Transistor modelling for mm-Wave technology pathfinding

B.Parvais[1], R. ElKashlan[1], H. Yu, A. Sibaja-Hernandez, B. Vermeersch, V. Putcha, P. Cardinael[2], R. Rodriguez, A. Khaled, A. Alian, U. Peralagu, M. Zhao, S.Yadav, G. Gramegna, J. Van Driessche, N. Collaert

imec, Kapeldreef 75, 3001 Leuven, Belgium

[1]also with Vrije Universiteit Brussels, 1050 Brussels, Belgium

[2]also with UCLouvain, Louvain-la-Neuve, Belgium

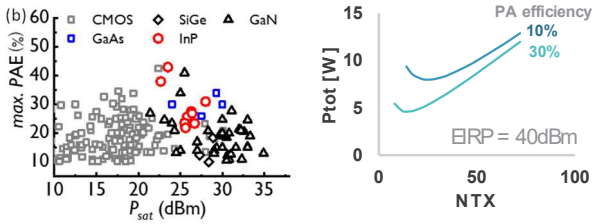
***Abstract*— A review of the modelling requirements to establish a Design-Technology Co-Optimization loop for mmWave Front-End Modules is presented. The example of GaN/Si technology is detailed, and recent modeling developments are explained.**

Keywords—mmWave, GaN, RF-DTCO.

I. INTRODUCTION

As initiated with the advent of 5G mmWave, the upcoming generations of mobile communications will operate at increased frequency [1], [2]. The D-band for instance, provides the bandwidth needed to enable extreme mobile broadband systems with >10 Gbps data rate. The overall system performance and power consumption of mm-wave transceiver systems is determined by the power amplifier (PA) in the front-end module (FEM) [2]. The generation of power at such high frequencies is indeed poorly efficient and the success of the next mobile radio generations will rely on the achievable power efficiency on a given form factor (Figure 1). Indeed, the limited area available for the radio module in mobile devices restricts the number of antenna paths available for beamforming. This, in turns, requires each PA to deliver a relatively large output power (P_{out}). The choice of FEM technology is therefore guided by its capability to realize power efficient transceivers. Among the traditional high-frequency device technologies, III-V HBTs and HEMTs are the only technologies that have demonstrated good power performance across the full mm-wave spectrum (Figure 1(a)). At medium power levels (20-30 dBm), these devices hold a significant power added efficiency (PAE) advantage over advanced CMOS and are competitive to GaN HEMTs, which are the preferred option for delivering very large power levels. However, most of the current III-V and GaN technologies are still limited to small size, expensive, non-Si substrates and use older generation processing. Migrating to a 200- or 300-mm Si platform and, manufacturing devices using standardized CMOS fab tools are critical steps toward the uptake of compound semiconductors for RF and mm-wave applications [4].

Migrating to a Si compatible platform impacts however the performance of the transistors, which suffer from a larger defect density when grown on Si. The BEOL is also adapted: Au is replaced by Cu, and planarization eases 3D hetero integration (CMOS + III/V), which will become crucial to enable 2D beam steering applications. Integrating CMOS functionality close to the FEM brings some key advantages as the ability to use digital calibration for the FEM. More



(a) (b)

Figure 1 (a) Maximum PAE for various power amplifier technologies in the above 40 GHz (data from [3]). For user held devices (20-30 dBm), III-V HBTs and HEMTs can be the technologies of choice. (b) Using efficient technology not only reduces the transmitter power, but also its area since less antenna elements are required for a given EIRP level.

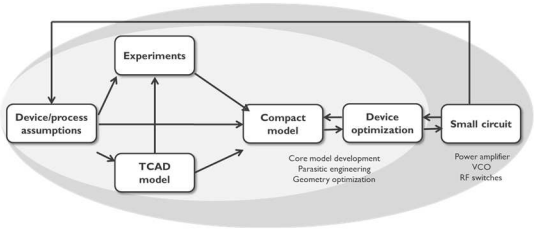


Figure 2 The RF-DTCO loop concept: from device modeling and exploration to benchmark circuits.

generally, the demand for integrating systems with higher complexity has grown over the past years and the heterogenous integration in package raises the challenge of being able to simulate at system level not only functionality and performance, but also reliability, thermal and mechanical aspects. Multidomain EDA tools and hybrid PDKs, including EM features, must be developed for co-optimization of IC and packaging technologies. In such a context, it is desirable to co-optimize technology and devices together with the design of circuits and systems, not only for optimal performance, but also for a fast deployment of the wireless technologies. The design-technology co-optimization (DTCO) methods originally developed in the context of logic circuits, can be ported in the RF domain provided that appropriate Figures-of-Merit (FOMs) are considered, and accurate models are available (Figure 2). This allows, for instance, to study the impact at circuit level of the introduction of any technology booster. The present paper reviews the transistor-level modelling requirements desired to establish a DTCO loop for mmWave application using compound semiconductors. A CMOS compatible GaN/Si technology [5] will be used to illustrate this.

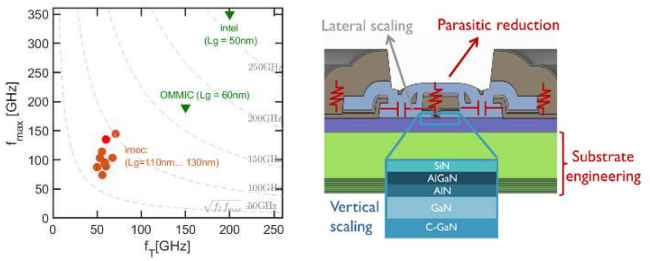
978-1-6654-0685-7/21/$31.00 ©2021 IEEE

II. Figures-of Merit and Scaling

Benchmarking various GaN/Si technologies in Figure 3(a) shows how important is gate length scaling to bring GaN in the mmWave application space. Lateral scaling is done in conjunction with vertical scaling to control short channel effects. A special attention is paid to parasitic reduction when scaling devices (Figure 3(b)): in CMOS compatible technologies, air bridges are replaced by dielectric filling and their larger permittivity leads to increased capacitances. Furthermore, the RF losses originated in the parasitic conduction layer at the interface between Si and buffer must be reduced as much as possible. At high frequencies, the parasitic interconnect capacitances and inductances cannot be neglected. Classically the cut off frequency of the current gain, f_T, and of the unilateral power gain, f_{max}, are the metrics used to benchmark the RF performance of transistors. In the context of PA, *fmax* is more relevant and depends on the layout (via gate resistance R_G):

$$f_{max} \approx \frac{f_t}{2\sqrt{g_d(R_g+R_s+R_i)+2\pi f_t R_g C_{gd}}} \qquad (1)$$

Given the medium to high power level involved in the FEM, the nonlinear characteristics of the devices must be reproduced accurately by the models. Not only the compression points (P1dB), but also the harmonic content (HD), and AM/PM must be known to predict the circuit EVM. Precise large signal evaluation helps designers to improve the performance by applying harmonic tuning or predistortion techniques. Compound semiconductors suffer from higher defectivity, compared to group IV materials. The trap levels created induce a time dependence of the device characteristics and, together with thermal effects, reliability must be carefully characterized.



(a) (b)

Figure 3 (a) Benchmark of the cut off frequencies, illustrating the importance of gate length scaling. (b) Engineering elements for improving the RF performance of GaN HEMTs.

III. Transistor Modelling

GaN HEMTs have been widely studied in the past decade and the TCAD model are available for device exploration. After calibration on our hardware, the benefits of barrier downscaling are projected in Figure 4, together with the improvement expected from adopting InAlN as barrier material. InAlN promises a larger 2DEG density, at the cost of a more negative threshold voltage [6].

Being more computation efficient, compact models are used for large signal evaluation and circuit simulation. For pathfinding, physics-based models (such as ASM [7] or MSVG [8]) are preferred for their predictivity. However, these compact models must be enhanced to capture technology elements:

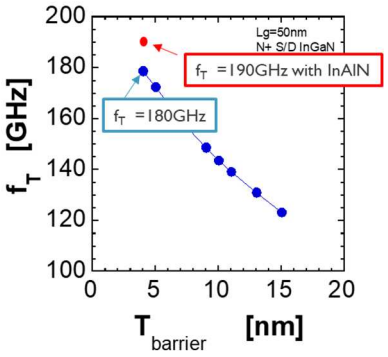


Figure 4 Impact of barrier thickness scaling on f_T, after TCAD.

1) *Effect of AlN interlayer*

The presence of the AlN interlayer increases the effective conduction band offset between AlGaN/GaN (Figure 5), which directly impacts the V_T. The additional ΔE_C provides more surface states of AlGaN above the Fermi level, leading to a larger 2DEG. The mobility is also impacted since the centroid charge shifts away from the interface.

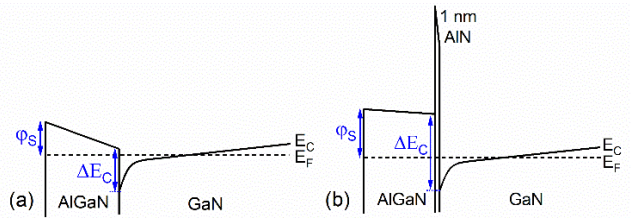


Figure 5 (a), (b) the presence of AlN increases the energy band offset ΔE_C at the barrier interface.

2) *Effect of channel thickness*

As shown in Figure 6, the C-GaN establishes a field F_{EX} in the GaN channel to confine the 2DEG. This enhances the gate control. When the channel is thinned down, which is desirable from a scaling perspective to reach higher frequencies, the 2DEG gets weaker because of the increased F_{ex} and increased ionized C acceptors in the C-GaN layer. Prediction from Poisson-Schrödinger solver in Figure 6 can be used to calibrate compact models.

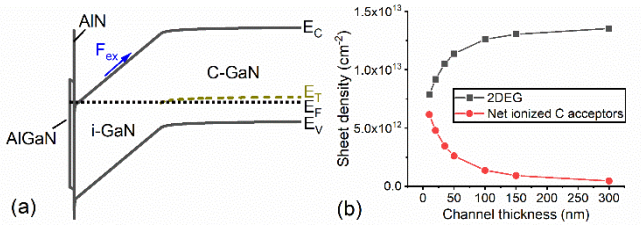


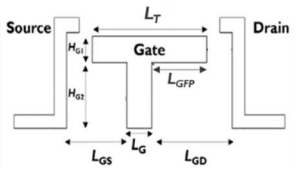
Figure 6 (a) Schematic energy band. (b) Simulated charge sheet density as a function of GaN channel thickness.

3) *Parasistics*

From (1), parasitic reduction helps to boost f_{max}. To study the layout impact, geometry scaling must be included in the parasitic models. In case of R_G for instance, the appropriate cross section of the T-shape gate must be considered (Figure 7). Our experiments further showed that R_{vert}, the transverse component of R_G, can be neglected in practical cases [9].

$$R_G = R_{ga} + \frac{1}{N_f \times L_T}\left(\left(\frac{R_{sh,h}}{3 \times N_{gc}^2} \times W_f\right) + \left(\frac{R_{vert}}{W_f}\right)\right)$$

$$\text{with } R_{sh,h} = \frac{\rho}{H_{G1} + H_{G2} \times \frac{L_G}{L_T}}$$



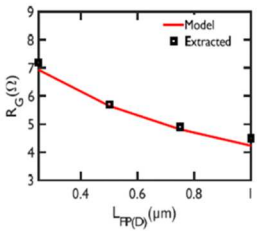


Figure 7 A gate resistance model that considers the cross-section of T-shape gates is required to predict f_{max} when the gate field plate is varied (bottom-right).

B. Dispersion modelling

GaN devices are subject to DC-RF dispersion (also named current collapse or dynamic R_{ON}), resulting from either defects in the barrier, surface traps, or bulk traps (Figure 8(a)). While surface traps can be controlled by field plates, bulk traps originate from the deep level dopants in the buffer and C-GaN layer, which is introduced to control leakage and short channel effects. We have previously reported the complex role of C/O/H impurities and the acceptor/donor energy levels associated with other buffer defects (Figure 8(b)), in the DC-RF dispersion observed for TLM devices subjected to buffer stress bias conditions [10].

In addition to the defects in the buffer layers, charge carrier interaction between 2DEG and defects in the barrier layer or surface interface states, also results in DC-RF dispersion. We propose to describe the complex kinetics associated with these defects using the Nonradiative-Multiphonon (NMP) theory [11], such that the temperature dependent inelastic tunnelling process (accounting for the interaction with lattice phonons) between the defect state and the GaN-conduction band is described by the *effective*-activation energy for charge capture (E_{Ac}) and emission (E_{Ae}). Consequently, the neutral and charged states of the defects can be represented as parabolic energy landscapes (Figure 8(c)) and the charge-trapping kinetics may be described by a distribution of activation energy barriers (E_{Ac} and E_{Ac}) that must be overcome for charge transfer between these states.

The charge-trapping kinetics associated with the surface/interface defects has previously been described with a time-dependent 'virtual-gate' extension model [12], such that the charge-injection from Gate metal into the surface defects is described using the SRH-theory [13] and the charge-emission from surface defects is described using the NMP-theory [11]. The interaction of the 2DEG with defects in barrier and bulk/buffer layers depends on the position of the Fermi-Energy level (E_F), and the temperature under the stress conditions. Therefore, it is essential to estimate the 2D potential profile in the barrier layer, such that the energy distribution of barrier defects can be accurately extracted under the required V_{GS}-V_{DS} operating bias conditions. The potential profile is seen to change considerably with distance from the Gate-edge (x = 0 nm to 48 nm) in the Drain-side access region (Figure 8(d)), while the shaded region of the defect distribution at each x-position illustrates the energy range of defects that will charge/discharge under a given V_{GS}-V_{DS} stress-bias condition.

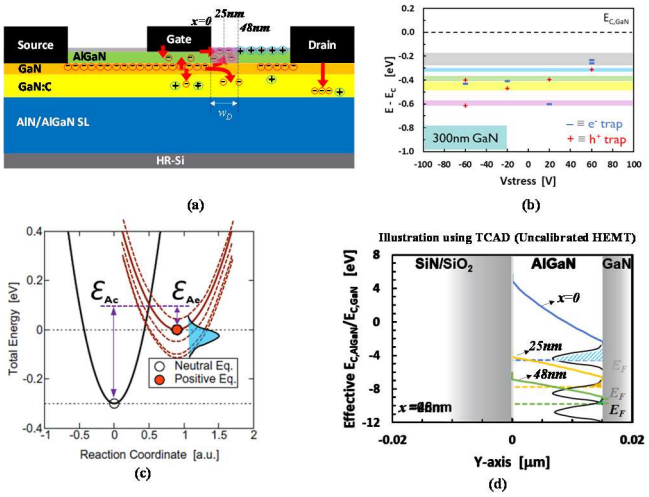


Figure 8 (a) Possible trapping centers in a GaN HEMT device include defects at the AlGaN/SiN interface, in the AlGaN barrier, and bulk/buffer layers, (b) various defect states in the buffer-stack characterized using buffer defect spectroscopy [10], (c) NMP-theory suggests a 2-state model to describe the neutral and charged states of a defect, such that E_{Ac} and E_{Ae} are the *effective* activation energies for charge capture and emission, respectively, (d) Potential profile in the AlGaN barrier in the drain-side access region (for x = 0nm to 48nm), along with the defect distribution that interacts with the 2DEG under a stress bias condition (V_{GS},V_{DS}) = (-1V, 10V), shown as shaded region.

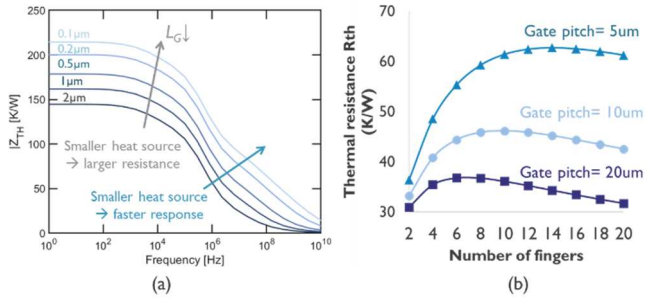


Figure 9 Impact of scaling on the thermal resistance simulated from BTE. (a) thermal impedance increases when gate length is reduced because of the smaller size of the heat source (width =50 μm). (b) Increase thermal impedance for short gate pitch; when the total width is fixed (Wtot=1mm, Lg=320nm), a maximum of *Rth* is observed for the 'squarest' layouts.

C. Thermal modelling

Because of the large power involved, the front-end modules heat up rapidly, and the performance of the technology can be limited by thermal effects. To account for thin film and interfacial filtering effects, we determine thickness-dependent in-plane and cross-plane GaN, AlGaN and AlN thermal conductivities from Monte Carlo BTE simulations with first-principles phonon dispersions and scattering rates [14]. These properties are then put into a 2.5D thermal model to assess self-heating for a variety of configurations. Device self-heating is a complex interplay of not only layer thicknesses and material properties but also surface layout of the active regions. The framework is used to study the impact of layout on thermal resistance (Rth) and temperature rise (Figure 9). For a given total width, the Rth peaks at the "squarest" layouts, i.e., those that minimise the aspect ratio and circumference of the active region. Such configurations

suffer most from finger cross-heating while benefiting least from lateral heat spreading along the sides and thus produce the highest Rth.

D. Substrate modelling

When grown on a Si platform, substrate modelling brings another challenge since lossy substrates affect the quality factor of passive components, but also the PA performance [15]. The small- and large-signal characteristics of substrates are studied from coplanar transmission lines (CPW). The direct relationship between RF losses and distortion observed in SOI substrates no longer holds in the case of GaN/Si stacks comprising a multi-layer semiconducting buffer, even if competing performance have been reported [16]. The strong hysteresis observed in Figure 10 is attributed to the long emission time constants of traps located inside the III/N buffer layers [17]. Therefore, the prediction of substrate nonlinearity as a function of DC bias is more challenging and requires a precise modelling of the buffer dynamic effects. While regular TCAD can be used to model the substrate nonlinearity [18], the method is limited to the case of TEM mode propagation.

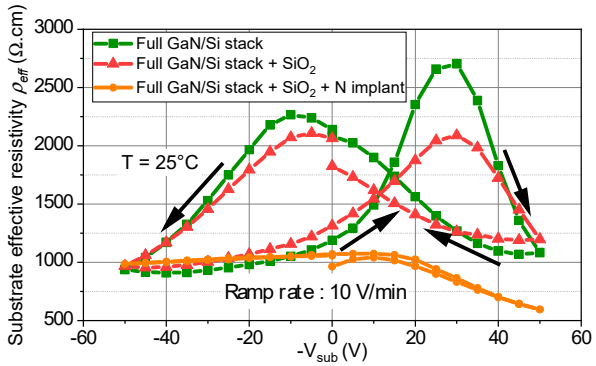


Figure 10 Effective substrate losses measured on CPW lines (50 µm separation) versus chuck bias for various GaN/Si stacks.

E. Large signal modelling

Crucial for the design of mmWave PA, the large signal modelling is however challenging because it combines the complications of high frequencies, nonlinear behaviour, and trap effects. Improvements in metrology in the past two decades permitted to validate nonlinear models over a wide frequency range. Nonlinear VNA can for instance be used to extract directly FOM such as AM/AM and AM/PM (Figure 11), but it is also useful in model calibration and validation.

As soon as the compact models incorporate the relevant effects listed above, they can be used to evaluate circuit FOM as a function of technology elements. This provides useful feedback to technology and feedforward to circuit design. The example of quantifying the impact of gate field plate on large signal metrics is illustrated in Figure 12.

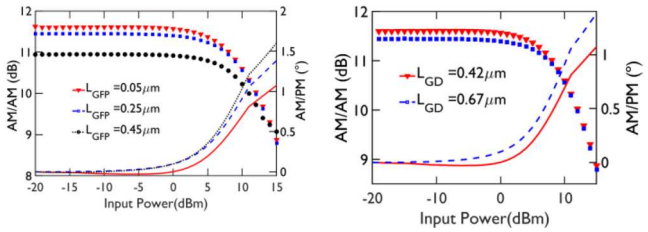


Figure 11 Nonlinear FOM extracted from NVNA on GaN HEMT with various gate field plates and gate-to-drain separations. LG=0.14um, Wf=25um, Nf=8, VDD=8V, ID=320mA/mm.

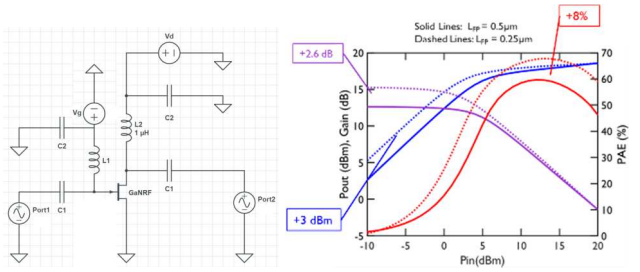


Figure 12 Simulating a simple PA stage with the pathfinding compact models. Smaller field plates are preferred for their reduced parasitic capacitance.

IV. Concluding Summary

Fast development of new technologies dedicated to wireless communication is based on the development of models capable of capturing effects over wide frequency scale (DC to mmWave) and large geometry range. Feedback to technology development is made possible when physics-based models are adopted. This paper illustrates the importance of modelling precisely geometry scaling, dispersion, thermal, and substrate in the case of GaN/Si technology.

References

[1] Nandana Rajatheva *et al.* , arXiv 2004.14247, 2020.
[2] J. Doré *et al*., 6G SUMMIT, 2020, pp. 1-5.
[3] Hua Wang, et al., https://gems.ece.gatech.edu/PA_survey.html.
[4] H. W. Then *et al.,* IEEE Trans. on Eletr. Dev., pp. 5306, Dec. 2020.
[5] B. Parvais *et al*., IEEE Int. Electr Dev. Meeting (IEDM), 2020.
[6] U. Peralagu *et al.*, IEEE Int. Electr Dev. Meeting (IEDM), 2019.
[7] A. Dasgupta *et al.*, IEEE Int. Conf. on EDSSC, 2015, pp. 495-498.
[8] U. Radhakrishna *et al.*, IEEE Trans. Electr. Dev., pp. 95-105, Jan. 2019.
[9] R. Y. ElKashlan *et al.*, IEEE Trans. Electr. Dev., pp. 4592, Nov. 2020.
[10] V. Putcha *et al.*, IEEE Int. Rel. Physics Symp. (IRPS), pp. 1-8, 2021.
[11] T. Grasser, *Bias Temp. Instability for Dev. and Circ.*, Springer, 2013, pp. 447–481.
[12] J. Kaushik *et al.*, *Semicond. science and techn.* Vol. 28, No. 1, 2012.
[13] D. K.Schroder, Semiconductor Material and Device Characterization 3rd edn (New York: Wiley), 2006
[14] J. Carrete *et al.*, Computer Physics Com., Vol. 220, pp. 351-362, 2017.
[15] Y. Yamaguchi *et al.*, IEEE CSICS, 2016, pp. 1-4.
[16] S. Yadav et al., IEEE Int. Electr. Dev. Meeting (IEDM), pp. 8.2.1, 2020.
[17] P. Cardinael *et al.*, IEEE ESSDERC, 2021 (accepted).
[18] M. Rack and J.-P. Raskin, IEEE Int. Microwave Symp, pp. 91-94, 2017.