FinFET Centric Variability-Aware Compact Model Extraction and Generation Technology Supporting DTCO

Xingsheng Wang, *Member, IEEE,* Binjie Cheng, *Member, IEEE,* Dave Reid, Andrew Pender, Plamen Asenov, Campbell Millar, *Member, IEEE,* and Asen Asenov, *Fellow, IEEE*

***Abstract*— In this paper we present a FinFET focused variability-aware compact model extraction and generation technology supporting design-technology co-optimization (DTCO). 14-nm CMOS technology generation SOI FinFETs are used as test-bed transistors to illustrate our approach. The TCAD simulations include long-range process-induced variability using a design of experiment (DoE) approach and short-range purely statistical variability (mismatch). The compact model extraction supports a hierarchical compact model approach including nominal compact model extraction, response surface compact model extraction and statistical compact model extraction. The accurate compact model generation technology captures the often non-Gaussian distributions of the key transistor figures of merit and their correlations preserving also the correlations between process and statistical variability. The use of the hierarchical compact model is illustrated in the simulation of FinFET based SRAM cells and ring oscillators.**

***Index Terms*—compact model, FinFET, interplay, process variation, ring oscillator, SRAM, statistical variability**

I. Introduction

VARIABILITY is the maker or breaker of advanced CMOS technology and design [1][2]. Understanding, controlling and mastering variability in technology development and accurately factoring variability in the design process is key to the commercial success of both foundries developing the technology, and the fabless companies designing circuits and systems for the broad range of electronic products. Various types of variability affect advanced CMOS technology and influence the design of corresponding chips accordingly [3][4]. The long-range process induced variability is on the increase due to the sub-lithography resolution size of transistors and interconnects - this is known as global variability (GV). The introduction of FinFETs steps up the challenge requiring accurate process control in three dimensions (3D). Purely statistical local variability (LV), also known as mismatch, is also on the increase, due to atomic scale sensitivity of the decananometer size transistor affected by random discrete dopants, material granularity and lithography imperfections [5]-[9]. In between the long-range (process-induced) variability and the short-range (statistical) variability there is also systematic layout induced variability. The challenge is that all variabilities are strongly inter-correlated, particularly in the FinFET technology case [10]. Treating these sources - and effects - as independent leads to significant pessimism, which can go a significant way towards canceling out the benefits of scaling.

Better and more accurate models that capture the different components of variability and their correlations are needed to enable more accurate and efficient design in contemporary and future CMOS technology generations. The traditional design approach of guaranteeing circuit integrity by simulating over all process corners can result in so much guard-banding that the design becomes severely compromised within the available constraints [11][12]. Improved models should support Design Technology Co-Optimization (DTCO), which has become 'must' in advanced CMOS [13]-[16]. They should also allow the yield to be factored in the analysis and verification and should enable performance/power/area/yield (PPAY) trade off and sign off [17] [18]. We reported the TCAD based study of global variation and statistical variability and their interplay in FinFETs [10], and TCAD based DTCO flow for nanoscale FinFET technologies [19], however, variability-aware compact model strategy employed in the advanced DTCO process, is not yet systematically published in detail. In this paper we present the details of the FinFET centric variability-aware compact modelling extraction and generation technology critical for the advanced DTCO process, accurately taking care of both global process-induced and local statistical variability and their correlation.

In this paper we present an advanced compact model (CM) extraction and generation technology aiming at high statistical accuracy and capturing the correlation between process and statistical variability. Although the approach is not inherently

Manuscript received 6 May 2015, revised July 10, 2015, aceepted 28 July 2015. This work was partially supported by Scottish Funding Council through project "StatDes" (HR10018).

X. Wang is with Device Modelling Group, School of Engineering at University of Glasgow, Oakfield Avenue, Glasgow G12 8LT, U.K. (phone: +44 141 330 2964; e-mail: xingsheng.wang@glasgow.ac.uk).

B. Cheng, D. Reid, A. Pender, P. Asenov, C. Millar are with Gold Standard Simulations Ltd., 11 Somerset place, Glasgow G3 7JT, U.K.

A. Asenov is with Gold Standard Simulations Ltd., 11 Somerset place, Glasgow G3 7JT, U.K; also Device Modelling Group, School of Engineering, University of Glasgow, Oakfield Avenue, Glasgow G12 8LT, UK (e-mail: asen.asenov@glasgow.ac.uk).

technology specific, here it is targeted at cutting edge FinFET nodes. TCAD data generated with the GSS 'atomistic' simulator GARAND [20] is used as the target for the compact model extraction, and to illustrate the FinFET specific correlations between long-range and short-range variability. The compact model extraction is carried out with the GSS compact model extractor MYSTIC [20]. The compact model generation technology is implemented in the GSS statistical circuit simulation engine RandomSpice [20].

In the next section we briefly introduce the TCAD simulation methodology and the test-bed FinFETs used to illustrate the compact model extraction and generation approach. Section III outlines the hierarchical compact model extraction approach, including parameter identification. The statistical compact model generation technology is described in Section IV. Circuit simulation examples are presented in section V, before conclusions are outlined in Section VI.

II. Test-Bed FinFETs and TCAD Simulation

TCAD simulations have become an essential part of compact model development. They improve the accuracy of the compact model at early stage of technology development. Carefully calibrated TCAD simulations avoid the pitfalls in identification of 'golden' transistor targets and allow physics based re-centering of the compact model during rapid technology development. TCAD is also essential in statistical compact model extraction since in many cases 'clean' statistical current-voltage characteristics are difficult or impossible to measure.

The test-bed transistors in this study are 14-nm CMOS technology generation SOI FinFETs, described in detail in [10]. They feature a nominal physical gate length of 20 nm and a high-k metal gate stack with equivalent oxide thickness of 0.8 nm. The nominal fin height and fin-width are 25 and 10 nm respectively. The channel is lightly doped with a concentration of 10^{15} cm^{-3}, while the source/drain have a maximum doping of 3×10^{20} cm^{-3}. The nominal FinFETs deliver drive currents of ~0.9/0.8 mA/µm and an off-current of 10 nA/µm (at T= 85˚C) for the n/p-channel FinFETs, with drain induced barrier lowering (DIBL) of 56/65 mV/V respectively. The GSS 'atomistic' TCAD simulator GARAND has been employed to generate the target current-voltage and capacitance-voltage characteristics used in the next section. The corresponding TCAD simulations are described in great detail in [10].

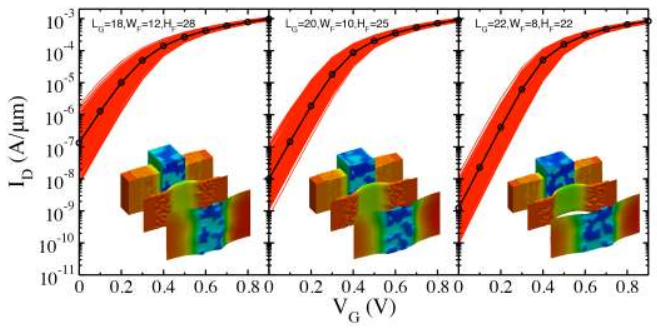


Fig. 1 Statistical target I_D-V_G characteristics at 3 nodes of the DoE space. The inset illustrates one microscopically different device from the sample, within which the front slice shows the potential across the gate oxide modulated by MGG, and the middle slice shows the electron density along the mid-channel.

Both long-range process induced GV and short-range random LV are taken into account when generating the compact model extraction targets. The correlations between the GV and LV is captured in a design of experiment (DoE) assuming 3σ process induced long-range variations of ΔL_G = ±2 nm; ΔW_{FIN} = ±2 nm; ΔH_{FIN} = ±3 nm forming a 3 × 3 × 3 = 27 node Cartesian product space. The global variations of L_G, W_{FIN} and H_{FIN} are 10%-20% of nominal values which are consistent with CD control while a relatively large fin-width variation is given due to the difficulty of fin patterning in the FinFET manufacturing. At each DoE point the characteristics of the "uniform" FinFET and 1,000 microscopically different atomistic devices subject to statistical variability sources of random discrete dopants (RDD), gate edge roughness (GER), fin edge roughness (FER) and metal gate granularity (MGG) were simulated. Fig. 1 shows statistical target I_D-V_G characteristics at 3 nodes of the DoE space. The inset illustrates one microscopically different device from the sample. The DoE and the generation of the full set of target I-V characteristics are fully automated using the GSS automation tool Enigma [20].

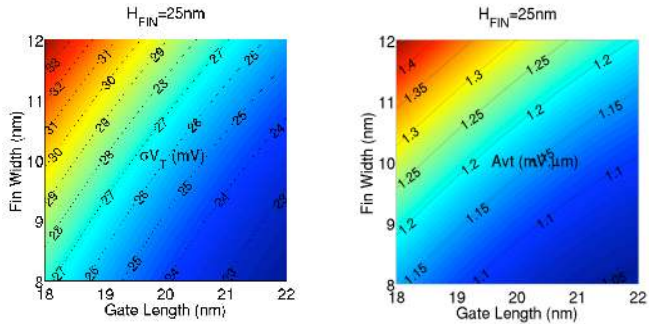


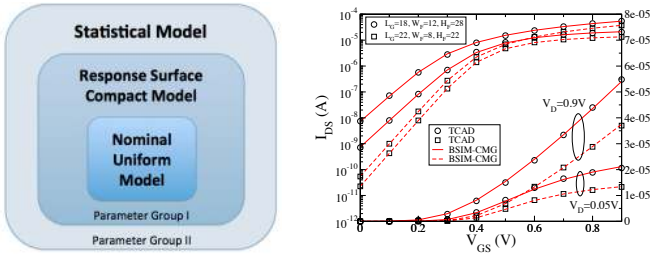
Fig. 2 Threshold voltage standard deviation σV_T and mismatching factor A_{VT} as a function of L_G and W_{FIN} at H_{FIN}=25nm [10].

FinFET performance varies significantly across the DoE and the statistical variability shows a large dependence on the long-range process variation leading to the deviations from Pelgrom's area law [21] as illustrated in Fig. 2 for the V_T mismatch. A high degree of automation, enabled by the GSS tool flow is essential for handling the large amount of data and cluster interaction associated with hierarchical compact model extraction. In this particular case, 27,000 statistical current voltage characteristics were generated to cover the DoE space. The data are automatically harvested in a common database and automatically accessed by the compact model extractor MYSTIC.

III. Hierarchical Compact Model Extraction

In order to handle the interplay between long-range GV and the short range LV, we have developed a hierarchical unified variability compact model strategy. It consists of three key extraction steps illustrated in Fig. 3(a): (i) extraction of full

nominal compact model; (ii) extraction of response surface GV compact model that covers the process variation space and; (iii) extraction of statistical LV models at each node of the DoE space. Following the accurate extraction of nominal uniform model, two minimal but separate groups of model parameters are employed to execute step 2 and step 3, capturing GV and LV respectively. Therefore, responding to the different requirements and objectives, this hierarchical strategy can provide variability models in various levels of complexity and accuracy in combinations.



(a) (b)

Fig. 3: (a) Schematic view of hierarchical variability compact modelling strategy including two groups of model parameters; (b) Models extracted at two extreme corners of the DoE space (L_G=18nm,W_{FIN}=12nm, H_{FIN}=28nm) and (L_G=22nm,W_{FIN}=8nm, H_{FIN}=22nm).

The surface potential based BSIM-CMG compact model for SOI FinFETs [22] is adopted in this study. First, a comprehensive model extraction procedure is applied to the designed nominal FinFET in the centre of the DoE space. The model is fitted against comprehensive TCAD simulation data of FinFET I_D-V_G and I_D-V_D characteristics and the relevant C-V characteristics for a range of operating temperatures. It precisely reproduces the transistor electrical characteristics including the sub-threshold characteristics, drain currents, and output resistance and drain-bias effects like DIBL. The overall fitting error is 1.86%.

To encompass the large range of GV, we deploy the first group of a few parameters to extract FinFET of corresponding geometry at every DoE point, extending the applicability of the nominal model. The reason for the use of geometry dependent parameters to handle the GV effects is the relative simplicity of BSIM-CMG, which cannot fit accurately the fin width dependent quantum confinement effect. A DoE macro model is then created by fitting a polynomial regression to each of the Group I model parameters, as a function of the DoE parameters. Fig. 4 shows the fitting error over DoE space, which is kept below 2%.

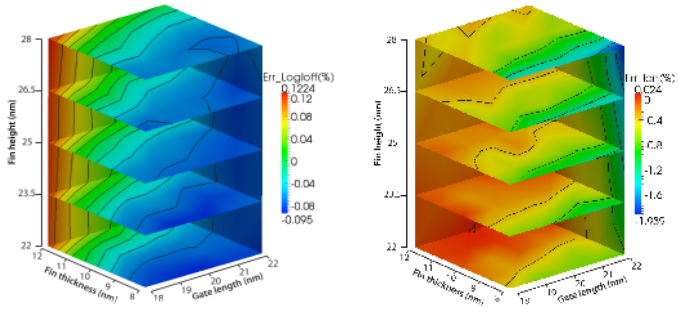
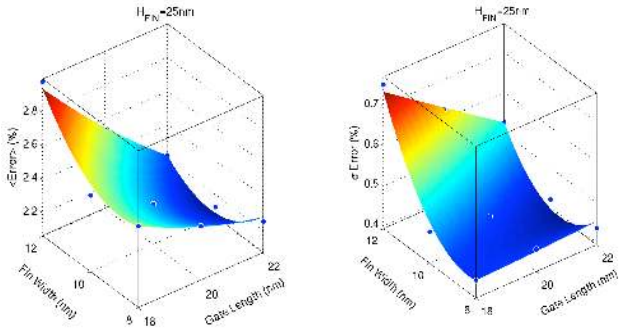


Fig. 4: Errors between the response surfaces for the off- and on-currents across the DoE space obtained from the TCAD simulation and the DoE extended response surface compact model of n-channel FinFET.

A second group of model parameters is used to extract statistical LV models of the simulated atomistic devices based on the previously obtained uniform model. The parameters are selected to capture essential variations in the key transistor figures of merit introduced by the different sources of statistical variability including random discrete dopants, line edge roughness and metal gate granularity. The same direct extraction procedure is applied to each point of the DoE, thus we have 27 sets of statistical compact models, for which extraction accuracy is closely monitored. As shown in Fig. 5(a), the average fitting error of the high/low drain I_D-V_G of 1,000 atomistic FinFETs is well controlled: below 3% across the entire DoE and the error standard deviation is below 0.75 %.



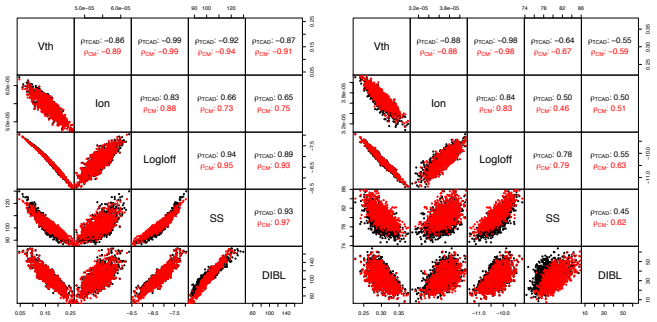
(a) (b)

Fig. 5: I-V fitting average error (a) and its standard deviation error (b) from the statistical compact model extraction across the DoE space.

Careful selection of CM extraction parameters is required in order to accurately capture variability effects in statistical compact models. Group I CM parameters are chosen to capture global variations over the DoE space. The instance parameter L and model parameters TFIN and HFIN are selected to represent corresponding physical variations in device dimensions, but the complex physics involved in the devices and the analytical simplifications necessary in a CM require the additional parameters to account for the full effect of global variation. These include the quantum correction factor QMFACTOR, threshold voltage parameter DVT1 and DIBL related parameter DSUB. RDSW is a source drain resistance model parameter and VSAT1 deals with velocity saturation. Group II statistical extraction parameters are utilized to accurately capture the impact of statistical Local Variations on electrical characteristics. PHIG is used to capture threshold voltage fluctuation, and EAT0 is for DIBL variation induced mainly by MGG and RDD. U0 and UA model low field transport variation, CDSC is responsible for variation in the subthreshold slope, and VSAT takes care of velocity saturation.

Fig. 6 illustrates the distributions and the correlations between the essential figures of merit obtained from the

TCAD "atomistic" simulations and from the response surface compact macro-model at two extreme DoE corners (L_G=18nm, W_{FIN}=12nm, H_{FIN}=28nm) and (L_G=22nm, W_{FIN}=8nm, H_{FIN}=22nm). Although the fitting at (L_G=22nm, W_{FIN}=8nm, H_{FIN}=22nm) could be the worst, both cases have achieved good matching between TCAD data and extracted CM.



(a) (b)

Fig. 6 Distributions and Correlations between the essential figures of merit obtained from the TCAD simulation and from the response surface compact macro-model at two extreme DoE corners (a) (L_G=18nm, W_{FIN}=12nm, H_{FIN}=28nm) and (b) (L_G=22nm, W_{FIN}=8nm, H_{FIN}=22nm).

IV. Compact Model Generation

The statistical circuit simulation engine RandomSpice is used to carry out the statistical compact model generation using GSS ModelGen technology, which handles non-Gaussian statistical parameter distributions with complex correlations [20]. The purpose of the statistical compact model generation is to circumvent the limitations associated with using finite input samples. Primarily, it avoids sub-sampling errors artificially distorting the output distribution. This is particularly of concern when evaluating the occurrence of rare events. As we will discuss in detail below, it also allows LV and GV to be continuously modelled across the entire DoE space.

Simpler methods, such as principal component analysis (PCA) and non-linear power method (NPM) [23]-[27], can be used for model generation, however these have intrinsic limitations. PCA assumes Gaussian marginal distributions and therefore introduces significant errors when modelling strongly asymmetric distributions. NPM models both non-Gaussian distributions and correlations, however it can suffer stability problems, and is limited in the range of distribution shape parameters it can capture. GSS ModelGen avoids all of these issues by employing a more flexible distribution fitting method and is therefore used exclusively in the remainder of this work. The full flow for combined GV and LV model generation is described in Fig. 7, in which GV selection is relevant to Fig. 8 and LV interpolation is relevant to Fig. 10.

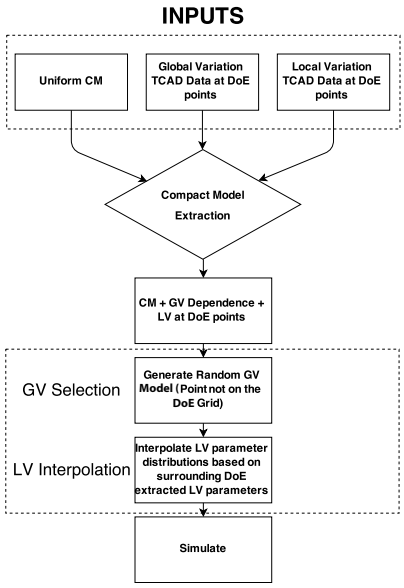


Fig. 7: Flow chart for GV and LV aware statistical model, ensuring accuracy across the whole GV space and capturing GV/LV correlation at arbitrary GV points.

We first consider GV in isolation. The method for generation of an accurate GV compact model is based on the response surface compact model. The Group I parameters used in the GV model extraction are all assumed to be dependent on length, width and height. Then, for a given circuit instance, we randomly select a process geometry for N- and P-FET. The associated compact model parameters for GV are then determined using the response surface model.

This concept is illustrated in Fig. 8. For simplicity, in this case H_{FIN} is fixed to 25 nm and only L_G and W_{FIN} are considered as random variables. Please note that the method handles arbitrary number of process parameters with arbitrary distributions and correlations.

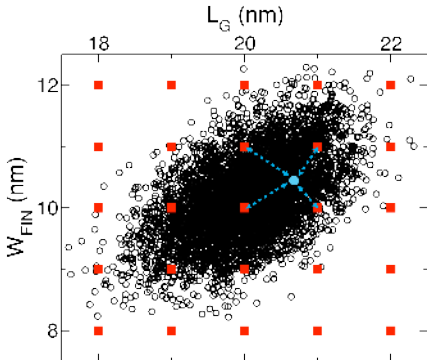


Fig. 8: GV model generation for a fixed H_{FIN} of 25 nm when L_G and W_{FIN} are considered as random variables. A single GV point is selected in blue. The arrows show how the nearby DoE points are used to calculate the GV model responding to the selection, using a response surface created based on the GV parameter distributions extracted at these points.

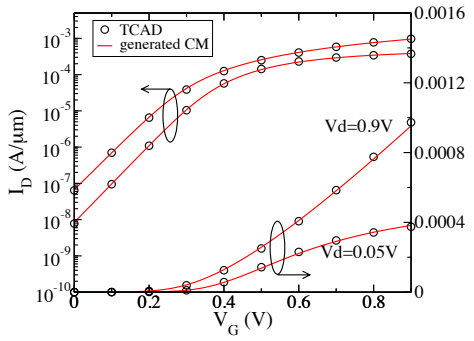


Fig. 9: Comparison between generated and TCAD simulated transistor characteristics with a random set of L_G (=18.5nm) and W_{FIN} (=11.5nm) with H_{FIN}=25nm.

As indicated above, L_G and W_{FIN} are randomly selected based on defined distributions. One such instance is illustrated with a blue dot in Fig. 8. As we first consider GV only, the same random instances selected for the N- and P-channel transistors are used throughout a given instance of the simulated circuit. In other words, GV is assumed to affect all transistors in a circuit instance the same way. Fig. 9 compares the generated compact model for one random instance of L_G and W_{FIN} and the results of the TCAD simulation for the same L_G and W_{FIN}.

We then include both GV and LV together. It should be noted that the model parameters associated with LV are dependent on the specified process variables (i.e. on the GV). The method for generation of accurate statistical compact models capturing LV at the selected point of the process variation space (shown in Fig 8) is illustrated in Fig. 10. Here L_G and W_{FIN} considered as random process variables. To determine the local variations at the randomly selected L_G and W_{FIN}, the extracted statistical parameter distributions at each of the nearest nodes of the DoE space (denoted by the blue arrows in Fig 10) and their correlations are interpolated to the selected process geometry. This information is used to generate an arbitrary number of statistical compact models that accurately reproduce the interpolated non-Gaussian statistical parameter distributions and their correlations. In contrast to GV alone, each model in a given instance will be distinct in this scenario.

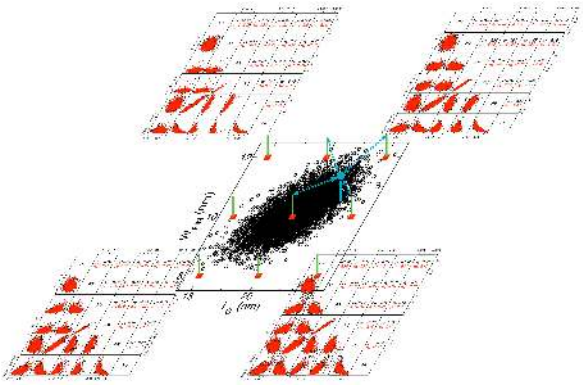
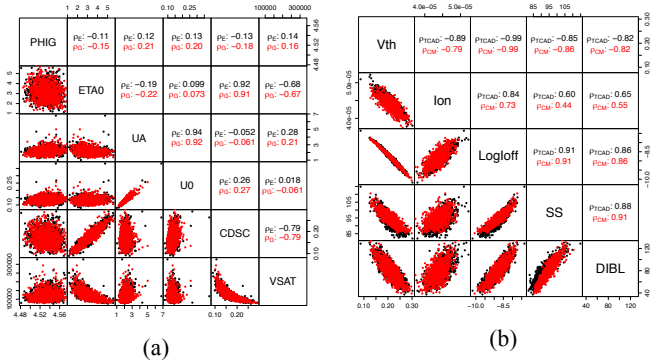


Fig. 10 Statistical compact model generation based on the extracted statistical compact model distribution at each node of the DoE space. Although the distributions in all 9 DoE nodes in the H_{FIN} =25 nm plane are used only 4 of them are illustrated in this picture for simplicity.

In order to validate the LV generation methodology, we have performed statistical TCAD simulations at a randomly generated point for example L_G=19nm, W_{FIN}=11nm, H_{FIN}=23.5nm, which is not among the statistical extraction grid points. Fig. 11 (a) illustrates the generated and the extracted distributions of the LV compact model parameters showing excellent agreement in distribution and in the shape of the correlations. In general correlation coefficients are accurately captured although there are small deviations of ~+/-0.05 in some cases. These are considered to be acceptable since given the intrinsic noise associated with the numerical calculation of correlation coefficients based on a fixed sample size and that the calculation involves two parameters with a standard deviation error of ~2.3% when generated from 1000 samples. Fig. 11 (b) illustrates the distribution and the correlations of the key FinFET figures of merit extracted from the statistically generated compact models at the randomly selected L_G and W_{FIN} and the corresponding TCAD simulations. The few FoM correlation deviations of around 0.1 are the cumulative result of CM parameter extraction, numerical calculation and generation errors.



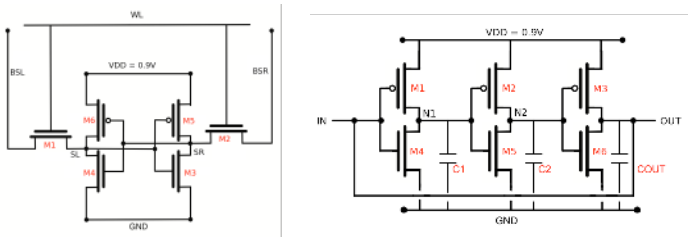
(a) (b)

Fig. 11: (a) The extracted (E) and generated (G) distributions of the LV compact model parameters at the randomly selected L_G and W_{FIN} (b) The distribution and the correlations of the key FinFET figures of merit extracted from the statistically generated compact model at the randomly selected L_G and W_{FIN} and the corresponding TCAD simulations.

V. Circuit Simulation Examples

In this section we will demonstrate the use of the developed compact model extraction and generation technology for statistical circuit simulations with RandomSpice that capture the individual and combined effect of GV and LV and their correlations. At the circuit level, the tracking of nFET to pFET correlation plays an important role in circuit variability [28], which can be naturally taken into account by our models and reproduced accurately at the compact model generation stage. As a test case study in this section, global L_G, W_{FIN} and H_{FIN} variations are perfectly correlated between nFinFET and pFinFET, and here we use the same global variation value of each L_G, W_{FIN} and H_{FIN} for both nFinFET and pFinFET. As a circuit simulation demonstrator we will use an ultra high-density 6T SRAM bitcell using a 1-1-1 fin sizing,

illustrated in Fig. 12 (a). The half-access static noise margin (SNM) serves as the metric of the bitcell stability, and the circuit simulation using nominal CM models shows that nominal 1-1-1 bitcell has an SNM of 160.9 mV. We also use a 3 stage ring oscillator (RO) as a demonstrator as shown in Fig. 12 (b). The oscillation frequency of the nominal RO is 3.09GHz, with a load capacitance of 3.0 fF.

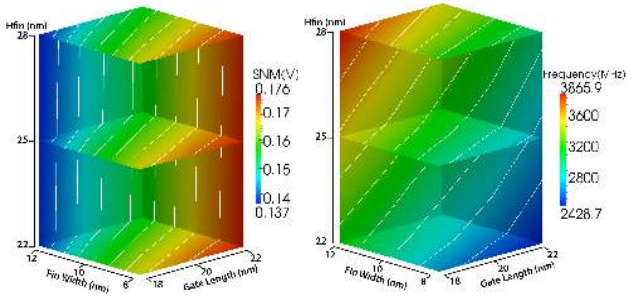


(a) (b)

Fig. 12: The schematic views of the 6T-SRAM bitcell (a) and the ring oscillator (b).

A. *Global Process Induced Variability (GV)*

Global process induced variation affects all transistors inside a circuit cell in a consistent way. The response surface compact model allows the handling of the process variability of the test circuit behavior. It is at first useful to understand the deterministic impact of process variations on the circuits. Fig. 13 illustrates the response of the SRAM bitcell static noise margin and RO frequency on L_G, W_{FIN}, H_{FIN} in the 3σ process variation space.

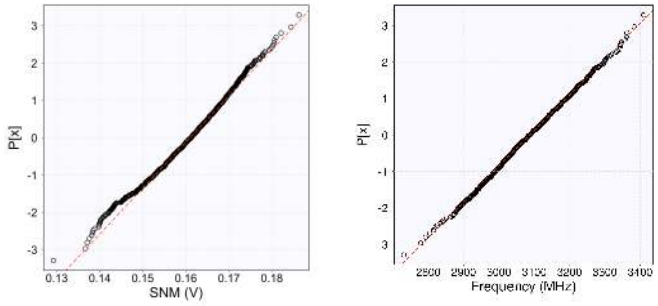


(a) (b)

Fig. 13: (a) Dependence of the SRAM static noise margin on L_G, W_{FIN}, H_{FIN} in the 3σ process variation space; (b) Dependence of the RO frequency on L_G, W_{FIN}, H_{FIN} in the 3σ process variation space.

It is clear that global process variation causes a significant reduction of SNM from 176mV at slow corner (L_G=22, W_{FIN}=8, H_{FIN}=22) to 137 mV at fast corner (L_G=18, W_{FIN}=12, H_{FIN}=28). Fig. 13 (b) illustrates the dependence of the RO frequency on L_G, W_{FIN}, H_{FIN} in the 3σ process variation space. On the contrary, the fastest RO is consisted of FinFETs with (L_G=18, W_{FIN}=12, H_{FIN}=28). These devices characterize to have the largest on-current and in return render the smallest skew delay of inverters. The RO frequency varies from 2.43 GHz to 3.87 GHz compared to the 3.09 GHz by the nominal design, rendering variations from -21% to 25%.

While exploring global variations in an explicit way is instructive in determining circuit performance, the study of how process-induced GV affects the statistical distribution of the circuit performance given a particular set of process conditions is beneficial to determine yield. For the purpose of this example, the distributions of L_G, W_{FIN}, H_{FIN} is modelled by a Gaussian distribution centered on the nominal process value and no correlations between the process parameters are taken into account. Fig. 14 (a) illustrates the GV induced distribution of the bitcell SNM while Fig. 14(b) illustrates the GV induced distribution of the RO frequency. SNM shows some deviations from Gaussian distribution while 3-stage RO frequency closely follows Gaussian distribution up to 3σ. The mean value/standard deviation of SNM is 8.25/160.6 mV (5%), and the mean value/standard deviation of RO frequency is 102/3089 MHz (3.3%).

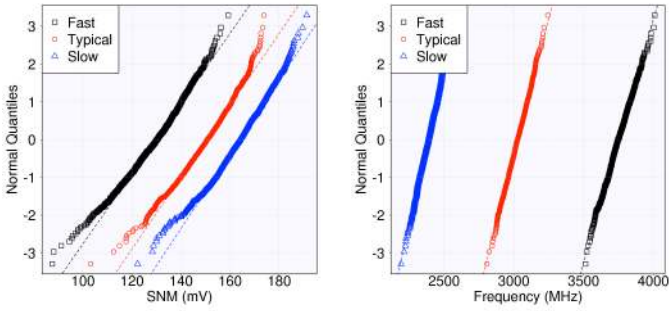


(a) (b)

Fig. 14 (a) GV induced distribution of the SRAM SNM. (b) GV induced distribution of the RO frequency.

B. *Local Statistical Variability (LV)*

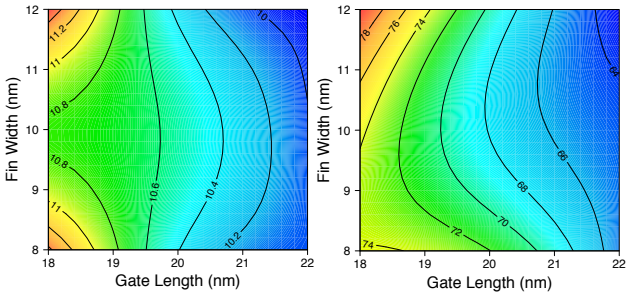
One of the main features of the developed hierarchical compact model extraction and generation technology is the accurate representation of the correlations between GV and LV. In order to illustrate this in Fig. 15(a) we present the LV induced statistical distribution of the SRAM SNM at the nominal (L_G=20nm, W_{FIN}=10nm, H_{FIN}=25nm) and the two extreme FinFET geometries of fast corner and slow corner. Different from the GV impact, shown in Fig. 15 (a) local statistical variability impacts each transistor differently, therefore it causes increased skew in the bitcell stable states, degrading SNM much more dramatically that GV alone. This is illustrated by the negative skewness of the LV based SNM distribution, while global variation of n/p-FinFETs entirely shifts and modulates SNM distributions. The sigma/mean values of SNM distributions at typical corner are 10.55/150 mV, and they are 11.13/129 mV at fast corner and 10.34/163 mV at slow corner respectively. Shown in Fig. 15 (b) RO frequency follows the Gaussian distribution, and sigma/mean values vary from 80/3752 MHz at fast corner, 68/3028MHz at typical corner, to 60/2388MHz at slow corner respectively.



(a) (b)

Fig. 15 (a) LV induced statistical distribution of the SRAM SNM at the nominal (L_G=20nm, W_{FIN}=10nm, H_{FIN}=25nm) and the two extreme FinFET geometries (L_G=18nm, W_{FIN}=12nm, H_{FIN}=28nm) and (L_G=22nm, W_{FIN}=18nm, H_{FIN}=22nm); (b) LV induced statistical distribution of the RO frequency at equivalent process conditions.

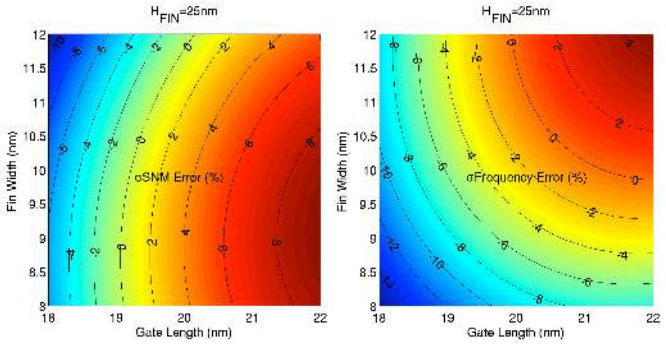
Fig. 16 (a) shows the distribution of the bitcell SNM in the L_G, W_{FIN} plane at H_{FIN}=25 nm, while Fig. 16 (b) show the corresponding RO frequency distribution. Although the distributions appear complicated, the global gate-length variation has the determining impact.



(a) (b)

Fig. 16 (a) Distribution of bitcell SNM sigma (in mV) in the L_G, W_{FIN} plane at H_{FIN}=25 nm. (b) The corresponding RO frequency sigma distribution in the unit of MHz in the L_G, W_{FIN} plane at H_{FIN}=25 nm.

The interplay of GV and LV produces a significant impact on circuit performance, power and yield. In order to demonstrate the importance of the accurate modeling of the interaction of GV and SV, we have simulated both SRAM SNM and 3 stage RO oscillation frequency applying the V_T-fluctuations obtained at the Typical-Typical design point to the entire DoE space. The statistics of two metrics are shown in Fig.17 for H_{FIN}=25nm. Assuming a fixed level of variation over the DoE space leads to the errors of -19% to 10% in the standard deviation of SNM and errors of -13% to 14% in the standard deviation of RO frequency with an additional small effect on the mean circuit performance.

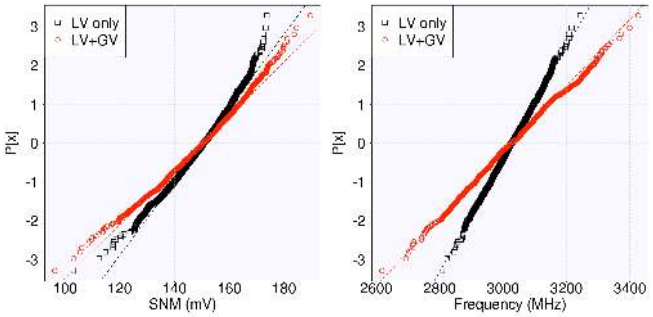


(a) (b)

Fig. 17. (a) Standard deviation errors of SRAM SNM (a) and RO frequency (b) across (L_G, W_{FIN}) at H_{FIN}=25 nm by applying threshold voltage fluctuation of nominal design to DoE space without including interplay of GV and LV.

C. *Combined Variability Impact*

Finally, we combine random global variations with local mismatch by turning on both the statistical LV and process GV input. Fig. 18(a) illustrates the combined GV and LV induced distribution of the bitcell SNM while Fig. 18(b) illustrates the combined GV and LV induced distribution of the RO frequency. It is clear with both global and local variability the SNM distribution become less skewed compared to only with statistical variability, but undergoes larger standard deviation of 13.91mV compared with 10.55mV. The variation of 3-stage RO frequency is simply boosted from 68 to 120 MHz when imposing additionally global variations.



(a) (b)

Fig. 18. (a) Combined GV and LV induced distribution of the SRAM SNM. (b) GV and LV induced distribution of the RO frequency.

In the analysis of the combined impact of random GV and LV on circuit performance, from the subsections A and B, it is clear that random GV has stronger impact on variation of RO frequency than LV, and the statistical LV produces more variation to SRAM SNM than GV. Thus, in order to reduce the impact on overall circuit performance, the reductions of LV and GV have corresponding benefits in the control of variations and yield of SRAM and RO frequency. From the design point of view the developed methodology provides a significant optimization tool by allowing designers to assess the balance between circuit performance, variations and yield in the DoE space.

VI. CONCLUSIONS

We have successfully developed a TCAD based variability-aware comprehensive compact model extraction and generation technology based on a hierarchical compact model. High levels of automation allow seamless extraction of all compact model parameters needed for the compact model generation process. We have demonstrated that accurate response surface model can be extracted using only a small number of compact model parameters. We have also demonstrated that accurate statistical compact models that represent the statistical distribution of the key transistor figures of merit can be extracted at any node of the DoE space. The extracted parameters are used to generate unlimited number of statistical compact models capturing the interplay between local and global variability. The accuracy of the generated statistical models is validated by performing TCAD simulations at the selected process variability point which is not present in the original DoE TCAD simulations. We have

illustrated the importance of the developed method in circuit simulations examples of SRAM and ring oscillators. By the demonstrative examples, the developed variability-aware compact model methodology and technology provide the tool kit for optimizing among circuit performance metrics and variations in the design space for the design technology co-optimization process.

REFERENCES

[1] C. Auth, C. Allen, A. Blattner, 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. Symp. VLSI Tech. Dig., 2012, pp.131-132.

[2] N. Planes, O. Weber, V. Barral, et al., "28 nm FDSOI technology platform for high-speed low-voltage digital applications," in Proc. Symp. VLSI Technol., Jun. 2012, pp. 133–134.

[3] In E. Karl, Z. Guo, Y.-G. Ng, J. Keane, U. Bhattacharya, K. Zhang, "The impact of assist-circuit design for 22nm SRAM and beyond," in *Proc. IEEE IEDM*, 2012, pp.561-564.

[4] B. Nikolic, J.-H. Park, J. Kwak, et al., "Technology variability from a design perspective," *IEEE Trans. Circuits and systems—I: regular papers*, vol.58 no.9, pp.1996-2009, Sept. 2011.

[5] A. Asenov, "Random dopant induced threshold voltage lowering and fluctuations in sub-0.1 μm MOSFETs: A 3D "atomistic" simulation study," *IEEE Trans. Electron Dev*. Vol. 45, No. 12, pp. 2505–2513, 1998.

[6] H. F. Dadgour, K. Endo, V. K. De, K. Banerjee, "Grain-Orientation Induced Work Function Variation in Nanoscale Metal-Gate Transistors—Part I: Modeling, Analysis, and Experimental Validation," IEEE Trans. Elec. Dev., vol.57 no.10, pp.2504-2514, Oct. 2010.

[7] X. Wang, A. R. Brown, N. M. Idris, S. Markov, G. Roy and A. Asenov, "Statistical Threshold-Voltage Variability in Scaled Decananometer Bulk HKMG MOSFETs: A Full-Scale 3-D Simulation Scaling Study," *IEEE Trans. on Electron Devices*, Vol. 58, No. 8, pp. 2293–2301, Aug. 2011.

[8] T. Matsukawa, Y. Liu, W. Mizubayashi, et al., "Suppressing Vt and Gm variability of FinFETs using amorphous metal gates for 14nm and beyond," in Proc. IEDM, 2012, pp.8.2.1-8.2.4

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

[10] X. Wang, B. Cheng, A. R. Brown, C. Millar, J. B. Kuang, S. Nassif, and A. Asenov, "Interplay between process-induced and statistical variability in 14-nm CMOS technology double-gate SOI FinFETs," *IEEE Trans. on Electron Devices*, Vol.60 No.8, pp.2485-2492, August 2013.

[11] S. Nassif, "Design for variability in DSM technologies," in Proc. ISQED, 2000, pp.451-454.

[12] T. Austin, V. Bertacco, D. Blaauw, and T. Mudge, "Opportunities and Challenges for better than worst-case design," in Proc. ASP-DAC, 2005, pp.I-1-7.

[13] S.-Y. Wu, C. Y. Lin, M. C. Chiang, et al., "An Enhanced 16nm CMOS Technology Featuring 2nd Generation FinFET Transistors and Advanced Cu/Low-k Interconnect for Low Power and High Performance Applications," in *Proc. IEEE IEDM*, 2014, pp.3.1.1-3.1.4.

[14] S. Natarajan, M. Agostinelli, S. Akbar, et al., "A 14nm Logic Technology Featuring 2nd-Generation FinFET transistors, Air-Gapped Interconnects, Self-Aligned Double Patterning and a 0.0588 m2 SRAM cell size," in *Proc. IEEE IEDM*, 2014, pp.3.7.1-3.7.4.

[15] C.-H. Lin, B. Greene, S. Narasimha, et al., "High Performance 14nm SOI FinFET CMOS Technology with 0.0174μm2 embedded DRAM and 15 Levels of Cu Metallization," in *Proc. IEEE IEDM*, 2014, pp.3.8.1-3.8.4.

[16] 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 *Proc. IEEE IEDM*, 2013, pp.818-821.

[17] S. Yang, L. Ge, J. Lin, et al., "High Performance Mobile SoC Design and Technology Co-Optimization to Mitigate High-K Metal Gate Process Induced Variations," in *Symp. VLSI Tech. Dig*., 2014, pp.1-2.

[18] X. Jiang, et al., "New Assessment Methodology Based on Energy-Delay-Yield Co-Optimization for Nanoscale CMOS Technology," IEEE Trans. Elec. Dev., vol.62 no.6, pp.1746-1753, 2015.

[19] A. Asenov, et al., "Variability aware simulation based design-technology co-optimization (DTCO) flow in 14 nm FinFET/SRAM co-optimization," *IEEE Transactions on Electron Devices*, Vol. 62 No. 6, pp.1682-1690, June 2015.

[20] The GSS tool chain. [online] http://www.goldstandardsimulations.com/products/

[21] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, "Matching properties of MOS transistors," IEEE J. Solid-State Circuits, vol. 24, no. 5, pp. 1433–1439, Oct. 1989.

[22] BSIM-CMG 106, [online] http://www-device.eecs.berkeley.edu/bsim/?page=BSIMCMG

[23] K. Takeuchi and M. Hane, "Statistical compact mode parameter extraction by direct fitting to variations," IEEE Trans. Elec. Dev., vol.55 no.6, pp.1487-1493. June 2008.

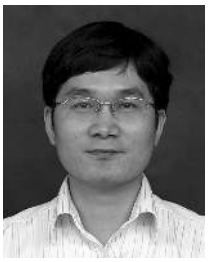
[24] B. Cheng, D. Dideban, N. Moezi, C. Millar, G. Roy, X. Wang, S. Roy and A. Asenov, "Statistical Variability Compact Modeling Strategies for BSIM4 and PSP," IEEE Design and Test of Computers, Vol. 27, No. 2, pp. 26–35, Mar./Apr. 2010.

[25] X. Wang, B. Cheng, A. R. Brown, C. Millar, C. L. Alexander, D. Reid, J. B. Kuang, S. Nassif and A. Asenov, "Unified Compact Modelling Strategies for Process and Statistical Variability in 14-nm node DG FinFETs," in Proc. 18th International Conference on Simulation of Semiconductor Processes and Devices, 2013, pp. 139–142.

[26] U. Kovac, D. Dideban, B. Cheng, N. Moezi, G. Roy and A. Asenov, "A Novel Approach to the Statistical Generation of Non-normal Distributed PSP Compact Model Parameters using a Nonlinear Power Method," in Proc. Simulation of Semiconductor Processes and Devices (SISPAD), 2010, pp. 125–128.

[27] X. Wang, B. Cheng, A. R. Brown, C. Millar and A. Asenov, "Accurate Simulations of the Interplay Between Process and Statistical Variability for nanoscale FinFET-based SRAM Cell Stability," in Proc. 44th European Solid-State Device Research Conference (ESSDERC), Venice Italy, Sept. 2014, pp. 349–352.

[28] J. Deng, et al., "SOI FinFET nFET-to-pFET Tracking Variability Compact Modeling and Impact on Latch Timing," IEEE Trans. Electron Devices, vol.62 no.6, pp.1760-1768, June 2015.



Xingsheng Wang (M'11) received the Master degree in mathematics from Tsinghua University, Beijing, China, in 2007, and the Ph.D. degree in electronics and electrical engineering from the University of Glasgow, Glasgow, U.K., in 2010. He is currently with the School of Engineering, University of Glasgow, where he has the research interest in nanoscale transistors, statistical variability and reliability, and device circuit co-design.

Binjie Cheng (M'11) received the B.S. and M.S. degrees in optical and physical electronic engineering, and the Ph.D. degree in electronic science and technology from Xi'an Jiaotong University, Xi'an, China, in 1994, 1997, and 2000, respectively. He is currently with Gold Standard Simulations Ltd., Glasgow, U.K., where he is involved in TCAD, statistic compact modeling, and design technology co-optimization.

David Reid received the B.Eng. (Hons) degree in Electronics and Electrical Engineering and the Ph.D. degree in semiconductor simulation from the University of Glasgow in 2007 and 2010, respectively. He joined GSS in 2010, where he is the lead developer of the circuit simulation engine RandomSpice and automation toolkit Enigma. He has 2 patents pending on statistical technology and has published over 40 papers.

Andrew Pender's bio and photo are unavailable.

Plamen Asenov completed a BEng in 2009, and a PhD in 2012 at the University of Glasgow. He spent 2.5 years as on-chip memory designer at ARM, with successful tape-outs from 28nm to 14/16nm FinFET with 1 patent. He currently works as a Senior Designer/Design Consultant at GSS.

Campbell Millar (M'09) received the B.Eng. degree in electronics engineering and music and the Ph.D. degree from the University of Glasgow, Glasgow, U.K., in 1999 and 2003, respectively. He was a Senior Research Fellow with the School of Engineering, University of Glasgow, before moving to Gold Standard Simulations Ltd., Glasgow, where he is currently the Chief Operations Officer and Head of Software Development.

Asen Asenov (M'96–SM'05–F'11) received the Ph.D. degree in solid-state physics from the Bulgarian Academy of Science, Sofia, Bulgaria, in 1989. He is currently the CEO of Gold Standard Simulations Ltd., Glasgow, U.K., and the James Watt Professor of Electrical Engineering with the University of Glasgow, Glasgow, UK.