POLITECNICO DI TORINO
Repository ISTITUZIONALE

Efficient TCAD Thermal Analysis of Semiconductor Devices

Original
Efficient TCAD Thermal Analysis of Semiconductor Devices / Catoggio, E., Guerrieri, S.D., Bonani, F.. - In: IEEE TRANSACTIONS ON ELECTRON DEVICES. - ISSN 0018-9383. - STAMPA. - 68:11(2021), pp. 5462-5468. [10.1109/TED.2021.3076753]

Availability:
This version is available at: 11583/2900732 since: 2021-10-26T09:31:33Z

Publisher:
IEEE

Published
DOI:10.1109/TED.2021.3076753

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
IEEE postprint/Author's Accepted Manuscript

©2021 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

12 October 2026

Efficient TCAD thermal analysis of semiconductor devices

E. Catoggio, *Student Member, IEEE*, S. Donati Guerrieri *Member, IEEE*, and F. Bonani, *Senior Member, IEEE*

***Abstract*— We present an efficient numerical technique for the temperature-dependent TCAD analysis of semiconductor devices. The approach is based on the linearization of the physical model around a nominal temperature operating condition, and exploits the Green's Functions (GFs) of the physical model to estimate the device characteristics at a different temperature. Self-heating can also be accounted for through coupling an external thermal circuit, with no numerical overhead. In the static (DC) case, the thermal analysis can be readily implemented in TCAD tools, using GFs customarily used for the electron device noise and variability analyses. The proposed technique is also extended to the dynamic case, to model the temperature dependency of electron devices operating in Large Signal conditions. We provide an extensive validation of the GF-based thermal analysis, including simulations of a 22 nm FinFET and of an InGaAs HEMT with 250 nm gate length. In the dynamic case, we demonstrate the efficient thermal analysis of a medium power amplifier in a FinFET technology.**

***Index Terms*— Semiconductor device modeling, Thermal analysis, Self-heating, Harmonic Balance analysis**

I. Introduction

ELECTRON device thermal modeling has evergrowing relevance, especially in the scenario of power devices (e.g. GaAs or GaN based HEMTs) and nanoscale devices (e.g. FinFETs) [1], [2]. Self-consistent physics-based electrothermal modelling, coupling the Fourier equation to the electrical device simulations, would be the ideal choice [3], [4], but often it represents a too demanding task for efficient and fast simulations. On the other hand, when the intrinsic device can be considered approximately isothermal at an equivalent lattice temperature T, electrical TCAD simulations varying T as a position-indepedent free parameter, allow to identify a T-dependent electrical device model, e.g. extending the X-parameter approach in [5], to be coupled to an external lumped thermal circuit to achieve self-consistency, and thus self-heating assessment. Repeated TCAD simulations at varying temperature or with a varying equivalent thermal circuit (e.g. a thermal resistance R_{th}), though, are still extremely demanding from the numerical standpoint, calling for efficient ways to calculate the device temperature response.

In this work we present a comprehensive framework to efficiently calculate the device response to temperature variations with respect to a nominal ("cold") thermal status. The modeling approach, based on the linearization of the physical model around the "cold" temperature, exploits the model Green's Functions (GFs). Such GFs can be calculated efficiently from the numerical standpoint, and are already one of the main tools for the device noise and variability analyses, where the device operation deviates from a nominal condition because of a deterministic or random variation of a technological or physical parameter. Here, the same technique is exploited, taking the device lattice temperature to be the varied parameter. With this approach, the numerical burden of the TCAD simulations is required only at one nominal temperature and the computational advantage scales with the number of temperatures or thermal resistances to be analyzed.

GFs are available in some commercial TCAD simulators (e.g. [6]) when the operating condition is time-invariant (DC), hence allowing only for a DC thermal analysis. Green's Functions-based analysis in the dynamic case requires instead TCAD tools not yet available at the commercial level. We exploit our in house drift-diffusion code, implementing the Harmonic Balance frequency domain analysis for the device Large-Signal (LS) multi-tone periodic or quasi-periodic simulations [7]. In our code, the Conversion Green's Functions (CGFs), extending the GF concept to the frequency (harmonic) domain, describe the linearized model around the steady-state periodic solution.

We report various examples to demonstrate the advantage of GF-based approach in T-dependent device simulations, including the self-heating analysis in varying embedding thermal circuits. Finally, we present, for the first time, a temperature dependent LS analysis of a FinFET-based RF medium power amplifier fully based on the proposed efficient TCAD analysis.

II. GF approach to T-dependent simulations

The Green's Function technique is a well known approach to device noise [7]–[10] and parametric variability analysis [11], [12], aimed at assessing the device response to the (small) variation of a physical parameter. In this work we focus on extending the GF technique [11], [12] to estimate the device response to temperature variations with respect to a nominal temperature T_0. All details of the GF numerical calculation can be found in [7], [9] and are not repeated here.

This work has been supported by the Italian Ministero dell'Istruzione dell'Università e della Ricerca (MIUR) under the PRIN 2017 Project "Empowering GaN-on-SiC and GaN-on-Si technologies for the next challenging millimeter-wave applications (GANAPP)"

Eva Catoggio (e-mail: eva.catoggio@polito.it), Simona Donati Guerrieri (E-mail: simona.donati@polito.it) and Fabrizio Bonani (Email: fabrizio.bonani@polito.it) are with the Electronics and Telecommunications Department, Politecnico di Torino, Torino, Italy

Let us consider a bipolar device physics-based model expressed in terms of a set of coupled partial differential equations $M^{(\alpha)}(\boldsymbol{\beta})$, e.g. the Poisson equation $M^{(\varphi)}$, the current continuity equations $M^{(n)}$ and $M^{(p)}$, the energy balance equations $M^{(T_n)}$ and $M^{(T_p)}$, and the quantum potential equations $M^{(q_n)}$ and $M^{(q_p)}$. Argument $\boldsymbol{\beta}$ includes all the relevant unknowns, e.g. the electrostatic potential φ, electron and hole concentrations n and p, carrier temperatures T_n and T_p, and quantum potentials q_n and q_p. The numerical solution of the model requires to discretize the differential equations on the device volume Ω, yielding, for each equation α

$$\mathbf{F}^{(\alpha)}(\boldsymbol{\beta}, \mathbf{v}_\mathrm{e}; T) = \mathbf{D}^{(\alpha)}\dot{\boldsymbol{\beta}}. \tag{1}$$

where $\boldsymbol{\beta}$ represents now the vector of unknown nodal samples and we have formally separated the time-invariant part of discretized equation $M^{(\alpha)}$, $\mathbf{F}^{(\alpha)}(\boldsymbol{\beta}, \mathbf{v}_\mathrm{e}; T)$, from the time-varying (memory) part, proportional to $\dot{\boldsymbol{\beta}}$ (the time derivative of $\boldsymbol{\beta}$) through the block diagonal matrix $\mathbf{D}^{(\alpha)}$ (notice that $\mathbf{D}^{(\alpha)}$ is null for memory-less equations, such as Poisson equation $\alpha = \varphi$). The external voltage source[1] vector $\mathbf{v}_\mathrm{e}$ sets the operating regime of the device, either static or dynamic.

In (1) we have highlighted the system temperature dependency, which takes place through a wide number of physical parameters and models, e.g. doping activation, mobility, diffusivity, thermal voltage etc. In temperature-dependent TCAD simulations, T is assumed to be uniform in the device and equal to the lattice temperature. Hence, in these simulations T is not a model unknown, but rather a model parameter.

A. Static (DC) analysis

In the static case the external voltage is constant (DC bias), and thus $\mathbf{D}^{(\alpha)} = \mathbf{0}$. The "cold" solution $\boldsymbol{\beta}_0$ is found solving (1) for each DC bias at the nominal lattice temperature T_0. When the temperature undergoes a variation $\Delta T = T - T_0$ with fixed electrical excitation, (1) is linearized around the nominal solution, yielding for the perturbation $\Delta\boldsymbol{\beta} = \boldsymbol{\beta} - \boldsymbol{\beta}_0$

$$\mathbf{J}^{(\alpha)}(\boldsymbol{\beta}_0, \mathbf{v}_\mathrm{e}; T_0)\Delta\boldsymbol{\beta} + \left.\frac{\partial \mathbf{F}^{(\alpha)}}{\partial T}\right|_{(\boldsymbol{\beta}_0, \mathbf{v}_\mathrm{e}; T_0)} \Delta T = 0 \tag{2}$$

where $\mathbf{J}^{(\alpha)}$ is the Jacobian matrix of $\mathbf{F}^{(\alpha)}$ evaluated at the nominal solution $(\boldsymbol{\beta}_0, \mathbf{v}_\mathrm{e}; T_0)$. Defining the equivalent source:

$$\mathbf{s}_0^{(\alpha)}(\boldsymbol{\beta}_0, \mathbf{v}_\mathrm{e}; T_0, T) = -\left.\frac{\partial \mathbf{F}^{(\alpha)}}{\partial T}\right|_{(\boldsymbol{\beta}_0, \mathbf{v}_\mathrm{e}; T_0)} \Delta T \tag{3}$$

equation (2) becomes

$$\mathbf{J}^{(\alpha)}(\boldsymbol{\beta}_0, \mathbf{v}_\mathrm{e}; T_0)\Delta\boldsymbol{\beta} = \mathbf{s}_0^{(\alpha)}(\boldsymbol{\beta}_0, \mathbf{v}_\mathrm{e}; T_0, T). \tag{4}$$

The term $\mathbf{s}_0^{(\alpha)}$, depending on both the nominal and varied temperature, represents a *distributed perturbation* for the linearized equation α, depending on the position inside the device (mesh node) through $\boldsymbol{\beta}_0$ and T. This source gives rise to a variation $\Delta I_k^T = I_k(T) - I_k(T_0)$ of the k-th device terminal current I_k that can be efficiently computed via the GF approach by the *discretized* spatial convolution integral

$$\Delta I_k^T = \sum_\alpha \mathbf{G}_k^{(\alpha)}(\boldsymbol{\beta}_0, \mathbf{v}_\mathrm{e}; T_0)\, \boldsymbol{\Omega}\, \mathbf{s}_0^{(\alpha)}(\boldsymbol{\beta}_0, \mathbf{v}_\mathrm{e}; T_0, T) \tag{5}$$

where the row vector of the Green's Function nodal values $\mathbf{G}_k^{(\alpha)}(\boldsymbol{\beta}_0, \mathbf{v}_\mathrm{e}; T_0)$ relates the observation variable I_k^T to a unit source in each mesh node of equation α and $\boldsymbol{\Omega}$ is a diagonal matrix containing the nodal Vornoi volumes.

The estimation of the source term (3) requires to compute the derivatives of the discretized physical equations with respect to T. Although this is in principle possible, it turns out to be quite complicated, as lattice temperature enters essentially in the analytical formulation of every physical model parameter and in the Bernoulli functions of the Scharfetter-Gummel discretization scheme of the current continuity equations [13]. A convenient workaround was proposed in [11] for a generic parametric variations. Accordingly, the analytical Jacobian in (3) is approximated through finite differences as

$$\mathbf{s}_0^{(\alpha)}(\boldsymbol{\beta}_0, \mathbf{v}_\mathrm{e}; T_0, T) \simeq -\mathbf{F}^{(\alpha)}(\boldsymbol{\beta}_0, \mathbf{v}_\mathrm{e}; T) + \cancelto{0}{\mathbf{F}^{(\alpha)}(\boldsymbol{\beta}_0, \mathbf{v}_\mathrm{e}; T_0)}$$

where the last term is null in the nominal solution. This effectively amounts to evaluating only the residual function $\mathbf{F}^{(\alpha)}$ at the varied temperature T. The computation is very fast and no matrix storage is required.

B. Dynamic, periodically time-varying analysis

In the dynamic case, time derivatives in (1) must be taken into account. For analog and power applications, we are often interested in the particular case of a periodic or quasi-periodic external source $\mathbf{v}_\mathrm{e}(t)$, so that the system steady-state becomes periodically time-varying. Such case is better treated in the frequency domain exploiting the Harmonic Balance (HB) technique [14]. Let us denote with $\tilde{\boldsymbol{\beta}}$ and $\tilde{\mathbf{F}}^{(\alpha)}$ the vector of the Fourier (harmonic) amplitudes of the time-varying functions $\boldsymbol{\beta}$ and $\mathbf{F}^{(\alpha)}$, truncated to the harmonic order n_H. The Fourier-transformed system (1) reads [7]

$$\tilde{\mathbf{F}}^{(\alpha)}(\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{v}}_\mathrm{e}; T) = \boldsymbol{\Omega}^{(\alpha)}\tilde{\boldsymbol{\beta}}. \tag{6}$$

where $\boldsymbol{\Omega}^{(\alpha)}$ is a block-diagonal matrix including the dependence on the fundamental frequency ω_0 and on its harmonics $\omega_n = n\omega_0$ $(n = 1, 2, \ldots, : n_\mathrm{H})$ and $\tilde{\mathbf{v}}_\mathrm{e}$ is the Fourier transform of $\mathbf{v}_\mathrm{e}(t)$. The solution of the above algebraic system yields the Fourier coefficients $\tilde{\boldsymbol{\beta}}_0$ of the large-signal (LS) steady-state periodic solution $\boldsymbol{\beta}_0$, with nominal temperature T_0.

When the temperature undergoes a variation, (1) must be linearized around $(\boldsymbol{\beta}_0, \mathbf{v}_\mathrm{e}; T_0)$, giving rise to a linear periodically time-varying (LPTV) system. LPTV systems can be analyzed in the frequency domain with the sideband frequency conversion analysis [15], a general framework to describe the response of a periodic system to a perturbation with a characteristic frequency ω: the variation is in general characterized by a set of *sidebands* of the LS harmonics $\tilde{\omega}_n = \omega_n \pm \omega$, but in the particular case of a time-invariant perturbation, such as the lattice temperature, we can set $\omega = 0$ and $\tilde{\omega}_n = \omega_n$, i.e. the sidebands collapse onto the unperturbed frequencies. In

[1] We consider here the case of a voltage-driven device. The extension to current driven terminals is obvious.

practice, a static perturbation of the LS state is characterized by the same set of frequencies as the unperturbed solution.

With this assumption, the linearized system can be expressed in the frequency domain as [7]

$$\tilde{\mathbf{J}}^{(\alpha)}(\tilde{\boldsymbol{\beta}}_0, \tilde{\mathbf{v}}_\mathrm{e}; T_0)\Delta\tilde{\boldsymbol{\beta}} - \boldsymbol{\Omega}^{(\alpha)}\Delta\tilde{\boldsymbol{\beta}} = \tilde{\mathbf{s}}_0^{(\alpha)}(\tilde{\boldsymbol{\beta}}_0, \tilde{\mathbf{v}}_\mathrm{e}; T_0, T) \quad (7)$$

where: $\Delta\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}_0$ is the effect of the temperature variation ΔT; $\tilde{\mathbf{J}}^{(\alpha)}$ is the jacobian conversion matrix, a Toeplitz matrix built with the Fourier coefficients of the time-periodic function $\mathbf{J}^{(\alpha)}(\boldsymbol{\beta}_0, \mathbf{v}_e; T_0) = \partial\mathbf{F}^{(\alpha)}/\partial\boldsymbol{\beta}$ calculated in the periodic solution $\boldsymbol{\beta}_0$ and at T_0. The source term $\tilde{\mathbf{s}}_0^{(\alpha)}$ collects the Fourier coefficients of

$$\mathbf{s}_0^{(\alpha)}(\boldsymbol{\beta}_0, \mathbf{v}_e; T_0, T) = -\frac{\partial\mathbf{F}^{(\alpha)}(\boldsymbol{\beta}, \mathbf{v}_e; T)}{\partial T}\,\Delta T \quad (8)$$

extending (3) to the periodic case. As in the static case, a finite difference approximation

$$\tilde{\mathbf{s}}_0^{(\alpha)}(\tilde{\boldsymbol{\beta}}_0, \tilde{\mathbf{v}}_\mathrm{e}; T_0, T) \simeq -\tilde{\mathbf{F}}^{(\alpha)}(\tilde{\boldsymbol{\beta}}_0, \tilde{\mathbf{v}}_\mathrm{e}; T) \quad (9)$$

avoids the explicit calculation of the temperature derivatives.

The temperature variation ΔT induces a perturbation of the terminal current at each device contact k, with spectral amplitudes $\Delta\tilde{\mathbf{I}}_k^T = \tilde{\mathbf{I}}_k(T) - \tilde{\mathbf{I}}_k(T_0)$. The Conversion Green's Functions (CGFs) [7] of the linearized system (7), matrices denoted as $\tilde{\mathbf{G}}_k^{(\alpha)}(\tilde{\boldsymbol{\beta}}_0, \tilde{\mathbf{v}}_\mathrm{e}; T_0)$, allow to compute such current variations through a spatial convolution integral, thus extending (5) to the large-signal case

$$\Delta\tilde{\mathbf{I}}_k^T = \sum_\alpha \tilde{\mathbf{G}}_k^{(\alpha)}(\tilde{\boldsymbol{\beta}}_0, \tilde{\mathbf{v}}_\mathrm{e}; T_0)\,\tilde{\boldsymbol{\Omega}}\,\tilde{\mathbf{s}}_0^{(\alpha)}(\tilde{\boldsymbol{\beta}}_0, \tilde{\mathbf{v}}_\mathrm{e}; T_0, T) \quad (10)$$

where $\tilde{\boldsymbol{\Omega}}$ is a block diagonal matrix containing the Vornoi volumes.

C. Advantages of the GF approach

The linearized approach provides a computational advantage over repeated T-dependent analysis. In fact, the Generalized Adjoint Approach (GAA) [7], [9], [10] enables the estimation of the relevant GFs with a negligible overhead with respect to the computation of the nominal solution. Thus, the simulation time with the GF approach is reduced roughly by a factor n_T compared to n_T repeated simulations. This advantage is especially appealing when the numerical burden of each individual solution is very high, such as for 3D simulations or in the large-signal analysis, where the number of equations to be solved is increased with respect to DC by a factor $2n_\mathrm{H}+1$ (usually $n_\mathrm{H} \geq 10$ for highly nonlinear RF applications).

Another clear advantage of the proposed method is its compatibility with other simulations exploiting the same GFs, for instance the parametric variability analysis. In particular, the *concurrent* temperature and variability analyses can be carried out with virtually no extra numerical burden.

Finally, the GF approach opens the way to device optimization through the analysis of the so-called *local variations*, i.e. the integrand functions of (5) and (10), and through the device *thermal sensitivity,* defined in the DC case as

$$S_{I_k}^T = \frac{\partial I_k^T}{\partial T} \approx \Delta I_k^T\big|_{T=T_0+1\ \mathrm{K}} \quad (11)$$

This parameter is easily extracted from (5), while the extension to large-signal operation through the *harmonic thermal sensitivities* (one for each harmonic) is an obvious generalization of (11) using (10).

The thermal sensitivity is also the key quantity for the efficient implementation of self-heating analysis, as discussed in next Sec. II-D.

D. Self-heating through device thermal sensitivity

Let us consider the device embedded into an external structure, through which heat is dissipated towards a heat sink at temperature T_0 [16]. In the representation adopted in this paper, the device is isothermal and the heat dissipation process can be represented by a lumped thermal resistance[2] R_th. With a DC external voltage $V_{0,k}$ at the k-th terminal, the device self-heating leads to the temperature increase

$$\Delta T = T - T_0 = R_\mathrm{th}P_\mathrm{diss} = R_\mathrm{th}\sum_k I_k\,(T_0, T)\,V_{0,k} \quad (12)$$

where T is the final device operating temperature, and P_diss the total dissipated power. The k-th terminal current can be approximated by a first order Taylor expansion around the "cold" condition ($T = T_0$, $R_\mathrm{th} = 0$, $I_k = I_{0,k}$):

$$I_k\,(T_0, T) = I_k\,(T_0, T_0 + \Delta T) \simeq I_{0,k}(T_0) + S_{I_k}^T\,(T - T_0) \quad (13)$$

where $S_{I_k}^T$ is the thermal sensitivity (11). Using (12):

$$I_k\,(T_0, T) \approx I_{0,k}(T_0) + S_{I_k}^T R_\mathrm{th}\sum_j I_j\,(T_0, T)\,V_{0,j} \quad (14)$$

Relation (14) represents a linear system of equations allowing to extract I_k at all terminals. In the simpler case where $V_{0,j} = 0$ or $I_{0,j} = 0$ for $j \neq k$ (a relevant example is the case of FETs where, neglecting gate leakage, the only dissipation is through the drain), we find from (14)

$$I_k\,(T_0, T) = \frac{I_{0,k}(T_0)}{1 - S_{I_k}^T R_\mathrm{th} V_{0,k}} \quad (15)$$

If T_0 is not simply the heat sink temperature, but rather a reference temperature conveniently selected to evaluate the thermal resistance (e.g., the temperature of the foot of the fin in a FinFET device), it can be in turn subject to fluctuations, either deterministic or statistical. To account for a variation ΔT_0, the "cold" current $I_{0,k}$ is further linearized with respect to T_0, exploiting again the thermal sensitivity (11):

$$I_{0,k}(T_0 + \Delta T_0) \simeq I_{0,k}(T_0) + S_{I_k}^T \Delta T_0 \quad (16)$$

Substituting into (15), we finally obtain:

$$I_k\,(T_0 + \Delta T_0, T) = \frac{I_{0,k}(T_0)}{1 - S_{I_k}^T R_\mathrm{th} V_{0,k}} + \frac{S_{I_k}^T \Delta T_0}{1 - S_{I_k}^T R_\mathrm{th} V_{0,k}} \quad (17)$$

Therefore, accounting for both *self-heating* and *reference temperature variations* only requires the computation of the device currents and of their thermal sensitivities at the "cold" temperature T_0.

[2]We consider only static self-heating: the extension to the dynamic case will be presented elsewhere

```
Physics {
        Temperature=300
        DeterministicVariation       (
          ParameterVariation "GreenT" (
          ( Model="DeviceTemperature" Parameter="Temperature" Value=@T@)
                    )
        )
}
```

Fig. 1. Example of the `sdevice` command file (*.des*) for the GF-based T-dependent analysis. The deterministic temperature `T` is provided externally from the Sentaurus Workbench. The nominal temperature is here 300 K.

E. Numerical Implementation

Green's Functions are routinely calculated in advanced device TCAD simulators. Some commercial tools allow sufficient flexibility for the end user to compute the device response to a *used-defined* variation. For example, `sdevice` by Synopsys [6] allows to vary the lattice temperature using built-in commands, originally dedicated to a generic deterministic parameter variation. As such, the proposed GF based thermal analysis can be directly implemented. Fig. 1 shows an example of the relevant lines to be added to the `sdevice` input command file. This method is adopted for all the simulations reported in Sec. III-A.

On the other hand, `sdevice` cannot be used to validate the large-signal Green's Function-based T-dependent analysis. We use, instead, our in-house simulator, which has been extended to allow for temperature dependent simulations. The temperature dependency of the physical parameters has been fine-tuned to match the one from `sdevice` for the relevant material systems, namely Silicon, AlGaAs/GaAs and AlGaN/GaN alloys.

To validate the GF-based self-heating analysis of Sec. II-D, the in-house code implements an *ad-hoc* modified TCAD analysis, where the drift-diffusion model is self-consistently solved with the additional equation (12).

III. Validation

A. T-dependent DC analysis

For the validation of the proposed T-dependent analysis in the static case, we first adopt `sdevice`, in order to exploit the accuracy of its built-in physical models and the 3D simulation capabilities. We consider two advanced device structures where thermal effects are expected to play a significant role.

In the first example, we consider the library AlGaAs/InGaAs/GaAs HEMT for RF applications[3] shown in Fig. 2. Here, temperature effects are important for typical RF power applications. The Al and In mole fractions are 0.3 and 0.75, respectively. The recessed gate is 0.25 μm long with 0.9 eV Schottky barrier. Physical simulations are carried out in 2D exploiting the electron energy balance transport model, i.e. including the electron temperature equation. Velocity saturation, doping dependent mobility and generation-recombination effects are included according to the default material parameters from the library example[3]. The device was tested in the $T \in [280 - 340]$ K temperature range. The GFs

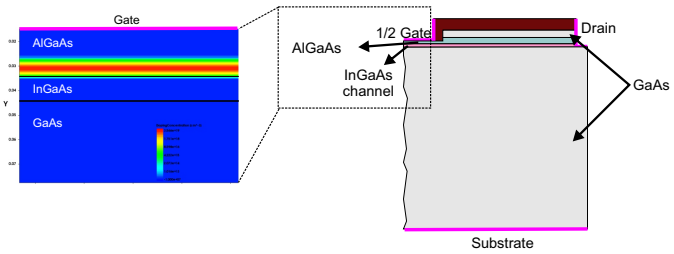

Fig. 2. Half structure of the HEMT (the device is symmetric). The AlGaAs supply layer is 20 nm and 10 nm thick. The InGaAs channel (undoped) is 10 nm thick. A 2 nm thick δ-doping layer with peak concentration of $\mathbf{3 \times 10^{19}}$ cm^{-3} is positioned 4 nm above the channel. The gate recess is 15 nm. The source and drain are implanted with a peak doping concentration of $\mathbf{10^{19}}$ cm^{-3} (not shown). In the inset: doping concentration under the gate, showing the δ-doping in the AlGaAs layer.

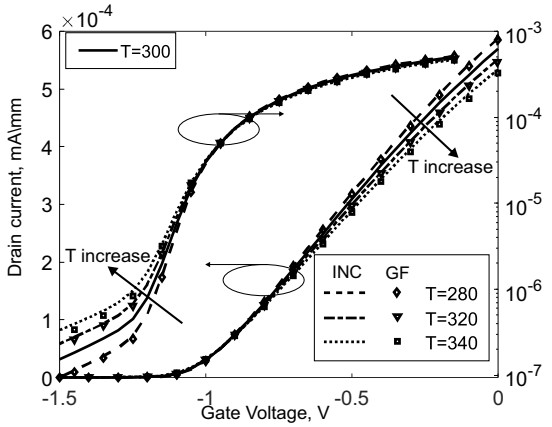

Fig. 3. HEMT transcharacteristics as a function of temperature. Left axis: linear scale; right axis: log scale.

are extracted at the nominal temperature $T_0 = 300$ K, and the GF based thermal analysis is validated against repeated simulations varying T (hereafter the incremental – INC – approach).

Fig. 3 and Fig. 4 show the transcharacteristic and output characteristics of the device, respectively. Above threshold, the drain current decreases as a function of the temperature, driven by the mobility degradation. Below threshold, the drain current has the opposite behavior, due to a complex intermix of bandgap temperature dependency and gate workfunction behavior. The GF approach always compares very well with the reference solution in the simulated temperature range. Furthermore, above threshold we verified it to be fairly accurate even up to 380 K (not shown in the figures), demonstrating the potential of the GF analysis.

In the second example, we address the 3D simulation of a 22 nm FinFET device, including the typical features of the FinFET technology, i.e. raised source/drain, high-k dielectric and oxide side-walls[4], see Fig. 5. Here, thermal effects are relavant because of the extreme miniaturization. Simulations account for doping dependent mobility, velocity saturation, bandgap narrowing, interface traps and quantum effects through the Density Gradient approach[4]. With roughly 120000 mesh nodes, the transcharacteristic simulation for this

[3]TCAD Sentaurus Version O-2018 installation, *Applications_Library/HETERO/HEMT_RF*

[4]TCAD Sentaurus Version O-2018 installation, *Applications_Library/Variability/FinFET_Variability_sIFM*

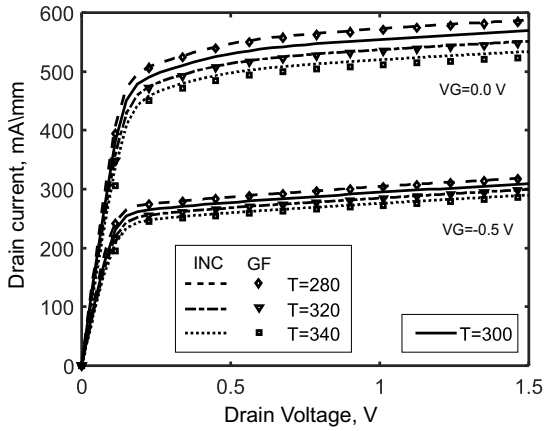

Fig. 4. HEMT output characteristics as a function of temperature.

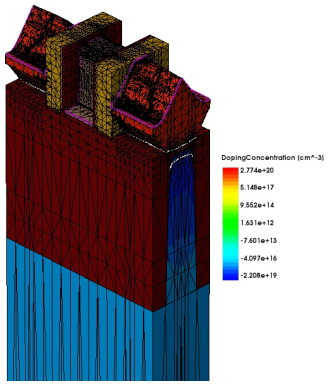

Fig. 5. 3D structure of the simulated 22 nm FinFET. Only the active (top) part of the simulated fin is shown.

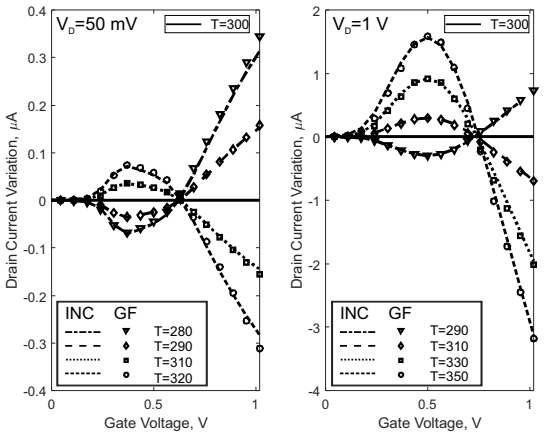

Fig. 6. FinFET drain current variation as a function of temperature. Left: linear region $\boldsymbol{V_D = 50}$ mV; right: saturation region $\boldsymbol{V_D = 1}$ V.

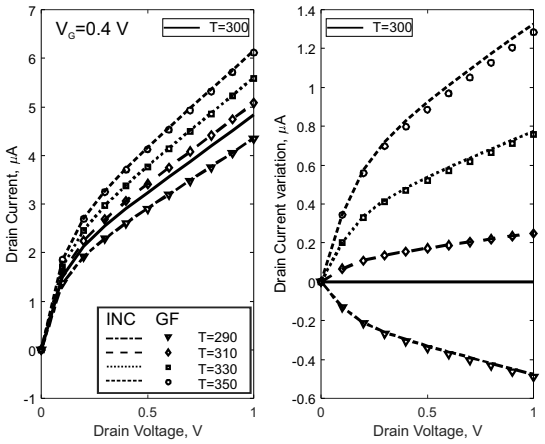

Fig. 7. FinFET drain current (left) and its variation (right) for $\boldsymbol{V_G = 0.4}$ V and varying temperatures.

device takes more than 5 hours on a PC with 64GB RAM and 8 core 3.6 GHz CPU (maximum number of threads equal to 4).

Fig. 6 shows the temperature dependency of the drain current, as a function of the gate voltage V_G. In the linear region ($V_D = 50$ mV) we observe an excellent agreement of the GF approach against the reference solution over a 40 K temperature interval, around the nominal value of 300 K; in the saturation region, the agreement is found in a an even wider range of 70 K, up to 350 K. In both cases the drain current first increases with temperature (up to $V_G = 0.6$ V), while at higher gate voltages the trend is opposite due to mobility degradation. Between the threshold ($V_{th} \approx 0.25$ V for this device) and $V_G = 0.6$ V the thermal sensitivity attains the highest value for any ON condition: in particular, Fig. 7 shows the spread of the output characteristics (left) for $V_G = 0.4$ V, along with the variation with respect to the nominal 300 K condition (right). At $T = 350$ K the drain current is more than 25% higher, but the GF analysis still compares in an excellent way with the INC approach, the latter requiring more than 40 hours of simulation time against the 8 hours of the proposed approach.

Fig. 8 highlights the FinFET behavior in the subthreshold regime, where the drain current dependency on T is exponential. Despite being a linearized approach, the GF method is still accurate vs. the INC approach, albeit on a slightly lower temperature range $T \in [290-320]$ K. The agreement is confirmed in Fig. 9, where the threshold voltage and off current are shown. The linear and saturation regions are both reported to highlight short channel effects. The typical degradation of the threshold voltage with T is accurately predicted by the GF model in all operating conditions.

B. DC Self-heating

The double gate device in Fig. 10 represents the simplified 2D cross section of a FinFET similar to the device in Fig. 5. The in-house simulator implementing the drift-diffusion model self-consistently solved with (12) has been used to validate the self-heating approach described in Sec. II-D, see (17). Fig. 11 shows the device output characteristics with $R_{th} = 0.3$ K/μW and $R_{th} = 1$ K/μW, showing a reduction of the current up to 17% due to self-heating. Fig. 12 shows the sensitivity to the external reference temperature T_0. With T_0 raised to 320 K and $R_{th} = 1$ K/μW, the device temperature is as high as 390 K for the largest dissipation. All simulations show that the GF approach, requiring a single simulation of the "cold" device with $T_0 = 300$ K and $R_{th} = 0$, compares very well with the reference simulations, that require instead multiple solutions with varying R_{th} and T_0.

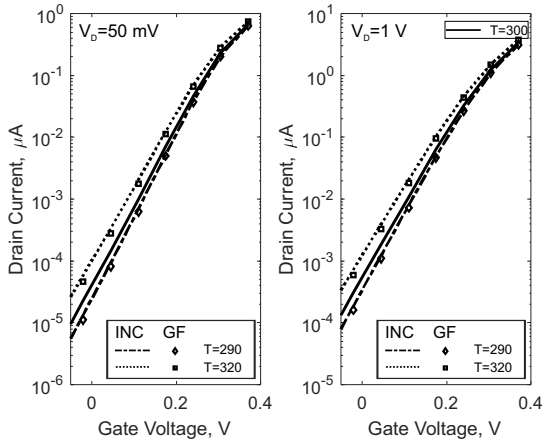

Fig. 8. FinFET transcharacteristic as a function of temperature in subthreshold and weak inversion conditions. Left: linear region $V_D = 50$ mV; right: saturation region $V_D = 1$ V.

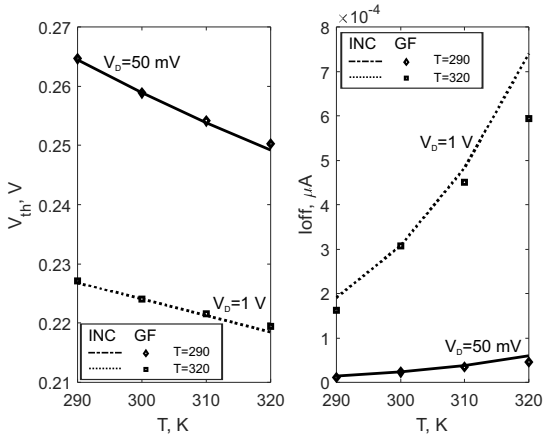

Fig. 9. Threshold voltage (left) and OFF current (right) as a function of temperature in linear and saturation condition. Lines: INC; symbols: GF

C. T-dependent Large Signal Analysis

FinFET technology is actively investigated for possible application in the RF 5G scenario [17], [18]. Small and medium power amplifiers are among the most challenging stages for these applications and, being especially prone to heating, their design calls for accurate temperature models. Here we propose the thermal analysis of a medium power class A amplifier operating at the frequency of 70 GHz, exploiting our in-house 2D device simulator applied to the double gate device in Fig. 10. The amplifier is designed assuming a multifinger device (10 fingers of 30 fins each) with a fin height of 25 nm, corresponding to a total gate periphery of 15 μm (only the two lateral channels for each fin are considered).

The DC bias of the class A power amplifier (PA) has

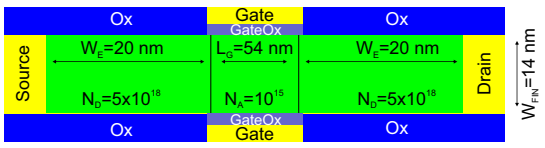

Fig. 10. Double Gate structure used for the validation of the in-house code for T-dependent self-heating and dynamic analysis.

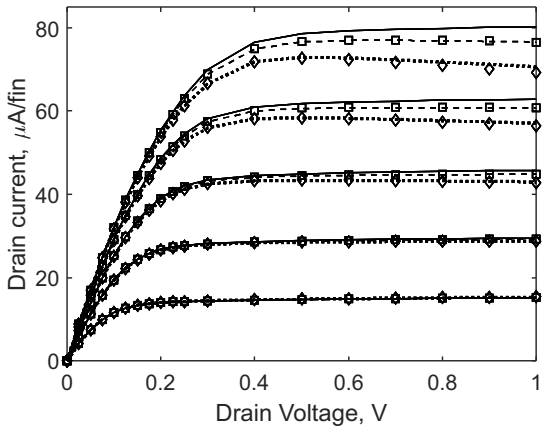

Fig. 11. FinFET output characteristics with gate voltage $V_G = [0.6, 0.7, 0.8, 0.9, 1]$ V. Solid lines: cold device ($T_0 = 300$ K); dashed lines & squares $R_{th} = 0.3\text{K}/\mu\text{W}$; dotted lines & diamonds $R_{th} = 1\text{K}/\mu\text{W}$. Lines: Drift-Diffusion coupled to (12). Symbols: GF analysis.

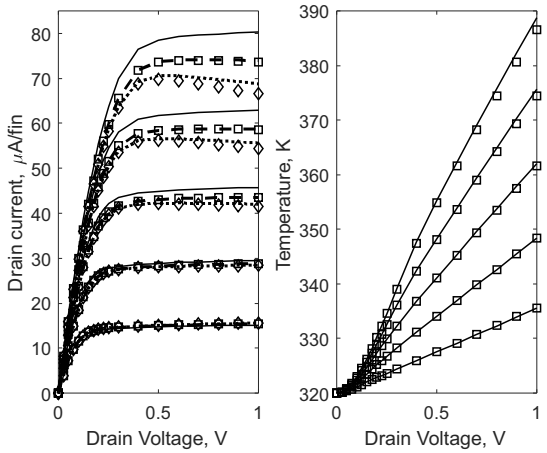

Fig. 12. Left: Output characteristics with gate voltage $V_G = [0.6, 0.7, 0.8, 0.9, 1]$ V. Solid lines: cold device ($T_0 = 300$ K, $R_{th} = 0$); dashed lines & squares $R_{th} = 0.3\text{K}/\mu\text{W}$ and $T_0 = 320$ K; dotted lines & diamonds $R_{th} = 1\text{K}/\mu\text{W}$ and $T_0 = 320$ K. Right: device temperature with $T_0 = 320$ K and $R_{th} = 1\text{K}/\mu\text{W}$. Lines: Drift-Diffusion coupled to (12). Symbols: GF analysis with (17).

been selected at $V_G = 0.675$ V and $V_D = 0.6$ V. The real part of the optimum load has been calculated according to the load-line approach, while the imaginary part tunes out the output admittance at the bias point and at the nominal temperature $T_0 = 300$ K. The optimum load turns out to be $Z_{opt} = 53 + j6\Omega$, while the drain harmonics are shunted by ideal tuners. The input port has been left unmatched and terminated with a 50 Ω/mm impedance. The large-signal analysis has been carried out with $n_H = 10$ harmonics, and increasing the input power driving the amplifier from back-off to approximately 2 dB gain compression. At each input power, the CGFs are calculated at $T_0 = 300$ K and the variation of the drain current harmonics with temperature is evaluated according to (10) for 5 temperatures ($T \in [310 - 350]$ K). GF results are compared to repeated LS analyses with varying temperature (incremental method, INC), always obtaining an excellent agreement between the two methods. Notice that the simulation time for each power sweep is roughly 4 hours, hence the GF method time saving is of about 20 hours.

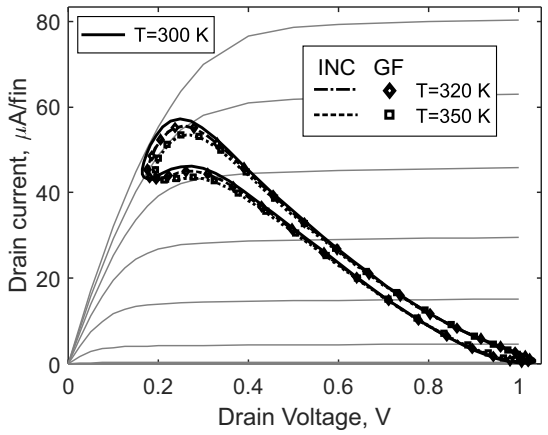

Fig. 13. Dynamic load lines as a function of temperature at 2 dB gain compression ($P_{disp} = -3$ dBm). Lines: incremental simulations. Symbols: GF approach.

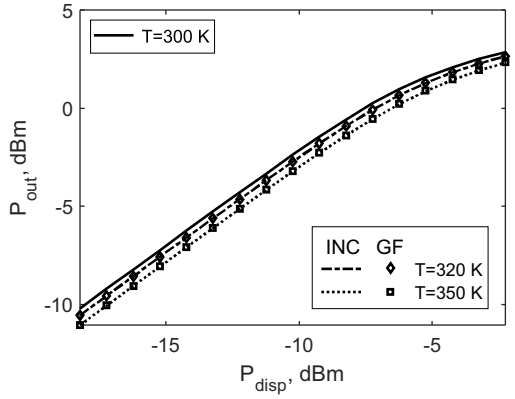

Fig. 14. $P_{in} - P_{out}$ plot for the class A PA as a function of temperature. Lines: incremental simulations. Symbols: GF approach.

Fig. 13 shows the dynamic load lines in the output characteristics plane: with increasing temperature, the drain current exhibits a reduced swing with respect to the "cold" device, essentially due to mobility degradation and to the increase of the knee voltage, suggesting that the output power is decreasing with T.

The $P_{in} - P_{out}$ curve in Fig. 14 confirms the thermal degradation, showing more than 1 dB output power reduction at $T = 350$ K. Due to the stage nonlinearity, though, the degradation depends on the input power. While the temperature sensitivity of the fundamental tone decreases at higher power, the harmonics are instead sharply increasing and also present the highest variation with T, see Fig. 15. The harmonic variations can be up to 10% with just 20 K of temperature increase. The linearized GF approach is always delivering a fast and accurate thermal analysis of the power stage.

IV. Conclusion

We have proposed and validated an efficient method based on Green's Functions for the TCAD temperature-dependent device simulation. The validation is carried out for realistic devices from different technologies in both the DC case, including self-heating, and the dynamic, periodically time-varying, regime. The technique is accurate even using advanced TCAD models, such as energy balance transport and the inclusion of quantum effects through the Density Gradient correction. For the first time, a T-dependent large signal analysis of a power amplifier is presented, showing the detailed degradation of the output power at increasing temperature.

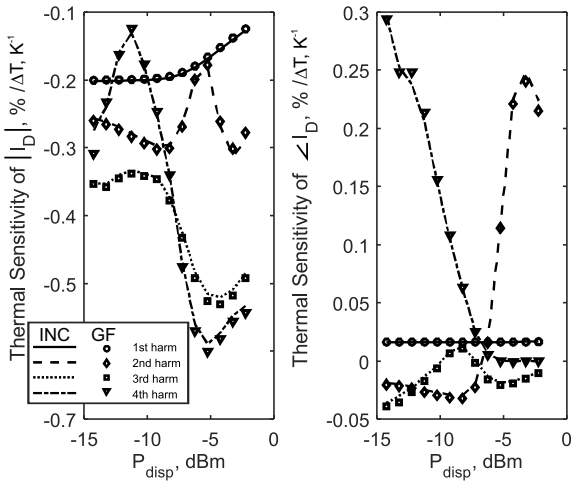

Fig. 15. Percentage variation per unit temperature of the magnitude and phase of the class A PA harmonics vs. input drive. Lines: incremental simulations. Symbols: GF approach.

References

[1] C. Prasad, "A review of self-heating effects in advanced CMOS technologies," *IEEE Transactions on Electron Devices*, vol. 66, no. 11, pp. 4546–4555, nov 2019.

[2] F. Bonani, V. Camarchia, F. Cappelluti, S. D. Guerrieri, G. Ghione, and M. Pirola, "When self-consistency makes a difference," *IEEE Microwave Magazine*, vol. 9, no. 5, pp. 81–89, oct 2008.

[3] A. Benvenuti, W. M. Coughran, and M. R. Pinto, "A thermal-fully hydrodynamic model for semiconductor devices and applications to III-v HBT simulation," *IEEE Transactions on Electron Devices*, vol. 44, no. 9, pp. 1349–1359, 1997.

[4] N. C. Miller, "Large-signal RF simulation and characterization of electronicdevices using fermi kinetics transport," Ph.D. dissertation, Michigan State University, 2017.

[5] S. Donati Guerrieri, F. Bonani, and G. Ghione, "Linking X parameters to physical simulations for design-oriented large-signal device variability modeling," in *2019 IEEE MTT-S International Microwave Symposium (IMS)*. IEEE, jun 2019.

[6] [Online]. Available: https://www.synopsys.com/silicon/tcad/device-simulation/sentaurus-device.html

[7] F. Bonani, S. Donati Guerrieri, G. Ghione, and M. Pirola, "A TCAD approach to the physics-based modeling of frequency conversion and noise in semiconductor devices under large-signal forced operation," *IEEE Transactions on Electron Devices*, vol. 48, no. 5, pp. 966–977, may 2001.

[8] W. Shockley, J. A. Copeland, and R. P. James, "The impedance field method of noise calculation in active semiconductor devices," in *Quantum Theory of Atoms, Molecules, and the Solid-State*, P.-O. Lowdin, Ed. Academic Press, 1966, pp. 537–563.

[9] F. Bonani, G. Ghione, M. Pinto, and R. Smith, "An efficient approach to noise analysis through multidimensional physics-based models," *IEEE Transactions on Electron Devices*, vol. 45, no. 1, pp. 261–269, 1998.

[10] F. Bonani, S. Donati Guerrieri, and G. Ghione, "Physics-based simulation techniques for small- and large-signal device noise analysis in RF applications," *IEEE Transactions on Electron Devices*, vol. 50, no. 3, pp. 633–644, mar 2003.

[11] 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 Transactions on Electron Devices*, vol. 63, no. 3, pp. 1195–1201, mar 2016.

[12] ——, "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 Transactions on Electron Devices*, vol. 63, no. 3, pp. 1202–1208, mar 2016.
[13] S. Selberherr, *Analysis and Simulation of Semiconductor Devices*. Springer Vienna, 2011.
[14] K. S. Kundert, J. K. White, and A. Sangiovanni-Vincentelli, *Steady-state methods for simulating analog and microwave circuits*. Boston: Kluwer Academic Publishers, 1990.
[15] S. Maas, *Nonlinear microwave and RF circuits*. Boston, MA: Artech House, 2003.
[16] A. M. Bughio, S. Donati Guerrieri, F. Bonani, and G. Ghione, "Physics-based modeling of FinFET RF variability," in *2016 11th European Microwave Integrated Circuits Conference (EuMIC)*. IEEE, oct 2016.
[17] J.-P. Raskin, "FinFET versus UTBB SOI — a RF perspective," in *2015 45th European Solid State Device Research Conference (ESSDERC)*. IEEE, sep 2015.
[18] A. M. Bughio, S. Donati Guerrieri, F. Bonani, and G. Ghione, "Multi-gate FinFET mixer variability assessment through physics-based simulation," *IEEE Electron Device Letters*, vol. 38, no. 8, pp. 1004–1007, aug 2017.