A 5800 \( \mu \text{m}^2 \) Process Monitor Circuit for Measurement of in-die Variation of \( V_{\text{th}} \) in 65nm

Liron Lisha, Ori Bass and Joseph Shor, Senior Member, IEEE

Abstract— A compact 5800 \( \mu \text{m}^2 \) process monitor circuit is demonstrated which can measure multiple threshold voltages \( (V_{\text{th}}) \) locally on a die. An accurate, process/voltage/temperature-independent current source is provided to measure \( V_{\text{th}} \) using the constant current method to an accuracy of \( \pm 3.6 \text{mV} \). The output is digitized by a sigma-delta modulator with a conversion time of 2ms. The circuit enables efficient and compact in-die variation monitoring of the key process parameters and is thus useful for the high-volume characterization of integrated circuit products. The circuit contains its own reference voltage, and thus can be used for calibration during testing or in the field.

Index Terms— Sensors, Process Monitor, Threshold Voltage, Sigma Delta, Switched Capacitor Circuits.

I. INTRODUCTION

In integrated circuits, the speed and performance of the products depend on the specific characteristics of the transistors. Systematic process variations between wafers and lots can alter the operating frequencies and performance of the IC’s, while random variations can affect the device speed across a single die. There have been process monitor circuits which utilize critical path monitors to detect the process, voltage and temperature (PVT) variation of CPU’s [1, 2]. These generally are either delay lines or ring oscillators which act as canary-bird circuits to detect when a critical speed path is violated. However, these circuits do not indicate individual devices parameters, but rather sense a combination of device parameters mixed together.

The most important device characteristics are threshold voltage \( (V_{\text{th}}) \) and mobility \( (\mu) \) since the following basic MOSFET equations are dependent on these parameters:

\[
\begin{align*}
\text{Linear region: } & \quad I_d = \mu C_{\text{ox}} \frac{W}{L} \left[ (V_{\text{gs}} - V_{\text{th}}) V_{\text{ds}} - \frac{1}{2} V_{\text{ds}}^2 \right] \quad (1) \\
\text{Saturation Region: } & \quad I_d = \frac{1}{2} \mu C_{\text{ox}} \frac{W}{L} \left[ (V_{\text{gs}} - V_{\text{th}})^2 \right] \quad (2)
\end{align*}
\]

where \( \mu \) = electron mobility, \( C_{\text{ox}} \) = oxide capacitance and \( V_{\text{th}} \) is the threshold voltage. \( W \) and \( L \) are the width and lengths of the device respectively. \( I_d, V_{\text{gs}}, V_{\text{th}} \) are the drain current, gate-source voltage, and drain source voltage. \( C_{\text{ox}} \) is a parameter which is very well controlled by the process and does not vary across corners or temperature. \( V_{\text{th}} \) and \( \mu \) both vary across corners and both are reduced as temperature increases.

There are a number of methods to measure \( V_{\text{th}} \), including the \( \text{gm}/I_d \) method, the 2nd derivative method, as well as the constant current (CC) method [3]. The \( \text{gm}/I_d \) and 2nd derivative method require a scan of \( I_d \) vs. \( V_{\text{gs}} \) and the peak in \( d(I_d)/dV_{\text{gs}} \) or \( d^2I_d/d^2V_{\text{gs}} \) are extracted. The CC method evaluates the gate voltage at which a single leg drives a constant current (usually 50-100nA) with the drain voltage fixed at either a low voltage (linear region) or the Vdd voltage (saturation region). Another version of the CC method applies the current to a diode connected device (Fig. 1(a)) while measuring the \( V_{\text{gs}} \). The CC method has found extensive use in the industry because of its inherent simplicity, since it only requires a single measurement and not a scan of \( I_d \) vs. \( V_{\text{gs}} \). In the \( I_d \) vs. \( V_{\text{gs}} \) curve, the transition between subthreshold and strong inversion is a gradual one. Essentially, the CC method allows the product designers to decide what level of device leakage is acceptable and utilize that point on the curve to determine \( V_{\text{th}} \). Usually semiconductor fabs place individual devices on the scribe lines between chips and measure them during wafer testing. However, these structures do not give an indication of the in-die variation within the product chips, especially if these chips are very large. Furthermore, the exact value of \( V_{\text{th}} \) is generally invisible to the customer unless a \( V_{\text{th}} \) detector is implemented on die. There have been several recent circuits which measure \( V_{\text{th}} \) for either process monitoring or sensors [4-9]. Some of these circuits utilize the \( V_{\text{th}} \) measurement to produce either a voltage or temperature reference [4,5,8,9]. These circuits generally measure only one type of \( V_{\text{th}} \), while in modern processes there can be multiple \( V_{\text{th}} \)’s, both for NMOS and PMOS devices. Most of these circuits produce an analog voltage which can only be digitized off-chip. There are also process monitors which use ring oscillators to monitor leakage [10], threshold voltage variation [11,12] as well as N/P ratio [13]. However, these ring oscillator circuits do not provide a direct and accurate measure of \( V_{\text{th}} \), but rather provide \( V_{\text{th}} \) variation measurements across the wafer. In this paper, we propose a miniaturized Process Monitor (PM) which is capable of monitoring \( V_{\text{th}} \), for multiple device types, both in testing and during regular operation. The PM contains an integrated Sigma-Delta (SD) analog-to-digital converter (ADC) and provides the data as a digital code. This circuit occupies 5800 \( \mu \text{m}^2 \) and can thus be placed in multiple locations across the die to monitor in-die variation. It also includes an integrated bandgap reference (BGREF) and

The authors are with the Faculty of Engineering, Bar Ilan University, Israel. This work was funded by the Hiper Consortium of the Israel Innovation Authority.
temperature sensor for multi-functionality, which are similar to [14].

![Image](image1.png)

Fig. 1. Elements to be measured (a) parallel legs of minimal sized devices, and (b) stacks of minimal devices. For the PMOS stacks, the bulks were connected to the sources.

![Image](image2.png)

Fig. 2. NMOSCAP and simulations of the capacitance in different corners.

II. PROCESS MONITOR CIRCUITRY

Figure 1 shows a method to measure $V_{th}$ and $\mu*C_{ox}$ using controlled current sources. With the CC definition, $V_{th}$ is the $V_{gs}$ of a single minimal-size MOS driven by 100nA. The purpose of this design is to detect the process corner as well as systematic $V_{th}$ shifts across the die or wafer. As such it is important to reduce the local mismatch between adjacent devices to a minimum. In order to reduce the local random variation to several mV, 100 minimal legs are driven by a precision 10 $\mu$A current source, and the $V_{gs}$ can be measured by an ADC (Fig 1a). There were also composite structures implemented which were stacks of minimal sized devices driven by the same current source (Fig 1b) in order to alter the current density. Simulation show that the stack composites behave very similarly to a single device at similar current densities. Intermediate nodes of the stacks could also be input into the ADC (Vdn and Vdp of Fig. 1b) to derive the $V_{dsat}$ voltage. This could enable measurement of $\mu*C_{ox}$, although this feature was not demonstrated in the present paper.

![Image](image3.png)

Fig. 3. Schematic of the process monitor and the DUT array.

The key to the success of this design is to provide an accurate current source which requires no external pins or references. In order to produce this, a known reference voltage can be utilized to charge switch capacitor elements. In the CMOS process, $C_{ox}$ is a parameter which is nearly constant across dies and wafers.

![Image](image4.png)

Fig. 4. Schematic of the Sigma Delta Modulator.

This is shown in Fig. 2, which exhibits the NMOSCAP, which is a n-device placed in an NWELL to form an accumulation capacitor. Simulations of this capacitor (Fig 2) show that for a constant bias of 800mV, the capacitance changes by $< \pm 4\%$ across process corners and temperatures. A simplified schematic of the PM is shown is Fig. 3. A reference voltage, $V_{ref}$, is provided to a voltage regulator which drives a switched-capacitor load. The reference voltage can either be connected to the BGREF, or to a divided $V_{cc}$ voltage. The current through M1* and M2* is regulated and obeys the following equation:

$$I = C_{fly}V_{fb}f$$  \hspace{1cm} (3)

All the parameters are well controlled, since $V_{th}$ is regulated, $f$ is a known reference frequency of 40MHz and $C_{cond}$ is based on the NMOSCAP, whose value is stable. The amplifier and output transistor M1* of the PM utilize self-biased cascodes. This technique takes advantage of the multiple $V_{th}$ which are available in the process. A composite cascode was used, with a low-$V_{th}$ device (LVT) connected in series with high-$V_{th}$ device (HVT). This enables both devices to be in saturation,
thus increasing the $R_{out}$ and gain. The $V_{ds}$ of the HVT device obeys the following equation:

$$V_{ds,HVT} = V_{gs,HVT} - V_{gs,LVT} > V_{dsat,HVT} \tag{4}$$

The use of the composite cascode to achieve higher gain is not a significant limitation, since most advanced process node have several $V_{th}$ levels. The PM amplifier (including the output stage M1*) has a nominal gain of 64dB. The current in M1* is mirrored to M2* which drives a MUX that selects the device under test (DUT). Since the active current of the enabled switch in the MUX is 10μA, the leakage currents of the disabled switches are negligible in comparison. Thus, the MUX can have many legs to measure multiple DUTs. There are 4 types of $V_{th}$ DUT, NMOS or PMOS standard $V_{th}$ (SVT) and low $V_{th}$ (LVT). There are also several composite stacks of NMOS and PMOS SVT. The output voltage of the MUX or the internal nodes of the stack are digitized by a first order Sigma Delta ADC (Fig. 4), which is identical to [14]. Inaccuracies in the current, due to mismatch of the amplifier and current source have a small effect on the $V_{th}$ measurements, since when $V_{gs} \approx V_{th}$, the DUT is still in subthreshold, and the $V_{gs}$ voltage is changed by the current in a logarithmic manner. The transition from subthreshold to the quadratic strong inversion operating point is a gradual one such that at $V_{gs} = V_{th}$ the behavior is still logarithmic. For example, simulations show that that an error of ±4% in the current, due to the corner spread of the NMOSCAP, could cause a small error of ±1.5mV in the $V_{th}$ measurement.

![Simulated SVT NMOS Vth vs Temperature](image1)

**Fig. 5.** Simulated SVT NMOS Vth – solid lines are analog voltages with an ideal 10μA current source, while the markers represent the simulated ADC output of the PM.

The ADC and BGREF, which occupy a combined area of 3000 μm² also form a temperature sensor as described in [14]. Although they were published and described in [14], some discussion is included here for comprehensiveness. The ADC (Fig. 4) is a first order, single bit, Sigma Delta modulator. It uses a common source amplifier as the integrator and is followed by a CMOS inverter amplifier which raises the signal to digital levels. An 11-bit counter is utilized as a sinc filter to decimate the bit-train and provide a digital code in 2ms. At this signal bandwidth, simulations show that the ADC has an equivalent number of bits (ENOB) > 11, which is sufficient for this application. The BGREF [14] exhibits a measured random variation sigma of 1.6mV and a temperature coefficient < 54ppm. Since it utilizes switched-capacitor elements, the BGREF is not sensitive to corners either. When the 0.8V BGREF reference is utilized, the LSB of the measurement is 0.4mV. The circuit also contains the capability of taking $V_{ref}$ from a divided Vcc.

![Simulated NMOS SVT Vth Monte Carlo](image2)

**Fig. 6.** Simulated NMOS SVT $V_{th}$ Monte Carlo, 200 splits, typical 27°C, (a) ideal current source and (b) Simulated PM current source.

![Measured Va’s of 4 Vth on a sensor vs temperature](image3)

**Fig. 7.** Measured $V_a$’s of 4 $V_{th}$ on a sensor vs temperature.

### III. SIMULATED AND MEASURED RESULTS

The PM circuit was analyzed using simulations and Silicon measurements (in 65nm) to validate that it could provide an accurate representation of the $V_{th}$ and the process corner. Figure 5 shows a simulation of the entire PM and ADC compared to an “ideal” $V_{th}$ measurement. The solid lines are an analog SVT NMOS $V_{th}$ simulation, whereby the DUT is charged by an ideal 100nA current source. The dotted markers are the digitized output representation of complete PM. The 3 process corners, TT, FF and SS are clearly distinguishable and differentiated by the simulation. The error produced by the analog circuitry, compared to the ideal measurement is nominally < 0.5%. This graph shows that the $V_{th}$ corner of the chip can be clearly identified by the PM. Monte Carlo simulations of the mismatch of the ideal $V_{th}$ measurement and the analog output $V_{DUT}$ of the PM are shown in Fig. 6. The mismatch of all the analog circuity is negligible compared to
the mismatch of the 100 parallel devices. In order to reduce this mismatch, more parallel devices would have to be charged. In this design the DUTs contained 100 parallel legs, charged by 10µA. This was chosen as a good tradeoff between accuracy and power consumption/area. Alternatively, it is also possible to reduce mismatch by sequentially measuring multiple DUT’s and averaging these measurements. Nevertheless, Figs. 5 and 6 demonstrate that this PM is capable of clearly distinguishing between different process corners. 

Figure 7 shows the Si measured $V_{th}$ vs. temperature of four different types of $V_{th}$ DUT. The SVT are clearly distinguishable from the LVT by > 50mV and the PMOS exhibited a higher $V_{th}$ than the NMOS. All of the $V_{th}$’s show a nearly linear negative temperature dependence, as expected. The measured $V_{th}$ of 5 sensor chips are shown for different devices in Fig. 8. Each chip had 4 PM’s, which measured each of the 4 $V_{th}$’s (NMOS/PMOS, SVT/LVT). The sigma variation within each die was < 2mV in most of the chips as shown for the PMOS SVT in Fig. 9. Between the die the average $V_{th}$ could vary by 3-9mV. Since the die came different parts of the wafer, this variation is to be expected. For the PMOS SVT, chips from wafers of two different lots were measured and plotted on the same graph (Fig. 8b and Fig. 9). The average $V_{th}$ between the 2 lot differs by 25mV, while the sigma variations within the lots are similar. The $V_{th}$’s are consistent with the TT corner in simulations and the sigma variations are consistent with the simulated values. The power supply sensitivity of the measurement is shown for several NMOS SVT measurements in Fig. 10. The measured Power Supply Rejection exhibited in this graph is -38dB. Figure 11 illustrates repeated measurements of the NMOS SVT using the BGREF as a reference. The worst RMS resolution is 168µV which indicates the resolution of the ADC.

IV. DISCUSSION AND CONCLUSIONS

Table 1 shows a comparison of this work to the recent prior-art. There is a very sparse amount of similar papers in the recent literature, so the PM is compared to circuits which utilize accurate $V_{th}$ measurements for other functions as well, such as voltage or temperature references. This is the only circuit in Table 1 which provides a digitized measurement of $V_{th}$. This PM can give absolute measurements of multiple $V_{th}$’s on a die. The circuit is small enough to place multiple instantiations of it on a product. This enables both temperature measurements [14] as well as measurements of the relative systematic variation of $V_{th}$ across the product die or the wafer – a function similar to the circuits reported in [10-13]. We intend to develop
additional capabilities in the circuit to measure mobility using the intermediate nodes of the composites. Using this architecture, it is also possible to measure \( V_{th} \) using the \( gm/I_d \) method or the 2nd derivative method by scanning \( I_d \). This can be done by varying the frequency of the switched capacitor current source and also altering the current densities by utilizing the stacks. The small size, flexible measurement, and digitized output enable this process monitor to be placed in multiple locations in a die, enabling accurate in-die-variation measurements of multiple \( V_{th} \)’s for both the FAB and product developers. A die photo as well as the layout of the PM (Fig. 3) is shown in Fig. 12. The PM circuit occupies 2800 \( \mu \)m\(^2\). The Sigma Delta circuit along with its biases and counter occupy roughly 700 \( \mu \)m\(^2\), while the BGREF requires \( \sim 2300 \) \( \mu \)m\(^2\).

**Fig. 12. Chip Photo + Layout of the Process Monitor.** (The BGREF and ADC, whose combined area is 3000\( \mu \)m\(^2\) [14], is not included)

### REFERENCES


[9] S. Kim and M. Seok, “A 30.1\( \mu \)m\(^2\), ±1.1°C-3eV, 0.4-to-1.0V temperature sensor based on direct threshold-voltage sensing for on-chip dense thermal monitoring,” *2015 IEEE Custom Integrated Circuits Conference (CICC)*, San Jose, CA, 2015, pp. 1-4.


