Performance Evaluation of FinFET based SRAM under Statistical VT Variability

Ahmed T. El-Thakeb[1], Hamdy Abd Elhamid[1], Hassan Mostafa[1, 2], Yehea Ismail[1]
[1]Center for Nano-electronics & Devices, American University in Cairo and Zewail City for Science and Technology, Cairo, Egypt
[2]Electronics and Communications Engineering Department, Cairo University, Giza 12613, Egypt
Emails :{ ahmed.t.elthakeb, habdelhamid, hmoustafa, and y.ismail}@aucegypt.edu

***Abstract*—FinFET devices are the most promising solutions for further technology scaling in the long term projections of the ITRS. The performance of extremely scaled FinFET-based 256-bit (6T) SRAM is evaluated with technology scaling for channel lengths of 20nm down to 7nm showing the scaling trends of basic performance metrics. In addition, the impact of threshold voltage variations on the delay, power, and stability is reported considering die-to-die variations. Significant performance degradation is found starting from the 10nm channel length and continues down to 7nm.**

Keywords — Nano-scale FinFET; 6T SRAM; Technology scaling; threshold voltage variations.

I. INTRODUCTION

Tri-gate (TG) FinFET has been deployed as the first winning successor of the conventional planar transistor for the sub 22 nm technology node due to its superior electrostatics and sub-threshold leakage control [1-3]. For the long term projections by the International Technology Roadmap of Semiconductors (ITRS - 2013)[4], several emerging new devices are on the horizon but the multi-gate structures are the most viable solution, so far, to effectively scale the channel length with controlling the short channel effects. However, how far FinFET can go is still an open issue. So, further investigation of the performance of Tri-gate FinFET along the scaling roadmap for basic circuit blocks is essential [5].

Static Random Access Memory (SRAM) occupies a significant portion of all system-on-chips and microprocessors as an efficient embedded memory block [4]. As a result of the increasing demand for higher performance and integration, higher density SRAM cells are designed with the minimum size transistors in a given technology node.

Increased process variability, and device reliability issues increase the necessity for performance evaluation of SRAM design methodologies and topologies with technology scaling.

On one hand, shrinking the channel length significantly increases the short channel effects (SCEs) which in turn degrade the basic cell metrics such as the leakage power. On the other hand, emerging novel devices such as FinFETs poses new challenges by adding new variability sources such as the fin thickness variations as a result of increased line edge roughness. In addition to new design issues such as width quantization which limits the design optimization [6]. However, having new geometry parameters such as the fin thickness, the quantized number of fins, and even surface orientation opens the way for new design optimization techniques [7]. On top of the challenges of scaling of SRAM on the design level as specified by the ITRS-2013, is to maintain adequate noise margins and control key instabilities and soft-error rates in the presence of random threshold voltage (V_T) fluctuations.

Table 1: The simulated device parameters

Device	TG-FinFET				
L (nm)	20	16	14	10	7
T_{fin}(nm)	15	12	10	8	6.5
H_{fin} (nm)	28	26	23	21	18
V_{DD}(V)	0.9	0.85	0.8	0.75	0.7
Fin ratio (N_{fin}) (PU:PD:PG)	(1 : 3 : 2)				

In addition to the difficulty in keeping the leakage current within tolerable limits.

Some studies have discussed the FinFET SRAM performance at the nano-scale. For instance, a simulation study for 14 nm SOI FinFET technology has been reported showing the impact of the relevant sources of variability and reliability on the cell stability [8], [9]. In [7], the FinFET SRAM design space is discussed, under different fin thicknesses and fin heights, to optimize stability, delays and leakage current but at constant channel length.

In this study, we report the conventional (6T) SRAM cell operation's limits within a given range of threshold voltage variations along with different technology nodes starting from 20 nm down to 7 nm. The variations are considered die-to-die variations. The operation's limits are determined through the evaluation of read/write static noise margins (RSNM/WSNM) as an indication for the cell's stability, read and write delays, active and leakage powers.

This paper is organized as follows; in Section II the simulation methodology and the design parameters are discussed. The Read/Write delays, power consumption and cell stability are reported, and the overall performance is discussed in Section III. Conclusions are drawn in Section V.

II. SIMULATION METHODOLOGY

In this paper, predictive technology model (PTM-MG) files [10] for Multi-gate devices (TG-FinFET in our case) are used from 20 nm down to 7 nm technology nodes for low-standby power devices (LSTP) with the BSIM-CMG compact models. A scaling strategy is adopted according to the PTM models which involves, besides the scaling of the channel length (L), scaling of the supply voltage (V_{DD}), fin thickness (T_{fin}), and fin height (H_{fin}). The used device parameters are reported in Table.1. Tri-gate FinFET structure is used such that the effective channel width is ($W_{eff} = 2H_{fin} + T_{fin}$).

Regarding the performance metrics, the read delay is calculated as the time period from the 50 % point of the word line (WL) low-to-high transition to a 10 % difference point developed between the BL and BLB.

978-1-4799-8153-3/14/$31.00 ©2014 IEEE

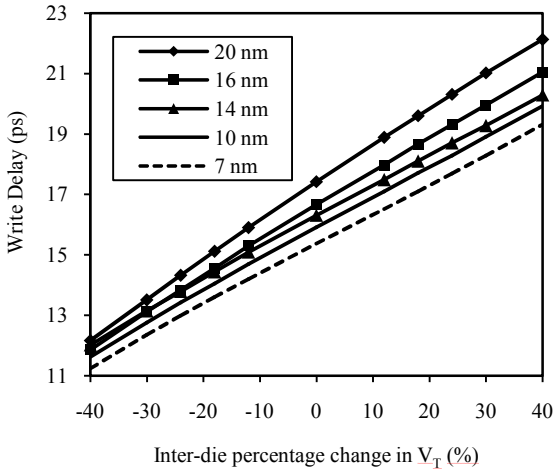

Fig. 1. SRAM Read delay sensitivity to threshold voltage inter-die variations range of +/-40 % at various technology nodes from 20nm down to 7nm node.

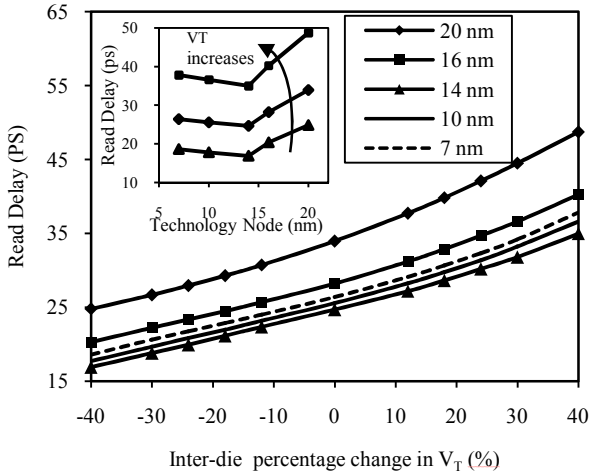

Fig. 2. SRAM Write delay sensitivity to threshold voltage inter-die variations range of +/-40 % at various technology nodes from 20nm down to 7nm node.

The read/write static noise margins (RSNM/WSNM) are used as a measure for the read/write operations stability of the SRAM cell respectively, and are defined as the maximum absolute DC voltage around the half-supply pre-charged bit-lines (BL, BLB) that causes the stored state of the cell not to flip during the read operation, or the maximum absolute DC voltage below V_{DD} for BL and above '0' for BLB that changes the state of the cell for a successful write operation.

The leakage power consumption is calculated for the SRAM cells in the idle mode; when the access transistors are cut-off and the bit-lines are left floating.

III. Results and Discussions

A. Read/Write Delays

Read/ Write delays are key parameters in evaluating the performance of SRAM cell. Fig.1, 2 show the sensitivity of the SRAM read/write delays with technology scaling and V_T variations (from - 40 % to 40 % of the nominal value). As it can be noticed, with increasing the threshold voltage the delay increases as a result of decreasing the overdrive voltage hence reducing the transistors' currents. For write operation, the delay encounters a variations of around +/- 25 % over the +/- 40 % threshold variations, and for the read operation, the variation in the delay is around +/- 35 % which is quite higher, with respect to the delay value at the nominal V_T.

On the other side, observing the behavior with the technology scaling. For the write operation, the delay is continuously decreasing with scaling down the technology as a result of shrinking the channel length despite the scaling of the supply voltage which usually leads to increase in the delay. However for the read delay it is quite different, since degradation is observed starting from the 10 nm node down to the 7 nm as it can be seen in the inset of Fig.2. To understand this behavior, we plot the read current at each technology node as shown in Fig.3.

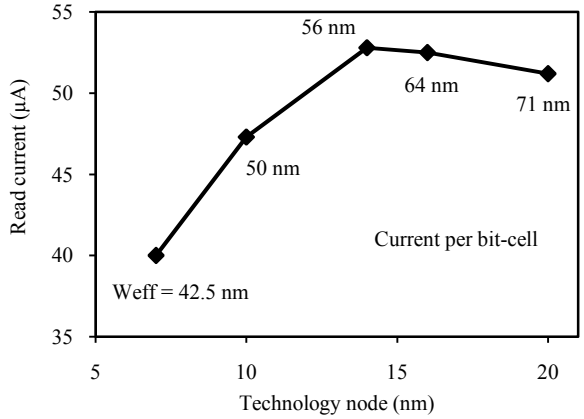

Fig. 3. Device current per bit-cell with technology scaling from 20nm to 7nm node, where Weff = 2Hfin+Tfin, and Wtot = Nfin Weff.

As it was discussed in the first section, with technology scaling, other parameters are scaled besides the scaling of the channel length such as the supply voltage, the fin thickness, and the fin height which is basically to compensate for the increased SCEs associated with such extreme scaling. This in fact has an adverse effect on the read current as it can be seen in Fig.3; the current is almost constant (slightly increasing) as we go from the 20 nm to 16 nm and 14 nm nodes, however it drops at the 10 nm and further decreases reaching the 7 nm node. So despite the fact that with technology scaling the current value per unit width is expected to increase, the current per bit-cell is decreasing as a result of the adopted scaling strategies to keep SCEs under control, since scaling both T_{fin}, and H_{fin} reduces the effective channel width.

Consequently, this raises a serious challenge for SRAM design in extremely scaled technology nodes, since this fact implies that to retrieve this loss of performance, keeping SCEs under control, some cell devices generally will need to be sized up which contradicts the trend of higher density SRAM arrays with scaling.

978-1-4799-8153-3/14/$31.00 ©2014 IEEE

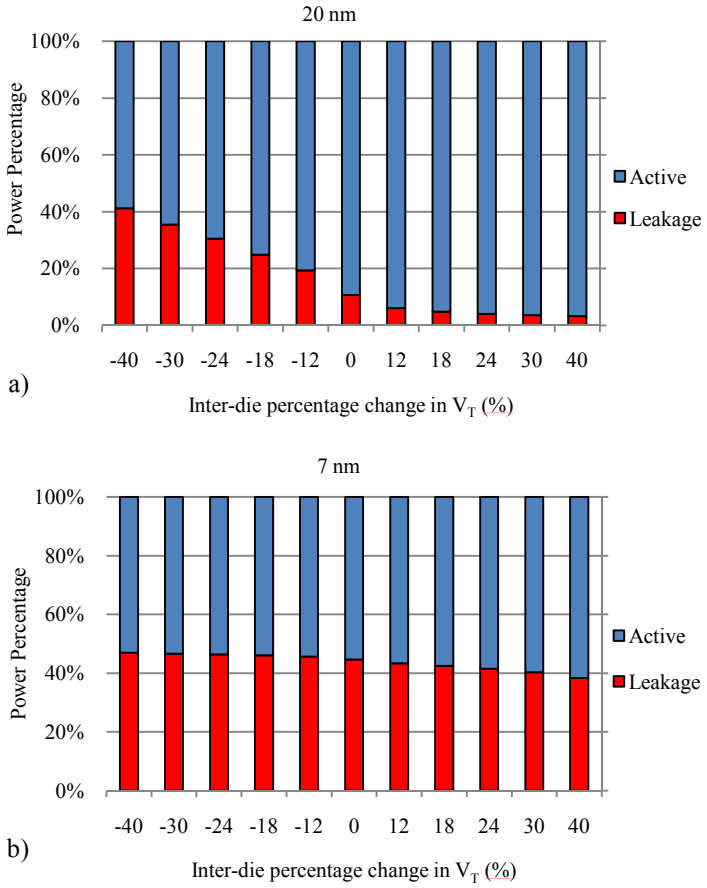

Fig. 4. Sensitivity of the percentage leakage power to the active power with threshold voltage variations; a) 20nm node, b) 7nm node.

B. *Power consumption*

Power consumption is one of the critical metrics for any logic circuits and analyzing the scaling trends of both the active and leakage components is of special concern. From one hand, as the technology scales, all sources of leakage power increase. Shrinking the channel length increases the sub-threshold leakage component and scaling the oxide thickness severely affects the gate tunneling current which is another component of the total leakage current. From the other hand, the increased variability sources with scaling and the resulting effect on the threshold voltage spread significantly impacts the leakage power due the exponential dependence on V_T. Fig.4 shows the percentage of the leakage power component to the active power component and its sensitivity with the threshold voltage variations at both 20nm and 7nm nodes. As it can be seen, for the 20nm node, as V_T decreases the amount of the leakage power increases and contributes significantly a larger portion of the total power consumption. While, for the 7nm, the leakage power already occupies a significant portion of the total power and changing the threshold voltage has a minor impact on the relative percentage.

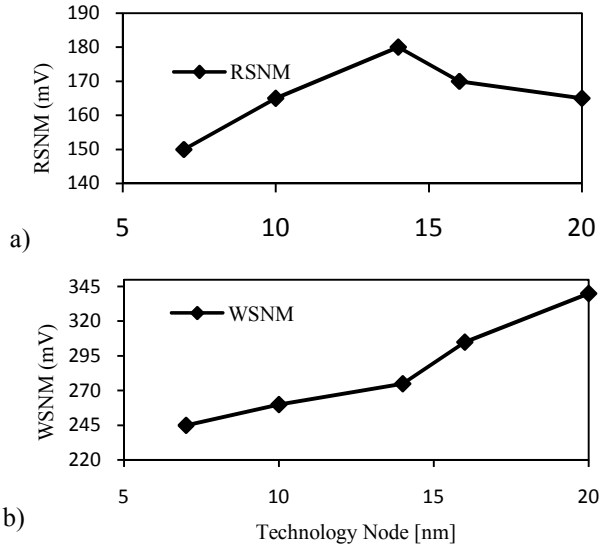

Fig. 5. Read and write static noise margins with technology scaling

It can be concluded that, as the technology scales, the leakage power component increases and occupies a significant portion of the total power consumption, however the variability of the leakage power to the V_T variation significantly reduces. This fact has further implications on other performance metrics as it will be discussed in the next section.

C. *Static Noise Margins*

Fig.5 shows the read and write static noise margins (RSNM, WSNM) with technology scaling. As it can be seen in Fig.5.a) the RSNM shows the same behavior of the read delay with a degradation starting from the 10nm node to the 7nm. This is also can be attributed to the degradation of the read current as discussed in the above section which affects the read operation as a whole from both the delay and stability point of view. For the WSNM, a clear degradation of 28% at 7nm with respect to its value at the 20nm node can be shown in Fig.5.b), which is primarily as a result of scaling the supply voltage. Fig.6 shows the sensitivity of the RSNM and WSNM with the V_T variations at the 20nm and 7nm technology nodes. First it can be seen in Fig.6.a) that with deceasing V_T the RSNM is degraded for both the technology nodes, since reducing V_T increases the leakage current which in turn increases the voltage of the node to be read (assuming read '0') leading to an increase in the probability of destructive read operation. In addition, reducing V_T affects the VTC of the inverters which affects the trip point making it easier for the '0' storage node to flip to '1'. Second, the degradation in the RSNM for the 20nm node is around +/- 25 % and for the 7nm is just about +/- 10 % with respect to the value at the nominal threshold voltage. This behavior can be explained as a result of less sensitivity of the leakage current to the V_T variations with technology scaling compared to that at the 20nm as discussed in the above section.

978-1-4799-8153-3/14/$31.00 ©2014 IEEE

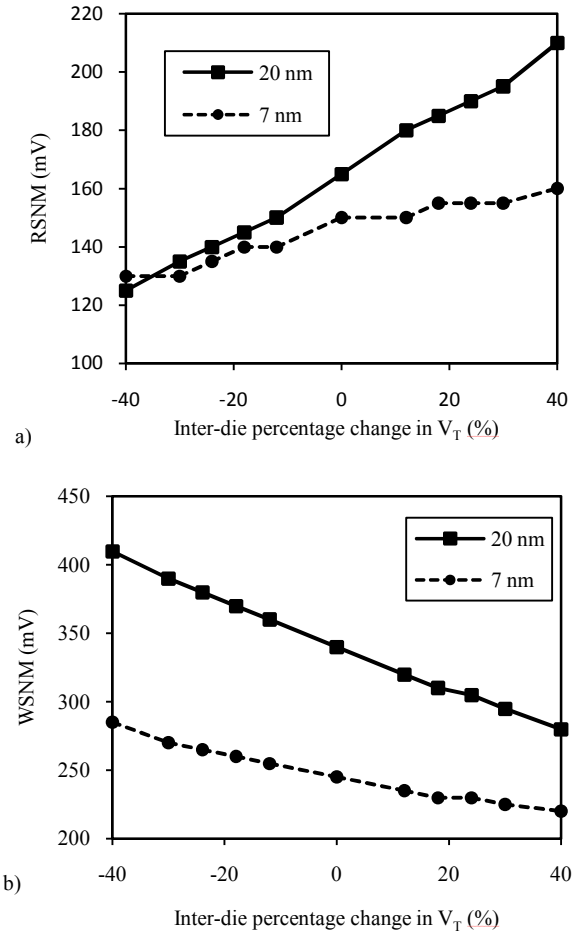

Fig. 6. Sensitivity of the read and write noise margins to the threshold voltage variations for 20nm and 7nm technology nodes; a) RSNM, b) WSNM

Fig.6.b) shows the sensitivity of the WSNM to the V_T variation showing the opposing response to the RSNM as it enhances with decreasing V_T. In addition the percentage change in WSNM for both the technologies is quite closer as compared to the RSNM.

IV. Acknowledgment

This research was partially funded by Zewail City of Science and Technology, AUC, the STDF, Intel, Mentor Graphics, MCIT and the Natural Sciences and Engineering Research Council of Canada (NSERC).

V. Conclusion

The performance of FinFET 6T SRAM of 256-bit cell is evaluated with technology scaling. The impact of a given range of threshold voltage variations on basic performance metrics is reported. The results show that, starting from the 10nm node and down to the 7nm, clear performance degradation is observed in the read operation impacting both the delay and stability metrics. The degradation of the read current per bit-cell with technology scaling as a result of scaling other parameters besides the channel length was seen to be the main reason behind the observed degradation in the read operation.

The study also shows that, with technology scaling, the leakage power occupies larger portion of the total power consumption, however the sensitivity of the leakage to threshold variations is reduced with scaling down the technology.

VI. References

[1] Auth , C. Allen , A. Blattner , D. Bergstrom , M. Brazier , M. Bost , M. Buehler , V. Chikarmane , T. Glassman , R. Grover , W. Han , D. Hanken , M. Hattendorf , P. Hentges , R. Heussner , J. Hicks , D. Ingerly , P. Jain , S. Jaloviar and R. James, "A 22 nm high performance and low-power CMOS technology featuring fully-depleted tri-gate transistors, self-aligned contacts and high density MIM capacitors," in *VLSI Symp. Tech. Dig.*, 2012.

[2] C.-H. Jan, U. Bhattacharya, R. Brain, S.- J. Choi, G. Curello, G. Gupta, W. Hafez, M. Jang, M. Kang, K. Komeyli, T. Leo, N. Nidhi, L. Pan, J. Park, K. Phoa, A. Rahman, C. Staus, H. Tashiro, C. Tsai, P. Vandervoorn, L. Yang, J.-Y. Yeh, P. Bai , "A 22nm SoC Platform Technology Featuring 3-D Tri-Gate and High-k/Metal Gate, Optimized for Ultra Low Power, High Performance and High Density SoC Applications ," in *IEEE International Electron Devices Meeting IEDM* , Dec. 2012, pp. 3.1.1 - 3.1.4.

[3] E. Karl, Y. Wang, Y.-G. Ng, Z. Guo, F. Hamzaoglu, U. Bhattacharya, K. Zhang, K. Mistry, M. Bohr, "A 4.6 GHz 162 Mb SRAM design in 22 nm tri-gate CMOS technology with integrated active VMIN-enhancing assist circuitry," in *IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, Febraury 2012, pp. 230-232.

[4] "International Technology Roadmap of Semiconductors ," http://www.itrs.net/Links/2013ITRS/Home2013.htm, 2013.

[5] Kevin Zhang, Eric Karl, and Yih Wang , "SRAM Design in Nano-Scale CMOS Technologies," in *Symp. VLSI Technol.*, 2012, pp. 85-86.

[6] Seid Hadi Rasouli, Hamed F. Dadgour,Kazuhiko Endo, Hanpei Koike, and Kaustav Banerjee., "Design Optimization of FinFET Domino Logic Considering the Width Quantization Property," *IEEE Trans. Electron Devices*, vol. 57, no. 11, pp. 2934-2943, Nov. 2010.

[7] Mingu Kang, S. C. Song, S. H. Woo, H. K. Park, M. H. Abu-Rahma, L. Ge, B.M. Han,J. Wang, G. Yeap, and S. O. Jung., "FinFET SRAM Optimization With Fin Thickness and Surface Orientation," *IEEE Trans. Electron Devices*, vol. 57, no. 11, pp. 2785-2793, Nov. 2010.

[8] A. Asenov, B. Cheng, X. Wang, A. R. Brown, D. Reid, C. Millar, C. Alexander, "Simulation Based Transistor-SRAM Co-Design in the Presence of Statistical Variability and Reliability," in *Tech. Digest of IEDM*, 2013, pp. 33.1.1-33.1.4.

[9] Xingsheng Wang, Binjie Cheng, Andrew R. Brown, Campbell Millar, Jente B. Kuang, Sani Nassif, Asen Asenov, "Impact of Statistical Variability and Charge Trapping on 14 nm SOI FinFET SRAM Cell Stability," in *ESSDERC*, Bucharest, 2013, pp. 234-237.

[10] Predictive Technology Model (PTM). http://ptm.asu.edu/.