Simulation of Semiconductor Processes and Devices 2016
Edited by E. Bär, J. Lorenz, and P. Pichler

Process Informed Accurate Compact Modelling of 14-nm FinFET Variability and Application to Statistical 6T-SRAM Simulations

Xingsheng Wang[1*], Dave Reid[2], Liping Wang[1], Campbell Millar[2], Alexander Burenkov[3], Peter Evanschitzky[3], Eberhard Baer[3], Juergen Lorenz[3], Asen Asenov[4]

[1] Formerly with University of Glasgow, Glasgow, UK, now with Synopsys, Glasgow, UK
[2] Gold Standard Simulations Ltd. (part of Synopsys), 11 Somerset Place, Glasgow G3 7JT, UK
[3] Fraunhofer IISB, Schottkystrasse 10, 91058 Erlangen, Germany
[4] School of Engineering, University of Glasgow, Glasgow, UK, also with Synopsys, Glasgow, UK
* E-mail: Xingsheng.Wang@synopsys.com

***Abstract*—This paper presents a TCAD based design technology co-optimization (DTCO) process for 14nm SOI FinFET based SRAM, which employs an enhanced variability-aware compact modeling approach that fully takes process and lithography simulations and their impact on 6T-SRAM layout into account. Realistic double patterned gates and fins and their impacts are taken into account in the development of the variability-aware compact model. Finally, global process induced variability and local statistical variability and their impacts are evaluated at the transistor and SRAM levels.**

Keywords—design technology co-optimization; FinFET; process; process variation; SRAM; statistical variability

I. INTRODUCTION

Three dimensional (3D) FinFET technology has been adopted for mass production of advanced applications. Compared with its predecessor, the bulk planar MOSFET, FinFETs provide considerable performance and variability advantages [1]. However, the 3D nature of this device architecture imposes strict requirements on patterning. Double patterning is introduced to enable finer pattern pitches in FinFET technology, but continues to challenge process quality control of critical dimensions (CD) such as gate length and fin width. Long-range global CD variation across cells becomes prominent due to the strong geometrical dependence of the device performance. The inherent discreteness of charge and granularity of matter in individual nanoscale transistors stubbornly continues to cause local statistical variability of transistor characteristics. Modelling the impact of both global variability (GV) and statistical variability (SV) on the performance of FinFETs and circuits is essential to optimal design for yield, and accurate compact models at the early stages of technology development are a must.

In this paper we present the design technology co-optimization (DTCO) process for 14nm SOI FinFET based SRAM, starting from comprehensive process/lithography simulations in the context of SRAM layouts. The key enabler development is an enhanced hierarchical variability aware SPICE compact modeling methodology capable of capturing global and local variation in a 14nm SOI FinFET technology as well as the interactions between them. The capabilities of the resulting compact model are illustrated in SRAM simulations.

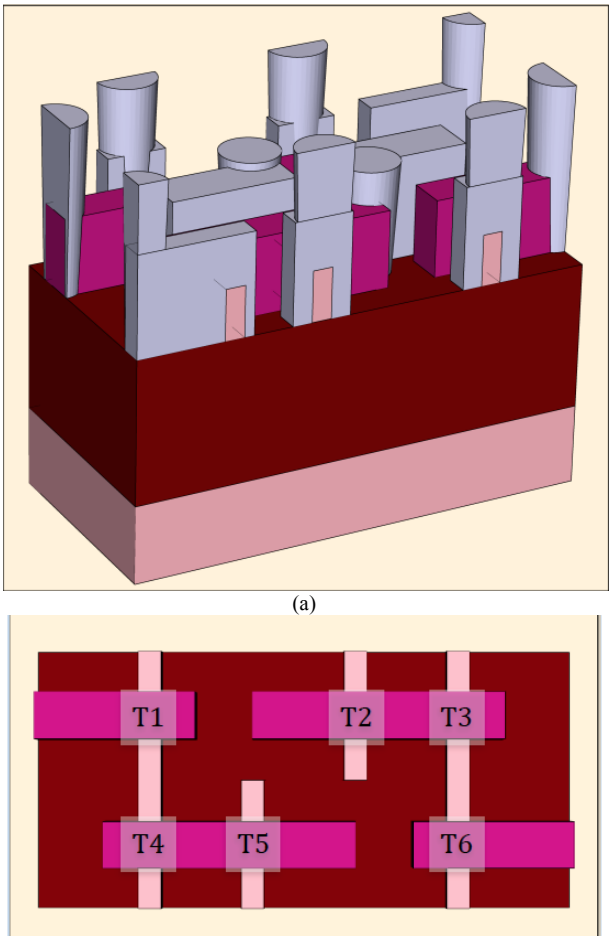

Fig.1. (a) The 6T-SRAM cell built with FinFETs on SOI wafer. The contact pads and plugs below metal 1 (M1) level are shown. (b) The layout of the SRAM cell.

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007 – 2013) under grant agreement no. 318458 SUPERTHEME, and also received funding from European Union's Horizon 2020 research and innovation programme under grant agreement No 687931.

978-1-5090-0818-6/16/$31.00 ©2016 IEEE

II. Process Simulations of FinFETs in SRAM Layout, and Statistical Variability

The FinFET process under consideration has been simulated by Fraunhofer IISB and is described in [2]. A 3D 6T-SRAM cell with a 1-1-1 fin configuration forms the basis of this study and is shown in Fig. 1(a), internally connected with contact pads and plugs under the first metal layer. From the top-view in Fig. 1(b), the top and bottom red horizontal lines are gate lines 1 and 2, while the two pairs of vertical lines are fins. Each cross of two vertical and horizontal lines, generated by two serial processes (fin patterning and gate deposition/patterning, respectively), forms a single fin transistor in the layout. In the 6T-SRAM cell layout shown in Fig. 1(b) T1 and T6 are pass-gate transistors, and T2 and T3 vs T5 and T4 are two pairs of inverters while T2 and T5 are pull-up transistors and T3 and T4 are pull-down transistors.

The patterning process of the FinFETs is done by Self-Aligned Double Patterning for the fins and Litho-Etch-Litho-Etch for the gate, as discussed elsewhere [2]. The nominal design features are listed in Table I while other process parameters and device simulation results are reported in [3]. These geometry parameters, in particular gate-length and fin-width, are subject to process variations, due to factors such as lithography de-focus and neighboring effect, which cause global process variations (GV). Following the process simulation [2], the distributions of device critical dimensions (CD) are examined in Quantile-Quantile plots in Figs. 2&3. Note that gate line 1 (of T1, T2 and T3) and gate line 2 (of T4, T5, and T6) are rendered from two separate Litho-Etch-Litho-Etch patterning sequences. This leads to a difference between gate length Lg 1 and gate length Lg 2. The average of Lg 1 is 0.33nm less than that average Lg 2; while the corresponding standard deviation is 0.09 nm less. The inner spacer defined fin is subject to a small variation with an average of 8.6nm, as shown in Fig. 3, while outer fin width is a constant of 11.3nm.

Table I. The nominal design features of 14nm SOI FinFETs under this study.

L_G (nm)	W_{FIN} (nm)	H_{FIN} (nm)	Equivalent Oxide Thickness (nm)
20	10	25	0.8

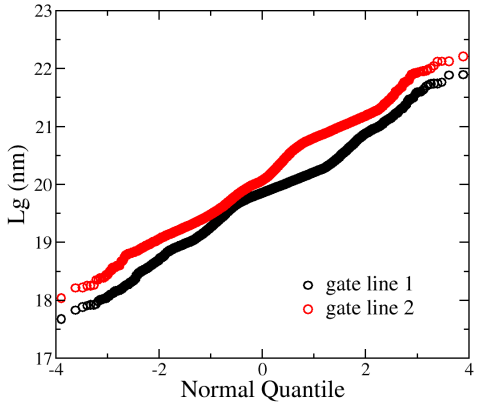

Fig. 2. The gate line width variation of gate line 1 (T1, T2, T3) vs gate line 2 (T4, T5, T6), among different SRAM cells, which results from Litho-Etch-Litho-Etch processes. The sigma of Lg 1 is 0.51nm and the mean is 19.80nm, and the sigma of Lg 2 is 0.60nm and the mean is 20.13nm.

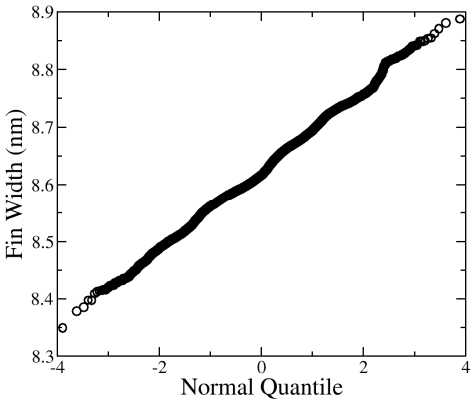

Fig. 3. Fin width (of T2 and T5) variation which is induced in the double patterning of SADP. The mean is 8.62nm, and the sigma is 0.07nm.

Furthermore, individual devices inside the cell are subject to local statistical variability (LV), which arises from the effects of random discrete dopants, line edge roughness (LER) and metal gate granularity (MGG). In the simulations, LER for gate and fin edges is parameterized by 3Δ=2.0nm and correlation length Λ=30nm. MGG is modeled with two work-functions differing by 0.2eV occurring with probability 0.4 and 0.6. As a result, the carrier distribution profiles (and transport) become irregular, and differ from device to device in the cell, as illustrated in Fig. 4. To capture the combined effect of statistical and process variability, we performed a design-of-experiment (DoE) study for (L_G, W_{FIN}, H_{FIN}) (Table II) where max and min deviations are ±10%~20% of corresponding nominal values required by CD control in the process, and at each DoE point 1,000 statistical variability samples were simulated using Garand [4].

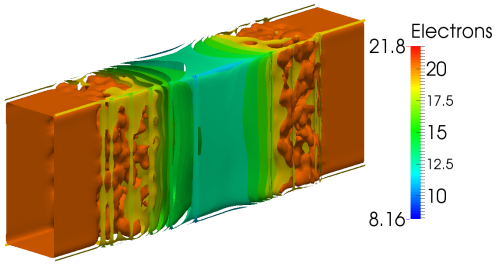

Fig. 4. The electron contour in the fin of an nFinFET, subject to statistical variability sources.

Table II. The design of experiments of physical simulations for global variation.

Parameter	Min	Max
L_G (nm)	18	22
W_{FIN} (nm)	8	12
H_{FIN} (nm)	22	25

III. Hierarchical Variability Macro Compact Models

Variability-aware compact modeling (CM) is a critical step (Fig. 5) in the DTCO flow. It is able to accurately embed the device variability into circuit design and verification. We use the BSIM-CMG SPICE model v106 to extract the model card

of the corresponding 14 nm SOI FinFET technology transistors [5]. The variability aware CM extraction process consists of the extraction of a comprehensive nominal uniform model (Fig. 6) with no variability present, CM extractions over the CD space defined by the DoE and statistical extraction of atomistic devices using orthogonal sets of compact model parameters as described in [3]. Thereafter, as shown in Fig. 7 and 8, the extracted parameters and their correlations are used to construct response surface models over the DoE space and can be used to generate process aware statistical compact models using ModelGen technology [4] The resulting macro CM can respond to process global variation and can generate as many corresponding statistical model cards as required.

Utilizing the process simulation results described in section II, especially the CDs of gate-length and fin-width for cells, each set of CD values for SRAM cell transistors is fed to the macro CM, and 100 statistical samples are simulated with statistical variability for each transistor. With combined global and local variation, the correlation of V_T between n and p devices varies considerably depending on the particular combination (e.g. T4 vs T5 in Fig. 9), bringing significant complexity in circuit simulation and making it necessary to revisit the definitions of traditional corners (e.g. FF, SS, FS and SF). Here since statistical variability induced V_T fluctuation is much larger than that from global variation, SV leads to the decorrelation of n-p devices in contrast to the strong correlation due to GV alone.

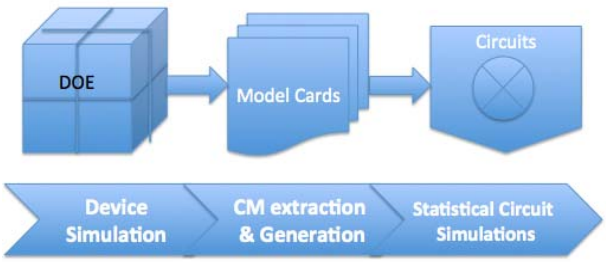

Fig. 5. The workflow of DTCO from device simulations, compact model extraction and generations, to statistical circuit simulations.

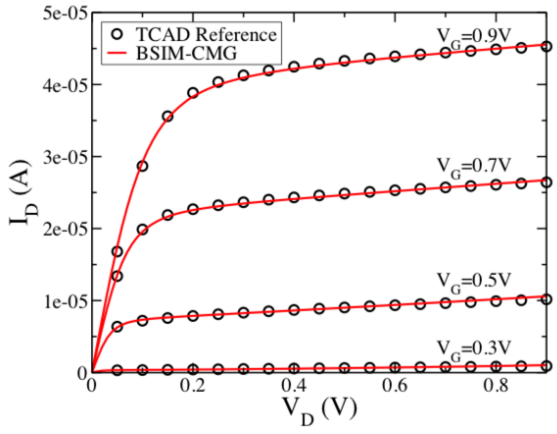

Fig. 6. The nominal uniform compact model accuracy is monitored. Here it is showing the comparison between TCAD simulation and SPICE simulation of compact model in terms of I_D-V_D characteristics.

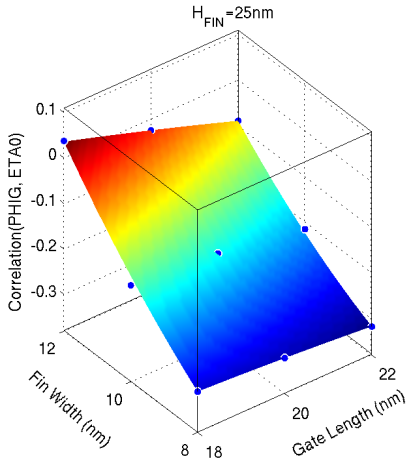

Fig. 7. The response surface of correlations of statistical CM parameters.

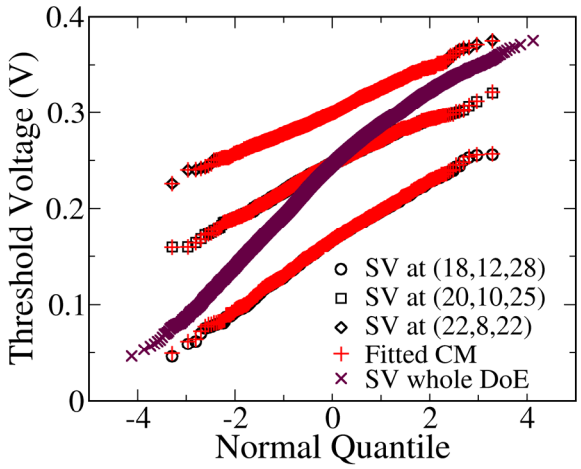

Fig. 8. The comparison between distributions of figures of merit (FoM) from statistical CM and TCAD over several corners in DoE. It also shows the statistical V_T distribution over whole DoE.

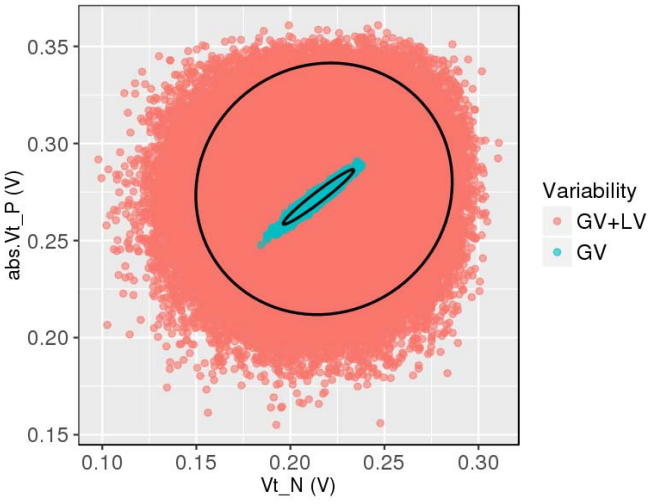

Fig. 9. The saturation V_T correlation between T4 (NMOS) and T5 (PMOS) subject to global process variation and combined global and statistical variability. Ellipses indicate 1 sigma.

IV. FinFET Based SRAM Simulations

The 1-1-1 6T-SRAM cells in a large array are used to illustrate the capabilities of this methodology. On top of each of 10^4 CD sets (long-range global variations), 100 microscopically statistically different cells are configured, thus in total 10^6 cells are simulated. The static noise margin (SNM) is simulated under the impact of both process-induced global and random local variations. For example, the butterfly curves for 100 microscopically different SRAM cells with one global process condition are shown in Fig.10. The results of SPICE simulation in the presence of global variations are presented in Fig.11. Since gate line 2 (gate length) is on average larger than gate line 1, left side SNM is correspondingly 2mV less than the other side SNM. However, this global asymmetry is overwhelmed by the impact of statistical variation when these effects are combined (Fig. 12), as V_T variation introduced by statistical variability is significantly larger than that from process variation (Fig.9). The statistics of the distribution of SNM are summarized in Table III where it is clear that combined GV+LV produces significantly more SNM variation than the individual sources. It is clear that GV causes the systematic mismatch in the read noise margin of two store nodes, and LV further increases the risk of failure of SRAM because of randomness of each store node.

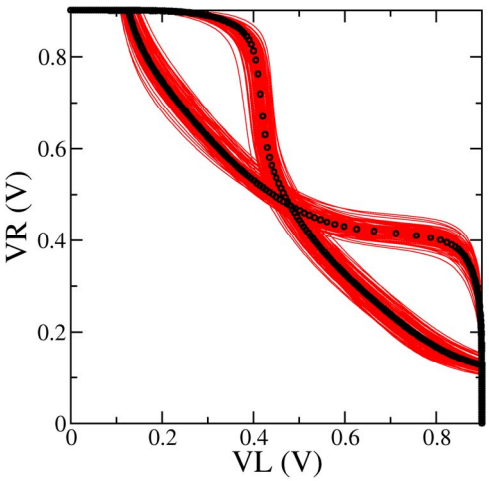

Fig. 10. The statistical butterfly characteristics of two store nodes due to local variability in one case of global (L_G, W_{FIN}, H_{FIN}=25) inputs.

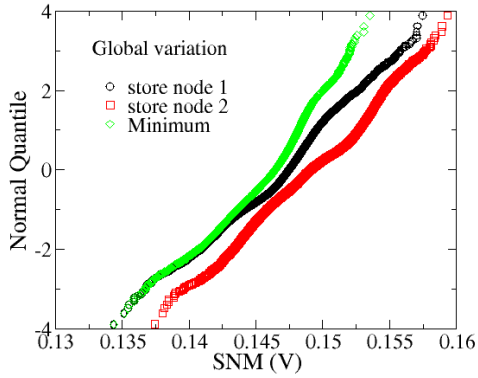

Fig. 11. The Q-Q plot of SNM distributions subject to global process variation.

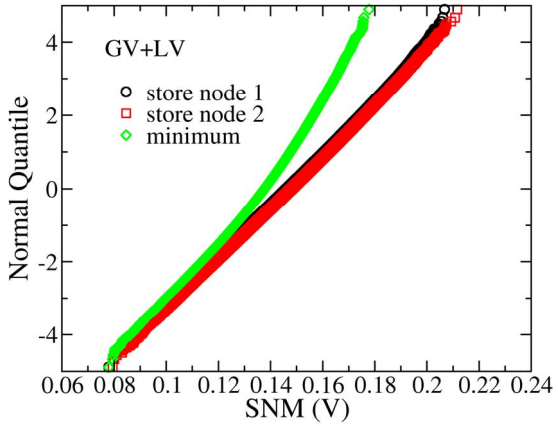

Fig.12. The Q-Q plot of SNM distributions due to the total global and statistical variability.

Table III. The statistical results of SNM extracted from circuit simulations of 6T SRAM associated with global variation and total variability.

SNM (mV)	GV	GV+LV
Mean_left	147.11	145.80
Mean_right	149.22	147.80
Mean_min	146.03	136.95
Stdev_left	2.83	14.46
Stdev_right	3.14	14.41
Stdev_min	2.36	10.72

V. Conclusions

This paper investigated the impact of realistic process (e.g. lithography) and statistical variability on SRAM read stability through the developed hierarchical macro compact model technology. The DTCO flow demonstrated here using the variability-aware macro compact model is used to successfully evaluate and differentiate the distinct impact of global CD and statistical variability on SRAM SNM. Global CD asymmetry can cause systematic mismatch, upon which the statistical variability of each transistor significantly skews the n-p device corrections and increases the SRAM failure risk.

References

[1] C. Auth, et al. “A 22nm High Performance and Low-Power CMOS Technology Featuring Fully-Depleted Tri-Gate Transistors, Self-Aligned Contacts and High Density MIM Capacitors,” in *Proc. Symposium on VLSI Technology*, 2012, pp.131.

[2] E. Baer, A. Burenkov, P. Evanschitzky, J. Lorenz, “Simulation of process variations in FinFET transistor patterning,” in *Proc. 21st International Conference on Simulation of Secmiconductor Processes and Devices (SISPAD)*, 2016. Accepted.

[3] X. Wang, B. Cheng, D. Reid, A. Pender, P. Asenov, C. Millar and A. Asenov, “FinFET Centric Variability-Aware Compact Model Extraction and Generation Technology Supporting DTCO,” *IEEE Transactions on Electron Devices*, Vol. 62, No. 10, pp. 3139–3146, Oct. 2015.

[4] GSS tool chain. http://www.goldstandardsimulations.com/products/

[5] BSIM-CMG, http://www-device.eecs.berkeley.edu/bsim/