Physics-based Analysis to Address Critical Aspects of FinFET Mm-wave Applications: Variability and Thermal Management

Physics-based analysis to address critical aspects of FinFET mm-wave applications: variability and thermal management

Simona Donati Guerrieri[1], **Fabrizio Bonani**[1], **Giovanni Ghione**[1]

[1]Dipartimento di Elettronica e Telecomunicazioni, Politecnico di Torino, Torino, Italy

Abstract— A numerically efficient TCAD approach for the sensitivity analysis of electron devices is exploited for the assessment of the critical aspects of FinFET technology for mm-waves applications. 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 FinFET thermal management. The AC variability analysis is also extensively used, to assess how doping or geometric device variations impact on the insertion loss and bandwidth of a 70 GHz high gain amplifier with transformer matching.

1. INTRODUCTION

The development of advanced components for mm-wave applications has led to new device concepts, also based on non conventional materials. Among the possible options, silicon devices, originally introduced for digital applications, have always called the attention for possible analog high-frequency applications due to the potential development of Systems on Chip (SoCs). Despite the reduced gate length, the MOSFETs structure is often not optimized for high frequency applications and their cut-off frequencies are usually lower than expected due to device parasitics. Moreover, technological variability is ever increasing with the development of the new technology nodes, and its effects can be mitigated more easily in digital than in analog circuits. All these limitations are exacerbated in the latter form of silicon technology, i.e. FinFETs. In fact, their peculiar 3D structure brings along a set of parasitics, especially in terms of fringing and mutual capacitances and of access resistances. The development of FinFET based circuits for next generation communication networks is thus the object of many efforts. Recently, INTEL [1] has developed an *ad hoc* 22 nm FinFET process specifically targeting analog application, showing a cut-off frequency higher than 200 GHz and $f_{\max}$ as high as 4 THz. The same paper also highlights the challenges to be tackled before the successful deployment of FinFETs in communication networks: severe self-heating, parasitics due to the scaled interconnects and the diminished device drive capability due to its reduced size and peculiar scalability. Higher parasitics result in variations of the input and output device quality factors Q (ratio of the real and imaginary part of Y_{11} and Y_{22}, respectively), thus making wideband matching more difficult with detrimental effects in all possible stages, from low-noise to medium power amplfiers and mixers. Nevertheless, several demonstrators of FinFET-based systems are already available, mainly up to 70-80 GHz [2]. Innovative schemes of analog stages exploiting the peculiar features of the multigate technology are also an open and promising research field [3].

In this contribution we will address specifically the FinFET variability analysis from various perspectives, all based on the numerically efficient Green's Function approach (briefly reviewed in the following Sec. 2), in order to benchmark this technology for mm-wave analog applications and explore the possible ways to mitigate variations, especially in terms of the correct choice of the bias point. Commercial software (Synopsys [4]) already allows for the DC variability analysis in 3D, while the variability analysis of the AC parameters are at present limited to an in-house 2D implementation. In Sec. 3, Synopsys is exploited, taking advantage of the accurate model library for the 22 nm FinFET technology: an example of temperature dependent variability analysis is provided, with relevant implications especially for medium power amplifiers. Sec. 4 is instead dedicated to the analysis of the RF performance of a high gain amplifier for mm-wave applications. Starting from TCAD simulations, the sensitivities of the AC Y-matrix of a double fin FinFET to the variation of selected geometric features are extracted, as a function of frequency and bias, and imported into the Agilent ADS circuit simulator allowing for a circuit-level variability-aware design of a high gain amplifier at 70 GHz: the insertion loss (IL) and the bandwidth (BW) of the input and output matching networks are examined as a function of the variations of the input and output device terminations, due to the device variability. We demonstrate that the choice of the device bias point must include, besides the best compromise between gain and bandwidth, a careful exploration of the bias-dependent device variability.

2. TCAD VARIABILITY ANALYSIS

Modelling variability, along with RF performances, is a primary need to successfully design mm-wave stages. TCAD analysis is an ideal tool to link the physical or geometric device variation (variability) to the spread of the device analog performances. Despite this, TCAD analyses focusing on RF variability are still limited, due to the numerical burden of the physics-based analysis, especially when including frequency dependent AC simulations and possibly multi-harmonic (Harmonic Balance) nonlinear analysis. The problem is exacerbated for 3D devices, such as FinFETs, and also when one needs to account for deterministic or random concurrent variations of multiple physical parameters (e.g. the gate length and workfunction, or the substrate and source/drain doping). Recently, advanced tools [5]-[6] have been demonstrated for the numerically efficient TCAD variability analysis, exploiting the Green's Function formulation, making it possible to address not only a parametric sensitivity analysis, but a full statistical device simulation. At present, a commercial software (Synopsys [4]) allowing for 3D device analysis, already implements the advanced variability techniques [5]-[6], but limited to the DC case only. Hence, the variability of AC and multi-frequency performance must be implemented resorting to several repeated analyses, with multiple physical parameters selected on a proper statistical ensamble (MonteCarlo analysis), with the corresponding numerical burden. On the contrary, in-house implementation of the techniques of [5]-[6], despite limited to 2D analysis, make the frequency-dependent variability analysis feasible with limited numerical effort, via the Harmonic Balance algorithm. In this way multi-bias, frequency dependent, AC variability can be easily performed simultaneously with conventional AC and LS analysis [7], addressing typical high frequency performances as f_{T}, $f_{\max}$, Q_{in}, Q_{out}, or power performances like the output power or intermodulation, as a function of the main device physical parameters [8].

3. 22 nm FINFET TEMPERATURE-DEPENDENT VARIABILITY

A preliminary investigation of the techniques allowing for an efficient variability analysis as a function of the device lattice temperature has been presented in [9], limited to a 2D test-case. A primary concern behind thermal analysis is to avoid TCAD self-consistent simulations of the electrical and thermal device model, especially for 3D devices such as FinFETs, requiring usually around 100000 grid nodes. Hence the thermal model is approximated by the device thermal resistance while the device is modelled at an equivalent (lattice) temperature T_E different from the sink temperature. To fix the ideas, we will denote the dissipated power P_{diss}, as

$$P_{\mathrm{diss}} = V_{\mathrm{DD}}\, I_{\mathrm{DC}} = V_{\mathrm{DD}} \times (I_{\mathrm{DC},0} + \Delta I_V) \tag{1}$$

where V_{DD} is the supply voltage, $I_{\mathrm{DC},0}$ is the DC current of the nominal device (without variability) and ΔI_V is the DC current spread due to technological or physical parameters' variations. Both $I_{\mathrm{DC},0}$ and ΔI_V are a function of temperature. At first order we neglect self-heating due to device variations (i.e. neglecting - or averaging- ΔI_V), i.e.

$$P_{\mathrm{diss}} \simeq V_{\mathrm{DD}} \times I_{\mathrm{DC},0} \tag{2}$$

which can be iteratively (or simultaneously) solved with the thermal resistance model

$$T = T_0 + R_{\mathrm{th}} \times (V_{\mathrm{DD}} \times I_{\mathrm{DC},0}) \tag{3}$$

where R_{th} denotes the device thermal resistance, which can be extracted from independent thermal simulations,[1] and T_0 is the thermal sink temperature. This yields the equivalent device temperature T_E and the corresponding DC nominal current. To account for variability, we still need to compute $\Delta I_V(T_E)$: from the modelling standpoint the most efficient solution is to extract $\Delta I_V(T)$ as a function of the temperature by means of independent variability analysis for each parameter undergoing technological variations, e.g. by repeated Green's Function analyses at varying temperature (hereafter this approach will be referred to as Method 1).

To further simplify the thermal model, we will also linearly approximate the drain current temperature dependency:

$$I_{\mathrm{DC},0} = I_{\mathrm{DC},0}\,(T = T_0) + \Delta I_T \tag{4}$$

[1] Notice that the thermal resistance R_{th} may also be affected by geometrical device uncertainties, e.g. fin width variations: in this case the equivalent device temperature may result into a random variable itself. The proposed approach, though, remains valid.

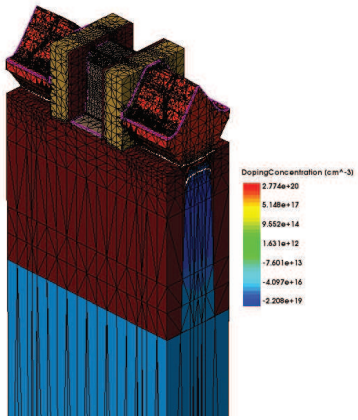

Figure 1: 3D structure of the simulated 22 nm FinFET (after [10]).

and

$$\Delta I_V = \Delta I_V \left(T = T_0\right) + \Delta I_T \tag{5}$$

From the TCAD standpoint this method (hereafter referred to as Method 2), is appealing especially for 3D devices like FinFETs, due to the huge reduction of computation time following the *double linearization* (i.e. in terms of the technological variations and in terms of the temperature), since it does not require repeated variability analyses with varying temperature. According to [9], $\Delta I_V \left(T = T_0\right)$ and ΔI_T can be instead simultaneously calculated through the efficient Green's Function approach, and then superimposed. Despite Method 2 is most appealing, the underlying approximations need careful validation.[2] As an example, we address a device from the 22 nm technology node, including all the typical features of the FinFET technology, like the raised source and drain extensions, the high-k dielectric and the fin oxide side-walls. The structure is taken from the Synopsys Library and the device structure is presented in Fig. 1. Simulations account for accurate physical models, including Fermi statistics, mobility degradation, bandgap narrowing, interface traps and quantum effects through the Density Gradient approach (see [10] for details). The variability analysis addresses concurrent deterministic temperature variations and random spread of the device workfunction due the gate metal granularity with average grain size of 5 nm and a workfunction distribution accounting for metal grain orientation [10]. Method 1 and 2 have been compared at varying gate voltage and at two different drain bias (linear and saturation regions). The linear region, at higher gate bias, is especially relevant e.g. for the assessment of the knee voltage variations in a power amplifier stage, while the saturation region at lower gate voltages is relevant e.g. for the quiescent condition of class A power amplifiers or high gain amplifiers.

Fig. 2 shows the statistical distribution of the FinFET DC drain current above the threshold, here around 0.25 V, in the linear region ($V_\mathrm{D} = 0.05$ V, left) and in saturation ($V_\mathrm{D} = 1$ V, right). Method 2 compares well with Method 1, with a noticeable reduction of simulation time (around 80% less). The drain current temperature dependency has opposite trends for gate voltage below 0.6 V, where the carrier exponential temperature dependency dominates, or above 0.6 V, where mobility degradation with temperature dominates. More in detail, the analysis of the linear condition shows that for lower gate voltage the overall variability is lower and the temperature dependency moderate. For intermediate gate bias (around $V_\mathrm{G} = 0.6$ V), the variability is higher (the gaussian has maximum variance) but nearly insensitive to temperature (i.e. device self-heating). For higher gate bias, the current becomes more temperature dependent. In other words, despite the overall drain current spread is somewhat lower than at intermediate bias, the temperature sensitivity severely affects the device knee voltage even for moderate temperature increase (e.g. 20 K). In saturation Method 1 and Method 2 still compare very well, but on a more limited temperature range (±20 K). The temperature sensitivity is nearly null around $V_\mathrm{G} = 0.8$ V, roughly corresponding to a possible class A operation, but the overall spread is always significant, unless the operating condition is chosen very close to the threshold voltage.

[2] In case the assumption of (4) are not well verified, a more refined approach would be through multi-point linear expansion

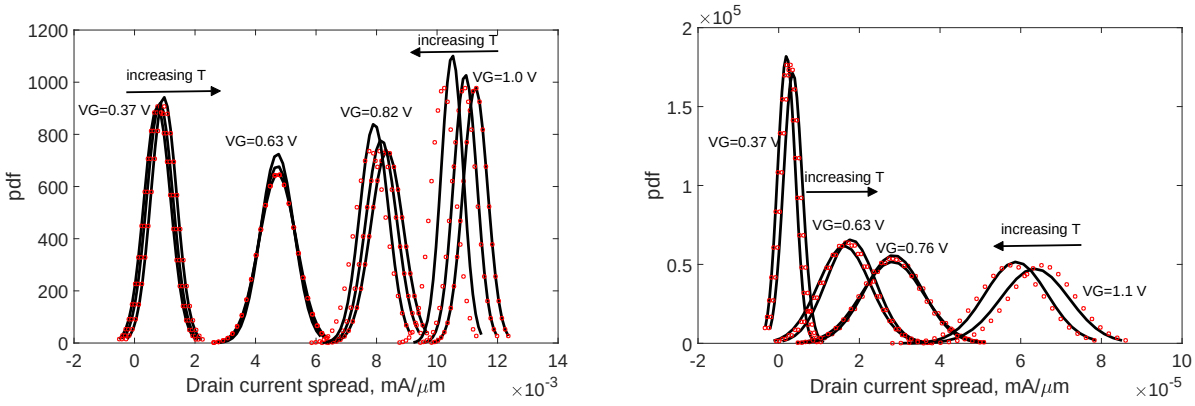

Figure 2: Probability density function of the drain current resulting from WF statistical variations. Left: linear region (V_D = 0.05 V) and temperatures of T=300, 320 and 350 K. Right: saturation region (V_D = 0.1 V) and temperatures of T=280 and 320 K. Black lines: Method 1; Red symbols: Method 2.

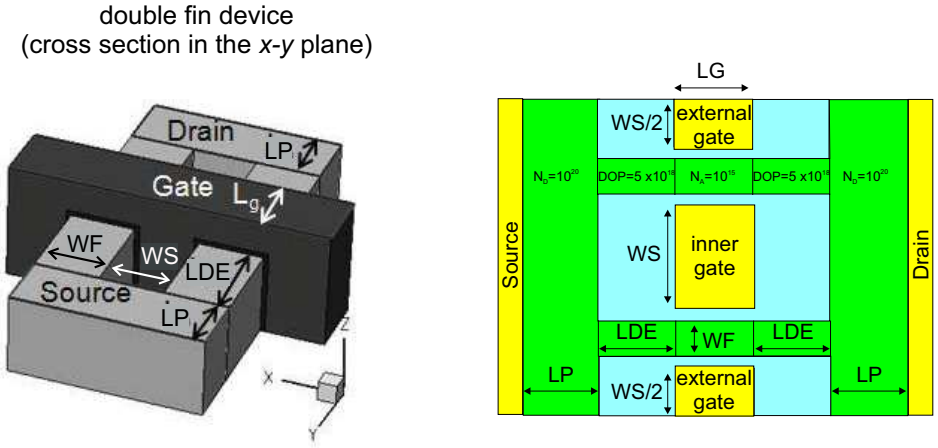

Figure 3: Two-fin structure and 2D cross-section used for the AC variability analysis.

4. FINFET HIGH GAIN AMPLIFIER VARIABILITY

A 2D cross-section of the two fin structure of Fig. 3 [11] is used as the core of a multi-fin device for RF applications. The main features of the multi-fin structure are captured by the given template, including inter-fin capacitances, fringe capacitances, drain to gate capacitance due to raised source/drain and parasitic resistances of the source/drain extensions. Periodicity conditions on non-contacted external boundaries are enforced in TCAD analysis trough discretization, hence making this structure an ideal core for fin parallelization. The AC **Y** matrix is extracted from the given structure and further scaled to a total gate periphery of 60 μm. The analysis has been carried out as a function of the gate bias (from threshold to V_G = 1 V) keeping V_D = 1 V (saturation), with the aim of investigating the proper bias condition of a high gain small signal amplifier for mm-wave applications. The center frequency of the design has been chosen to be 70 GHz, while the TCAD analysis covers the frequency interval 60-80 GHz to monitor the design bandwidth.

We also perform the AC variability analysis with an in house simulator implementing the efficient Green's Function approach based on the Harmonic Balance algorithm. The variations of the elements of the **Y** matrix are calculated, over the same bias and frequency range, accounting for variations of the source/drain doping (DOP), source/drain extensions (LDE), fin width (WF) and fin separation (WS), see Fig. 3. Such parameters have been chosen for their impact on RF performances, especially via the gate parasitic resistance [13]. Denoting with $N_b = 9$ the number of bias points, $N_f = 5$ the number of frequencies and $N_p = 4$ the number of parameter values ($\pm 5\%$ and $\pm 10\%$ variation with respect to the nominal device), the overall dataset accounts for $(4 \times N_b \times N_f \times N_p)$ values for each element of the Y matrix (the factor 4 accounts for the 4

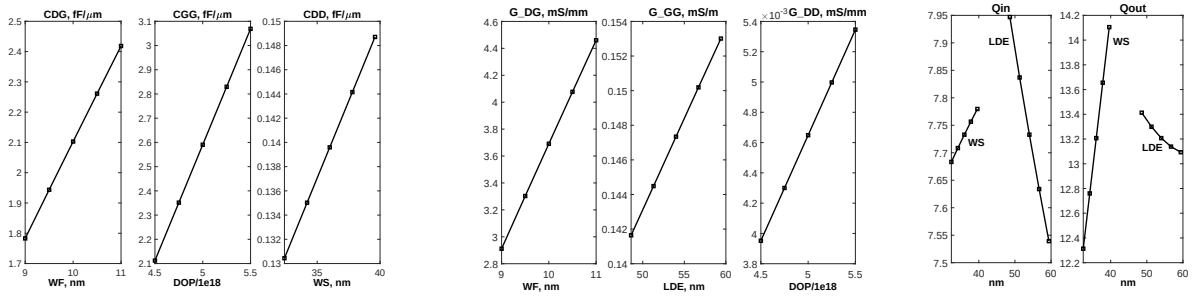

Figure 4: Variation of selected elements of the AC $\mathbf{Y}$ matrix as a function of selected technological parameters at $V_{\mathrm{G}} = 0.8$ V, $V_{\mathrm{D}} = 1$ V and $f = 70$ GHz. The bias point chosen is close to peak transconductance, which in this device occurs around $V_{\mathrm{G}} = 0.7 - 0.8$ V. Left: imaginary parts normalized to the angular frequency (capacitances). Middle: real parts. Right: Q factors.

parameters considered, i.e. OM, DOP, LDE and WS). The input and output quality factors

$$Q_{\mathrm{in}} = \frac{\Im\left(Y_{11}\right)}{\Re\left(Y_{11}\right)} \qquad\qquad Q_{\mathrm{out}} = \frac{\Im\left(Y_{22}\right)}{\Re\left(Y_{22}\right)} \tag{6}$$

have also been extracted as a function of bias, frequency and parameter variations.

As an example of the obtained results, Fig. 4 shows the variations of some elements of $\mathbf{Y}$ and Q as a function of selected technological parameters. Since variations are quasi linear with respect to paramter variations, the AC sensitivities are calculated as

$$S^{\Re}_{P_{i,j}} = \frac{\mathrm{d}\Re\left(Y_{i,j}\right)}{\mathrm{d}P} \qquad\qquad S^{\Im}_{P_{i,j}} = \frac{\mathrm{d}\Im\left(Y_{i,j}\right)}{\mathrm{d}P} \tag{7}$$

where P is any of OM, DOP, LDE or WF. Finally, the nominal values of the elements $Y_{i,j}$ along with their sensitivities are stored in a *Citifile* data format as a function of bias and frequency and imported into the circuit simulator Keysight ADS. The device input and output quality factors their sensitivities are also imported in ADS.

We can now proceed to the variability aware design of analog circuits, taking as an example the design of a small-signal high gain amplifier, with the specific aim of discussing the design of the input and output matching networks in terms of variations. In CMOS-like technologies, single-ended matching networks of analog stages are generally designed (on the basis of the nominal device parameters) with a transformer whose primary and secondary quality factors Q_{P} and Q_{S} guarantee tuning out the device input or output capacitances, while providing the correct transformation ratio of the real parts of the input and output device impedance to the desired source and load terminations. The device Q factors become the most relevant design parameters: in fact, while higher values usually correspond to higher available gain, they also make wideband matching more difficult, because of the non-ideal transformer behavior, i.e. the finite values of Q_{P} and Q_{S}. In general, higher values of Q_{in} and Q_{out} are not desirable for matching, hence the device bias point is chosen at a compromise between peak transconductance and bandwidth. Turning to the variability analysis, Q_{in} and Q_{out} undergo significant spread with technological variations (see Fig. 4), which will directly impact the insertion loss and bandwidth of the matching networks. In fact, when the device undergoes variations, the transformer will be de-tuned with respect to the nominal operating frequency. Furthermore, if Q increases, the insertion loss (IL) of the matching networks also degrades.

To investigate such effects we consider the circuit of Fig. 5, that was implemented into ADS, using the non ideal transformer (Two Coupled Resistive Coils): following [1], the source and load Q factors mimic the input and output device quality factors, while the transformer quality factor is considered to be fixed, at least at the design center frequency. We will analyze three cases: 1) $Q_{\mathrm{S}} = Q_{\mathrm{L}} = Q_{\mathrm{in}}$ to test the effect of input port variation with respect to the nominal design; 2) $Q_{\mathrm{S}} = Q_{\mathrm{L}} = Q_{\mathrm{out}}$ to test the effect of output port variability; 3) $Q_{\mathrm{S}} = Q_{\mathrm{in}}$ and $Q_{\mathrm{L}} = Q_{\mathrm{out}}$ which mimic the effects of a possible inter-stage matching, i.e. matching the device input port to the output port of a previous stage. The values of Q_{S} and Q_{L} are always extracted from the *Citifile*, see Fig. 5,

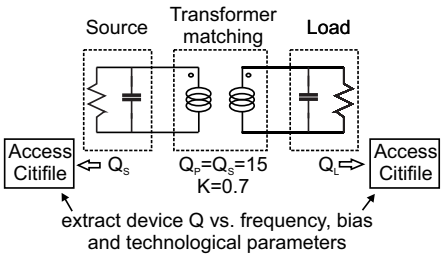

Figure 5: Circuit used to test the effect of the variability on the matching network IL and bandwidth.

hence the source and load terminations accurately reproduce the behavior of the input and output device impedances, including bias and frequency dependency, and variations of all the considered technological parameters. The values of Q_{P} and Q_{S} are reasonable according to the FinFET INTEL technology [1] and have been considered equal for this preliminary analysis. Concerning the transformer, we consider two cases: in Method a) the primary and secondary inductance are chosen to resonate with the device capacitance, following its variations with technological parameters, while the resistance is chosen to be *fixed* (i.e. not frequency dependent) so as to provide the required Q_{P} and Q_{S} only at the nominal frequency of 70 GHz; in Method b) the primary and secondary inductances are chosen to resonate with the device capacitance only at the value of the *nominal* device capacitance and the resistance is chosen as in Method a). Despite case b) is somewhat more realistic, since the transformer parameters are designed for a nominal device at the nominal operating frequency and considered to be fixed with frequency (as expected for a wideband matching network), case a) is also significant since the variations of the device Q factors are limited to the insertion loss, while the resonance frequency is maintained at the nominal value of 70 GHz; in case b), instead, a significant degradation of matching is also due to frequency de-tuning. Notice that, contrary to [1], we do not consider the transformer G_{max} constant over the band, since this would also require a frequency dependent parasitic resistance of the primary and secondary leads, which seems unreasonable at this design level and in any case would strongly depend on the physical implementation of the transformer (e.g. through skin effects). As an example of the results obtained, Fig. 6 shows the a MonteCarlo simulation performed within ADS with DOP values statistically distributed with a gaussian standard deviation of 5% of the doping nominal value. The cases 1) 2) and 3) are compared using Method a) for two bias points close to the maximum transconductance condition. The S_{21} scattering parameter of the circuit in Fig. 5 yields the insertion loss of the matching network. The edges of the 3 dB bandwidth are also monitored for variability effects. It is clear that, while case 1) is most sensitive to variations at the lower bias, case 2) results in a considerable spread of the matching condition at higher bias. The insertion loss varies as much as 1 dB with the given parameter spread, while the lower and upper band edges exhibit more than 2 GHz variations. Notice, though, that such variations are correlated. Similar results can be found varying all other technological parameters: while WF variations also induce significant matching uncertainty, DOP variations shown in Fig. 6 are in any case one of the dominant contributions.

Finally, Fig. 7 shows simulations carried out with Method b). Here, deterministic variations ($\pm$ 5%) of the fin width WF with respect to the nominal value induce a significant de-tuning of the matching network and, consequently a degradation of the insertion loss at the nominal operating frequency as high as 2 dB. The above results show how effectively TCAD simulations can be translated into circuit simulators to assist the RF stage design.

5. CONCLUSIONS

Various examples of efficient techniques amenable for the assessment FinFET variability from TCAD simulations have been presented, focusing on the critical aspects that have to be considered for a successful application of such technology to mm-waves. The obtained results can be readily imported into circuit simulators for variability-aware RF design.

REFERENCES

1. S. Callender, W. Shin, H.-J. Lee, S. Pellerano, C. Hull, "FinFET for mm Wave - Technology and Circuit Design Challenges", *Proc. BCICTS 2018*, pp. 168–173, 15-17 Oct. 2018, San Diego,

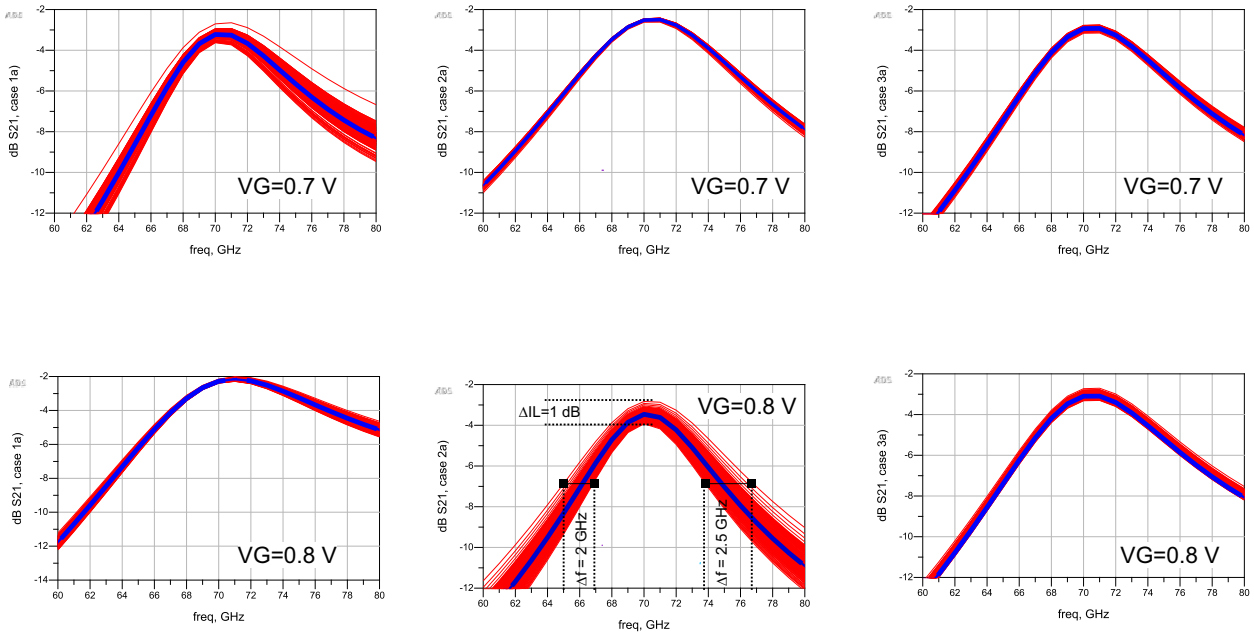

Figure 6: MonteCarlo simulation with random distribution of doping DOP at $V_{\mathrm{G}} = 0.7$ V and $V_{\mathrm{G}} = 0.8$ V for the cases 1) 2) and 3) using Method a). In order to investigate the effect of the insertion loss due to the variations of the Q factors only, scattering parameters have been been calculated with input(output) port impedance equal to the real part of Q_{S} (Q_{L}).

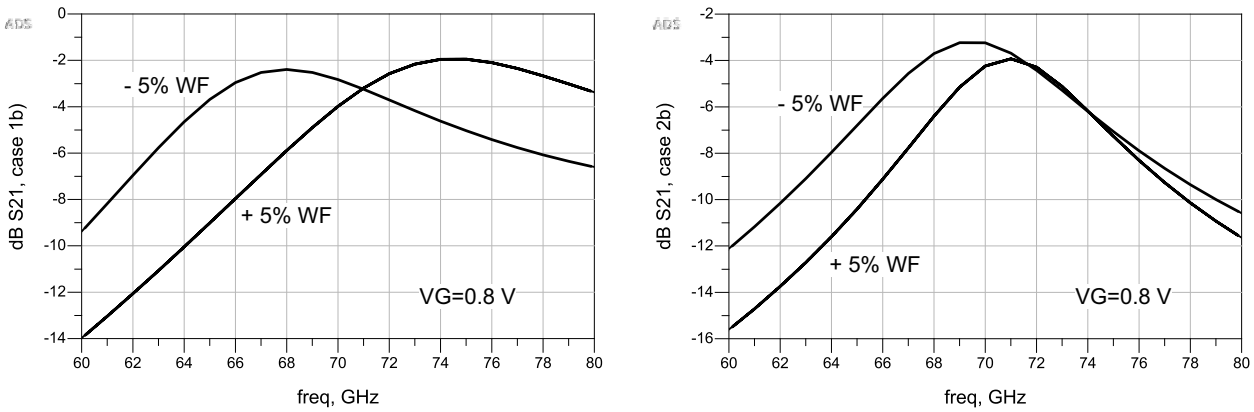

Figure 7: Simulations of the matching condition vs. $\pm$ 5 % variation of the fin width WF for the cases 1) and 2) and using Method b). $V_{\mathrm{G}} = 0.8$ V.

CA, USA.
2. P. Sagazio, S. Callender, W. Shin, O. Orhan, S. Pellerano, C. Hull, "Architecture and Circuit Choices for 5G Millimeter-Wave Beamforming Transceivers", IEEE Communications Magazine, Vol. 56, N. 12, Page(s): 186-192, December 2018.
3. A. M. Bughio, S. Donati Guerrieri, F. Bonani and G. Ghione, "Multi-gate FinFET Mixer Variability assessment through physics-based simulation", IEEE Electron Device Lett., Vol. 38, No. 8, 1004-1007, 2017.
4. see online: https://www.synopsys.com/silicon/tcad/device-simulation/sentaurus-device.html
5. S. Donati Guerrieri, F. Bonani, F. Bertazzi, and G. Ghione, "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. Electron Devices*, vol. 63, no. 3, pp. 1195–1201, 2016.
6. S. Donati Guerrieri, F. Bonani, F. Bertazzi, and G. Ghione, "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. Electron Devices*, vol. 63, no. 3, pp. 1202–1208, 2016.
7. S. Donati Guerrieri, M. Pirola, F. Bonani, "Concurrent Efficient Evaluation of Small-Change Parameters and Green's Functions for TCAD Device Noise and Variability Analysis", *IEEE Trans. El. Dev.*, Volume: ED-64, No. 3, pp. 1269–1275, March 2017.
8. F. Bertazzi, F. Bonani, S. Donati Guerrieri, G. Ghione, "Physics-based SS and SSLS variability assessment of microwave devices through efficient sensitivity analysis", *Int. Workshop on Integrated Nonlinear Microwave and Millimetre-wave Circuits (INMMIC 2012)*, 3-4 Sept. 2012, Dublin, Ireland
9. 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.
10. Available from TCAD Sentaurus Version O-2018 installation, go to *Applications_Library/Variability/FinFET_Variability_sIFM*.
11. A.M. Bughio, S.Donati Guerrieri, F. Bonani, G. Ghione, "Physics-based modeling of FinFET RF variability", *2016 European Microwave Integrated Circuits Conference (EuMIC)*, 3–7 Oct. 2016, London, UK.
12. A. M. Bughio, S. Donati Guerrieri, F. Bonani and G. Ghione, "Variability of FinFET AC parameters: A physics-based insight ", International Journal of Numerical Modelling: Electronic Networks, Devices and Fields, Vol. 31, No. 3, May/June 2018.
13. H.-J. Lee, S. Rami, S. Ravikumar, V. Neeli, K. Phoa, B. Sell, Y. Zhang, "Intel 22nm FinFET (22FFL) Process Technology for RF and mm Wave Applications and Circuit Design Optimization for FinFET Technology", International Electron Devices Meeting (IEDM 2018), p.14.1.1-14.1.4, 1-5 Dec. 2018, San Francisco, CA, USA.