Threshold Voltage Mismatch and Intra-Die Leakage Current in Digital CMOS Circuits

José Pineda de Gyvez, Member, IEEE, and Hans P. Tuinhout

Abstract—Due to device and voltage scaling scenarios for present and future deep-submicron CMOS technologies, it is inevitable that the off-state current ($I_{off}$) of MOSFET transistors increases as the technology minimum dimensions scale down. Experimental evidence shows that the leakage current distribution of modern deep-submicron designs not only has a higher mean value but it also presents a larger variability as well. In this paper, we investigate the impact of threshold voltage mismatch as one plausible source for this increased variability. In digital circuit design, it is commonly assumed that the threshold voltage difference (mismatch) of static CMOS cells is negligible. However, threshold voltage mismatch ($\Delta V_t$) has a two-sided effect on the off-state current. Namely, the total cell’s current can increase or decrease depending upon the direction of the $V_t$ mismatch shift. This effect can be so severe that $I_{off}$ can increase by more than one order of magnitude with respect to its nominal value due only to $V_{to}$ mismatch. We further show through experimental results that the $V_{to}$ mismatch of paired transistors working in the subthreshold regime can be worse by a factor of two as compared to transistors working in the saturation or linear regions. A factor of two larger spread is obviously quite devastating in terms of area, speed, and power consumption, should it be desired to attain high parametric yield [1]. In fact, due to the use of submicron transistor dimensions, modern circuits become quite sensitive to random local $V_t$ fluctuations are often not considered during the circuit design process. Moreover, while the leakage current of modern deep-submicron circuits has increased as expected, we have also noticed that the variance to mean ratio of measured leakage current distributions has increased as well. This is a clear indication of the impact of process variability on modern circuit designs. Unfortunately, many CAD tools or digital design flows cannot take intra-die fluctuations into account. For designs based on low-voltage low-power premises, this can result in unexpected clock skews, excessive leakage currents, out of spec critical-path delays, etc.

In this work, we study the above-mentioned leakage current variability in relation to threshold voltage mismatch. Most of the published work on $V_{to}$ mismatch consists of device and technology characterization and is intended primarily for analog and mixed-signal circuits and systems [2]–[4]. Worth stressing is that while the detrimental effects of $V_t$ mismatch in analog circuits are well described [11], [12], little is known in comparison for digital circuits. Previous works on digital-design $V_t$ variability focus primarily on global effects [13]–[16], i.e., they consider a common augmenting or decreasing value of the nominal $V_t$. This means that fundamental device limitations such as $V_t$ differences (mismatch) among transistors are not accounted for [6]–[10]. A statistical approach for intra-die $V_t$ variability for critical-path delay analysis was performed in [17], a study of the impact of process variability on leakage current levels was done in [5] and a statistical estimation of a chip’s global leakage current was shown in [31]. Here, we carry out a thorough analytical investigation to demonstrate the effect of $V_t$ mismatch on the leakage current of a single digital cell. Our results explain why for some cases, despite the presence of $V_t$ mismatch, the leakage current can be less than the nominal expected value, or why the leakage current behavior appears to be random. Also, we show that in the presence of mismatch the off-state current can span more than 1.5 orders of magnitude and in some cases the total off-state current of a cell simulated with nominal global parameters can be comparable to that of the same cell employing fast-process devices. The latter result is particularly important for the design and testing of low-power circuits and systems and partly explains the excessive variability of measured leakage distributions.

Since the transistor’s off-state (leakage) current is due to its operation in the subthreshold domain, it makes sense to investigate the corresponding $V_t$ mismatch in this region as well. The subject of matching of MOS transistors in the subthreshold region has been addressed several times in the literature [18]–[22]. Forti and Wright [20] presented the first fairly qualitative study on drain current mismatch fluctuations for N- and P-MOSTs from four different CMOS processes. Pavasovic et al. [18] demonstrated that mismatch characterization in the subthreshold regime can be used as an accurate and powerful technique for CMOS process evaluation. Chen et al. [21] studied drain current mismatch for MOSFETs in weak inversion as a function of gate voltage and back-gate bias. They derive a model that relates the drain current mismatch fluctuations in weak inversion to fluctuations of the flat-band

Manuscript received September 27, 2002; revised July 5, 2003.

The authors are with Philips Research Laboratories, 5656 AA Eindhoven, The Netherlands (e-mail: jose.pineda.de.gyvez@philips.com; hans.tuinhout@philips.com).

Digital Object Identifier 10.1109/JSSC.2003.820873

1To simplify the notation, we will henceforth use $V_t$ instead of $V_{to}$.
voltage and the body effect parameters. Denison et al. [22] proposed a prediction of weak inversion drain current mismatch fluctuations based on threshold voltage fluctuations as observed in the strong inversion regime. One of the justifications for this approach is that they claim that measuring mismatch in weak inversion is “a difficult exercise.” They show however, a perfect correlation between calculated weak inversion drain current mismatch fluctuations as based on strong inversion threshold voltage mismatch observations when compared to actual measured drain current mismatch variances. Bastos [19] also observed a good correlation between the \( V_t \) mismatch as observed in the weak inversion and the saturation regimes and concluded that one model or characterization technique is sufficient to describe the mismatch behavior of a transistor in all regions of operation. In Section VI we demonstrate, however, that this correlation does not necessarily hold for MOS transistor device architectures in deep-submicron CMOS technologies. By correlating strong inversion \( V_t \) mismatch measurements with gate voltage mismatch measurements in deep subthreshold, we demonstrate that off-state current fluctuations can even be significantly larger than what is predicted using conventional modeling approaches.

II. SOURCES OF THRESHOLD VOLTAGE VARIABILITY

In essence, it is possible to identify two sources of \( V_t \) variations. One of them due to global manufacturing variations and the other due to local random fluctuations. Let us first illustrate how \( V_t \) varies as a function of the transistor size and global process fluctuation. Fig. 1 shows a typical example plot of threshold voltage versus channel length for an arbitrary CMOS 0.18-\( \mu \)m technology. The three curves correspond to nominal, fast and slow \( V_t \) values. Basically, Fig. 1 depicts the typical behavior of the state-of-the-art of deep-submicron devices where the traditional \( V_t \) reduction due to short channel effects (SCEs) is partly compensated by local channel implants (pockets) that result in an increase of \( V_t \) (reverse short channel effect, or RSCE). In long channel transistors \( V_t \) variations are mainly due to global variations on ion implantation dose, gate dielectric thickness, etc. For short channel transistors, on the other hand, \( V_t \) variability is predominantly determined by the transistor’s geometries, in particular \( L_{eff} \). Although this is a deterministic phenomenon (physically well understood and properly modeled), for circuit simulation purposes this can be treated as a global variation as well.

Let us now consider a pair of transistors. Experimental results have shown that there can be random \( V_t \) (mismatch) differences between two closely placed “identical” transistors of the order of about 100 \( \mu \)V to 100 mV [22]. This is a local effect that becomes extremely important in deep-submicron transistors as the mismatch magnitude is generally observed to be inversely proportional to the square root of the transistor’s area. This local effect is random and is among other things due to a statistical fluctuation of dopant atoms per unit volume [20], [21]. Usually the standard deviation of the \( V_t \) mismatch scales with the inverse square root of the active device area. The proportionality constant \( A_{\Delta V_t} \) in \( mV \mu m \) units is a characteristic parameter for a technology. A simple rule of thumb to estimate threshold voltage mismatch is to assume a variation of 1 \( mV \mu m \) of square root of active area per nanometer of gate-oxide thickness. For a technology with 4 nm of oxide thickness and minimum channel length and width of 0.18 \( \mu m \), this corresponds to a threshold voltage standard deviation of approximately 22 mV. Experimental measurements of threshold voltage due to (global) inter-die variations show that it can be considered as normally distributed with a mean \( \mu_G \) and standard deviation \( \sigma_G \). Threshold voltage mismatch observations based on transistor pairs can be described as well with a normal distribution with mean \( \mu_\Delta \) and standard deviation \( \sigma_\Delta \). We therefore let the latter terms describe the intra-die variations. If we assume that global and intra-die variations are independent, it follows that the total \( V_t \) variation of a single transistor can be calculated as [2]

\[
(\sigma_{V_t})^2 = (\sigma_G)^2 + 0.5(\sigma_\Delta)^2
\]

(1)

with

\[
\sigma_\Delta = \frac{A_{\Delta V_t}}{\sqrt{WL}}
\]

(2)

where \( A_{\Delta V_t} \) is a technology conversion constant (in \( mV/\mu m \)), and \( WL \) denotes the product of the transistor’s active area. The 0.5 factor in (1) arises because only one transistor from the pair is considered at a time. This paper presents results of a 0.18-\( \mu \)m CMOS process that yields an \( A_{\Delta V_t} \) value of 7 \( mV/\mu m \) for the NMOS transistors provided they do not operate in the subthreshold regime. Thus, for NMOS transistors with minimum dimensions, e.g., \( W/L = 0.28 \mu m/0.18 \mu m \), the estimated statistical \( V_t \) mismatch is \( \sigma_\Delta = 31.18 \) mV. It is worth
pointing out that $\sigma_{G}$ is of the order of 30 mV. It follows then that threshold voltage mismatch is very important for deep-submicron technologies.

Fig. 2 illustrates the typical inter-die threshold voltage distributions of large and small area NMOS transistors in CMOS18. The spread of $V_t$ is obvious for small area transistors.

III. OFF-STATE CURRENT MODEL BASED ON $V_t$ MISMATCH

The typical schematics of a digital cell consist of either a P-section of parallel connected p-channel transistors with a common node tied to an N-section of series connected n-channel transistors, or vice versa (see Fig. 3). In the absence of transistor tapering, it is expected that the dimensions of the transistors are equal in their corresponding N or P sections. Moreover, in a typical layout, these transistors are closely spaced from each other and hence can be supposed to be matched. Naturally, it is also expected that their intrinsic $V_t$‘s are the same in either the N or P sections. Also, for purposes of high speed and maximum packaging density, these cells are designed typically with minimum channel length transistors.

Let us now investigate the role of threshold voltage mismatch on the off-state current of a typical digital CMOS cell. Assume without loss of generality a cell with two inputs. Let us further assume a local random variation in $V_t$‘s of two transistors M1 and M2, e.g., one arising from fluctuation of random dopant diffusion, dopant clustering, interface states, etc. Since $V_{t1} \neq V_{t2}$, their $V_t$ mismatch and average values can be expressed as in (3) and (4), respectively:

$$\Delta V_{t21} = V_{t2} - V_{t1}$$

$$V_{to} = \frac{V_{t1} + V_{t2}}{2}.$$  

(3)

(4)

After some algebraic manipulation, it is possible to express individual $V_{to}$‘s as a function of their mismatch and average values as indicated in (5) and (6):

$$V_{t1} = V_{to} - \frac{\Delta V_{t21}}{2}.$$  

(5)

$$V_{t2} = V_{to} + \frac{\Delta V_{t21}}{2}.$$  

(6)

Based on this definition of mismatch and a common description of subthreshold conduction (after [25]) we can derive an analytical estimation of the leakage current of a CMOS cell, as will be shown next. Without loss of generality, let us consider a series-connected (stack) of NMOS transistors as shown in Fig. 4(a). They could, for instance, be the NMOS transistors of a two-input NAND cell.

For the circuit of Fig. 4(a), let us further assume that the path from $V_{dd}$ to ground is open and that a current equal to the off-state current of the series-connected transistors is flowing. This off-state current is primarily formed by the subthreshold current since the junction and gate leakage currents are mostly negligible in 0.18-$\mu$m MOS circuits. Let us denote the subthreshold current in the drain of M1 as $I_1$, and correspondingly, $I_2$ for M2. Making use of a qualitative subthreshold current model based on BSIM [25] and introducing now the $V_t$ mismatch between M1 and M2 as found in (3) and (4), we can express $I_1$ and $I_2$ as follows:

$$I_1 = I_{01} \exp \left( \frac{V_{GS1} - (V_{to} - \frac{\Delta V_{t21}}{2}) + \eta V_{DS1} - \gamma V_{SB1}}{nU} \right) \times \left( 1 - \exp \left( \frac{-V_{DS1}}{U} \right) \right)$$  

(7)

$$I_2 = I_{02} \exp \left( \frac{V_{GS2} - (V_{to} + \frac{\Delta V_{t21}}{2}) + \eta V_{DS2} - \gamma V_{SB2}}{nU} \right) \times \left( 1 - \exp \left( \frac{-V_{DS2}}{U} \right) \right)$$  

(8)

in which $\eta$ is the drain-induced barrier lowering (DIBL) coefficient, $\gamma$ is the linearized body effect factor, $n$ is the so-called...
subthreshold slope ideality factor, $U$ is $kT/q$, and $I_o$ is the so-called saturation current given as [25]

$$I_o = \frac{W}{L} \frac{\mu}{L} C_{ox} U^2$$

(9)

where $\mu$ is the electron mobility and $C_{ox}$ the gate-oxide capacitance per unit area. Typical values for the previous parameters in a CMOS 0.18-$\mu$m technology are given in Table I for $W/L = 10 \mu$m/0.18 $\mu$m. Before proceeding any further, it is worth noting that (7) and (8) predict an off-state current that has high and low values depending upon the shift in $V_t$ mismatch. For equal $V_C$, one can see that if $\Delta V_{D21} > 0$ then $I_1 > I_2$ and vice versa.

A. Analysis of $I_{off}$ for $V_t$-Mismatched Stacked Transistors

Let us return now to the circuit of Fig. 4(a) and consider the following cases: 1) $V_1 = 0$, $V_2 = 0$; 2) $V_1 = V_{dd}$, $V_2 = 0$; and 3) $V_1 = 0$, $V_2 = V_{dd}$. Fig. 5(a) shows simulated results of the off-state current as a function of $V_t$ mismatch obtained through PSTAR, which is an internal SPICE-like simulator that uses Philips’ MOS-9 compact transistor model [26]. Furthermore, simulations were carried out using a full CMOS 0.18-$\mu$m process description. Transistor sizes used in the simulation were $W/L = 1.4 \mu$m/0.18 $\mu$m. For comparison, Fig. 5(b) shows the same qualitative behavior based on (7) and (8). The difference between the two plots arises from the fact that (7) and (8) are simplified models that do not consider all short channel effects.

Fig. 5 shows the known behavior of many digital cells, i.e., there is an obvious dependence of the leakage current on the cell’s inputs. For nominal conditions, e.g., $\Delta V_t = 0$, the input state that shuts off all the transistors of a digital cell is the one that renders the lowest leakage current; for other input states the off-state current level is simply higher. Putting this aside, what is really new about this plot is the dependence of the leakage current on the $V_t$ mismatch between the cell’s transistors. In the presence of mismatch, i.e., $\Delta V_t \neq 0$, the cell’s leakage current has a significant spread of approximately 1.5 orders of magnitude. This situation is not uncommon in modern digital ICs since a voltage mismatch of $\Delta V_t = 90$ mV amounts to an expected $3\sigma$ tolerance window from the fabrication process, and this is just considering only intra-die variations. Needless to say, the larger this current is, the larger the chip’s power consumption will be. This shows the relevance of threshold voltage mismatch in modern deep-submicron low-power designs.

The focus of this work is, thus, to understand the leakage behavior of digital cells in the presence of $V_t$ mismatch, such as the one of Fig. 5. In the rest of this section, we will present simulation results based on the full CMOS process description and qualitatively explain the results using (7) and (8). Subsequently, we will show simplified analytical solutions of $I_{off}$ as a function of $V_t$ mismatch for each case study to explain the observed results.

### Table I

<table>
<thead>
<tr>
<th>$\eta$</th>
<th>$\gamma$</th>
<th>$U$</th>
<th>$V_C$</th>
<th>$n$</th>
<th>$t_{ox}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.034</td>
<td>0.33</td>
<td>26mV</td>
<td>0.51V</td>
<td>1.37</td>
<td>4.2nm</td>
</tr>
</tbody>
</table>

**Fig. 5.** Off-state current for different voltages at the inputs of two series-connected transistors simulated (a) using a complete process description and (b) using simplified models that do not consider all short channel effects.

**Case 1:** $V_1 = 0$ and $V_2 = 0$: Observe first that $V_{CS1} = -V_{DS1} = -V_{DS2}$, $V_{DS1} = V_{dd} - V_{DS2}$, and $V_{CS2} = V_{SS} = 0$. Notice also that since $V_1 = V_2 = 0$, there is no channel formed and thus the impedance between drain and source is large, rendering a $V_D > U$. For $V_{DS} > 3U$, we have that $\exp(-V_{DS}/U) \approx 0$. Now making $I_1 = I_2$ and solving for $V_{DS2}$, we have that

$$V_{DS2} = \frac{1}{1 + 2\eta + \gamma} \left[ \Delta V_{21} + \eta V_{dd} - nU \ln \left( \frac{I_{02}}{I_{01}} \right) \right].$$

(10)

Notice that this model is valid only for $\Delta V_{21} + \eta V_{dd} > 0$, otherwise $V_{DS2}$ becomes negative. Assuming that geometrical variations are negligible, i.e., neglecting the $V_{dd}$ term, and noting that $\Delta V_{21}$ is in the order of $\eta V_{dd}$, this equation shows that $V_{DS2}$ is highly sensitive and directly proportional to $\Delta V_{21}$. Inserting (10) into (8), we obtain

$$I_{off} = I_{02} \exp \left( -\frac{V_{to} + \Delta V_{21}}{nU} \right) \times \left( 1 - \exp \left( \frac{-\Delta V_{21} + \eta V_{dd}}{U(1 + 2\eta + \gamma)} \right) \right) \approx I_{02} \exp \left( -\frac{V_{to} - \Delta V_{21}}{nU} \right) \left[ 1 - \exp \left( -\frac{\Delta V_{21} + \eta V_{dd}}{U(1 + 2\eta + \gamma)} \right) \right].$$

(11)
when $\Delta V_{121} < \eta V_{dd}$ and $(\Delta V_{21} + \eta V_{dd})/(1 + 2\eta + \gamma) < U$, and noting that the expression within curly brackets can be approximated as a term asymptotically converging to a very small value given by $(\Delta V_{21} + \eta V_{dd})/(U (1 + 2\eta + \gamma))$. Because of its small value, this expression dominates the other exponential. Consequently, the total off-state current remains more or less constant for $\Delta V_{21} < \eta V_{dd}$. On the other hand, for $(\Delta V_{21} + \eta V_{dd})/(1 + 2\eta + \gamma) > U$ the exponential within curly brackets approaches zero and thus its effect is negligible, i.e., we see that the total off-state current is dominated by the $V_t$ mismatch in the other exponential. Despite this $V_t$ mismatch and because of the exponential decaying behavior of the current, the off-state current is lower than the nominal value without mismatch. This is one of the rare situations in which a circuit can benefit from a usually undesired mismatch.

Case 2: $V_1 = V_{dd}$ and $V_2 = 0$: In this case, observe that $V_{GS1} > V_{T1}$ and $V_{DS1} \approx 0$ V because there is an inversion layer underneath the gate of M1. Since $V_{GS1} = V_{dd} - V_{DS2}$, it follows that $V_{dd} = V_{DS2} \geq V_{T1}$. Thus, we have that $V_{DS2} \leq V_{dd} = V_{T1}$. The off-state current of this series connection is primarily due to M2. Substituting $V_{DS2}$ into (8), we have

$$I_{\text{leak}} \approx I_2 \approx I_{02} \exp \left( -\frac{(V_{to} + \frac{\Delta V_{21}}{2})}{n U} + \eta V_{dd} \right) \times \left( 1 - \exp \left( -\frac{V_{dd} - V_{T1}}{U} \right) \right).$$

(12)

The off-state current presents a pure decaying exponential behavior and is largely dependent on the DIBL effect $\eta$. This nonnegligible $\eta V_t$ contribution causes the asymmetry between cases 2 and 3. Due to the negative sign before $\Delta V_{21}$, we will find that the off-state current increases as the $V_t$ mismatch shifts from negative to positive values (see Fig. 5). Consequently, we can see that depending upon the $V_t$ mismatch shift, the cell’s off-state current can be found in a very wide range covering more than 1.5 orders of magnitude. This wide range can be deleterious for low-power systems, not to mention the difficulty of attaining such a design.

Case 3: $V_1 = 0$ and $V_2 = V_{dd}$: Since $V_2 = V_{dd}$, we have that there is an inversion layer under the gate of M2 and thus its output resistance is low. In this case, observe that since $V_{GS2} = V_{dd}$, it is expected that $V_{DS} \ll U$ as there is a direct path to ground. Thus, we have now that for practical purposes $V_{DS2} \approx 0$ and most of the power supply voltage appears across $V_{GS1}$. The current flowing in the circuit is due to M1, as follows:

$$I_{\text{leak}} \approx I_1 \approx I_{01} \exp \left( -\frac{(V_{to} - \frac{\Delta V_{21}}{2})}{n U} + \eta V_{dd} \right) \times \left( 1 - \exp \left( -\frac{V_{dd}}{U} \right) \right) \approx I_{01} \exp \left( \frac{-V_{to} + \frac{\Delta V_{21}}{2} + \eta V_{dd}}{n U} \right).$$

(13)

This expression is similar to (12). However, we can see that the off-state current behaves oppositely due to the positive $\Delta V_{21}$ (see Fig. 5).

B. Analysis of $I_{\text{leak}}$ for $V_t$-Mismatched Parallel Transistors

Consider now the parallel connected transistors of Fig. 4(b). Let us assume that both $V_2$ and $V_2$ are at 0 V. Then the total leakage current is the sum of the leakage current of each transistor. Observe that $V_{GS1} = V_{GS2} = 0$ and $V_{DS1} = V_{DS2} = V_{dd}$. Then using (7) and (8), the total leakage current is obtained as

$$I_{\text{leak}} = I_1 + I_2 \approx \left( I_{01} \exp \left( -\frac{(V_{to} + \frac{\Delta V_{21}}{2})}{n U} + \eta V_{dd} \right) \right)$$

$$+ \left( I_{02} \exp \left( -\frac{(V_{to} - \frac{\Delta V_{21}}{2})}{n U} + \eta V_{dd} \right) \right) \times \left( 1 - \exp \left( -\frac{V_{dd}}{U} \right) \right).$$

(14)

Observe from (14) that when $\Delta V_{21} < V_{T1}$, M1 has a dominant effect because of the exponential nature of the current. The same holds for M2 when $\Delta V_{21} > 0$. The net result is that parallel networks always exhibit an increase in leakage current when there is a threshold voltage mismatch. Fig. 6 shows the corresponding simulation results.

IV. IMPACT OF INTER- AND INTRA-DIE $V_t$ VARIATIONS ON $I_{\text{leak}}$

To evaluate the implications of the theory derived in Section III-A, it is interesting to compare the results of intra- and inter-die off-state currents for stacked transistors. In-house inspection of wafer maps reveals that devices close to the wafer’s edge have lower $V_t$’s and exhibit an obvious die-to-die variation. Therefore, we used a “fast” process as reference for inter-die variations. In our technology, the “fast-process” NMOS transistors have $V_{to} = 0.48$ V. The top and bottom windows of Fig. 7 show the results of the nominal and fast processes, respectively. The curves are labeled as 1, 2, and 3 to correspond with each case study of Section III-A. We denote the intersection of the dotted vertical line with the curves of the fast process, at $\Delta V_{21} = 0$ V, as the (worst-case) inter-die off-state current. This is shown for each case through point A in Fig. 7. A couple of straightforward observations can be made from this plot. For case 1, the difference in off-state currents between the fast and nominal processes is about a factor of 20% at maximal negative mismatch. It can also be seen that this difference is greatly reduced for positive $\Delta V_{21}$. Case 2 shows a difference of about 50% between inter- and intra-die off-state currents at maximal negative mismatch, and so does case 3.

The aforementioned results are obvious and were expected. Let us now shift our attention to the plot of Fig. 8(a), which shows the intra-die variations of case 2 for both fast and nominal processes. The plot shows a current of about 150 pA for no $V_t$ mismatch for the fast process (inter-die off-state current reference); see point A. Following the horizontal trace, one can see that this current is equivalent to the off-state current of a nominal process with 85 mV of $V_t$ mismatch (about $\sigma_{\Delta}^t$, calculated
using $A_{\Delta Vt} = 7 \text{ mV} \mu\text{m}$, based on $W/L = 1.4 \mu\text{m}/0.18 \mu\text{m}$; see point $B$. Thus, the level of $I_{\text{off}}$ in a nominal process is quite severe if $V_t$ mismatch is taken into account. Although not shown, a similar behavior was observed for case 3. Let us focus now on the fast process only and take into account the positive $V_t$ mismatch. The plot shows that the off-state current doubles (from 150 pA to about 300 pA) for mismatches equal to 42 mV ($3\sigma\Delta$), calculated using $A_{\Delta Vt} = 7 \text{ mV} \mu\text{m}$. In other words, if no intra-die variations are considered the typical worst-case inter-die off-state current significantly underestimates the true value. Finally, let us investigate now the case of minimum transistor dimensions. Fig. 8(b) presents simulation results using NMOS transistors of $W/L = 0.28 \mu\text{m}/0.18 \mu\text{m}$. Except for the fact that the subthreshold current is smaller, the same trends as for a transistor of larger width can be observed. We can also see that when $\Delta V_t = 90 \text{ mV}$, the off-state current of the nominal process is comparable to the one of the fast process. However, observe that in this situation, 90 mV corresponds to a deviation of merely $2.9\sigma\Delta$, calculated using $A_{\Delta Vt} = 7 \text{ mV} \mu\text{m}$, rather than the previous $6\sigma\Delta$. In other words, the impact of $V_t$ mismatch on $I_{\text{off}}$ has almost doubled compared to the one of a transistor with dimensions $W/L = 1.4 \mu\text{m}/0.18 \mu\text{m}$.

The above examples illustrated the strong dependence of $V_t$ mismatch on modern deep-submicron designs.

V. STATISTICAL AND EXPERIMENTAL VERIFICATION OF THE OFF-STATE CURRENT

To substantiate the analytical results of Section III-A and Fig. 5, statistical simulations were carried out to verify the impact of other parameters on the off-state current based on the same circuit of Fig. 3(a) for $W/L = 1.4 \mu\text{m}/0.18 \mu\text{m}$. These simulations involve parameter variations of up to $4\sigma$ of $V_t$, $L$, $\beta$, and some sheet resistances. Results are shown in Fig. 9. Since the scatter plots follow the trends of Fig. 5, we can conclude that $V_t$ variability is indeed a dominant factor. The histograms show the distribution of off-state currents; worth noticing is a common current spread of about 40% with respect to the mean value. These large spreads make it more difficult for designers to keep the IC performance target within a $3\sigma$ tolerance window.

A. Experimental Verification of Intra-Die Variation

Leakage current measurements were carried out on six identical DSP-like cores within the same test chip. The DSPs have approximately 60,000 gates for a round total of 240,000 transistors. The experiments were done using the standard current measurement features of the Agilent 93000 tester, which has a resolution of 100 nA for 100-µA measurements. The measurement flow is simple: for each DSP, apply 75 input vectors and log the corresponding leakage current per vector; repeat the procedure 50 times to minimize measurement errors. This flow
allows us to extract an average leakage current per vector and per DSP with minimized instrumentation error. Two independent sets of 50 measurements were performed for various dies on the wafer.

Fig. 10(a) shows the normalized leakage current, sorted in ascending order, per vector and per DSP. The normalization is with respect to the average leakage current per DSP. The min and max boundaries were obtained from measuring four fault-free distinct dies. From this figure, one can observe the state dependence of the leakage current, e.g., a different current for each input vector, and a “tolerance window” with the intra-die leakage variations of the DSPs. Fig. 10(b) depicts a die for which, possibly, an excessive mismatch is present in various cells, giving origin to higher leakage currents. The excess in current is small enough to discard the possibility of low-resistance shorts. Observe that the magnitude of the outlying points is comparable to the width of the tolerance band. This excess can be attributed to a mismatch in the cells, such as the one hereby described, and from cell to cell. One can also notice points above and below the tolerance band. This can be explained by recalling that the off-state current of a mismatched cell increases or decreases depending upon the $V_T$ mismatch shift and the input state. Let us carry out some simple calculations to infer the number of mismatched cells that would result in a current out of the tolerance window. Let us assume a normalized average leakage current per cell equal to 1/60 000. Let us also assume that there is a group of cells that have a $\sigma$ deviation from their mean value, e.g., a hundredth of the average value, and also that they have the same input state. Let us now consider an outlying point with a normalized current of 1.004. A simple arithmetic calculation results in 239 gates with mismatch. This can be interpreted as an equivalent of 239 cells whose effect did not average out with all other cells and that yielded a current deviation of 0.004. The calculated number of cells actually defies the law of probabilities in the sense that if a strict normal distribution of $\Delta V_T$ is considered along with a strict uniform distribution of input states, then the number of cells equal or in excess of a $3\sigma$ deviation is only 162. Nevertheless, the doubt to consider is the possibility of abnormal process shifts, or a nonuniform switching state of the cells.
VI. EXPERIMENTAL CHARACTERIZATION OF \( V_t \) MISMATCH IN THE SUBTHRESHOLD REGIME

So far, it has been shown that the off-state current variations of a digital cell are strongly affected by random threshold voltage fluctuations. This section demonstrates that \( L_{\text{eff}} \) fluctuations in deep-submicron transistors can even be significantly larger than what has been modeled and simulated so far. This is done by actually measuring \( V_{gs} \) offset fluctuations in the subthreshold regime (near \( L_{\text{eff}} \)) as opposed to basing them on extrapolations from \( V_t \) mismatch measurements in the strong inversion regime. This approach circumvents extrapolation errors due to limitations or imperfections of the models. Moreover, it can yield additional insights into the transistor behavior as the transistor current fluctuations are not necessarily caused by the same physical phenomena in the different operation regions of the device. As mentioned in Section I, this is in contradiction to some subthreshold MOSFET matching studies reported in the literature [19], [22], but in this section we show that certain device architectures can give rise to significantly larger fluctuations in subthreshold than those measured in strong inversion.

Experimental mismatch fluctuation results, in both strong and weak inversion, were obtained using MOSFET matched-pair test structures that were fabricated in an industrial 0.18-\( \mu \)m CMOS technology. Ten populations (90 samples each) of matched pairs with a range of device dimensions (see Table II) were measured on a single wafer. Obviously, it is well accepted that any threshold voltage is a rather arbitrary physical quantity. In fact, there is no such thing as a "physical" threshold voltage. The reason the threshold voltage has become such a popular device parameter is that it can be incorporated quite conveniently into threshold-voltage-based compact models and does indeed allow quite acceptable description of the macroscopic (dc \( I-V \)) device behavior. In the subthreshold regime, however, there is certainly no sensible definition of a threshold voltage, hence we chose to characterize the mismatch fluctuation near \( L_{\text{eff}} \) by determining the \( V_{gs} \) offset between the two transistors of a pair.
TABLE II
THRESHOLD VOLTAGE MISMATCH STANDARD DEVIATIONS IN STRONG INVERSION ($\sigma_{\Delta V_t}$) AND SUBTHRESHOLD $V_{th}$ MISMATCH ($\sigma_{\Delta V_{th}}$ at $I_{dr} = 10$ pA/(W/L)) AND THE CORRELATIONS BETWEEN THE MISMATCH OBSERVATIONS FOR A RANGE OF TRANSISTOR DIMENSIONS

<table>
<thead>
<tr>
<th>drawn W (µm)</th>
<th>drawn L (µm)</th>
<th>$\sigma_{av}$ (strong inversion) (mV)</th>
<th>$\sigma_{\Delta V_{th}}$ [I=10pA/(W/L)] (mV)</th>
<th>Correlation factor $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>4</td>
<td>0.7</td>
<td>2.1</td>
<td>0.07</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>1.1</td>
<td>3.0</td>
<td>0.05</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>1.7</td>
<td>4.5</td>
<td>0.03</td>
</tr>
<tr>
<td>0.4</td>
<td>10</td>
<td>2.5</td>
<td>5.1</td>
<td>0.24</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>5.4</td>
<td>10</td>
<td>0.09</td>
</tr>
<tr>
<td>0.32</td>
<td>4</td>
<td>4.0</td>
<td>6.6</td>
<td>0.32</td>
</tr>
<tr>
<td>2</td>
<td>0.2</td>
<td>12</td>
<td>27</td>
<td>0.26</td>
</tr>
<tr>
<td>0.4</td>
<td>1</td>
<td>8.9</td>
<td>12</td>
<td>0.25</td>
</tr>
<tr>
<td>0.4</td>
<td>0.24</td>
<td>23</td>
<td>38</td>
<td>0.32</td>
</tr>
<tr>
<td>0.32</td>
<td>0.2</td>
<td>27</td>
<td>44</td>
<td>0.45</td>
</tr>
</tbody>
</table>

at a fixed drain current of 10 pA/(W/L). The relatively low target drain current results in a $V_{gs}$ that is quite close to $I_{off}$. Hence, the $V_{gs}$ offset fluctuations that are derived from these measurements are more representative for $I_{off}$ calculations than those based on extrapolations from strong inversion $V_t$ mismatches.

In this work, a standard fixed $V_{gs}$ linear region three-point $V_t$ extraction technique was used to characterize the strong inversion $V_t$ mismatch. This method generally provides a well-defined threshold voltage indicator [22], [23]. Subsequently, the same populations of matched pairs were remeasured using a current-based subthreshold $V_{gs}$ offset characterization algorithm. In this algorithm, we applied two gate voltages, respectively, 60 mV above and 60 mV below a target gate voltage that would yield approximately 10 pA/(W/L) drain current (see Fig. 11). The two measured currents with their respective gate biases above and below the target current are subsequently interpolated to a $V_{gs}$ at $I_{target}$ using logarithmic interpolation. By executing exactly the same procedure for both transistors in a matched pair, a $V_{gs}$ offset at $I_{dr} = 10$ pA/(W/L) can be determined. All measurements (in strong as well as in weak inversion) were performed with 0 V on the substrate.

The measurement system that was used is based on a standard high-precision semiconductor parameter analyzer (HP4156A) connected to a semi-automatic wafer prober (Cascade-Microtech 12k). The system is capable of characterizing transistor currents down to a level of about 10 fA without any special precautions. The short-term repeatability standard deviation for the strong inversion threshold voltage mismatch measurements is typically better than 200 µV. This is more than sufficient to assure that the threshold voltage mismatch fluctuation observations as reported in this paper are not seriously affected by measurement noise. Contrary to what is suggested in [22], the weak inversion mismatch measurement algorithm proves quite easy and performs with good repeatability on our system. The short-term repeatability standard deviation of the mismatch determined using this approach is significantly below 100 µV, which also means that the obtained weak inversion $V_{gs}$-mismatch standard deviations are not significantly affected.

**Fig. 11.** Example of 2/1 NMOSFET subthreshold characteristics. $I_{off} = I_{dr}$ at $V_{gs} = 0$ V. Filled data points are used for the interpolation. Dashed line identifies $V_{gs}$ for $I_{dr} = 10$ pA/(W/L) (20 pA for this transistor). This $V_{gs}$ is used in this work to calculate the subthreshold mismatch because it is more appropriate to quantify $I_{off}$ fluctuations compared to the $V_t$ mismatch.

**Fig. 12.** Mismatch area scaling graph. Diamonds: strong inversion (linear region) $V_t$ mismatch. Triangles: subthreshold $V_{gs}$ mismatch (at 10 pA/square). The 6 mV/µm and 12.5 mV/µm lines are estimates for the corresponding area scaling factors for the strong and weak inversion mismatch standard deviations, respectively.
by the measurement system. The resulting standard deviations for the mismatch as measured in strong inversion as well as in subthreshold for the experiment described above are listed in Table II. For all device dimensions, it is observed that the \( V_{gs} \)-mismatch fluctuations as encountered in the weak inversion regime (\( \sigma_{\Delta V_{gs}} [I_{d}=30 \text{ pA}/(W/L)] \)) are substantially larger than what would have been predicted using the strong inversion measurements. This is also depicted in Fig. 12.

The effective mismatch area scaling constants according to (2) were estimated using the devices geometries listed in Table II. For this case (n-channel devices; in a 0.18-\( \mu \)m CMOS technology), we found a quite acceptable strong inversion \( A_{\Delta V_{gs}} \)-factor of slightly below 6.5 mV\( \mu \)m. However, Fig. 12 also shows that when the same populations of transistor pairs are characterized in the subthreshold regime, the corresponding mismatch area scaling factor \( A_{\Delta V_{gs}} \) turns out to be about a factor of two larger! Another noteworthy observation from Table II is associated with the remarkably low correlation factors \( R^2 \) as given in the last column. These correlation factors are the outcome of calculating the (point-by-point) correlation between the observed \( \Delta V_T \)'s and the corresponding \( \Delta V_{gs} \)'s from all pairs. Whereas Bastos [19] reported a correlation factor of over 0.8, we found correlations that range from practically uncorrelated \( (R^2 < 0.2) \) for the larger transistor dimensions to weakly correlated \( (0.2 < R^2 < 0.5) \) for the shortest and narrowest transistors. Some examples of these correlations are depicted in Fig. 13. This figure shows scatter plots of the strong inversion threshold voltage mismatches \( \Delta V_T \) versus the corresponding \( V_{gs} \) offsets in weak inversion for the \( W/L = 0.32/0.2, 2/0.2, 2/1 \) and 10/1 populations of matched pairs. Note that the axes have different scales for each population. The vertically stretched shapes of the four clouds clearly illustrate the larger standard deviations for the \( \Delta V_{gs} \) offset observations as compared to the strong inversion \( V_T \) mismatch fluctuations. Moreover, all four clouds clearly demonstrate the weak, down to even nonexistent, correlation between the two different types of mismatch observations. As the measurement repeatability for these measurements is better than a fraction of a millivolt, we cannot but come to the conclusion that the mismatch in strong inversion is (at least partly) due to another physical cause than the one in weak inversion!

As mentioned in Section I, Denison et al. [22] and Bastos [19] observed no major differences between mismatch in the strong and weak inversion regions. In itself, this makes sense from a physical standpoint. Most models for both the threshold voltage and the subthreshold region behavior include the same physical device construction elements such as the channel doping, gate dielectric thickness, effective device dimensions, and mobility. Whereas these lead to fluctuations of the flat-band voltage and the body effect that combine to (strong inversion) threshold voltage fluctuations, the same flat-band voltage fluctuations, combined with fluctuations of the subthreshold swing (also determined by dielectric thickness and body effect), must be held responsible for the \( V_{gs} \) offset fluctuations in weak inversion [22]. One could argue, though, that microscopic device architec-
nature disturbances such as dopant fluctuations do not necessarily have exactly the same quantitative impact on the mismatch as observed in the different operation regions, but then one would at least expect a fair correlation between the two if they are caused by the same physical fluctuation cause. This, in fact, must be the reason why previous studies indeed found such good correlations between mismatch encountered in the subthreshold and strong inversion regimes. It is not the intention of this paper to suggest that these earlier works were wrong. In fact, we have observed similar agreements between subthreshold and strong inversion mismatch standard deviations for several device types in different process generations. On the other hand, it was also not the first time we encountered enhanced subthreshold voltage $V_{gs}$ offset fluctuations, as shown in Table II and Fig. 12. As this paper specifically addresses the impact of $I_{off}$ fluctuation in a 0.18-μm CMOS technology that did show these additional fluctuations for the n-channel devices, we decided to pursue this phenomenon further.

The question thus arises as to which additional device architecture fluctuation can be responsible for the much larger $V_{gs}$ offset fluctuations in the subthreshold regime compared to strong inversion $V_i$ fluctuations in the case discussed in this paper. Although this cannot be proven completely using the information provided in this paper, we believe that these significantly larger fluctuations can be attributed to an effect in modern deep-submicron technologies that is referred to as the “subthreshold hump” [23]. The subthreshold hump can often be distinguished in the MOSFET subthreshold transfer characteristic (see Fig. 14). Instead of a smooth transition from the superthreshold region to the deep-subthreshold region, an additional current contribution can be distinguished. This additional current is generally attributed to the formation of side channels along the shallow-trench-isolation (STI) edges [24]. As exemplified in Fig. 14, these side channels are usually much more visible when a substrate bias is applied to the transistor. Although the hump is not so easily discernible at a substrate bias of 0 V (where the $I_{off}$ is determined), we think that the $V_{gs}$-offset fluctuations as encountered in this work suggest that the side channels are nevertheless there and, moreover, that they fluctuate significantly! In our opinion, this also explains why we are dealing with an entirely different mismatch mechanism in subthreshold compared to the (channel dopant fluctuation dominated) conventional $V_i$ mismatch, since the contribution of the side channels to the total current is practically negligible in strong inversion.

The reported experimental subthreshold $V_{gs}$ offset enhancements imply that all discussions on the impact of $I_{off}$ fluctuations can in practice be severely aggrivated due to this additional fluctuation mechanism. It must be realized that due to the square-root area behavior shown in (2), a factor-of-four larger device area is required if a mismatch standard deviation must be reduced by a factor of two. A factor-of-two larger for $A_{\Delta V_i}$ is quite devastating in terms of area, speed, and power consumption, should the same $I_{off}$ level be required as for a transistor characterized conventionally in strong inversion. In principle, if the enhanced $V_i$ mismatch in subthreshold is indeed due to the suggested side-channel fluctuations, one could argue that the situation could even be worse. If one would assume that these side channels are independent of the channel width (as might be expected, though not conclusively confirmed in this work), their contribution to the $V_{gs}$ offset fluctuation would rather result in a $(\sqrt{L})^{-1}$ dependence. This implies that it would not help to increase the width to reduce the $I_{off}$ fluctuation. Although the standard deviations reported in Table II and Fig. 12 may suggest some enhanced $L$ dependence (particularly, the differences between the 2/0.2 and 0.4/1 populations are remarkable), there is not enough evidence in these data to substantiate this suggestion. The exact physical explanation of the observed effects is still a subject of further study. What stands firm, however, are the measurements as summarized in Fig. 12: the $V_{gs}$ offset fluctuations are definitely larger than the $V_i$ mismatch fluctuations!

**VII. Conclusion**

Through experimental and simulation results, we have evaluated the role of threshold voltage mismatch on the off-state current of digital cells. Experimental results show that $V_i$ mismatch characterized in the subthreshold regime can be significantly worse compared to other regions of operation. From the analytical results, the following can be concluded for a stack of transistors: 1) local $I_{off}$ variations of a fast process render results that are worse than the typical worst-case global variations estimated at $\Delta V_{i21} = 0$ V; 2) $I_{off}$ in a nominal process can reach the levels of its counterpart in a fast process depending upon the amount of $V_i$ mismatch; and 3) downscaling transistor sizes yields a higher dependence of $I_{off}$ on $V_i$ mismatch ($\sigma_{\Delta}$. Parallel networks will always present an increment in leakage current.

**Acknowledgment**

The authors would like to thank B. Kruseman for providing them with the experimental data on leakage current mismatch, and R. van Veen and K. van Kaam for doing the DSP leakage current measurements.
REFERENCES


José Pineda de Gyvez (S’88–M’90) received the Ph.D. degree from the Eindhoven University of Technology, Eindhoven, The Netherlands, in 1991. From 1991 to 1999, he was a Faculty Member in the Department of Electrical Engineering, Texas A&M University, College Station. Since 1999, he has been a Principal Scientist with Philips Research Laboratories, Eindhoven. He was a co-editor of IC Manufacturability: The Art of Process and Design Integration (New York: IEEE Press, 1975). His research interests are in the general areas of design for manufacturability and analog signal processing.

Dr. Pineda has been an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—PART I and an Associate Editor for Technology of the IEEE TRANSACTIONS ON SEMICONDUCTOR MANUFACTURING.

Hans P. Tuinhout received the M.Sc. degree in electrical engineering from Delft University of Technology, Delft, The Netherlands, in 1980.

Since then, he has been with Philips Research, Eindhoven, The Netherlands, on process and device characterization for silicon CMOS and BiCMOS integrated circuit technologies. His current research activities are focused on high precision de I–V measurements for characterization of device mismatch.