Open access • Journal Article • DOI:10.1109/JSSC.2006.883344

Read Stability and Write-Ability Analysis of SRAM Cells for Nanometer Technologies — Source link

Evelyn Grossar, Michele Stucchi, Karen Maex, Wim Dehaene

Institutions: Katholieke Universiteit Leuven

Published on: 30 Oct 2006 - IEEE Journal of Solid-state Circuits (Institute of Electrical and Electronics Engineers)

Read Stability and Write-Ability Analysis of SRAM Cells for Nanometer Technologies

Evelyn Grossar, Michele Stucchi, Karen Maex, *Member, IEEE*, and Wim Dehaene, *Senior Member, IEEE*

***Abstract*—SRAM cell read stability and write-ability are major concerns in nanometer CMOS technologies, due to the progressive increase in intra-die variability and V_{dd} scaling. This paper analyzes the read stability N-curve metrics and compares them with the commonly used static noise margin (SNM) metric defined by Seevinck. Additionally, new write-ability metrics derived from the same N-curve are introduced and compared with the traditional write-trip point definition. Analytical models of all these metrics are developed. It is demonstrated that the new metrics provide additional information in terms of current, which allows designing a more robust and stable cell. By taking into account this current information, V_{dd} scaling is no longer a limiting factor for the read stability of the cell. Finally, these metrics are used to investigate the impact of the intra-die variability on the stability of the cell by using a statistically-aware circuit optimization approach and the results are compared with the worst-case or corner-based design.**

***Index Terms*—Intra-die V_{th} variations, N-curve, read stability and write-ability of the SRAM cell, statistically-aware design optimization, V_{dd} scaling.**

I. Introduction

SRAM cell design has to cope with a stringent constraint on the cell area to achieve high integration density in modern system-on-chips (SoCs). This leads to choosing minimal width-to-length ratios for the SRAM cell transistors. As dimensions scale down to nanometer regime, the variations in CMOS transistor parameters, e.g., the threshold voltage (V_{th}), increase steadily [1] due to random dopant density fluctuations in channel, source and drain. Therefore, two closely placed, supposedly identical transistors, have important differences in their electrical parameters as V_{th} and make the design of the SRAM less predictable and controllable. Moreover, the stability of the SRAM cell is seriously affected by the increase in variability and by the decrease in supply voltage (V_{dd}). In the past there has been considerable effort in understanding and modeling the stability of the SRAM cell. Several analytical models of the static noise margin (SNM) have been developed to optimize the cell design, to predict the effect of parameter changes on the SNM [2] and to assess the impact of intrinsic parameter variations on the cell stability [3]. Furthermore, new SRAM cell circuit designs have been developed to maximize the cell stability for future technology nodes [4]. Little work has been published on an alternative definition of cell stability based on the SRAM cell N-curve [5]. In this paper, we analyze and model this N-curve definition and we compare it with the SNM. We demonstrate that the N-curve contains information both on the read stability and on the write-ability, thus allowing a complete functional analysis of the SRAM cell with only one N-curve (Section II). To our knowledge, this extension in using the N-curve for the write-ability is reported here for the first time. Analytical models of the N-curve metrics for the read stability and write-ability of the cell are derived in Section III by using a classical deep-submicron (DSM) transistor model. We also describe in this section the possible tradeoffs between the different N-curve metrics. Finally, these N-curve metrics are used in Section IV to investigate the impact of V_{th} variability on the cell [6], [7]; we derive new design criteria for the SRAM cell affected by intra-die V_{th} variations, based on a statistically-aware optimization approach [13]. The variability analysis is based on both 130-nm and 65-nm technology nodes.

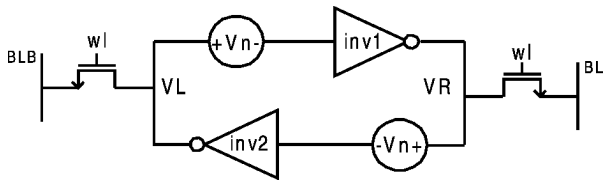

Fig. 1. The standard setup for the SNM definition is shown. The two DC noise voltage sources V_n are placed in series with the cross-coupled inverters and with worst-case polarity at the internal nodes V_r and V_l of the SRAM cell.

II. Read Stability and Write-Ability of the SRAM Cell

A. The SRAM Cell Read Stability

Data retention of the SRAM cell, both in standby mode and during a read access, is an important functional constraint in advanced technology nodes. The cell becomes less stable with lower supply voltage (V_{dd}), increasing leakage currents and increasing variability, all resulting from technology scaling. The stability is usually defined by the SNM [2] as the maximum value of DC noise voltage (V_n) that can be tolerated by the SRAM cell without changing the stored bit. In Fig. 1, the equivalent circuit for the SNM definition is shown. The two DC noise voltage sources (V_n) are placed in series with the cross-coupled inverters and with worst-case polarity at the internal nodes of the cell. Locating the smallest square between the two largest ones delimited by the eyes of the butterfly curve determines graphically the SNM (Fig. 2). When V_n is equal to the SNM, the VTCs move horizontally and/or vertically [8] until the stable point A

Manuscript received February 28, 2006; revised June 14, 2006.

E. Grossar and K. Maex are with IMEC, Leuven, B-3001 Belgium, and also with Katholieke Universiteit Leuven, Leuven, B-3001 Belgium (e-mail: grossar@imec.be; Evelyn.Grossar@imec.be; Karen.Maex@imec.be).

M. Stucchi is with IMEC, Leuven, B-3001 Belgium (e-mail: Michele.Stucchi@imec.be).

W. Dehaene is with Katholieke Universiteit Leuven, Leuven, B-3001 Belgium (e-mail: Wim.Dehaene@esat.kuleuven.be).

Digital Object Identifier 10.1109/JSSC.2006.883344

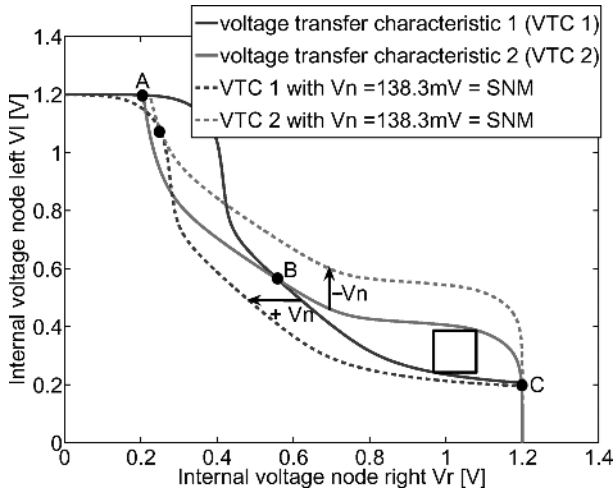

Fig. 2. The static voltage transfer characteristics (VTCs) of the two cross-coupled inverters during read access of the cell are represented by the solid curves. When the worst-case static noise is applied, the VTCs move horizontally and/or vertically until point A and B coincide (dotted curves). With more noise applied, the VTCs have only one common point C and the cell content is flipped.

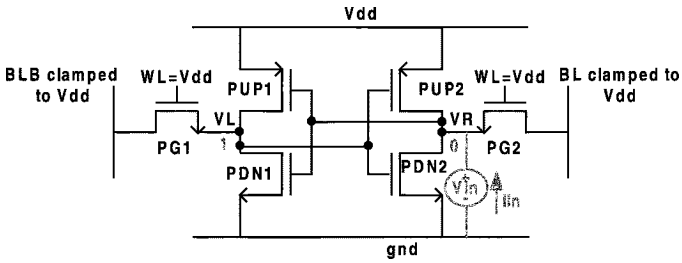

Fig. 3. For extracting the N-curve during read operation, the bit-lines are clamped at V_{dd} and the word-line is activated. Next, a voltage sweep $V_{\rm in}$ from 0 V to V_{dd} is applied at "0" internal storage node V_r to obtain the corresponding current $I_{\rm in}$.

and the meta-stable point B coincide. The cell is most vulnerable to noise during a read access since the "0" internal storage node (V_r) rises to a voltage higher than ground [2]. Due to this voltage division on V_r, the SNM is primarily determined by the ratio of the pull down (PDN) to pass gate (PG) transistor, known as the cell ratio [2]. In an ideal case, each of the two cross-coupled inverters in the SRAM cell has an infinite gain. As a result, the butterfly curves delimit a maximal square side of maximum $0.5V_{dd}$, being an asymptotical limit for the SNM. Therefore, V_{dd} scaling limits the stability of the cell. An additional drawback of the SNM is the inability to measure the SNM with automatic inline testers [5], due to the fact that after measuring the butterfly curves of the cell the static current noise margin (SINM) still has to be derived by mathematical manipulation of the measured data. An alternative definition for the SRAM read stability is based on the N-curve of the cell [5], which is measurable by inline testers. The combined voltage and current information provided by the N-curve allows to overcome the limitations of V_{dd} scaling described for the SNM, as shown in the following paragraphs. For extracting the N-curve (Fig. 3), the bit-lines are both clamped at V_{dd} and the word-line is activated to put the cell in read operation mode. A voltage sweep $V_{\rm in}$ from 0 V to V_{dd} is applied at the "0" internal storage node V_r and the corresponding current $I_{\rm in}$ is measured. In three points A, B and C of

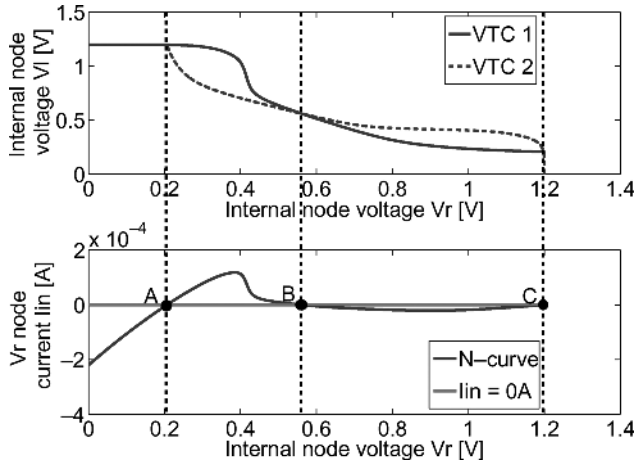

Fig. 4. The N-curve and the butterfly curve of the cell, both obtained by simulation, are shown. The three points A, B, and C correspond to the two stable points A and C and the meta-stable point B of the butterfly curve. The voltage in A is determined by the PDN to PG ratio or cell ratio while the voltage in B is related to the PDN to PUP ratio and PG of the cell. The voltage in C is defined by the PUP to PG ratio or the pullup ratio of the cell.

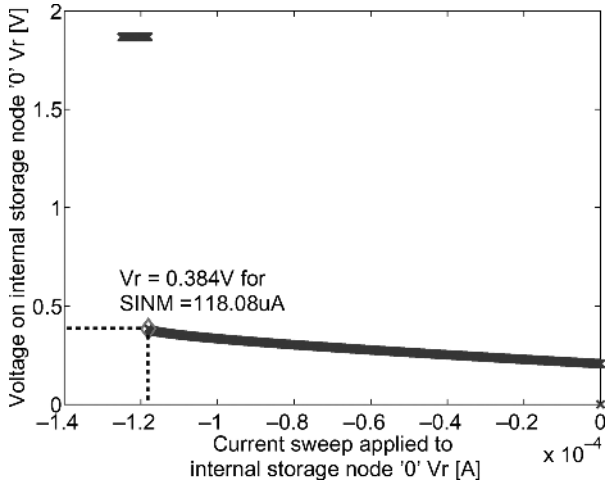

Fig. 5. A sweep of current values is applied to "0" internal storage node V_r to demonstrate that the SINM the maximal DC noise current is that can be tolerated by the SRAM cell before it changes state.

the N-curve (Fig. 4), the current injected in V_r is zero; A and C correspond to the two stable points of the butterfly curve while B corresponds to the meta-stable point. When points A and B coincide, the cell is at the edge of stability and a destructive read can easily occur. The voltage difference between point A and B indicates the maximum tolerable DC noise voltage at V_r of the cell before its content changes. This voltage metric is the static voltage noise margin (SVNM). The additional current information provided by the N-curve, namely the peak current located between point A and B, can also be used to characterize the cell read stability. This current metric is the static current noise margin (SINM). It is defined as the maximum value of DC current that can be injected in the SRAM cell before its content changes (Fig. 5). By using the combined SVNM and SINM, the read stability criteria for the cell are defined properly. For example, a small SVNM combined with a large SINM will still result in a stable cell since the amount of required noise charge to disturb the cell is large. Therefore, a comparison between the

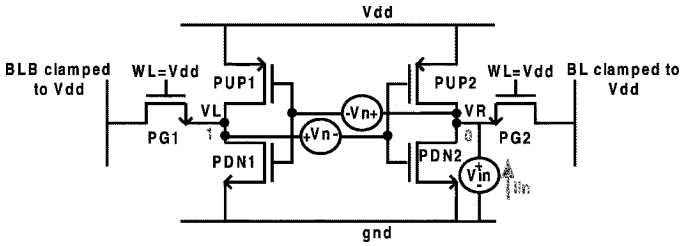

Fig. 6. The setup for comparing the usual SNM definition and the N-curve stability metrics SVNM and SINM.

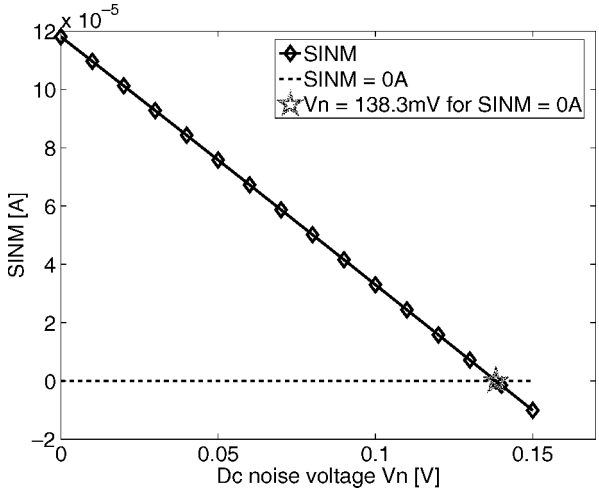

Fig. 7. The SINM is simulated for a sweep of different DC noise voltages V_n (Fig. 6). It can be seen that the SINM is equal to 0 A when V_n is equal to 138.3 mV, which is the SNM of that particular SRAM design. Similar reasoning can be done for the SVNM definition.

usual SNM definition and the N-curve metrics is imperative. For this purpose, the following experimental simulation setup has been used (Fig. 6). Since V_n is the maximum tolerable DC noise voltage before the cell changes state, the SINM is equal to 0 A when V_n is equal to the SNM value of that particular cell design, which is confirmed by simulation (Fig. 7). Both the SNM and SINM thus represent the same stability criteria for the SRAM cell. Similar reasoning can be done for the SVNM definition. However, two different SRAM cells can have identical values for both the SNM and SVNM, but this does not mean that they are equally stable. To verify this statement, a comparison based on the N-curve read stability metrics and the SNM definition is made in Table I for two different designs in 130-nm IMEC platform technology. The designs use very long and wide transistors to avoid deep-submicron effects. The relative ratio between the transistors in the SRAM cell is kept the same for the two cell designs, but for case 2 the widths are sized up with a factor 2. The transistors of the second SRAM cell have an improved on current and therefore this cell should be more stable. This difference is only visible in the factor 2 improvement of the SINM of cell 2 with respect to cell 1, while the SNM and SVNM values are the same for both cell designs. This confirms that wrong conclusions can be drawn for the read stability of the cell, if no current information is taken into account. Furthermore, V_{dd} scaling no longer limits the SRAM cell stability to the ideal value of $0.5V_{dd}$. In fact, the SINM can be improved by

TABLE I
COMPARISON OF THE USUAL SNM AND THE N-CURVE READ STABILITY METRICS SVNM AND SINM FOR TWO DIFFERENT CELL DESIGNS

	SNM [mV]	SVNM [mV]	SINM [mA]
Case 1	253.1	429	0.296
Case 2	253.5	428.8	0.586

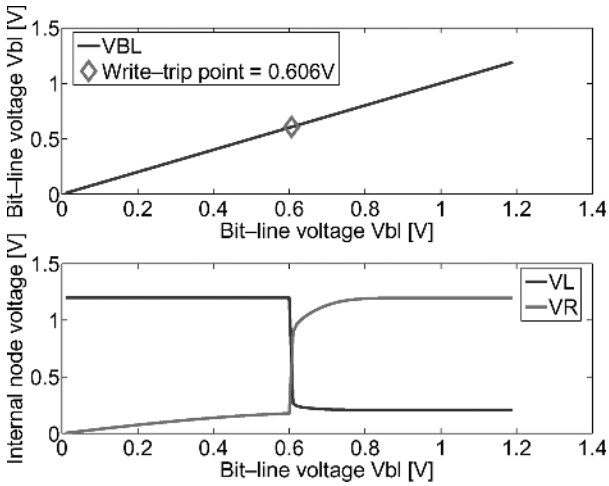

Fig. 8. The usual write margin of a SRAM cell is defined by the write-trip point. This is the maximum bit-line voltage, needed to flip the state of the cell.

upsizing the transistors in the cell, thus allowing to compensate for the detrimental effect of V_{dd} scaling. In conclusion, a correct analysis of the read cell stability requires both the N-curve metrics SVNM and SINM.

B. The Write-Ability of the SRAM Cell

Besides the read stability for the SRAM cell, a reasonable write-trip point [10] is equally important to guarantee the write-ability of the cell without spending too much energy in pulling down the bit-line voltage to 0 V. The write-trip point defines the maximum voltage on the bit-line, needed to flip the cell content (Fig. 8). The write-trip point is mainly determined by the pull-up ratio of the cell while the read stability is determined by the cell ratio of cell; this results in the well-known conflicting design criteria [9]. The SRAM N-curve can also be used as alternative for the write-ability of the cell, since it gives indications on how difficult or easy it is to write the cell. As discussed, the simulated N-curve in Fig. 4 refers to the read operation of the cell. For the write operation, pulling down the bit-line to ground discharges the "1" internal node V_l. Therefore, the N-curve is now analyzed from right to left starting from point C where the internal storage node V_r is 1. The negative current peak between point C and B or the write-trip current (WTI) is the amount of current needed to write the cell when both bit-lines are kept at V_{dd}. This is the current margin of the cell for which its content changes (Fig. 9). The ability to write a cell with both bit-lines clamped at V_{dd} results actually in a destructive read operation; therefore, the absolute value of WTI should be large enough to cope with the read stability requirement. On the other hand, the lower the absolute WTI is, the higher the write-trip point of the cell. Similarly, the voltage difference between point C and B or the write-trip voltage (WTV) is the voltage drop needed to flip the internal node "1" of the cell with both the bit-lines clamped

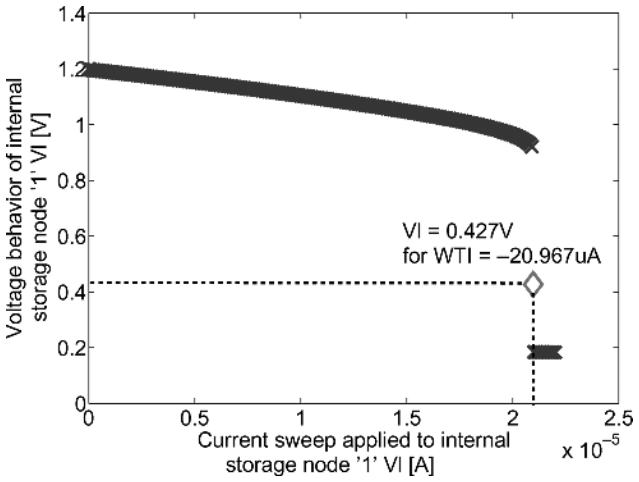

Fig. 9. To demonstrate that the write-trip point definition and the new N-curve write-ability metrics of the cell both define the same write margin of the cell, a sweep of current values is applied to "1" internal storage node V_l. It is shown that at the WTI value the cell changes state.

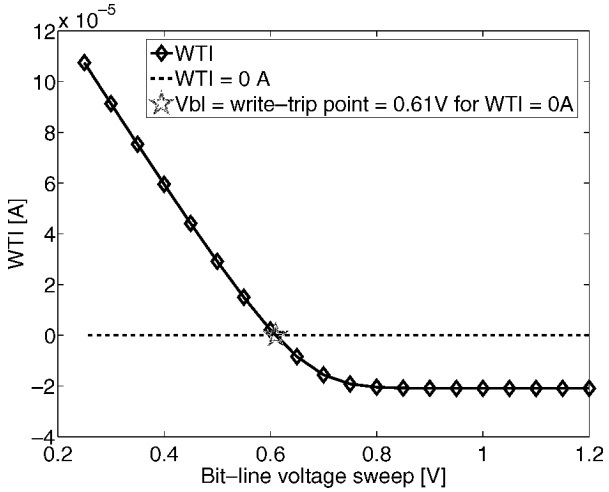

Fig. 10. The WTI is simulated for a sweep of bit-line values (Fig. 3). It can be seen that the WTI is equal to 0 A when the bit-line voltage is equal to the write-trip point. Similar reasoning can be done for the WTV definition.

TABLE II
COMPARISON OF THE USUAL WRITE-TRIP POINT DEFINITION AND THE NEW N-CURVE WRITE-ABILITY METRICS WTV AND WTI FOR TWO DIFFERENT CELL DESIGNS

	Write-trip point [mV]	WTV [mV]	WTI [uA]
Case 1	366.5	630.4	-93.8
Case 2	366.6	631	-186.1

at V_{dd}. Again, both metrics should be equivalent. According to the write-trip point definition, WTI should be zero when the bit-line is pulled down to the write-trip point value. This is verified in simulation by pulling down the bit-line voltage to perform a write operation. Fig. 10 confirms that WTI is equal to 0 A when the bit-line voltage drops to the write-trip point (Fig. 8). Similar reasoning can be done for the WTV definition. By considering the two different SRAM cells used for comparing the read stability metrics, the same write-trip point and the same WTV but a different WTI is found (Table II). Determining the write-ability requires both the WTV and WTI. Increasing the transistor widths of cell degrades the WTI while it improves the SINM, thus confirming the conflicting constraints between the read and write operation of the cell.

III. ANALYTICAL DERIVATION OF THE N-CURVE METRICS

A. Assumptions and Analytical Expressions

The N-curve of the SRAM cell can be expressed analytically by solving Kirchoff's current law at both V_l and V_r nodes of the SRAM cell. During the input voltage sweep at V_r the operation regions of the SRAM cell transistors are changing continuously, resulting in the N-curve. For $V_{\rm in}$ equal to 0 V, PG2 is in velocity saturation region and its drain current I_{pg2} is therefore larger than PDN2, which is in linear region. According to Kirchoff's current law, the rest of the PG2 current flows into the input voltage source current to keep V_r at 0 V. In point A of Fig. 4, the currents of PG2 and PDN2 are equal, thus resulting in $I_{\rm in} = 0$ A. Between point A and the SINM, the operation regions of the devices are not changed, but as $I_{\rm in}$ increases, the drain current in PG2 is now smaller than in PDN2, as indicated by the change in sign of $I_{\rm in}$. At SINM, PDN2 moves from linear to velocity saturation region. Between SINM and WTI, PUP2 is now active and the working regions of PG2, PDN2, and PUP2 all move to saturation region. At WTI, both the PG2 and PUP2 are in linear region while PDN2 moves from active to non-active working region. With this analysis, the analytical expressions of $I_{\rm in}$ in the neighborhood of SINM and WTI can easily be derived.

1) Analytical Expression for the N-Curve: An explicit expression of $I_{\rm in}$ is then formulated at both V_l and V_r, respectively:

$$I_{pg1} + I_{pup1} = I_{pdn1} \tag{1}$$

$$I_{\rm in} + I_{pg2} + I_{pup2} = I_{pdn2}. \tag{2}$$

The subscripts correspond to the transistors in Fig. 3. This notation will be used in the remainder of this paper to identify any device parameter relative to a specific transistor of the SRAM cell. Equation (1) yields the voltage behavior on V_l, which can then be plugged into (2) to yield $I_{\rm in}$. The equations of the classical DSM transistor model are used [9] with some additional changes. The drain-induced-barrier lowering (DIBL) effect is included in all the operation regions of the transistor and represents the dependency of V_{th} on V_{dd} [11]. On the other hand, the body effect is neglected, since in nanometer technologies the V_{th} control by the back gate bias is no longer effective [12]. Additionally, smoothing of the transition regions is implemented to remove discontinuities. Due to their complexity, the equations are solved numerically. Fig. 11 shows a good agreement between the HSPICE simulations [23] and the correspondent analytical solutions, obtained with transistor models calibrated in 130-nm IMEC platform technology.

2) Analytical Expression for the SINM and the WTI: In the neighborhood of the SINM (Fig. 4), the gate-source voltage V_{gs} of PG1 is below V_{th}, so I_{pg1} can be neglected in (1), while PUP1 operates in linear region. Equation (1) is in this way simplified to (3). PDN1 changes from subthreshold to velocity saturation region. In (2), I_{pup2} is neglected since V_{gs} is approximately equal to V_{th}. PG2 operates in velocity saturation and PDN2 moves

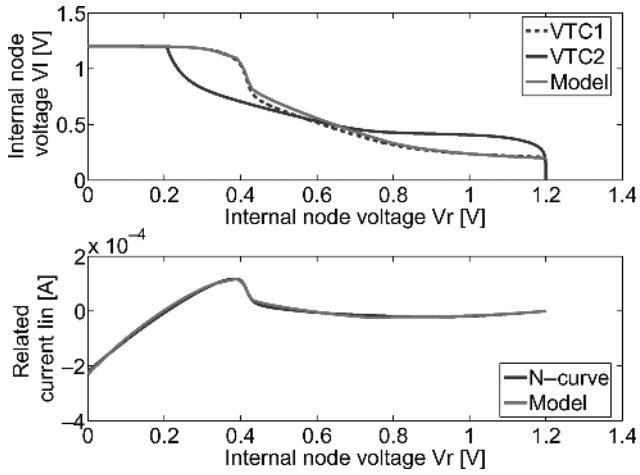

Fig. 11. Good agreement is observed between the HSPICE simulations and the analytical N-curve model, which is solved numerically due to the complexity of the classical DSM transistor model. To avoid discontinuities, smoothing of the transition regions is implemented in DSM transistor model.

from linear to velocity saturation region. Equation (2) is then simplified to (4):

$$I_{pup1} \cong I_{pdn1} \tag{3}$$

$$I_{\rm in} + I_{pg2} \cong I_{pdn2}. \tag{4}$$

Similar reasoning in the neighborhood of the WTI results in the following simplified equations:

$$I_{pg1} \cong I_{pdn1} \tag{5}$$

$$I_{\rm in} + I_{pup2} \cong 0. \tag{6}$$

For first-order analytical modeling of SINM and WTI, the approximations assumed are reasonable and the equations of the classical DSM transistor model are used [9], but with some simplifications. The channel length modulation effect (CLM) is always considered in velocity saturation region and only for saturation and linear region when a analytical solution is feasible. Additionally, no curve smoothing is implemented with the purpose of simplifying the analytical solution. Two solutions for (3) are found for the SINM. Solution I (Vl_{1a}) corresponds to PDN1 in subthreshold region and solution II (Vl_{1b}) corresponds to PDN1 in velocity saturation. (See the first set of equations at the bottom of the page.) The correspondent analytical solutions in the neighborhood of SINM then become

$$I_{in1a} = I_{ds,pdn2a} - I_{ds,pg2} \tag{7}$$

$$I_{in1b} = I_{ds,pdn2b} - I_{ds,pg2} \tag{8}$$

$$\begin{aligned} I_{ds,pdn2a} = {} & \beta_{pdn2a}\left(Vl_{1a} - (V_{th,pdn2} - \eta_{pdn2}V_{\rm in}) - 0.5V_{\rm in}\right) \\ & \times V_{\rm in}(1 + \lambda_{pdn2}V_{\rm in}) \end{aligned} \tag{9}$$

$$\begin{aligned} I_{ds,pg2} = {} & \beta_{pg2}\left((V_{dd} - V_{\rm in}) - (V_{th,pg2} - \eta_{pg2}(V_{dd} - V_{\rm in})\right. \\ & \left. -0.5V_{dssat,pg2}\right) V_{dssat,pg2} \\ & \times (1 + \lambda_{pg2}(V_{dd} - V_{\rm in})). \end{aligned} \tag{10}$$

$I_{ds,pdn2b}$ is obtained by replacing β_{pdn2a} and Vl_{1a} in (9) with β_{pdn2b} and Vl_{1b}, respectively. The SINM is then obtained by setting the derivative of $I_{\rm in}$ to $V_{\rm in}$ to zero:

$$\frac{\partial I_{\rm in}}{\partial V_{\rm in}} = 0.$$

By solving the simplified analytical equations (5) and (6), the voltage solution is as shown in the second set of equations at the bottom of the page. The analytical solution in the neighborhood of WTI becomes

$$\begin{aligned} I_{in2} = {} & -I_{ds,pup2} \\ I_{ds,pup2} = {} & \beta_{pup2}\left((V_{dd} - Vl_2) - (V_{th,pup2} - \eta_{pup2}(V_{dd} - V_{\rm in})\right. \\ & \left. -0.5(V_{dd} - V_{\rm in})\right)(V_{dd} - V_{\rm in}). \end{aligned}$$

$$\begin{aligned} Vl_{1a} = {} & V_{dd} \\ Vl_{1b} = {} & \frac{\left(A + B - \sqrt{(A-B)^2 - C\begin{pmatrix} \beta_{pup1}V_{dd}\left(V_{dd}(1+2\eta_{pup1}) - 2(V_{\rm in} + V_{th,pup1})\right) \\ -2\beta_{pdn1}V_{dssat,pdn1}(V_{\rm in} + V_{dssat,pdn1} - V_{th,pdn1}) \end{pmatrix}}\right)}{-0.5C} \\ A = {} & -2\beta_{pup1}(V_{\rm in} - V_{dd}\eta_{pup1} + V_{th,pup1}) \\ B = {} & 2\beta_{pdn1}\eta_{pdn1}V_{dssat,pdn1}(1 + V_{\rm in} - V_{th,pdn1} - 0.5V_{dssat,pdn1}) \\ C = {} & -4\left(-\beta_{pup1}(1+2\eta_{pup1}) - 2\beta_{pdn1}\eta_{pdn1}\lambda_{pdn1}V_{dssat,pdn1}\right) \end{aligned}$$

$$\begin{aligned} Vl_2 = {} & \frac{\left(A + B + \sqrt{(-A-B)^2 - 4C\begin{pmatrix} -2V_{dd}\beta_{pg1}V_{dssat,pg1}(1+\eta_{pg1}) \\ +\beta_{pg1}V_{dssat,pg1}(V_{dssat,pg1} + 2V_{th,pg1}) \end{pmatrix} \cdot (1 + \lambda_{pg1}V_{dd})}\right)}{2C} \\ A = {} & -\beta_{pg1}V_{dssat,pg1}\left((1+\eta_{pg1})(1 + 4V_{dd}\lambda_{pg1}) + \lambda_{pg1}(V_{dssat,pg1} + 2V_{thpg1})\right) \\ B = {} & 2\beta_{pdn1}(V_{th,pdn1} - V_{\rm in}) \\ C = {} & -\beta_{pdn1}(1 - 2\eta_{pdn1}) - 2\beta_{pg1}\lambda_{pg1}V_{dssat,pg1}(1 + \eta_{pg1}) \end{aligned}$$

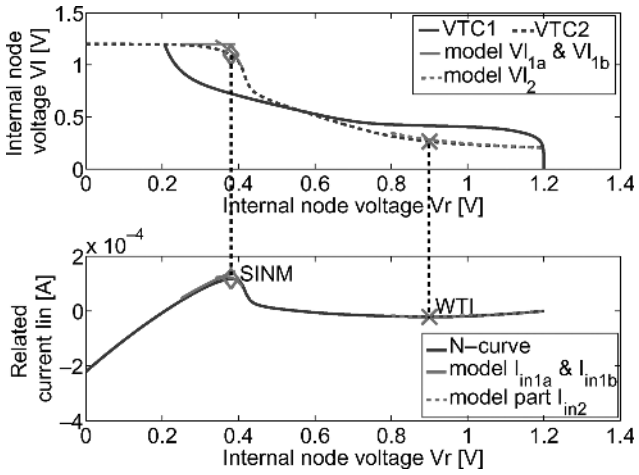

Fig. 12. Analytical modeling of the N-curve is shown in the neighborhood of the SINM and WTI.

Fig. 12 shows the comparison between HSPICE simulation of the N-curve in the neighborhood of the SINM and WTI and the analytical model. The maximum deviation between the model and the simulation is smaller than 10%. The SVNM and WTV can be modeled in a similar way by approximating (1) and (2) in the neighborhood of points A and B and points B and C, respectively.

B. Analysis of N-Curve Metrics With Respect to V_{dd}, Cell Ratio, and Pull-Up Ratio

When looking at the N-curve definitions for the read stability and write-ability of the SRAM cell, some interesting general conclusions can be drawn with respect to V_{dd}, cell ratio, and pull-up ratio of the cell. First, as shown in Fig. 13(a) and (b), the read stability N-curve metrics degrade with decreasing V_{dd}. Therefore, the stability of the cell is still limited by V_{dd} scaling, but, as mentioned above, V_{dd} is no longer the limiting factor. The SINM can be improved for lower V_{dd} by increasing the transistor widths, which, of course, is at the expense of area. Second, Fig. 14(a) shows that an increase in cell ratio still positively affects the read stability. To avoid a destructive read operation, the SVNM value should be as large as possible. It is clear that for this particular SRAM design, the SVNM value is higher than $0.25V_{dd}$ for cell ratios higher than 1.2 [Fig. 14(a)]. For cell ratios below 1.2, the WTV is below $0.5V_{dd}$, which means that write-ability of the cell is higher, when both bit-lines are clamped to V_{dd}. In general, the increase in cell ratio degrades the write-ability of the cell due to the increase in WTV as shown in Fig. 14(a). The SINM improves significantly [~50%, Fig. 14(b)] due a cell ratio increase from 1.33 to 2. When comparing Fig. 14(a) and (b), it is evident that increasing the PG and therefore decreasing the cell ratio will affect the SINM more than the SVNM; moreover, this loss in SINM can be traded off with almost the same gain in read current of the cell. Additionally, Fig. 14(b) clearly shows the conflicting needs; while the SINM improves significantly, the WTI degrades strongly. For good write-ability of the cell, the internal storage node "1" should be pulled down below $0.5V_{dd}$. The smaller the WTV, the faster the cell is written. In this particular design, WTV will be

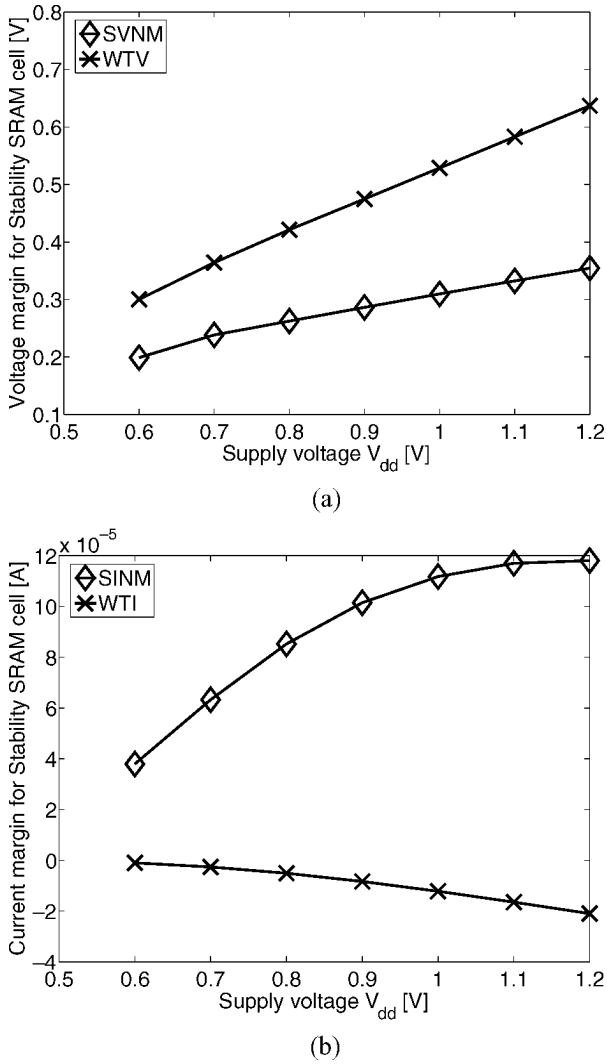

Fig. 13. (a) SVNM and WTV metrics of the SRAM cell versus the supply voltage V_{dd}. (b) SINM and WTI metrics of the SRAM cell versus the supply voltage V_{dd}. All the N-curve metrics degrade with lower V_{dd}.

lower than $0.5V_{dd}$ for pull-up ratios smaller than 1 [Fig. 14(c)]. When decreasing the pull-up ratio from 1 to 0.8, the WTI improves significantly (~30%), while the SINM slightly degrades [Fig. 14(d)].

C. Tradeoffs Between the Different N-Curves of the SRAM Cell

After investigating the N-curve definitions for the read stability and write-ability of the SRAM cell with respect to V_{dd}, cell ratio, and pull-up ratio, it is also valuable to look at the tradeoffs between these different metrics. For this purpose, a new composed metric is introduced, which includes both the voltage and current information for the read stability or write-ability of the cell. For the read stability, the area below the N-curve between point A and B (Fig. 4) is considered. Since it has formally a unit of power, it is called the static power noise margin (SPNM) [5]. Similarly, the write-ability metric of the cell can be defined as the write trip power (WTP) and is equal to the area above the curve between point B and C (Fig. 4). These new power metrics contain both the current and voltage margins, so they should be suitable to compare the stability and write-ability

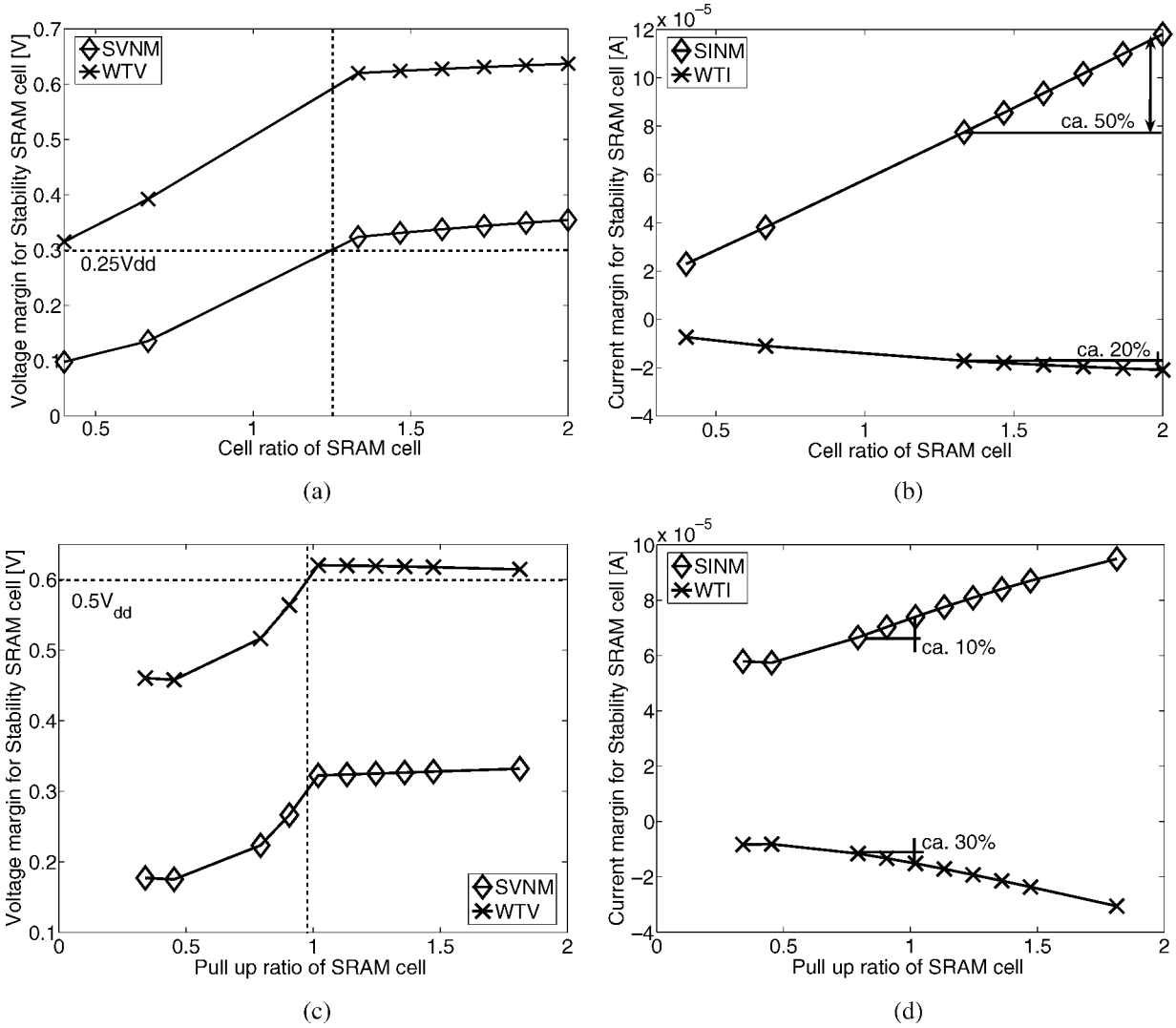

Fig. 14. (a)–(b) An increase in cell ratio improves the read stability of the cell, while the write-ability of the cell degrades. The positive impact of the cell ratio on the read stability of the cell is more significant for the SINM. (c)–(d) An decrease in pull-up ratio improves the write-ability of the cell but due to the conflicting read and write needs the read stability degrades.

of different SRAM cells. As mentioned before, between two different cell designs, having the same SVNM but different SINM, the one with the highest SINM ensures the most stable cell. The same reasoning can be done for the write-ability metrics of the N-curve. However, when two different cell designs have both a different SVNM and SINM, one should compare the unit of power SPNM for the read stability of the cell. For example, if cell 1 has a low SVNM and a high SINM, while cell 2 has a high SVNM and a low SINM, then it is the cell with the highest SPNM that is more stable. On the other hand, it is also important to look at the other performance metrics of the cell, since the cell with the highest SPNM can have a low SVNM metric, resulting in higher WTV metric and therefore degrading the write-ability of that cell. For the write-ability of the cell it is preferable to have a small WTV and a small absolute WTI, which means the cell design with the smallest WTP. It should be notified that the risk of obtaining an unstable cell rises by choosing the cell with the smallest WTP. Of course, the SPNM or the WTP metric could be the same for both cell designs. The circuit designer has to compare in this case the cell not only on the read stability and on the write-ability but also on the other performance parameters (e.g., read current, retention currents, etc.). Finally, the analysis of the N-curve metrics is completed in the next section by using them for statistically optimizing the cell in the presence of variability, as this is a growing concern in nanometer technologies.

IV. Statistical Optimization of the SRAM Cell

SRAM cell design uses minimum-sized transistors and the performance parameters depend heavily on matched transistors. In this paper, only V_{th} mismatch is considered because of the significant impact of this parameter on the stability [14], the standby leakage power [20], and the access time of the cell. Intra-die V_{th} variations are further worsened by random dopant variations in the channel region of the device [1]. The extended Pelgrom model with gate length and width dependency [21] is used here for the estimation of the magnitude of the V_{th} mismatch. Extension to any other transistor parameter can easily be done. The increasing variability and the increasing leakage currents, together with the conflicting read and write constraints and the limited cell area make the cell design very difficult. In

[13], we proposed a statistically-aware optimization approach to tackle this problem by optimizing the design statistically for a given yield target. This approach, originally proposed in [19] for digital circuits, uses the statistical information during the optimization of the circuit for speed, energy and yield. The probability density functions (pdfs) of the performance parameters are considered and statistical sensitivities are used to guide the optimization of the leakage power of the cell. The cell is optimized for minimal standby leakage power by using the dual-V_{th} optimization approach. Also in [13], we indicated that the statistically obtained results reduce the problem of over-design significantly, which is a major issue in the worst-case optimization approach, even in the enhanced worst-case optimization approach [24]. In this paper, we apply this approach both in 130-nm and 65-nm technology, taking the N-curve read stability and write-ability metrics as the functionality constraints. Since the read current ($I_{\rm read}$) of the weakest SRAM cell mainly limits the access time of the SRAM memory, as long as the bit-line capacitance is dominating, the read current is used as first approximation for the delay constraint during the read operation. The total leakage of the SRAM cell is the sum of different contributions [16], namely, the subthreshold leakage current $I_{d,sub}$, the gate leakage current, and the band-to-band-tunneling leakage current. Since $I_{d,sub}$ has an exponential relation to the V_{th} of the transistor [11], only this component is considered here. Next, two important steps of this statistically-aware optimization approach are applied.

A. Statistical Definition of the Performance Parameters of the SRAM Cell

Performance parameters are formulated statistically to take the variability into account. In other words, for each target value of the performance parameters (e.g., SINM, WTI, etc.) the corresponding performance yield target is formulated. The SINM of the SRAM cell depends on the variations of V_{th} of all six transistors. Each V_{th} is considered as an independent random variable with a Gaussian distribution defined by mean η and variance σ^2. Consequently, the mean $(\eta_{\rm SINM})$ and variance $(var_{\rm SINM})$ of the random variable SINM can be estimated by applying the Taylor series theorem [15]:

$$\eta_{\rm SINM} = {\rm SINM} + 0.5\left[\sum_{i=1}^{2}\left(\frac{\partial^2 {\rm SINM}}{\partial V_{th,pg_i}^2}\sigma^2_{V_{th,pg_i}} + \frac{\partial^2 {\rm SINM}}{\partial V_{th,pdn_i}^2}\sigma^2_{V_{th,pdn_i}} + \frac{\partial^2 {\rm SINM}}{\partial V_{th,pup_i}^2}\sigma^2_{V_{th,pup_i}}\right)\right] \quad (11)$$

$$var_{\rm SINM} = \sum_{i=1}^{2}\left[\left(\frac{\partial {\rm SINM}}{\partial V_{th,pg_i}}\right)^2\sigma^2_{Vth,pg_i} + \left(\frac{\partial {\rm SINM}}{\partial V_{th,pdn_i}}\right)^2\sigma^2_{Vth,pdn_i} + \left(\frac{\partial {\rm SINM}}{\partial V_{th,pup_i}}\right)^2\sigma^2_{Vth,pup_i}\right]. \quad (12)$$

The same reasoning is applied for the other performance parameters of the SRAM cell. The pdfs of the N-curve metrics are shown in Fig. 15(a)–(d) together with their distributions obtained by Monte Carlo simulations. The Gaussian distributions assumed for modeling the pdfs match well with the results of the simulations, which are based on 130-nm technology. Although it is clear that the Gaussian modeling is not equally accurate for all the N-curve metrics, it is good enough for designing the cell. For example, the WTI metric in Fig. 15(d) shows a long tail at the right side of the distribution. The samples in this tail represent SRAM cell designs that require more energy to write the cell due to a much lower write-trip point corresponding to the low negative WTI values. The intra-die V_{th} variations cause an asymmetrical behavior of the SRAM cell, resulting in different performance parameters with respect to V_l and V_r. The actual SINM is then the minimum of two SINMs obtained from the two N-curves of the cell at V_l and V_r, respectively. According to order statistics and assuming that the two SINMs are independent identically distributed random variables, in this case ${\rm SINM}_l$ and ${\rm SINM}_r$, the pdf $(f_{\min({\rm SINM}l,{\rm SINM}r)})$ for the minimum of the two SINMs distributions yields [17]

$$f_{\min({\rm SINM}l,{\rm SINM}r)} = 2f_{\rm SINM}(1 - F_{\rm SINM})$$

with $F_{\rm SINM}$ the cumulative distribution function (cdf) of either of the two SINMs. In reality, although ${\rm SINM}_l$ and ${\rm SINM}_r$ are normally distributed, they are not independent since they originate from the same six transistors. Therefore, any deviation in V_{th} for one of the transistors affects both SINMs (Fig. 16). The assumption of independence is a first-order approximation.

B. Optimization of the Cell Guided by Statistical Sensitivities

As well as considering the pdfs of the performance parameters, it is also useful to evaluate the improvement in performance yield, e.g., SINM yield of the cell, during the optimization. In fact, this statistical sensitivity gives information about the direction and the magnitude in which a design parameter, in our case the transistor width, has to move to improve the performance yield. The sensitivity of the performance yield to the transistor width is derived as the slope of the yield improvement graph (Fig. 17) [18]. Each point corresponds to the percentage of samples of the SINM distribution which fulfills the design target of 100 μA. For this particular SRAM design, this graph shows the different sensitivities of the SINM yield to the increasing widths of the PUP, PDN and PG transistors. In particular, an increase in PUP width increases the SINM yield to the maximum value, an increase in PDN width slightly increases the SINM yield and an increase in PG width will drastically decrease the SINM yield. This information can then be used to guide the optimization of the circuit design [7].

C. Results

The intra-die V_{th} variations used for the 130-nm results are estimated on data obtained from measurements on test structures. For 65 nm, the estimation of the variations is based on [22]. For the 130-nm technology, the analytical modeling explained above is used for the statistical results. Fig. 18(a) shows the optimized leakage power of the cell versus a range of SINM targets for a given SVNM and read current target.

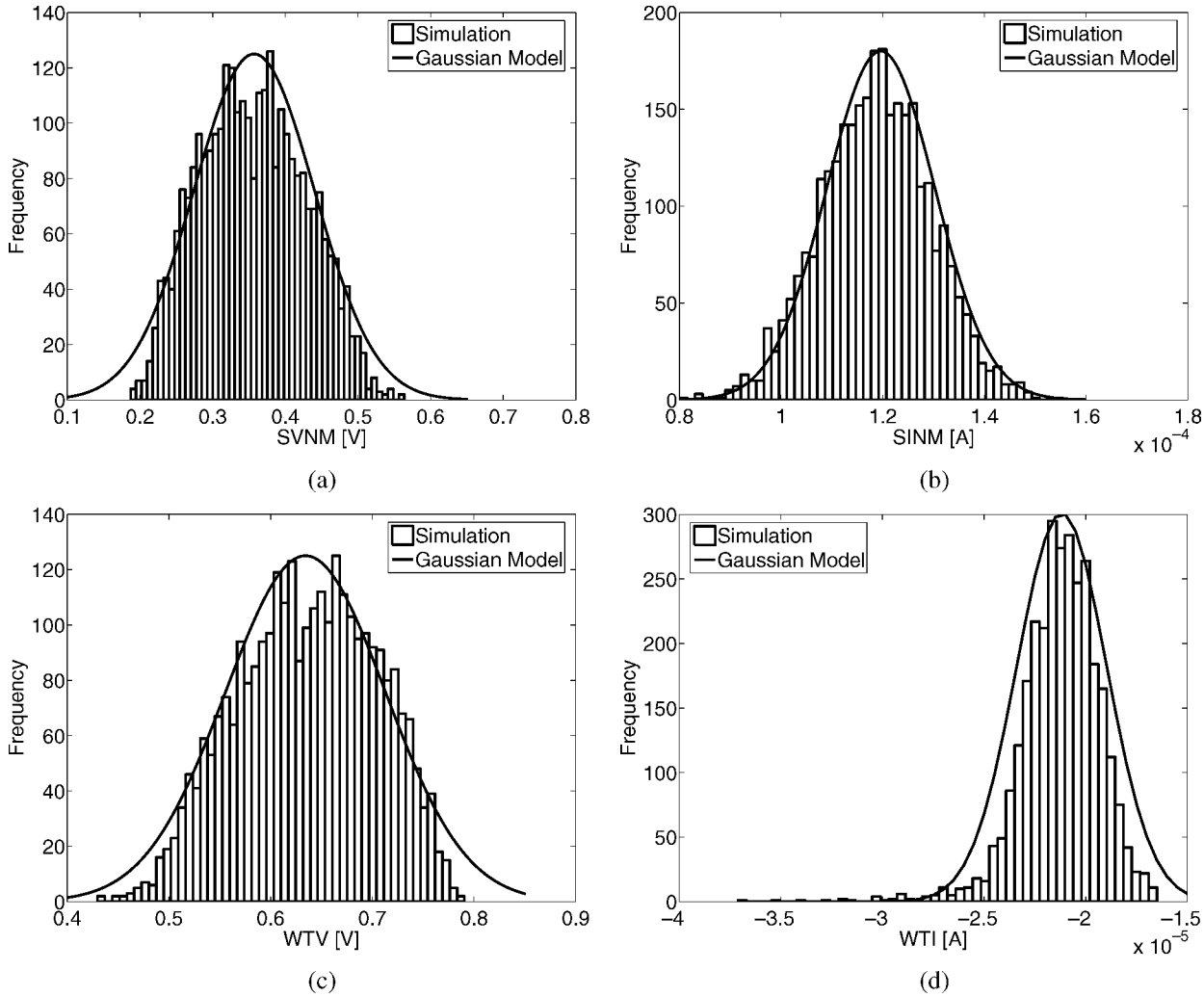

Fig. 15. The estimated Gaussian distributions match well with the simulated distributions. (a) Modeling of the SVNM distribution. (b) Modeling of the SINM distribution. (c) Modeling of the WTV distribution. (d) Modeling of the WTI distribution.

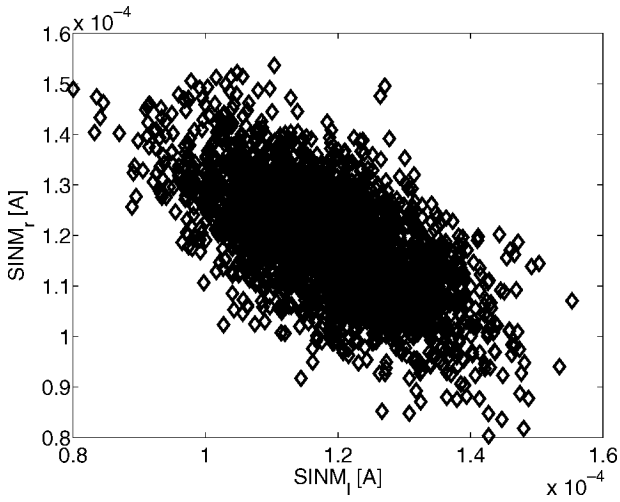

Fig. 16. Dependency between SINM_l and SINM_r of the SRAM cell.

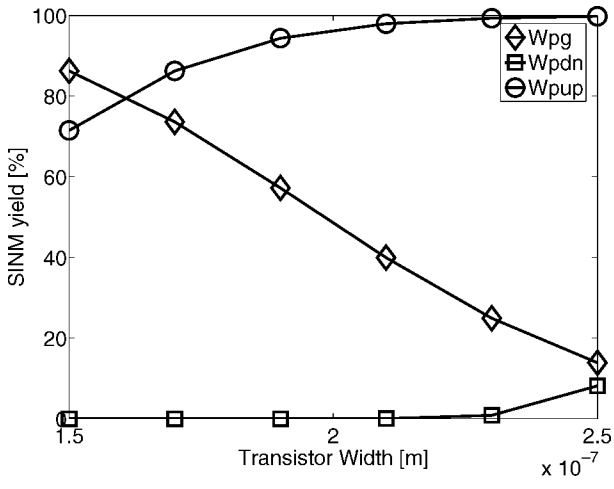

Fig. 17. The SINM yield improvement or degradation for a particular SRAM design is shown versus the transistor width. The increase in PUP width increases the SINM yield to 100% while the increase in PDN width slightly improves the SINM yield, and an increase in PG width degrades the SINM yield.

The correspondent nominal SINM and the area penalty are represented in Fig. 18(b) and (c), respectively. Fig. 19(a) includes the optimized leakage power results, simulated in 65-nm technology with supply voltage of 1 V, versus SINM targets for a SVNM target of 100 mV. The correspondent nominal SINM values and area penalty of the optimized results are shown in

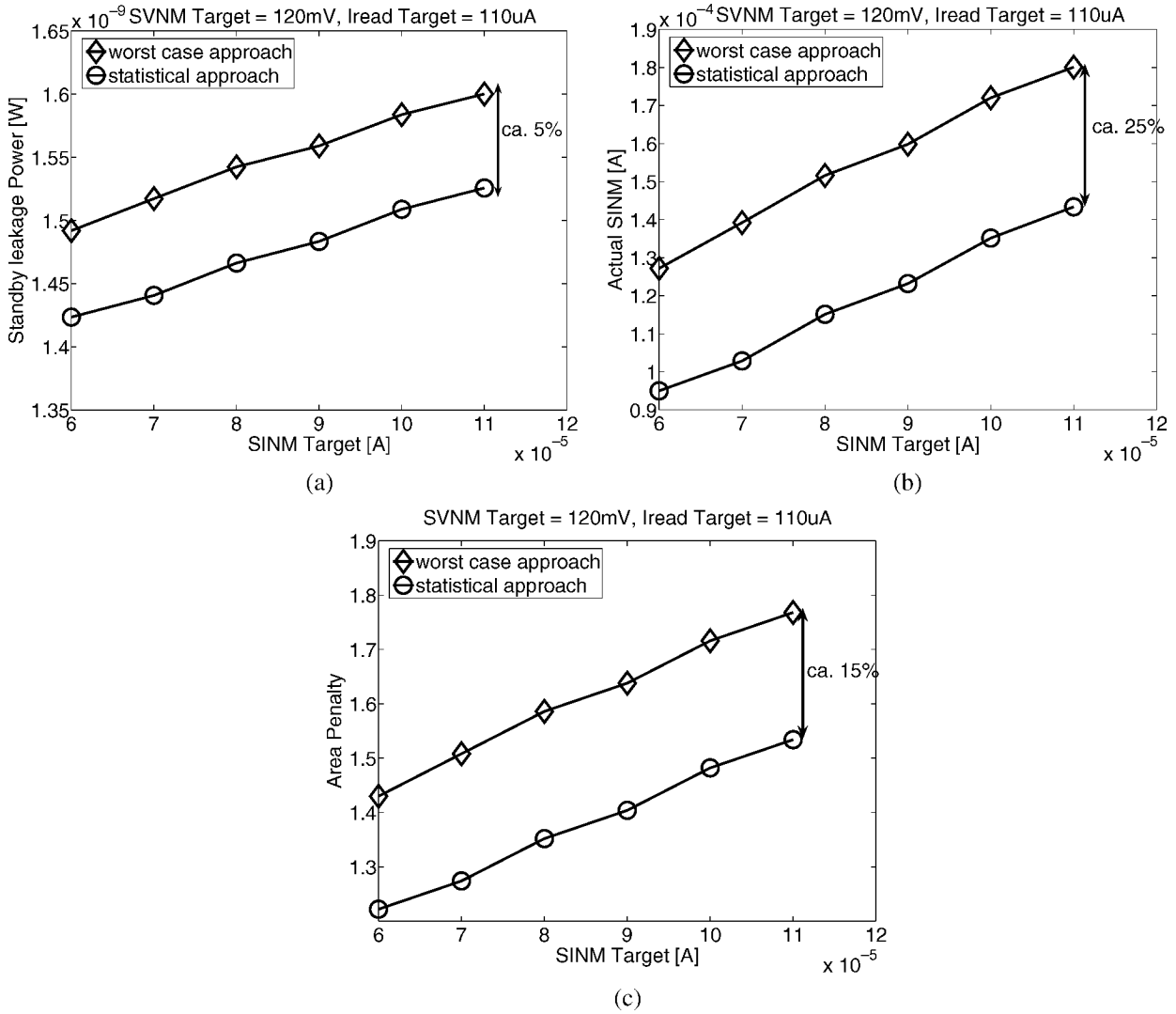

Fig. 18. SRAM cell optimization for minimum standby leakage power in 130 nm for both the statistical and the worst-case approach. (a) Standby leakage power versus a range of SINM targets. (b) Actual SINM value versus the SINM targets. (c) Area penalty versus the SINM targets.

Fig. 19(b) and (c), respectively. From Figs. 18(a) and 19(a), it is clear that the SRAM cell optimization allows gaining in leakage power with respect to the worst-case design approach while meeting the SVNM and read current targets. With the worst-case design approach, meeting the SINM constraint results in over-designing for the actual read stability (~25% for 130 nm [Fig. 18(b)] and ~40% for 65 nm [Fig. 19(b)]) and for the area (~15% for 130 nm [Fig. 18(c)] and ~26% for 65 nm [Fig. 19(c)]). The percentage of over-designing is drastically increased with respect to the 130-nm results in Fig. 18(a)–(c). Moreover, designing for SVNM and SINM targets with the worst-case approach is only feasible when drastically relaxing the WTV and even more the WTI. To meet the read stability target, the worst-case design approach yields WTI values, obtained from simulation, around -55 μA, well outside the desired range of WTI values located in the far left tail of the WTI distribution, which has a mean η_{WTI} of -21.5 μA [Fig. 15(d)]. Therefore, the minimum WTI constraint, e.g., -30 μA, for a particular SRAM design is violated strongly by the worst-case design, while this is not the case for worst-case values $\eta_{\mathrm{WTI}} - 3\sigma_{\mathrm{WTI}}$ (with $\sigma_{\mathrm{WTI}} = 1.9$ μ A) of the statistical approach. Consequently, with the worst-case design approach the SRAM cell cannot be optimized for the conflicting read and write constraints and the statistically-aware optimization approach is thus imperative.

V. Conclusion

In this paper, we introduced new N-curve metrics for the write-ability of the SRAM cell. For the first time, a comparison has been made between the N-curve metrics for the read stability and the usual SNM definition and between the N-curve write-ability metrics and the write-trip point of the cell. The N-curve current information is critical for designing a cell in nanometer technologies. Moreover, it allows overcoming the read stability limit of $0.5V_{dd}$. Finally, a statistical optimization approach is used to deal with the intra-die V_{th} variability of the SRAM cell. The obtained results show a gain both in leakage power and area with respect to the worst-case design approach. SRAM cell optimization with the worst-case design approach is even not feasible when considering intra-die V_{th} variations due

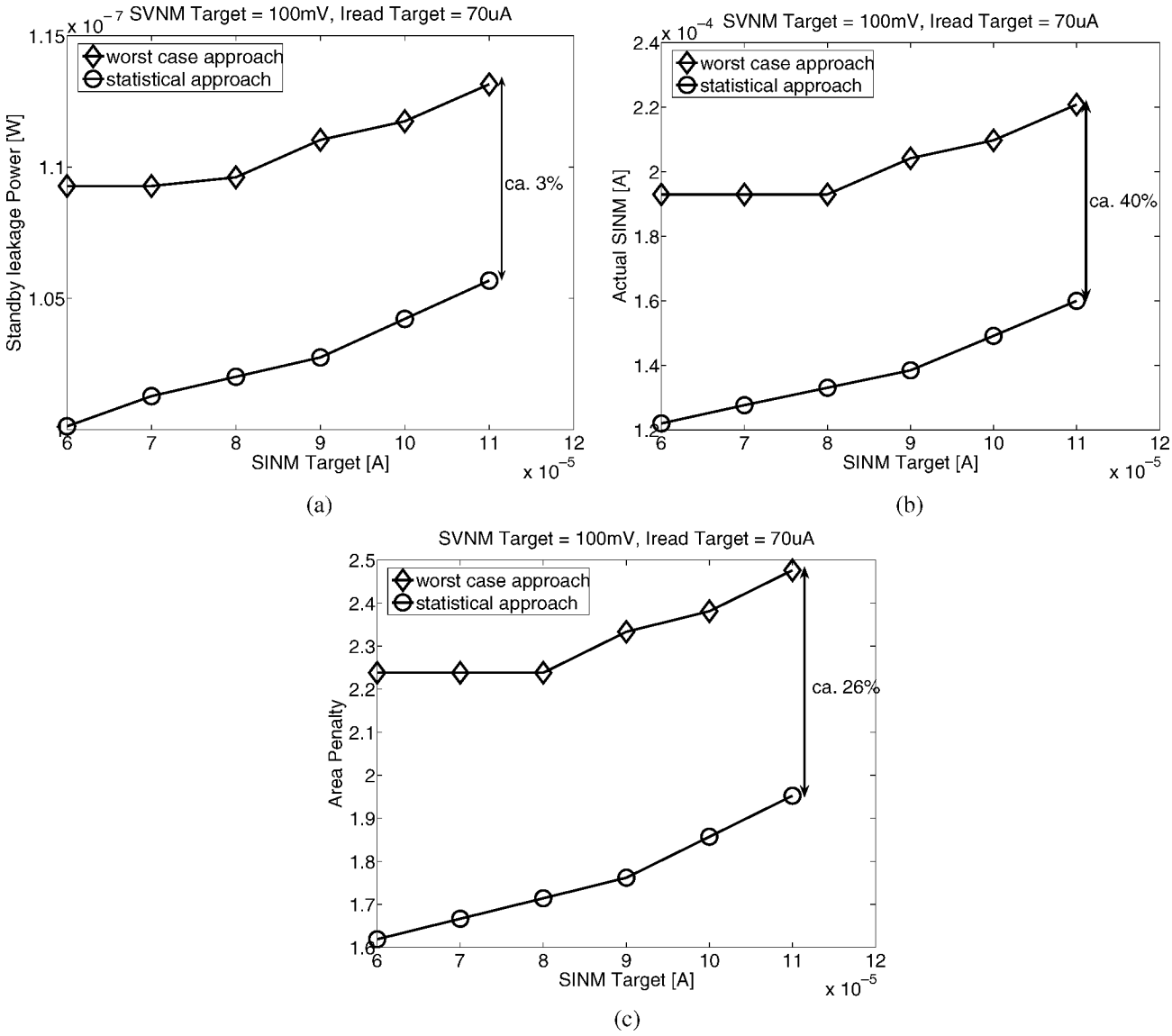

Fig. 19. SRAM cell optimization for minimum standby leakage power in 65 nm for both the statistical and the worst-case approach. (a) Standby leakage power versus a range of SINM targets. (b) Actual SINM value versus the SINM targets. (c) Area penalty versus the SINM targets.

to the conflicting read and write constraints of the cell. The increasing over-design and the hard-to-meet design criteria make the statistically-aware circuit optimization very promising for SRAM cell designs in future technology nodes.

ACKNOWLEDGMENT

The authors would like to thank P. Roussel for the statistical support and the IMEC SPDT/MSTI group of S. Decoutere for providing the 130-nm V_{th} mismatch data.

REFERENCES

[1] B. Cheng *et al.*, "The impact of random doping effects on CMOS SRAM cell," in *Proc. ESSCIRC*, Sep. 2004, pp. 219–222.

[2] E. Seevinck *et al.*, "Static-noise margin analysis of MOS SRAM cells," *IEEE J. Solid-State Circuits*, vol. SC-22, no. 5, pp. 748–754, Oct. 1987.

[3] A. J. Bhavnagarwala *et al.*, "The impact of intrinsic device fluctuations on CMOS SRAM cell stability," *IEEE J. Solid-State Circuits*, vol. 36, no. 4, pp. 658–665, Apr. 2001.

[4] L. Chang *et al.*, "Stable SRAM cell design for the 32 nm node and beyond," in *Symp. VLSI Technology Dig. Tech. Papers*, Jun. 2005, pp. 128–129.

[5] C. Wann *et al.*, "SRAM cell design for stability methodology," in *Proc. IEEE VLSI-TSA*, Apr. 2005, pp. 21–22.

[6] S. Mukhopadhyay *et al.*, "Modeling and estimation of failure probability due to parameter variations in nano-scale SRAMs for yield enhancement," in *Symp. VLSI Circuits Dig. Tech. Papers*, Jun. 2004, pp. 64–67.

[7] E. Grossar, "A yield-aware modeling methodology for nano-scaled SRAM designs," in *Proc. ICICDT*, May 2005, pp. 33–36.

[8] J. Lohstroh *et al.*, "Worst-case static noise margin criteria for logic circuits and their mathematical equivalence," *IEEE J. Solid-State Circuits*, vol. SC-18, no. 6, pp. 803–807, Dec. 1983.

[9] J. M. Rabaey, A. Chandrakasan, and B. Nikolic, *Digital Integrated Circuits: A Design Perspective (2nd edition)*. Englewood Cliffs, NJ: Prentice Hall.

[10] R. Heald *et al.*, "Variability in sub-100 nm SRAM designs," in *Proc. IEEE/ACM ICCAD*, Nov. 2004, pp. 347–352.

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

[12] J. T. Kao *et al.*, "A 175-mV multiply-accumulate unit using an adaptive supply voltage and body bias architecture," *IEEE J. Solid-State Circuits*, vol. 37, no. 11, pp. 1545–1554, Nov. 2002.
[13] E. Grossar *et al.*, "Statistically aware SRAM memory array design," in *Proc. IEEE ISQED*, San Jose, CA, 2006.
[14] K. Itoh, "Low-voltage embedded RAMs in the nanometer era," in *Proc. ICICDT '05*, pp. 235–242.
[15] A. Papoulis, *Probability, Random Variables and Stochastic Process*. New York: McGraw-Hill, 1991.
[16] R. W. Mann *et al.*, "Ultralow-power SRAM technology," *IBM J. Res & Dev.*, vol. 47, no. 5/6, pp. 553–566, Sep./Nov. 2003.
[17] C. Rose *et al., Mathematical Statistics With Mathematica*. New York: Springer-Verlag, 2002.
[18] J. Purviance *et al.*, "Statistical performance sensitivity–A valuable measure for manufacturing oriented CAD," in *IEEE MTT-S Dig.*, 1992.
[19] A. Srivastava *et al.*, "Statistical optimization of leakage power considering process variations using dual- V_{th} and sizing," in *Proc. IEEE DAC*, Jun. 2004, pp. 773–778.
[20] P. Geens *et al.*, "A small granular controlled leakage reductions system for SRAMs," *J. Solid State Electron.*, no. 49, pp. 1776–1782, Nov. 2005.
[21] J. A. Croon *et al.*, "An easy-to-use mismatch model for the MOS transistor," *IEEE J. Solid-State Circuits*, vol. 37, no. 8, pp. 1056–1064, Aug. 2002.
[22] B. Tavel *et al.*, "Thin oxynitride solution for digital and mixed-signal 65 nm CMOS platform," in *IEDM Tech. Dig.*, Dec. 2003, p. 27.6.(1-4).
[23] *"HSPICE Simulation and Analysis User Guide, Version W-2005.03,"* Synopsys, Mountain View, CA, 2005.
[24] Y. Tsukamoto *et al.*, "Worst-case analysis to obtain stable read/write DC margin of high density 6T-SRAM-array with local Vth variability," in *Proc. IEEE/ACM ICCAD*, Nov. 2005, pp. 398–405.

Evelyn Grossar was born in Tongeren, Belgium, in 1979. She received the degree of electrical engineering (Burgelijk Ingenieur), option micro-electronics, from the Katholieke Universiteit Leuven (K.U.Leuven), Belgium, in 2002. She is currently working toward the Ph.D. degree in electrical engineering at the K.U.Leuven.

She is with the Interuniversity MicroElectronics Center (IMEC), Leuven, Belgium. Her current research involves technology-aware design and modeling for deep-submicron SRAM circuits.

Michele Stucchi was born in Bari, Italy. After receiving the Masters degree in electrical engineering in 1988, he worked from 1989 to 1991 as metal and silicide process engineer for DRAM plant of Texas Instruments, Avezzano, Italy, and from 1991 to 1996 as IC failure analysis and electrical characterization engineer in the science park Tecnopolis, Bari, Italy. In 1996, he joined Imec Belgium as senior research engineer, involved in the electrical characterization and test structure design of silicides and advanced Cu-Lowk interconnects architectures. Since 2003, he coordinates the Technology-Aware Design program at IMEC, Leuven, Belgium, which aims at finding solutions to variability and low-power issues of future systems-on-chip by enhancing and establishing new links among technology, circuit, and system design.

Karen Maex (M'01) received the Master's degree in electrical engineering and the Ph.D. degree from the Katholieke Universiteit Leuven (K.U.Leuven), Belgium, in 1982 and 1987, respectively.

From 1982 to 1987, she was a Research Assistant of the Belgian National Fund for Scientific Research. She continued her research at the Interuniversity Microelectrics Center (IMEC) as a Research Director of the Fund for Scientific Research Flanders. She has been a Professor at the K.U.Leuven since 1989. She was the Director of the Interconnect Technology and Silicide Department at IMEC until December 2000, after which she became Coordinator of the Strategic Research within the Silicon Process and Device Technology Divrtision. She started the Master of Nanoscience and Nanotechnology program at KULeuven. She has authored or coauthored more than 200 publications. In 2001, she was appointed an IMEC Fellow. In 2005, she started her new role as Vice-Rector of the Group Exact Sciences at the K.U.Leuven.

Wim Dehaene (SM'04) was born in Nijmegen, The Netherlands, in 1967. He received the M.Sc. degree in electrical and mechanical engineering and the Ph.D. degree from the Katholieke Universiteit Leuven, Belgium, in 1991 and 1996, respectively. His thesis was entitled "CMOS integrated circuits for analog signal processing in hard disk systems."

After receiving the M.Sc. degree, he was a research assistant at the ESAT-MICAS Laboratory of the Katholieke Universiteit Leuven. His research involved the design of novel CMOS building blocks for hard disk systems. The research was first sponsored by the IWONL (Belgian Institute for Science and Research in Industry and agriculture) and later by the IWT (the Flemish institute for Scientific Research in the Industry). In November 1996, he joined Alcatel Microelectronics, Belgium. There he was a Senior Project Leader for the feasibility, design and development of mixed-mode systems on chip. The application domains were telephony, xDSL, and high-speed wireless LAN. In July 2002, he joined the staff of the ESAT-MICAS Laboratory of the Katholieke Universiteit Leuven where he is now an Associate Professor. His research domain is circuit level design of digital circuits. The current focus is on ultra-low-power signal processing and memories. He is teaching several classes on digital circuit and system design.