TCAD analysis of FinFET temperature-dependent variability for analog applications

S. Donati Guerrieri[1], F. Bonani[1], G. Ghione[1]
[1]Dipartimento di Elettronica e Telecomunicazioni, Politecnico di Torino
Corso Duca degli Abruzzi, 24, I-10129 Torino, ITALY

***Abstract*—The Green's Function based TCAD device variability analysis is extended to allow for temperature-dependent variability, with negligible overhead in terms of simulation time with respect to fixed temperature simulations. We provide temperature and bias-dependent 3D variability analysis of the DC current for a FinFET structure from the 22 nm node, showing how to predict and mitigate the effects of poor thermal management. Based on the quasi-stationary assumption, preliminary analysis of self-heating effects of a FinFET medium power amplifier is also presented.**

***Index Terms*—FinFET, Variability, Temperature modeling**

I. Introduction

FinFET technology is nowadays well established for digital applications, and its interest for analog RF applications is growing [1]. From the thermal standpoint, though, FinFETs are known to pose significant concern and require accurate modeling, leading to deep investigations of the heat dissipation mechanisms [2]. The inefficient dissipation through the substrate suggests FinFETs can be better described as single fin floating devices, embedded into thermal networks accounting for the heat dissipation through nearby fins and/or metallizations [3]. Modeling electrical FinFET behavior, including variability, as a function of the fin lattice temperature is therefore needed. Variability aware temperature-dependent models are relevant for compact modeling, used e.g. for the concurrent dynamic control of chip temperature, frequency and voltage in digital circuits [4]; and for the optimization of the digital circuit performance under concurrent process, temperature and voltage variations (PVT [5], [6]), where modeling self-heating along with electrical variability is relevant, e.g. for the development of medium power RF stages [1].

In this work we exploit a novel and computationally efficient method to incorporate temperature dependency into Green Function-based (GF) variability analysis [7]–[9], also referred to as the Impedance Field Method (IFM) [10], with negligible numerical overhead with respect to fixed-temperature simulations. The new technique was first introduced in [11], limited to simplified two dimensional double-gate structures. Here we exploit the new approach on a 22 nm FiNFET to further validate the technique and to investigate the sensitivity of the RF performance of a medium power amplifier. We demonstrate that limited temperature dependency is present on a class A-AB bias point, while the knee current is extremely sensitive to self-heating. High bias currents, despite leading to higher gain and better Q factors, need to be carefully evaluated in terms of variations and self-heating, since they also introduce a temperature dependency in the bias point, which is fully correlated to the maximum AC swing (output power).

II. Linearized approach to temperature-dependent variability

An effective approach to device variability through TCAD physics-based simulations can be carried out via a linearized approach, exploiting the Green's Functions (GFs) of the lienarized physical model, as described in [8], [9]. The same approach is also implemented in Sentaurus Synopsys, improperly referenced to as the Impedance Field Method (IFM), although limited to the variability analysis of the DC device performances only. In particular, the so called *statistical IMF* allows for statistical analysis, while for deterministic variations, the `ParameterVariation` command is explicitly dedicated to user-defined parameter variations. In this work, we propose to exploit this Sentaurus feature to investigate deterministic variations of the lattice temperature T_L (details can be found in [11]) in conjunction with the statistical technological variability, finally allowing for a *temperature-dependent variability analysis*. Based on the treatment of [8], when the device is subject to the variation of any parameter P with respect to a nominal value P_0, the device current I_k at each terminal k will undergo a variation which can be expressed as

$$\Delta I_k^P = \int_\Omega G_k(x_0, P_0)\, S_P(x_0, P)\, \mathrm{d}V$$

where the integral extends over the device volume Ω, x_0 collectively denotes the physical model solution with nominal parameters, G_k is the GF related to terminal k and the source term $S_P(x_0, P)$ accounts for the parameter variation ΔP. It can be calculated by the derivative of the physical model equations $F(x)$ with respect to P or in an approximated way by the same physical model calculated with the nominal solution but *varied parameter*, i.e.

$$S_P(x_0, P) = F(x_0, P) - F(x_0, P_0) = F(x_0, P)$$

since $F(x_0, P_0)$ is null by definition, being the residual of the physical model calculated in the nominal solution. Notice that the above source is accurate only for small variation of P and it is a linear (first order) perturbation in ΔP. We now add the lattice temperature dependency, extending the above formulas as

$$\Delta I_k^P = \int_\Omega G_k(x_0, P_0, T)\, S_P(x_0, P, T)\, \mathrm{d}V \qquad (1)$$

978-1-7281-0940-4/19/$31.00 © 2019 IEEE

We refer to this approach as Method 1 (M1). M1 requires the solution of the physical model and GFs for each temperature, with a linear increase of simulation time with the number of temperatures for which we require the analysis. The computational overhead for the source term $S_P(x_0, P, T)$ is instead negligible. Finally the overall current reads

$$I_k = I_k(T, P_0) + \Delta I_k^P$$

To further reduce the simulation time, though, we follow [11] to linearize the device current with respect to the temperature,

$$I_k(T, P_0) = I_k(T_0, P_0) + \Delta I_k^T$$

where

$$\Delta I_k^T = \int_\Omega G_k(x_0, P_0, T_0)\, S_T(x_0, P_0, T)\, \mathrm{d}V$$

where the source term is

$$S_T(x_0, P_0, T) = F(x_0, P_0, T) - F(x_0, P_0, T_0) = F(x_0, P_0, T)$$

Here the required solution of the physical model is only one (at nominal temperature and nominal parameter). Collecting all the variations

$$I_k(T, P) = I_k(T_0, P_0) + \Delta I_k^T + \Delta I_k^P$$

still requires the solution for ΔI_k^P at multiple temperatures. At first order with respect to the concurrent variations of P and T, though, we approximate (1) as:

$$\Delta I_k^P = \int_\Omega G_k(x_0, P_0, T_0)\, S_P(x_0, P, T_0)\, \mathrm{d}V \qquad (2)$$

hereafter denoted by Method 2 (M2) and sketched in Fig. 1, again requiring the solution only with nominal parameters and at a single temperature. To assess the consistency of this *double linearization* and to which extent neglecting second order terms limits the model accuracy, one should verify whether both the GFs (G_k) and the source term S_P show a weak T dependency. Notice, however, that the source term S_P is not directly available in Synopsys, while the GFs can be extracted. To this aim, we consider a FinFET device from the 22 nm technology node, including all the typical technology features, such as the raised source and drain extensions, a high-*k* dielectric and the fin oxide side-walls. The structure is taken from the Synopsys Library [12] and the device geometry is presented in Fig. 2. The device is simulated taking into account accurate physical models, including strain, interface traps and quantum effects through the Density Gradient approach (see [12] for details). Fig. 2 shows the behaviour of the drain terminal GF at room temperature $T_0 = 300$ K. To assess the T dependency, the same was calculated at $T = 320$ K: Figs. 3 and 4 show the nominal value and the variations of the GFs for the Poisson and electron current continuity equations as a function of the position along the channel, in the mid-fin point and at a different depth in the fin. Is evident that the GFs variations are extremely limited, always at maximum equal to 5% of the nominal value, thus validating the approximation in (2). The same trend was observed moving the cut-line C1 away from

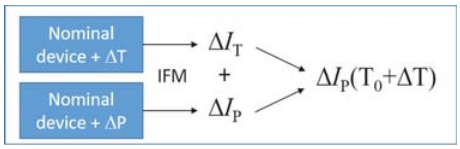

Fig. 1. Schematic representation of the proposed approach.

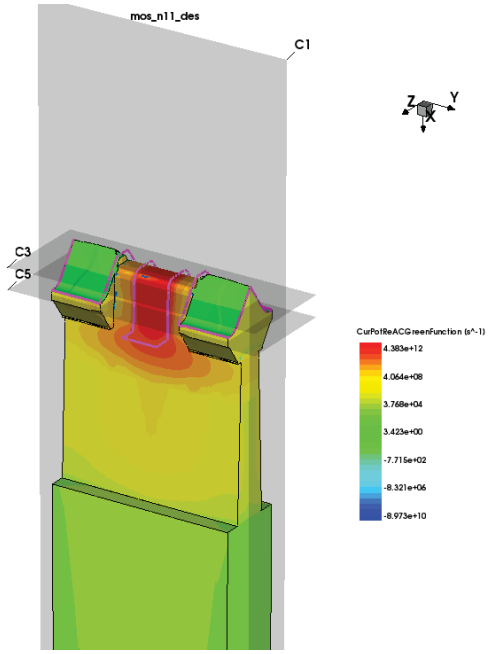

Fig. 2. 3D structure of the simulated 22 nm FinFET (after [12]). Oxide layers are not included in the figure, to allow for internal quantity inspection. In particular the plot shows the drain Green's Function of the Poisson equation at room temperature and the cut-lines used for the T dependent analysis.

the fin mid-point, i.e. towards the side gates. Concerning the source S_P, albeit it can't be directly inspected, notice that, even when the approximation of (2) is not verified, it may still be evaluated without approximations at the varied temperature T, still with negligible numerical overhead, but it would require *ad hoc* tools from Synopsys or in-house tools like [8].

III. 22 nm FinFET medium power amplifier variability

To fix the ideas, we refer to the standard design of a tuned load power amplifier (PA), whose schematic is shown in Fig. 5: the bias point and the load line are selected in order to find the best compromise in terms of output power, efficiency, gain and linearity. Therefore the bias point can be chosen at higher or lower gate voltages (drain currents) for a selected drain bias in saturation, while the maximum output power will be determined by the amount of current at the knee voltage. It is therefore significant to select these two operating conditions of the device (linear at the knee or saturation) and investigate how variability and self-heating affect the overall PA performance.

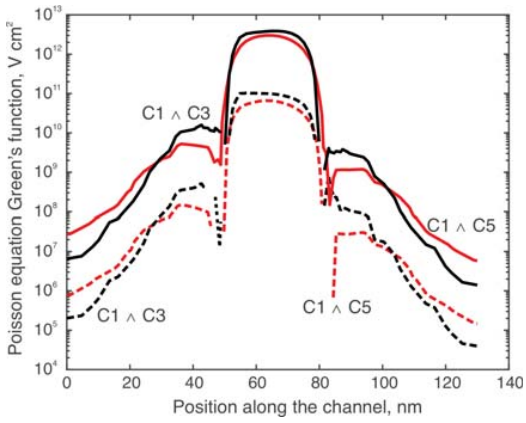

Fig. 3. Poisson drain GF at 300 K and its variation for $T = 320$ K. The cutlines are at the intersection of C1 and C3 and C1 and C5 shown in Fig. 2.

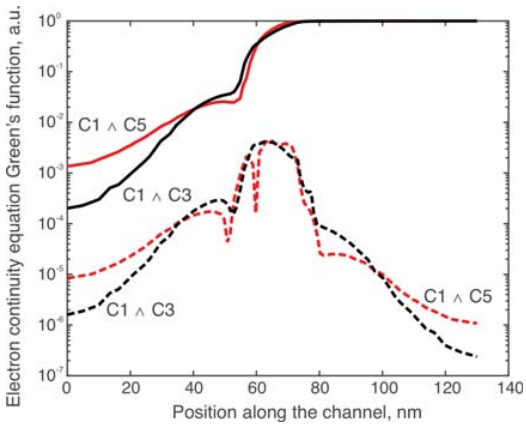

Fig. 4. Electron continuity equation drain GF at 300 K and its variation for $T = 320$ K. The cutlines are the same as Fig. 3.

We apply this concept to the analysis of a possible power amplifier (PA) built on the 22 nm FinFET analysed in the previous section. In this device, we limit the supply voltage to 1.1 V. Figs. 6–9 show the statistical distribution of the FinFET DC drain current above threshold for various values of the gate voltage. In particular, Figs. 6 and 7 show the linear region ($V_{\rm D} = 0.1$ V), to investigate the effect of the PA knee voltage. On the other hand, the saturation region (here, $V_{\rm D} = 0.6$ V), shown in Figs. 8 and 9, is important for the bias condition, usually at lower/intermediate gate voltages, e.g. $V_{\rm G} = 0.6 \div 0.8$ V (Fig. 5). The variability analysis addresses concurrent deterministic lattice temperature variations and random gate workfunction variations (WFV, average metal grain size 5 nm) and random doping fluctuations (RDF). These can be regarded as the main sources of variability [12]. The proposed approach M2 based on the *double linearization*, was compared to M1 showing always a good agreement, up to 50 K of heating, with a remarkable reduction of simulation time (about 80% less). Turning to the detailed analysis of variations, in all operating conditions the drain current temperature dependency has opposite trends as a function of the gate voltage: for gate bias below 0.6 V, where the carrier exponential temperature dependency dominates, the current is increasing with temperature while above 0.6 V, where mobility degradation with T dominates, it decreases. Notice also that the RDF spread is always higher than WFV, at all temperatures. In the linear region (Figs. 6 and 7), for lower gate voltage (just above the threshold, here 0.25 V), the overall spread is lower and the T-dependency very weak. At intermediate gate bias (around $V_{\rm G} = 0.6$ V), the spread is larger (maximum gaussian variance), but nearly insensitive to T (i.e. device self-heating). For large gate bias (amplifier maximum current and power), the current becomes more T dependent, even if the technological spread is reduced: in this case T severely affects the device knee even for a moderate temperature increase. Summarizing, the temperature sensitivity is increasing with $V_{\rm G}$, while the gate workfunction and RDF variability is decreasing with $V_{\rm G}$: since the knee voltage is usually exploited in a power amplifier for larger current values (see Fig. 5), we conclude that the knee voltage variations are usually dominated by self-heating. Turning to Figs. 8 and 9, at intermediate gate voltages, usually exploited for the class A PA design, the temperature sensitivity is nearly null, but the overall technological spread is always significant, unless the operating condition is chosen very close to the threshold voltage (class B). On the contrary, even higher gate voltages could be exploited for gain and bandwidth enhancement but would introduce T variability in the bias current, which would add to the knee variability, resulting in even more perturbations of the overall AC voltage swing.

A reasonable estimate of self-heating for this PA can be made supposing for the PA periphery 10 fingers of 30 fins with a thermal resistance $R_{TH} = 3$ K/mW [13]. When $V_{\rm G} = 0.76$ V, we obtain 15 mA overall current (~ 0.05 mA/fin), and a DC power consumption $P_{\rm DC} \approx 11$ mW. This is completely dissipated to the thermal sink if the PA is not delivering any power (or in back-off), causing a temperature increase of 33 K, while at peak power and 50% efficiency this amount is halved. The predicted self-heating is therefore well within the temperature range for which the analysis has been carried out.

IV. Conclusions

We have presented a numerically efficient approach for the temperature dependent variability analysis of electron devices, carefully exploiting TCAD simulations and the Green's Functions of the linearized physical model. This approach is seamlessly integrated with the well-known Impedance Field Methos used e.g. in Synopsys Sentaurus to account for device technological variations. This approach can be successfully applied in a variety of circuits where the device self-heating and variability impact the overall circuit performance. In particular, a preliminary analysis of a power amplifier based on the 22 nm FinFET technology as a function of the WF and RDF variations was successfully carried out.

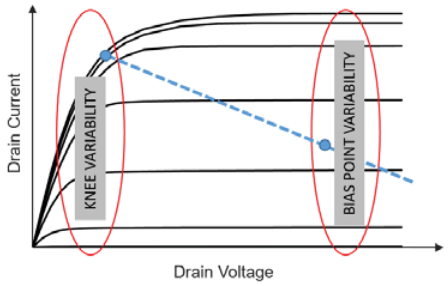

Fig. 5. Investigation of power amplifier variations as a function of current variations at the knee and saturation operating conditions.

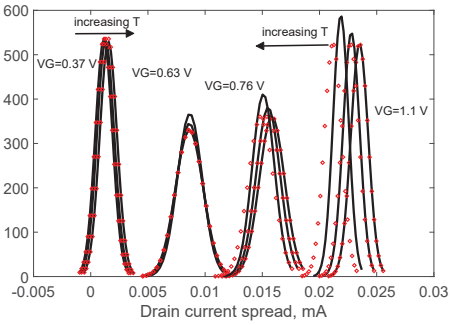

Fig. 6. WFV probability density function of the drain current in linear bias ($V_{\mathrm{D}} = 0.1$ V). T=300, 320 and 350 K. Black lines: M1; Red symbols: M2.

REFERENCES

[1] S. Callender,, et. al., "FinFET for mm Wave - Technology and Circuit Design Challenges", *Proc. BCICTS 2018*, pp. 168–173, 15-17 Oct. 2018, San Diego, CA, USA.

[2] J. Jeon, et. al., "Investigation of Electrothermal Behaviors of 5-nm Bulk FinFET", *IEEE Trans. El. Dev.*, Volume: 64, No: 12, pp. 5284–5287, Dec. 2017.

[3] B. Swahn, et. al., "Electro-Thermal Analysis of Multi-Fin Devices", *IEEE Trans. VLSI*, Volume: 16, No: 7, pp. 816–829, July 2008.

[4] W. Lee, et. al., "Dynamic Thermal Management for FinFET-Based Circuits Exploiting the Temperature Effect Inversion Phenomenon", *Proc. ISLPED 2014*, Aug. 2014, La Jolla, CA, USA.

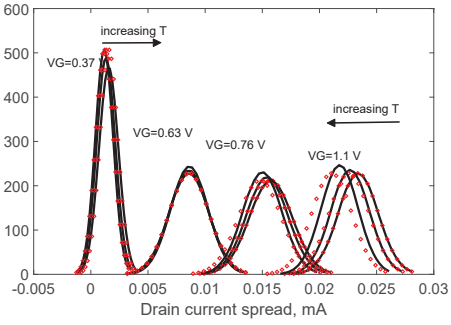

Fig. 7. RDF probability density function of the drain current in linear bias ($V_{\mathrm{D}} = 0.1$ V). T=300, 320 and 350 K. Black lines: M1; Red symbols: M2.

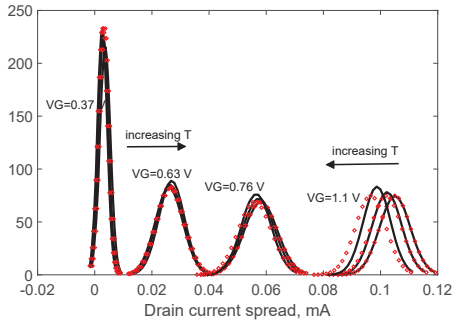

Fig. 8. WFV probability density function of the drain current in saturation ($V_{\mathrm{D}} = 0.6$ V, possible power amplifier bias). Temperatures of T=300, 320 and 350 K. Black lines: M1; Red symbols: M2.

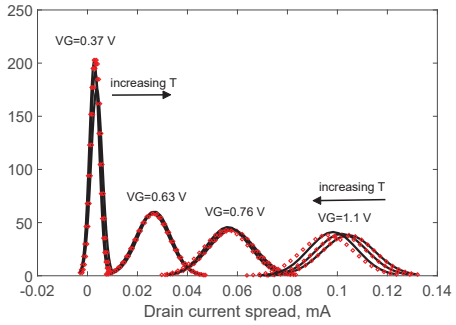

Fig. 9. RDF probability density function of the drain current in saturation ($V_{\mathrm{D}} = 0.6$ V, possible power amplifier bias). Temperatures of T=300, 320 and 350 K. Black lines: M1; Red symbols: M2.

[5] Y. Yang, et. al., "FinPrin: FinFET Logic Circuit Analysis and Optimization Under PVT Variations", *IEEE Trans. VLSI*, Vol. 22, No: 12, pp. 2462–2475, Dec. 2014.

[6] J. H. Choi, et. al., "The effect of process variation on device temperature in finFET circuits", *Proc. ICCAD 2007*, pp. 747–751, Nov. 2007, San Jose, CA, USA.

[7] F. Bertazzi et. al., "Physics-based SS and SSLS variability assessment of microwave devices through efficient sensitivity analysis", *Proc. INMMIC 2012*, 3-4 Sept. 2012, Dublin, Ireland

[8] S. Donati Guerrieri, et. al., "A Unified Approach to the Sensitivity and Variability Physics-Based Modeling of Semiconductor Devices Operated in Dynamic Conditions. Part I: Large-signal sensitivity", *IEEE Trans. El. Dev.*, Vol. ED-63, No: 3, pp. 1195–1201, March 2016.

[9] S. Donati Guerrieri, et. al., "A Unified Approach to the Sensitivity and Variability Physics-Based Modeling of Semiconductor Devices Operated in Dynamic Conditions. Part II—Small-Signal and Conversion Matrix Sensitivity", *IEEE Trans. El. Dev.*, Vol. ED-63, No: 3, pp. 1202–1208, March 2016.

[10] see online: https://www.synopsys.com/silicon/tcad/device-simulation/sentaurus-device.html

[11] S. Donati Guerrieri, F. Bonani, G. Ghione, "A novel TCAD approach to temperature dependent DC FinFET variability analysis", *2018 European Microwave Integrated Circuits Conference (EuMIC)*, 23-25 Sept. 2018, Madrid, Spain.

[12] Available from TCAD Sentaurus Version O-2018 installation, go to *Applications_Library/Variability/FinFET_Variability_sIFM*.

[13] B. González, et. al., "DC self-heating effects modelling in SOI and bulk FinFETs", *Microelectronics Journal*, Volume 46, Issue 4, April 2015, Pages 320-326