Hierarchical Variability-Aware Compact Models of 20nm Bulk CMOS

Xingsheng Wang[1*], Dave Reid[2], Liping Wang[1], Alex Burenkov[3], Campbell Millar[2], Juergen Lorenz[3], Asen Asenov[1,2]
[1] Device Modelling Group, School of Engineering, University of Glasgow, Oakfield Ave., Glasgow G12 8LT, UK
[2] Gold Standard Simulations Ltd., 11 Somerset Place, Glasgow G3 7JT, UK.
[3] Fraunhofer IISB, Schottkystrasse 10, 91058 Erlangen, Germany
* E-mail: Xingsheng.Wang@glasgow.ac.uk

***Abstract*—This paper presents a hierarchical variability-aware compact model methodology based on a comprehensive simulation study of global process variation and local statistical variability on 20nm bulk planar CMOS. The area dependence of statistical variability is carefully examined in the presence of random discrete dopants; gate line edge roughness; metal gate granularity; and their combination. Hierarchical variability-aware compact models have been developed, extracted and used to evaluate the impact of process variation and statistical variability on SRAM stability and performance.**

Keywords—mismatch, MOSFET, process variation, SRAM, statistical variability

I. INTRODUCTION

Statistical variability (SV) associated with the discreteness of charge and granularity of matter is a significant challenge to advanced transistor integration and circuit design [1][2]. The dominant statistical variability sources identified via experimental measurement and TCAD simulation include random discrete dopants (RDD), gate line edge roughness (LER), and metal gate granularity (MGG) [3]. Although FinFETs deliver performance advantages and reduced statistical variability, bulk planar MOSFETs remain in mass production including the 20 nm CMOS technology, partly due to their low manufacturing cost. Bulk planar MOSFETs are subject to significant random statistical variability, due to high channel doping. Moreover, lithography imperfections and process deviations can cause global long-range process variations, which interact with statistical variability and complicate the issue. Although there are numerous studies of individual sources of variations, the interplay of global process variation and statistical variability in the bulk technology has not been comprehensively investigated. In addition, variability-aware compact models are required to accurately capture transistor variability and to facilitate variability aware circuit design [4][5]. In this paper, we utilise the Sentaurus Process [6] and the GSS tool chain [7] to systematically study both global process variation and purely statistical local variability in 20 nm bulk MOSFETs. Hierarchical variability-aware compact models based on TCAD simulations are extracted, and used to accurately evaluate the impact on SRAM cells.

II. TCAD SIMULATION METHOD

Process simulations of the example 20 nm bulk planar CMOS transistors used in this study are carried out using Sentaurus process, and then transferred to GSS device simulator Garand [7]. The transistors have a 23.5nm physical gate length and 33 nm channel width. The realistic transistor structure includes shallow trench isolation (STI), incorporated into the simulation domain, which is critical in order to accurately capture narrow width effects. Simulations of NMOS and PMOS transistors are benchmarked against industrial 20 nm bulk CMOS MOSFET technology [8], with the NMOS result shown in Fig. 1.

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.

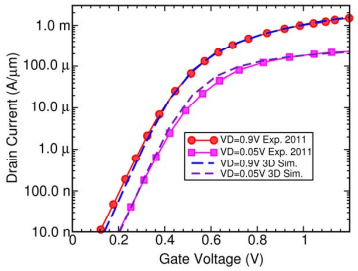


Fig. 1. The benchmark of Id-Vg characteristics against to industry 20nm bulk MOSFETs.

Table I. DoE for Global Variations of Gate Length and Channel Width.

L (nm)	17	20.25	23.5	26.75	30
W (nm)	24	28.5	33	37.5	42

In order to capture the gate-length L and channel-width W designs in the layout and their long-range process-induced critical dimension (CD) deviations due to lithography imperfections we use a design of experiments (DoE) approach. A Cartesian DoE includes discrete L and W values and serves as an input for device simulation. As the DoE L and W values are listed in Table 1 covering the large range of possible process and layout dependent CD variation space. Therefore, a total of 25 transistors with different dimensions are simulated to cover the DoE space.

Based on each uniform transistor in the DoE space, 1000 "atomistic" microscopically different transistors are simulated using GARAND. The dominant statistical variability sources including RDD, LER, and MGG are simulated in combination. In the simulations, 2.0 nm LER with 30 nm correlation length is used to model gate edge roughness [9], and two types of metal grains with different orientations, work function

978-1-4673-7860-4/15/$31.00 ©2015 IEEE

variation difference of 0.2eV, 0.4/0.6 occurrence, and 5nm average grain size account for the TiN/high-k gate stack granularity [10]. An example of one of these devices in the presence of RDD, LER and MGG is shown in Fig. 2.

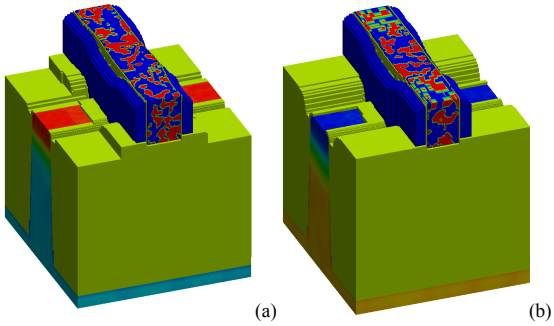


Fig. 2. The atomistic simulations of n-channel (a) and p-channel (b) 20nm bulk MOSFETs. It includes random dopants, gate line edge roughness, and metal gate granularity.

The full I_D-V_G characteristics are simulated, and the corresponding figures of merit (FoM) are extracted for each point in the DoE space. Fig. 3 shows the sets of I-V characteristics including statistical variability at different (L, W) corners. Off-state leakage currents can vary by several orders of magnitude, and on-current can change by a factor of more than 100 %.

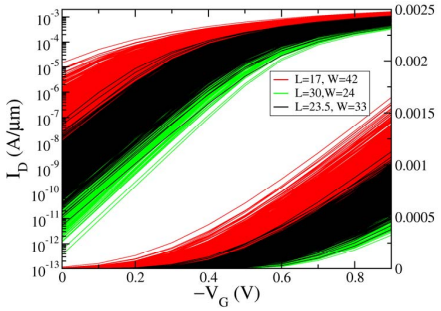


Fig. 3. Examples of "atomistic" simulations of I_D-V_G characteristics at high drain bias at typical, fast and slow corners of PMOS.

To quantitatively examine the impact of statistical variability sources, the average values and standard deviations of key FoM are extracted and examples are shown in Fig. 4 and Fig. 5. Over the DoE process variation space the threshold voltage (V_T) roll-off is prominent. As expected, V_T is reduced with a decrease in channel width [11][12]. Due to the response of V_T to CD variations, I_{ON}/W also varies significantly. The standard deviations of FoM directly reflect the impact of statistical variability. It is clear that with the reduction of channel length the standard deviation of threshold voltage (σV_T) increases, while its width dependence is more complicated. σV_T ranges from 38mV to 68.5mV over the CD process variation space. However, for σI_{ON} the change is significant along channel length with the largest σI_{ON} at (L=17, W=42) nm, equivalent to nearly 15% of the corresponding average value.

If we consider the individual contributions of different statistical variability sources to σV_T (as shown in Fig. 6), RDD is obviously the major source, while MGG and LER contribute almost equally to threshold voltage fluctuations. It is also important to note that SV-induced V_T variations do not always follow a Gaussian distribution. For example, LER-induced V_T variations have a distribution that is negatively skewed.

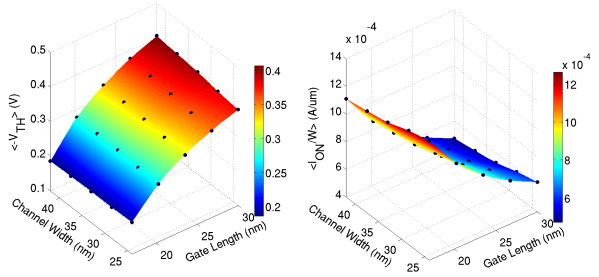


Fig. 4. The average figures of merit of I_D-V_G characteristics over (L, W) for PMOS.

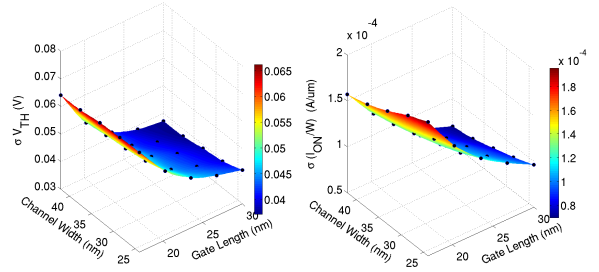


Fig. 5. The standard deviations of figures of merit of I_D-V_G characteristics over (L, W) for PMOS.

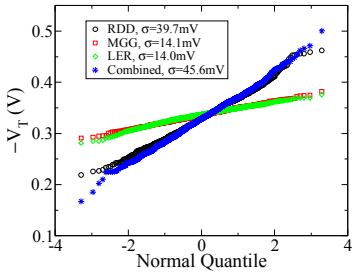
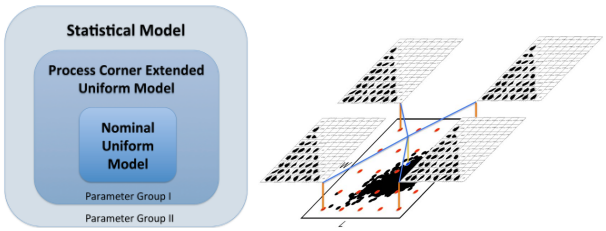


Fig. 6. Normal quantile-quantile plot of threshold voltages due to various statistical variability sources for the nominal dimension PMOS.

III. Hierarchical Variability-Aware Compact Models

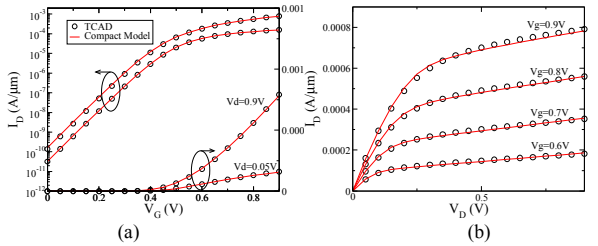
In order to accurately capture statistical variability, a hierarchical variability-aware compact modelling approach is adopted to extract variability information. As illustrated in Fig. 7(a) the extraction strategy involves three key steps: 1) extraction of a comprehensive nominal uniform model; 2) extension of response surface process variation model using a minimum group of parameters; 3) statistical extraction of statistical "atomistic" sample at corresponding process

variation corner using a second, different group of model parameters. Shown in Fig. 7(b) the L, W are inputs of the hierarchical variability-aware compact models, and the global variation is covered by the DoE points, upon which the statistical local variation information is extracted. The continuous information is obtained by interpolation from the nearby discrete DoE points. As shown in Fig. 8, the nominal I_D-V_G transfer characteristics are fitted accurately, and match the key FoMs, such as threshold voltage, subthreshold slope, DIBL, drain currents.



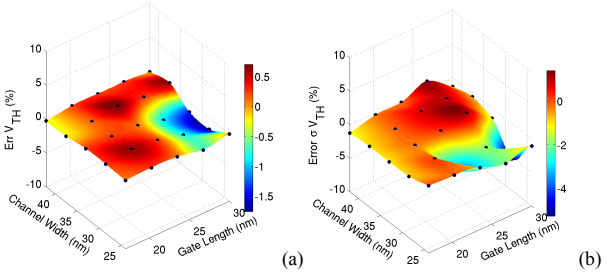
(a) (b)

Fig. 7. The hierarchical variability-aware compact modelling extraction and generation.



(a) (b)

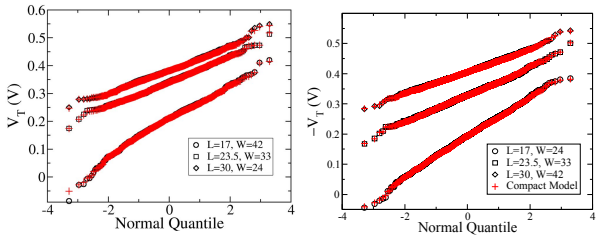
Fig. 8. The compact model extraction of nominal design n-MOSFET. (a) is the I_D-V_G characteristics; (b) is the I_D-V_D characteristics.



(a) (b)

Fig. 9. Compact model threshold voltage extraction errors in process variation (a) and statistical variability (b), compared with TCAD.

The I_D-V_G characteristics of transistors with global process variation are also matched well, for example the error in V_T over (L, W) space is between 2% and 6% as shown in Fig. 9(a). The statistical variability is also accurately captured. As illustrated in Fig. 10, it compares the distribution of V_T obtained from statistical compact models to that obtained from the 'atomistic' TCAD simulations at different nodes of the DoE space. Our hierarchical variability compact model accurately captures the local statistical variability, the global CD design and process variation, and their interplay. The realistic area-dependence shown in Fig. 5 is also captured by the extracted models.



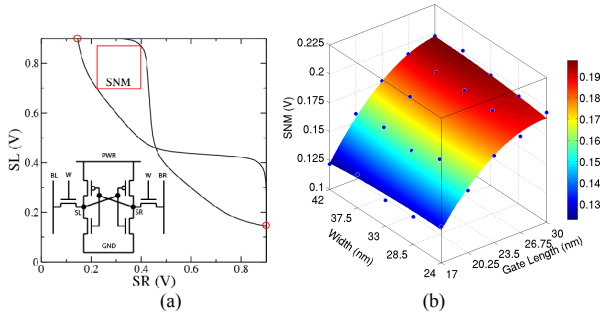
(a) (b)

Fig. 10. Statistical compact model extraction at different process variation corners for nMOS (a) and pMOS (b).

IV. Circuit Examples

A. Static noise margin of 6T-SRAM

We use the developed hierarchical variability aware compact models to simulate a 6T-SRAM test circuit, since SRAM is the most vulnerable circuit to variability, as it generally uses minimum sized transistors. As shown in Fig. 11(a) it is composed of two complementary inverters and two access NMOS transistors. The butterfly characteristics are characteristic for the read mode operation. Static noise margin (SNM) for one node (state) is the lateral length of the corresponding maximum box inside the immune region, for example the SNM for SL=high is shown in Fig. 11(a). The cell SNM is usually defined as the minimum of SNM for the two storage conditions. For the demonstration we assume that the 6 n/pMOSFETs in the SRAM cell have identical size. We at first study the geometry dependence of the SNM as shown in Fig. 11(b). It is clear that the larger L and W change SNM from 122mV to 200mV.

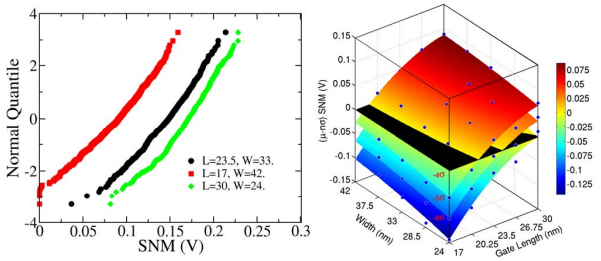


(a) (b)

Fig. 11. (a) The 6T-SRAM butterfly characteristics, with an inset of the schematic view of SRAM under test; (b) Impact of global process variation on SRAM SNM.

In the SPICE circuit simulations SRAM cell transistors are subjected to global systematic process variation in L_G and W, and then each resulting transistor is subjected to statistical variability. It is known that statistical variability causes mismatch of the two inverters in the SRAM cell, therefore SNM is defined as the minimum SNM of the two nodes. The

SNM distribution is known to have an elongated left tail (negative skew) [13][14] due to the min operation in the SNM calculation, as is shown in Fig. 12(a), but in this study we use a measure of a single 'eye' in the butterfly curve to aid statistical analysis. At first global process CD variations shift the mean of SNM and modulate the standard deviation from 29.8mV to 32.6mV to 36.5mV from the (L=30,W=24) nm to typical design to (L=17, W=42) nm. In Fig. 12(b), at each process variation point, the -4σ, -5σ, -6σ value of the distribution of the one-sided SNM is plotted as a function of L and W. A zero plane is shown for reference, as zero SNM indicates complete cell read failure, and therefore yield loss. At L=17-20.25nm, there is a large yield loss even at -4 sigma. Therefore the lithography/process improvement to control CD process variation is required to increase the yield.

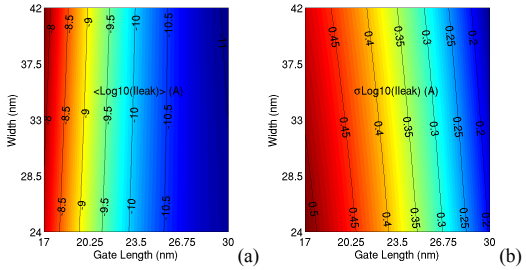


(a) (b)

Fig. 12. (a) The Q-Q plots of the distribution of the minimum SNM obtained from statistical compact models. (b)The -4σ, -5σ, -6σ of one-node SNM over (L, W) variation space.

B. *Leakage currents*

In the retention period the current leaks from the power supply to the ground through the SRAM cell, which accounts for a large proportion of the static power dissipation. Due to local statistical variability of the cell transistors, cell leakage currents display randomness even with the same global CD variation. We obtained the leakage current distributions of corresponding SRAM cells associated with L and W global variations using variability aware models in SPICE simulation. The distributions of the log of the leakage current follow a close to Gaussian distribution, and their statistics calculated for (L, W) CD process variations are shown in Fig.13. Cell leakage current variation is dominated by gate-length variation, and it spans 3 orders of magnitude from ~10^{-11} A at L = 30 nm to ~10^{-8} A at L = 17 nm on average, while the standard deviation has the largest value of 0.51 at (L=17, W=24) nm and the smallest value of 0.16 at (L=30, W=42) nm. From the point of view of reducing static power dissipation, process deviation, especially those leading to shorter gate-length should be tightly controlled.



(a) (b)

Fig. 13. (a) The average and (b) standard deviation of leakage currents of SRAM cells with different global CD variations.

V. CONCLUSIONS

In this paper, using Synopsys Sentaurus Process and the GSS tool chain, we have studied the statistical variability in the presence of global process variations for 20nm bulk planar MOSFETs. The area dependence of statistical variability has been thoroughly examined. Hierarchical variability aware compact models has been extracted to accurately capture the area-dependence of statistical variability in order to evaluate SRAM yield and power dissipation in the presence of global process variation.

REFERENCES

[1] B. H. Calhoun, A. P. Chandrakasan, "Static Noise Margin Variation for Sub-threshold SRAM in 65-nm CMOS", IEEE Journal of Solid-State Circuits, Vol. 41, No. 7, July 2006

[2] K. Agarwal, S. Nassif, "The Impact of Random Device Variation on SRAM Cell Stability in Sub-90-nm CMOS Technologies", IEEE Trans. On VLSI Systems, Vol. 16, No. 1

[3] X. Wang, et al., "Simulation Study of Dominant Statistical Variability Sources in 32-nm High-κ/Metal Gate CMOS," IEEE Electron Device Letters, Vol. 33, No. 5, pp. 643–645, May 2012.

[4] B. Cheng, et al., "Statistical Variability Compact Modeling Strategies for BSIM4 and PSP," IEEE Design and Test of Computers, Vol. 27, No. 2, pp. 26–35, Mar./Apr. 2010.

[5] A. Asenov, B. Cheng, X. Wang, A.R. Brown, C. Millar, C. Alexander, S. M. Amoroso, J.B. Kuang, and S. Nassif, "Variability aware simulation based design-technology co-optimisation (DTCO) flow in 14 nm FinFET/SRAM co-optimisation," IEEE Transactions on Electron Devices. Vol. 62 No. 6, pp.1682-1690, June 2015.

[6] Sentaurus Mannual, Synopsys.

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

[8] H. -J. Cho et al., "Bulk Planar 20nm High-K/Metal Gate CMOS Technology Platform for Low Power and High Performance Applications," in Proc. IEDM, 2011, pp.350-353.

[9] A. Asenov, S. Kaya and A. R. Brown, "Intrinsic parameter fluctuations in decananometer MOSFETs introduced by gate line edge roughness," IEEE Transactions on Electron Devices, Vol. 50, No. 5, pp. 1254–1260, 2003.

[10] X. Wang, et al., "Statistical Threshold-Voltage Variability in Scaled Decananometer Bulk HKMG MOSFETs: A Full-Scale 3-D Simulation Scaling Study," IEEE Transactions on Electron Devices, Vol. 58, No. 8, pp. 2293–2301, Aug. 2011.

[11] X. Wang, et al., "Impact of STI on Statistical Variability and Reliability of Decananometer MOSFETs," IEEE Electron Device Letters, Vol. 32, No. 4, pp. 479–481, Apr. 2011.

[12] J. Zhou, et al., "The impact of inverse narrow width effect on sub-threshold device sizing," in Proc. ASP-DAC, 2011, pp.267-272.

[13] P. Asenov, et al., "The Effect of Compact Modelling Strategy on SNM and Read Current variability in Modern SRAM," in Proc. SISPAD, 2011, pp.283-286.

[14] X. Wang, et al., "Statistical Variability in 14-nm node SOI FinFETs and its Impact on Corresponding 6T-SRAM Cell Design," in Proc. ESSDERC, 2012, pp. 113–116.