

Variability impact on on-chip memory data paths

E. Amat^{*}, A. Calomarde, R. Canal and A. Rubio

Electronic Engineering Department, Universitat Politècnica de Catalunya (UPC), Barcelona, Spain

^{*}currently at CEA-LETI (Grenoble, France)

e-mail: esteve.amat@upc.edu

Abstract— Process variations have a large impact on device and circuit reliability and performance. Few studies are focused on their impact on more complex systems, as for example their influence in a memory data path. In our study, the impact of variations in the cell block is the largest measured, as it is usually designed with the minimum device dimensions. Moreover, we observe a significant influence of the device type (p/nMOS) used to implement the memory cell in terms of delay and variability robustness.

Keywords—variability, DRAM, delay, temperature.

I. INTRODUCTION

The continuous device dimensions reduction has entailed an improvement of the electronics performance (faster, smaller and lower power consumption). But, it has also entailed significant reliability challenges among other caused by device process variations. The main consequence is a relevant device and circuit malfunction, due to a non-expected threshold voltage (V_T) shift [1], especially for sub-45nm nodes. Additionally, the circuit performance is also highly affected for device variability, and a relevant worsening of its behavior is observed, e.g. enlargement of the circuits' delay and memory cell instability. There are several sources of device variability, but the main one is random dopant fluctuation (RDF), which is related with the large amount of device doping, what is introduced with the objective to reduce short channel effects. This results in a significant variation of V_T at small device dimensions, and consequently a variation on the device behavior. Moreover, there exist other variability sources, as well, with impact on device performance, as for instance line edge roughness (LER) and work-function fluctuation (WFF) [2][3], but with lower relevance in conventional planar devices. For this, as the technology node continuously reduces their dimensions to achieve better device behavior, the variability levels increase and become unacceptable. For this, reliability must be considered as a key design issue.

Currently, in order to continue the device scaling and enhance device performance beyond 22nm, multi-gate devices (FinFETs) have emerged as a feasible option. FinFETs are nowadays the best positioned candidate to substitute planar devices, which are highly affected by several reliability threats (e.g. variability and leakage currents). One of FinFETs main advantages is the lower impact of the short channel effect, due to their better channel control and electrostatics. This allows the device designers to reduce the channel doping, and even an undoped one is feasible. As a consequence, the overall multi-gate device variability is significantly reduced [3], and

consequently better devices and circuits performance could be achieved [4].

In this context, the community has shown a big interest in analyzing the circuit performance when considering device variability. Device process variation has become one of the main detrimental factors for device and circuit reliability at technology nodes beyond 45nm. Their impact concerns extend to the more difficult performance predictability of whole system behavior. Usually, the memory cell is the only focus of circuit analysis, as it is designed by using the minimum feature size for density reasons. At a circuit level, variability translates to performance instabilities, as for instance at memory cell level in the SRAM case SNM instabilities are observed [5]; or reduction of the retention time in the dynamic memories (DRAM) [6]. Nevertheless, a microprocessor is not only implemented by memory cells, and for example a memory access involves also other logics blocks, e.g. sense amplifier, multiplexor. So then, it is important to consider the contribution of the rest of circuits in a data path to better analyze the impact of variations on the system. Notice that a conventional memory data path is usually designed with different logic blocks, as for instance sense amplifiers (SA), decoders (DEC), multiplexors (MUX) and flip-flops (FF), which have been analyzed stand-alone under variability [7], but in few occasions the whole path is evaluated. The main impact of the process variations on the data path is observed in a delay of the overall system, and consequently, a slowdown in its performance.

In this work, we study the variability influence in a conventional memory data path, when using FinFET devices. Therefore, the rest of the paper is organized as follows. Section II depicts the simulation framework, where the introduced level of variability, and the definition of the data path implemented are determined, as well. In Section III, we show the results for several process variability scenarios and environmental conditions. Finally, Section IV concludes the work.

II. SIMULATION FRAMEWORK

In this section, we describe the methodology and the analysis performed

A. Memory data path descriptor

Throughout all this study, we considered a standard data path configuration, formed by different logic blocks. In this context, Figure 1 presents the data path analyzed. The blocks in Figure 1.a are plotted in the order a read access is done. The first one is a decoder used to select the desired memory row to perform a read or write operation. Figure 1.b shows the

implementation of the final decoder stage (i.e. wordline generator). The next block is the memory array where data is stored. For our study we considered a 3T1D-DRAM cell (dynamic memory cell) [8], depicted in Figure 1.c. 3T1Ds have shown to be a feasible candidate to substitute the SRAM cells in data caches, which are highly affected by variability problems at small technology nodes. Afterwards reading from the memory cells in the array, data passes through the sense amplifiers, the column decoders (or muxes) and eventually get latched (before they are used in a subsequent pipeline stage). We have simulated the most simple single-ended sense amplifier (SA) based only on an inverter configuration, as Figure 1.d depicts. Afterwards, a multiplexor is required to select the appropriate data from the row, as rows may hold several 32-bit or 64-bit words. We just analyze the data path, thus the activation (i.e. decoding) of the mux is not shown in Figure 1.e. The last system block is a flip-flop where data is stored. We simulate a D-type flip-flop. Figure 1.f represents the implemented structure, which is a classical implementation [9].

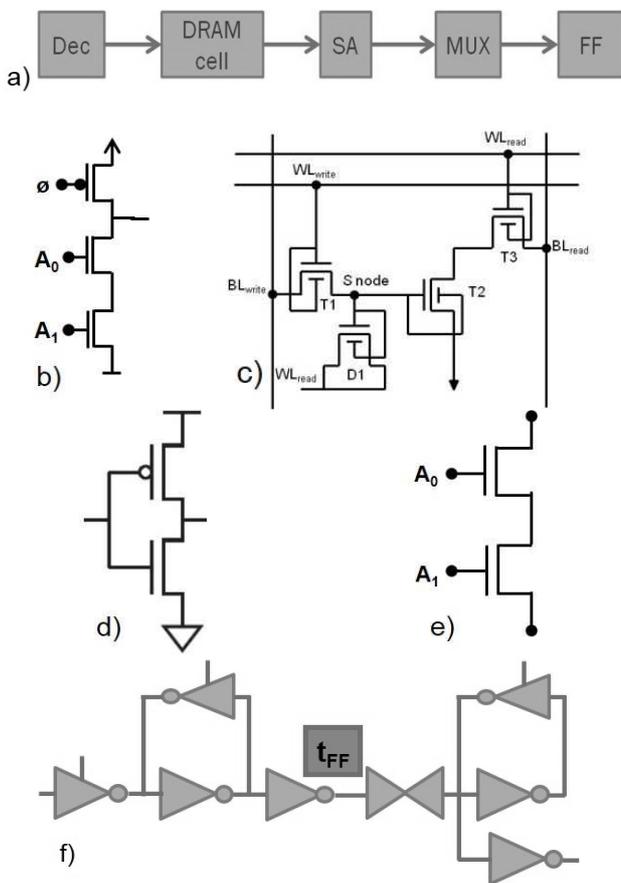


Fig. 1 a) Schematic of the analyzed data path formed by different blocks, b) decoder, c) memory cell (3T1D-DRAM), d) sense amplifier (inverter), e) multiplexor and f) D flip-flop (DFF).

Furthermore, in order to obtain a realistic behavior of the data path we must size the transistors correctly. In this context, we contemplate two different scenarios. On one hand, we have regarded the design of the complete system where the propagation delay is optimized, i.e. the delay is equally

distributed along the data path. With this matter, we adjusted the device dimensions used for the chain of blocks taking into account an optimum delay design. For instance, as a reference system block, we use the 3T1D-DRAM cell with minimum device dimensions, since for density reasons the memory cell should be as small as possible. From this, we determined that the following blocks must have larger dimensions. Indeed, we optimize the logical effort at each stage to obtain a more optimum system delay [9], i.e. decoder (2x), memory cell (1x), sense amplifier (2x), multiplexor (4x) and D Flip-Flop (8x). On the other hand, the industry could be also interested to reduce the overall area overhead as much as possible. For this, and as a comparison, we also simulated the data path using always the minimum device dimensions that corresponds to two fins for each FinFET as a device width, and their nominal channel length. Next, we analyze the impact of variability in both configurations, i.e. delay optimized and minimum size systems, and determine the benefits and drawbacks for both strategies.

B. Performance analysis

First, we study the behavior of the data path as a function of the device type used to implement the memory cell. Initially, the DRAM memory cell used in this contribution is based on nMOSFETs [8]. But, recently, it is observed that a cell implementation based on pMOSFETs shows a significantly better performance [10] in front of their nMOS counterparts, i.e. larger retention time. This is important as it determines the refresh interval[6]. So then, in this study we compared both implementation types (n/pMOS-based) as a function of supply voltage (0.4-1V); and for different environment temperatures (25°-125°C). Notice that for both analyses, we evaluate the system performance regarding always the system delay. We measure it as the time between the access starts and until it is read in the flip-flop. This time is called flip-flop delay (t_{FF}), additionally the previous delays, after the sense amplifier (t_{SA}) and the multiplexor (t_{MUX}), are also recorded to observe their delay contribution.

C. Variability impact study

In the context of the process variation analysis, we should take into account that different variability levels are usually regarded for pMOS or nMOS devices [11], due mainly to the larger doping required for the former ones. Moreover, we should remember that we simulate this system using FinFET devices based on High Performance Predictive Technology Models (HP PTM), provided by Arizona State University [12]. For this, we consider a lower variability impact in comparison the usual bulk devices [3]. So then, we stated a variation level of 20%, reflected by a V_T -shift of the 7nm FinFET. Additionally, it is also well established that the process variation impact on the device depends on their size dimensions [13], and when larger area are used smaller device variability we should introduce. Finally, to perform the variability analysis we carried out 10000 Monte-Carlo simulations, and the variability relevance is evaluated by a statistical distribution with mean (μ) and standard deviation (σ), obtaining the $3\sigma/\mu$ ratio factor, expressed in percentage as impact factor.

III. RESULTS

First of all, we analyzed the system performance when the supply voltage and the environment temperature are modified. In this scenario, we simulated two data path scenarios where the difference is regarded on the device type employed to implement the memory cell, since we used p or nMOS for each case. Then, we recorded the system behavior by analyzing the several delays defined in the previous section. Note that in order to avoid redundant information, we establish the optimum delay configuration and the delay at the flip-flop, as the reference ones.

A. Performance analysis

Figure 2 shows how the delay of the system behaves when the supply voltage (V_{DD}) is modified. As it is expected as the biasing of our system is reduced, we observe a relevant increase in delay. First, we observe that pMOS-based cells always have a larger delay in comparison to the nMOS ones. This is caused by their lower channel mobility. While both configurations present similar delay enlargement ($\sim 10x$) along the analyzed V_{DD} range, our simulations show a different evolution of the system delay for both systems. Systems based on nMOS memory cells show an almost proportional rise, whereas for the pMOS ones two different regions are clearly depicted. While above 0.6V, the delay increase can be considered moderate ($\sim 2x$), at lower V_{DD} the delay values worsen significantly ($6x$). This behavior could be caused by the different threshold voltage levels of nMOS and pMOS devices.

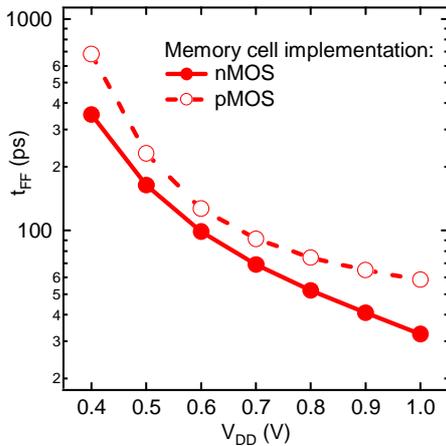


Fig. 2 Study of the influence of the supply voltage modification, from 0.4 to 1V for an optimum delay data path based on a nMOS or pMOS memory cell. The delay is recorded at the end of our data path in the D flip-flop.

Next analysis focuses on the impact of the environment temperature on the system behavior. In this context, Figure 3 shows the system delay evolution as the temperature shifts from 25° to 125°C. Again, the device type (p/nMOS) makes a difference. Though both configurations show a decrease in delay, we can easily see different slopes and magnitudes. Whereas for pMOS cell systems, we observe a delay reduction around 40% during all the analyzed range, nMOS counterparts show a more significant evolution of the delay reduction ($3x$). This poor behavior of the nMOS cell is caused by the larger

impact of the leakage currents, which grow exponentially as environment temperature becomes higher.

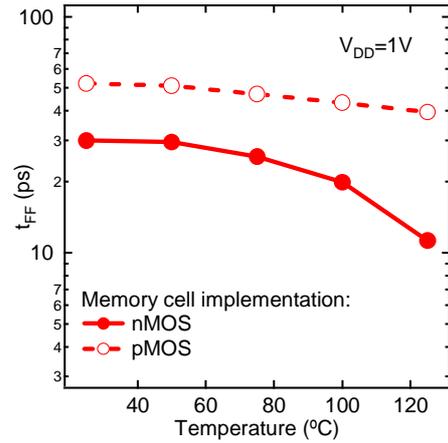


Fig. 3 Influence of the modification of the environment temperature, from 25° to 125°C, for an optimum delay data path based on a nMOS or pMOS memory cell. The supply voltage used is 1V. The delay is again recorded at the end of our data path in the D flip-flop.

We should note that for both previous studies we only focus on the flip-flop delay. We have observed insignificant differences for the other two delay (t_{SA} and t_{MUX}). In terms of the results of the optimum delay or minimum device dimensions configurations, we have obtained again similar trends. But, when we compare the optimum delay and the minimum device dimensions designs, one interesting result that we observe from the inter-medium delays is to measure how the delay optimization strategy improves the overall system behavior. Although, the minimum device dimensions design obtains a smaller system area, Figure 4 shows a more stable (almost constant) and small system delay for the optimum delay strategy (line). In contrast, for the minimum device size configuration (dash), we observe a significant increase of the system delay in function of the delay location in the analyzed system.

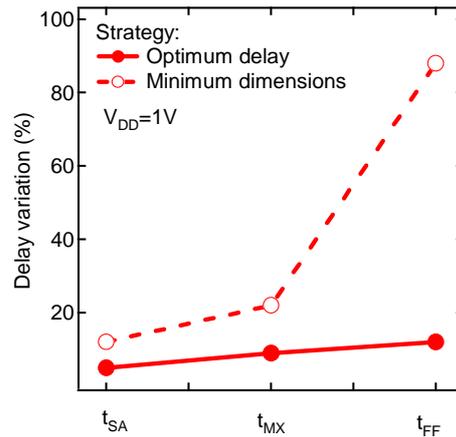


Fig. 4 Variation of the system delay as a function of where it is measured, at the sense amplifier (t_{SA}), at the multiplexor (t_{MUX}) or at the flip-flop (t_{FF}). Two different data path strategies are considered: an optimum delay (line) or a minimum device dimensions (dash).

B. Variability impact

The variability influence is really significant when the technology node of the used devices is below 45nm. Along this section, we study the process variation impact in the analyzed system. Note that we simulate our system using 7nm HP PTM FinFET devices. Moreover, we should remember that we designed the system with two different size configurations: one where a strategy regards on the delay optimization (different device dimensions as a function of the system block), and the other with the smaller area possible (i.e. minimum device dimensions for all the system blocks, that we stated at two fins). Furthermore, we also analyze the influence of the memory cell implementation by using n or pMOS devices, in order to reduce the leakage current and increase the retention time of our dynamic memory cells. Likewise, variability location along the data path is also analyzed. For this, we considered process variation in all system blocks (total), and when it is introduced only at a concrete part of the overall system, e.g. memory cell, SA.

Figure 5 points out the variability impact (stated as a percentage of the $3\sigma/\mu$ ratio) in both scenarios and as a function of their location, as well. Figure 5a shows how the optimized delay configuration presents smaller susceptibility (around 17% and 25% for data paths based on n/pMOS memory cells, respectively) of the process variation than the minimum dimensions option, as it is depicted in Figure 5b. This is due to the larger device sizes used to obtain an optimum system delay. Additionally, we observe how the larger variability contribution could be mainly attributed to the memory cell, as it could be expected, since it is designed with the minimum dimensions. Another relevant aspect in our study is the influence of the device type used for the memory cell implementation. In this context, the data path where the memory cell is implemented with pMOS devices presents a larger variability (~25%), in contrast to their nMOS counterparts. In the case of the minimum device dimensions strategy (Fig. 5b), we observe a more distributed variability. While the memory cell still presents the highest delay variation due to the process variation, the rest of the system blocks show a relevant increase of their weakness in front of the process variation.

Summarizing, for both configurations (optimum delay and minimum dimensions) the memory cell is clearly the main contributor to the overall system variability, almost 80%. The contribution of the rest of the system blocks is significantly lower. Indeed, adding all the individual variability contributions we could observe that it exceeds the total variability obtained, and this could be attributed to the compensation of the variability for each system block.

IV. CONCLUSIONS

We have performed a design space exploration of a conventional memory data path implemented with 7nm PTM FinFETs. We have considered: (1) different supply voltages, (2) different environment temperatures, (3) 20% of variability reflected in a V_T -shift, (4) two different system configurations (optimum delay and minimum device dimensions) and (5)

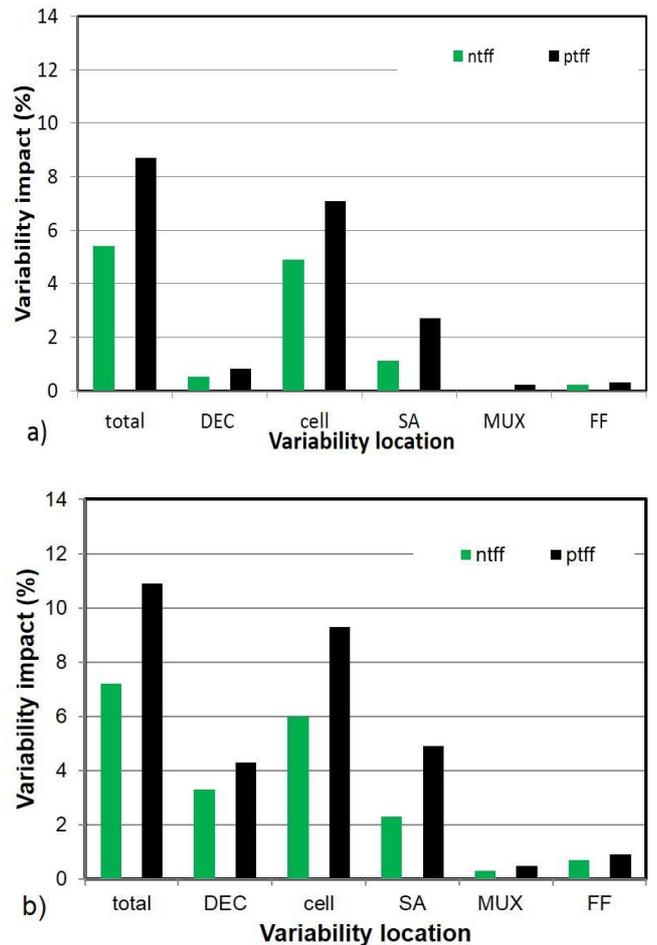


Fig. 5 Process variation relevance regarding the data path delay at the flip-flop block and the device type used to implement the memory cell, nMOS (ntff) or pMOS (ptff). Two scenarios are considered to simulate the data path: a) Optimum delay resizing, and b) minimum device dimensions. Moreover, the relevance of the variability location is also contemplated.

different device type to implement the memory cell. This last consideration has entailed as a relevant factor to take into account, since in both reliability analysis (performance and variability) has presented a relevant impact. In terms of V_{DD} and temperature impact, we observe a different behavior as a function of the device type. As pMOS shows lower impact of temperature and a less lineal dependence on the supply voltage in contrast to their nMOS counterparts. Finally, in terms of variability impact, we have observed that the pMOS-based memory cell systems always present larger variability impact. At the data memory path level and for all the cases tested, the main contribution to the overall system variability impact is obtained for the memory cell.

ACKNOWLEDGMENTS

This work was supported by the Spanish MINECO (JCI-2010-07083, TIN-2010-18368) and partially supported by 2012 Intel Early Career Faculty Honor program.

REFERENCES

- [1] B. Cheng, S. Roy, G. Roy, F. Adamu-Lema, and A. Asenov, "Impact of intrinsic parameter fluctuations in decanano MOSFETs on yield and functionality of SRAM cells," *Solid. State. Electron.*, vol. 49, no. 5, pp. 740–746, May 2005.
- [2] A. Asenov, S. Kaya, and A. R. Brown, "Intrinsic parameter fluctuations in decananometer MOSFETs introduced by gate line edge roughness," *IEEE Trans. Electron Devices*, vol. 50, no. 5, pp. 1254–1260, 2003.
- [3] T. Matsukawa, Y. Liu, S. O'uchi, K. Endo, J. Tsukada, H. Yamauchi, Y. Ishikawa, H. Ota, S. Migita, Y. Morita, W. Mizubayashi, K. Sakamoto, and M. Masahara, "Comprehensive analysis of Ion variation in metal gate FinFETs for 20nm and beyond," *IEEE Int. Electron Devices Meet.*, pp. 23.5.1–23.5.4, 2011.
- [4] E. Amat, A. Calomarde, C. G. Almudever, N. Aymerich, R. Canal, and A. Rubio, "Impact of FinFET and III-V/Ge technology on logic and memory cell behavior," *IEEE Trans. Device Mater. Reliab.*, 2013.
- [5] B. Cheng, S. Roy, and A. Asenov, "CMOS 6-T SRAM cell design subject to 'atomistic' fluctuations," *Solid. State. Electron.*, vol. 51, no. 4, pp. 565–571, Apr. 2007.
- [6] E. Amat, C. G. Almudever, N. Aymerich, R. Canal, and A. Rubio, "Variability mitigation mechanisms in scaled 3T1D DRAM memories to 22nm and beyond," *IEEE Trans. Device Mater. Reliab.*, vol. 13, no. 1, pp. 103–109, 2013.
- [7] M.-L. Fan, V. P.-H. Hu, Y.-N. Chen, P. Su, and C.-T. Chuang, "Variability Analysis of Sense Amplifier for FinFET Subthreshold SRAM Applications," *IEEE Trans. Circuits Syst. II Express Briefs*, vol. 59, no. 12, pp. 878–882, 2012.
- [8] W. K. Luk, J. Cai, R. H. Dennard, M. J. Immediato, and S. Kosonocky, "A 3-Transistor DRAM Cell with Gated Diode for Enhanced Speed and Retention Time," *Symp. VLSI Circuits*, pp. 184–185, 2006.
- [9] N. Weste and D. Harris, *CMOS VLSI design*. Ed. Addison-Wesley.
- [10] W. Zhang, K. C. Chun, and C. H. Kim, "Variation aware performance analysis of gain cell embedded DRAMs," *ACM/IEEE Int. Symp. Low-Power Electron. Des.*, pp. 19–24, 2010.
- [11] T. Chiarella, L. Witters, A. Mercha, C. Kerner, M. Rakowski, C. Ortolland, L.-Å. Ragnarsson, B. Parvais, A. De Keersgieter, S. Kubicek, A. Redolfi, C. Vrancken, S. Brus, A. Lauwers, P. Absil, S. Biesemans, and T. Hoffmann, "Benchmarking SOI and bulk FinFET alternatives for PLANAR CMOS scaling succession," *Solid. State. Electron.*, vol. 54, no. 9, pp. 855–860, Sep. 2010.
- [12] S. Sinha, G. Yeric, V. Chandra, C. B, and Y. Cao, "Exploring sub-20nm FinFET designs with Predictive Technology Models," *Des. Autom. Conf. Proc.*, pp. 283–288, 2012.
- [13] M. Pelgrom, A. Duinmaijer, and A. Welbers, "Matching properties of MOS transistors," *IEEE J. Solid-State Circuits*, vol. 24, no. 5, pp. 1433–1439, 1989.