	65 nm
Design	Custom	Cell-based	Custom	Custom	**Cell-based**
Calibration (Point)	Voltage (1)	Thermal (1)†	Thermal (1)‡	Thermal (1)	Thermal (1)
Temperature range [°C]	$-55 \sim 130$	$0 \sim 100$	$0 \sim 100$	$40 \sim 90$	$5 \sim 75$
Error [°C]	$\pm 0.15(3\sigma)$	$-0.7 \sim 0.6$	$\pm 1.95(3\sigma)$	< 3.1	$-1.3 \sim 1.4$
Area [mm^2]	0.08	N/A⋆	0.0010	0.0012	**0.0045**
Power [μW]	5.1	175	100	N/A	**0.076**

† Master curve used. ‡ Thermal diode used. ⋆ FPGA implementation.

TABLE IV
COMPARISON WITH STATE-OF-THE-ART ON-CHIP PROCESS MONITORS

Reference	[7]	[9]	[27]	This work
Technology	65 nm	90 nm	65 nm	65 nm
N/P monitoring	Yes	No	Yes	Yes
Global Variation	Yes	Yes	Yes	Yes
Local Variation	Yes	No	No	Yes
Supply voltage [V]	0.76–1.5	1.2	1.2	**0.5–1.0**
Area [mm^2]	0.41	0.0061	0.0033	**0.0045**
Power [μW]	N/A	660	N/A	**0.076**

remove the voltage dependency parameters from (5). Assuming variation in sub-threshold swing coefficient n and DIBL coefficient λ to be negligible, delay ratio can be expressed as follows:

$$\Delta \ln(T_{\rm osc}) = \ln\left(T^i_{\rm osc}\right) - \ln\left(T^j_{\rm osc}\right) = \frac{\Delta V_{\rm th}}{n v_T}. \qquad (24)$$

Here, $T^i_{\rm osc}$ and $T^j_{\rm osc}$ are oscillation periods for the ith and jth stage respectively. i and j need to be chosen such that $\Delta V_{\rm th}$ gives us sufficient resolution of temperature monitoring. Equation (24) does not have supply voltage related term and only depends on temperature. Thus, the proposed sensor provides opportunities to implement various methods by exploiting the capability of monitoring multiple transistors. We plan to investigate on various algorithms to monitor robust wide-voltage-range temperature monitoring and their on-chip implementation methods in the future.

E. Comparison With State-of-the-Art

A state-of-the-art comparison is shown in Tables IV and V. Table IV shows a comparison for on-chip process variability monitors. In order to measure nMOSFET local variability and pMOSFET local variability separately, large number of samples are required which makes the monitor area large [7]. Analog approach to measure leakage current consumes large current [9]. Digital approach using standard cells fails to give us area-efficiency to implement local variability monitors [27]. Monitors utilizing actual logic behaviour [7], [27] can be useful in making design choices, but lack from providing us information for precise variability modeling. The proposed sensor provides digital output and gives us $V_{\rm th}$ variability information for both the nMOSFET and pMOSFET. The area required for 124 sample size for nMOSFET and 124 sample size for pMOSFET is only 0.0045 mm^2. Our sensor has achieved a wide range of supply operation down to 0.5 V without any tuning required.

Table V shows a comparison for on-chip temperature sensors. Highly accurate temperature sensor is proposed in [12] which uses ADCs and charge-balancing schemes. Sensors utilizing MOSFET leakage current as a parameter to sense temperature change lacks in high accuracy but provides high area-efficiency. An easy to design methodology such as cell-based design is proposed in [15] which is implemented in an FPGA. However, power consumption remains an issue for these sensors. The proposed sensor uses MOSFET leakage current as a parameter and monitors temperature variation. The proposed sensor is area-efficient and consumes the lowest power. Furthermore, the sensor can monitor multiple transistors providing further opportunity in achieving higher accuracy and low-cost implementation. Last but not the least, our sensor is an integrated on-chip process and temperature sensor together.

VIII. CONCLUSION

Normally, different sensors are required for monitoring MOSFET leakage current variation, process variation, temperature variation, and so on. The sensor design becomes more complex if we want to monitor nMOSFET and pMOSFET separately. We in this paper proposed a sensor architecture by which a single sensor will provide us leakage variation, process variation and temperature variation. Furthermore, parameters for nMOSFET and pMOSFET can be monitored without any area over-head. We proposed a reconfigurable leakage monitor cell for our sensor circuit and showed that leakage current variation can be effectively used for monitoring process and temperature variations. An on-chip demonstration sensor circuit is designed and fabricated in a 65 nm process. Our circuit is fully digital and cell-based design automation is performed for our implementation. Total area of the circuit is only 4500 μm^2. The circuit operates at wide supply voltage range down to 0.5 V. We successfully measured leakage variation of nMOSFET and pMOSFET and observed that within-die leakage current between nMOSFET and pMOSFET has no significant correlation, whereas die-to-die leakage current has relatively higher correlation. Leakage current variation remains almost constant over wide supply range. Thus the circuit can be ported to applications targeting high performance or low power without any tuning. Temperature dependency of nMOSFET and pMOSFET leakage current has also been measured. Our circuit provides a large number of transistor measurement thus we can develop an on-chip accurate temperature sensor. The sensor circuit is thus useful for various applications, such as

optimum back-gate biasing for leakage energy reduction and process variation compensation. The sensor can also be used for tracking chip power and temperature for energy and thermal management.

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their thorough reviews and critical comments which were essential for this paper. The VLSI chip in this study was fabricated in the chip fabrication program of the VLSI Design and Education Center (VDEC), University of Tokyo, in collaboration with e-Shuttle Inc. and Fujitsu Semiconductor Ltd.

REFERENCES

[1] S. Borkar, T. Karnik, S. Narendra, J. Tschanz, A. Keshavarzi, and V. De, "Parameter variations and impact on circuits and microarchitecture," in *Proc. Design Automation Conf.*, 2003, pp. 338–342.

[2] C. R. Lefurgy *et al.*, "Active management of timing guardband to save energy in POWER7," in *Proc. IEEE/ACM Int. Symp. Microarchitecture*, 2011, pp. 1–11.

[3] J. Howard *et al.*, "DVFS in 45 nm CMOS A 48-Core IA-32 message-passing processor with DVFS in 45 nm CMOS," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, 2010, vol. 9, no. 2, pp. 58–59.

[4] J. Tschanz *et al.*, "Adaptive frequency and biasing techniques for tolerance to dynamic temperature-voltage variations and aging," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, 2007, pp. 292–604.

[5] S. Saxena *et al.*, "Variation in transistor performance and leakage in nanometer-scale technologies," *IEEE Trans. Electron Devices*, vol. 55, no. 1, pp. 131–144, Jan. 2008.

[6] A. M. Islam, J. Shiomi, T. Ishihara, and H. Onodera, "Wide-supply-range all-digital leakage variation sensor for on-chip process and temperature monitoring," in *Proc. IEEE Asian Solid-State Circuits Conf.*, 2014, pp. 45–48.

[7] F. Klass, A. Jain, G. Hess, and B. Park, "An all-digital on-chip process-control monitor for process-variability measurements," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, 2008, pp. 408–623.

[8] S. Mukhopadhyay and K. Kim, "An on-chip test structure and digital measurement method for statistical characterization of local random variability in a process," *IEEE J. Solid-State Circuits*, vol. 43, no. 9, pp. 1951–1963, Sep. 2008.

[9] C. Kim, K. Roy, S. Hsu, R. Krishnamurthy, and S. Borkar, "On-die CMOS leakage current sensor for measuring process variation in sub-90 nm generations," in *Symp. VLSI Circuits Dig.*, 2004, pp. 250–251.

[10] 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," *IEEE Trans. Very Large Scale Integration (VLSI) Syst.*, vol. 14, no. 6, pp. 646–649, Jun. 2006.

[11] M. Fujii, H. Suzuki, and H. Notani, "On-chip leakage monitor circuit to scan optimal reverse bias voltage for adaptive body-bias circuit under gate induced drain leakage effect," in *Proc. Eur. Solid-State Circuits Conf.*, 2008, vol. 2, pp. 258–261.

[12] K. Souri *et al.*, "A CMOS temperature sensor with a voltage-calibrated inaccuracy of 0.15°C (3σ) from -55°C to 125°C," *IEEE J. Solid-State Circuits*, vol. 48, no. 1, pp. 292–301, 2013.

[13] P. Ituero, J. L. Ayala, and M. Lopez-Vallejo, "A nanowatt smart temperature sensor for dynamic thermal management," *IEEE Sensors J.*, vol. 8, no. 12, pp. 2036–2043, Dec. 2008.

[14] P. Ituero and M. López-Vallejo, "Ratio-based temperature sensing technique hardened against nanometer process variations," *IEEE Sensors J.*, vol. 13, no. 2, pp. 442–443, 2013.

[15] P. Chen, S. Chen, Y. Shen, and Y. Peng, "All-digital time-domain smart temperature sensor after one-point calibration," *IEEE Trans. Circuits Syst.*, vol. 58, no. 5, pp. 913–920, 2011.

[16] L.-T. Pang and B. Nikolic, "Measurements and analysis of process variability in 90 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 44, no. 5, pp. 1655–1663, 2009.

[17] K. Roy, S. Mukhopadhyay, and S. Member, "Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits," *Proc. IEEE*, vol. 91, no. 2, pp. 305–327, 2003.

[18] N. Kim, T. Austin, D. Baauw, and T. Mudge, "Leakage current: Moore's law meets static power," *Computer*, vol. 36, no. 12, pp. 68–75, 2003.

[19] K. J. Kuhn, "Moore's law past 32 nm: Future challenges in device scaling," in *IEEE Int. Workshop on Computational Electronics*, May 2009, pp. 1–6.

[20] R. Rao and A. Srivastava, "Statistical analysis of subthreshold leakage current for VLSI circuits," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 12, no. 2, pp. 131–139, 2004.

[21] N. H. E. Weste and D. M. Harris, *CMOS VLSI Design: A Circuits and Systems Perspective*, 4th ed. Boston, MA, USA: Addision-Wesley, 2011.

[22] S. Narendra, S. Borkar, V. De, D. Antoniadis, and A. Chandrakasan, "Scaling of stack effect and its application for leakage reduction," in *Proc. Int. Symp. Low Power Electronics and Design (ISLPED)*, 2001, pp. 195–200.

[23] A. M. Islam, T. Ishihara, and H. Onodera, "Reconfigurable delay cell for area-efficient implementation of on-chip MOSFET monitor schemes," in *Proc. IEEE Asian Solid-State Circuits Conf.*, 2013, pp. 125–128.

[24] A. M. Islam and H. Onodera, "Area-efficient reconfigurable ring oscillator for device and circuit level characterization of static and dynamic variations," *Jpn. J. Appl. Phys.*, vol. 53, no. 4S, pp. 04EE08-1–04EE08-8, 2014.

[25] T. Tsunomura *et al.*, "Analyses of 5σ Vth fluctuation in 65 nm-MOSFETs using Takeuchi plot," in *Symp. VLSI Technology Dig.*, 2008, pp. 156–157.

[26] S. Fujimoto, A. K. M. M. Islam, T. Matsumoto, and H. Onodera, "Inhomogeneous ring oscillator for within-die variability and RTN characterization," *IEEE Trans. Semicond. Manufact.*, vol. 26, no. 3, pp. 296–305, 2013.

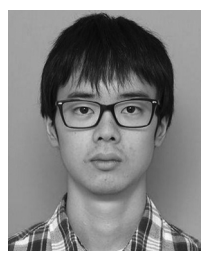
[27] T. Iizuka and K. Asada, "All-digital PMOS and NMOS process variability monitor utilizing shared buffer ring and ring oscillator," *IEICE Trans. Electronics*, vol. E95-C, no. 4, pp. 627–634, 2012.

[28] G. R. Chowdhury and A. Hassibi, "A 0.001 mm^2 100 μW on-chip temperature sensor with ± 1.95°C (3σ) inaccuracy in 32 nm SOI CMOS," in *Proc. IEEE Int. Symp. Circuits and Systems*, 2012, no. 1, pp. 1999–2002.

[29] E. Saneyoshi, K. Nose, M. Kajita, and M. Mizuno, "A 1.1V 35μm$\times 35\mu$m thermal sensor with supply voltage sensitivity of 2°C/10%-supply for thermal management on the SX-9 supercomputer," in *Symp. VLSI Circuits Dig.*, 2008, pp. 138–139.

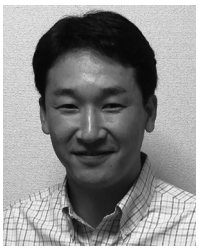
A. K. M. Mahfuzul Islam (S'11–M'14) received the B.E. degree in electrical and electronics engineering in 2009, the M.E. degree in communications and computer engineering in 2011, and the Ph.D. degree in informatics in 2014, all from Kyoto University, Kyoto, Japan.

He was a Research Fellow of Japan Society for the Promotion of Science from 2013 to 2015 at Kyoto University. In 2015, he joined the Institute of Industrial Science, University of Tokyo, Japan, as a Research Associate. His current research interests include development of low-cost and area-efficient on-chip sensor circuits, variability characterization and modeling, and ultra low-power design techniques.



Jun Shiomi (S'15) received the B.E. degree in electronic engineering from Kyoto University, Kyoto, Japan, in 2014. He is currently working toward the M.E. degree in communications and computer engineering from Kyoto University.

His research interest includes computer-aided design for low power and low voltage system-on-chips.



Tohru Ishihara (M'03) received the B.E., M.E., and Dr.E. degrees in computer science from Kyushu University, Fukuoka, Kyushu, Japan, in 1995, 1997, and 2000, respectively.

From 1997 to 2000, he was a Research Fellow of the Japan Society for the Promotion of Science. For the next three years, he worked as a Research Associate in the VLSI Design and Education Center at the University of Tokyo, Tokyo, Japan. From 2003 to 2005, he was with Fujitsu Laboratories of America as a research staff of an advanced CAD technology group. From 2005 to 2011, he was with the System LSI Research Center, Kyushu University, as an Associate Professor. In 2011, he joined Kyoto University, Kyoto, Japan, where he is currently an Associate Professor in the Department of Communications and Computer Engineering. His research interests include low power design and methodologies for embedded systems.

Dr. Ishihara has served as an editorial board member of the *Journal of Low Power Electronics* since 2008. He has also served on the technical program committees of several IEEE/ACM conferences including DATE, CODES+ISSS, and ISLPED. He is also a member of IPSJ, IEICE, and ACM.

Hidetoshi Onodera (M'87–SM'12) received the B.E., and M.E., and Dr. Eng. degrees in electronic engineering, all from Kyoto University, Kyoto, Japan.

He joined the Department of Electronics, Kyoto University, in 1983, and is currently a Professor in the Department of Communications and Computer Engineering, Graduate School of Informatics, Kyoto University. His research interests include design technologies for digital, analog, and RF LSIs, with particular emphasis on low-power design, design for manufacturability, and design for dependability.

Dr. Onodera has served as Program Chair and General Chair of ICCAD and ASP-DAC. He was Chairman of the IPSJ SIG-SLDM (System LSI Design Methodology), the IEICE Technical Group on VLSI Design Technologies, the IEEE SSCS Kansai Chapter, the IEEE CASS Kansai Chapter, and the IEEE Kansai Section. He served as Editor-in-Chief of the IEICE *Transactions on Electronics* and IPSJ *Transactions on System LSI Design Methodology*.