When Self-Consistency Makes a Difference

Fabrizio Bonani, Vittorio Camarchia, Federica Cappelluti, Simona Donati Guerrieri, Giovanni Ghione, and Marco Pirola

Thermal management of semiconductor devices and integrated circuits [from digital to analog and power radio frequency (RF) and microwave circuits] is a well-known critical issue in modern electronic design. Technology advances, such as device downscaling to increase the maximum operating frequency and the use of wide-bandgap semiconductors [such as silicon carbide and (SiC) and gallium nitride (GaN)] with breakdown voltages one order of magnitude larger than in conventional III-V compounds, have significantly increased power densities in compound semiconductor microwave and mm-wave transistors, thus making effective thermal design a key point for successful technology development. Heating is an issue also in Si-based RF technologies, as in lateral double diffused metal oxide semiconductor (LDMOS) devices; however, despite the impressive total RF power such devices exhibit, the power density is in fact much lower, of the order of 1 W/mm against 10 W/mm or more in GaN high electron mobility transistor HEMTs; gallium arsenide GaAs FETs have similar or somewhat larger power densities than LDMOS, but with poorer substrate thermal conductivity. Self-heating not only is a major limitation to the device reliability (through thermal instabilities, hot spot formation, and thermal runaway), but also affects [sometimes in a subtle way, e.g., when long-term thermal memory affects the device linearity (see the "Dynamic Thermal Modeling" section)] the device performance. In many applications, a critical role is also played by the transient thermal response. Examples are pulsed-mode high-power amplifiers (HPAs) (RF and microwave). RF power circuit design

Fabrizio Bonani, Vittorio Camarchia, Federica Cappelluti, Simona Donati Guerrieri, Giovanni Ghione, and Marco Pirola are with the Electronic Department, Politecnico di Torino, Corso Duca degli Abruzzi, 24, 10129 Torino, Italy.

Digital Object Identifier 10.1109/MMM.2008.927638

therefore increasingly needs to be supported by reliable electrothermal semiconductor device models, whose development requires accurate nonlinear dynamic thermal models, which may ultimately be derived from complex three-dimensional (3-D) multiphysics analysis techniques, to be coupled with temperature-dependent electrical compact models, suitable for the integration in circuit computer aided design (CAD) environments. In fact, even though fully-coupled 3-D or even quasi-two-dimensional (2-D) physics-based electrothermal models would accurately reproduce the dynamic electrothermal interaction, the related computational burden most often prevents their use in simulating the device performance under practical operating conditions, also accounting for the interaction with other active or passive devices.

The Thermal Model

Semiconductor device thermal analysis can be finalized to different goals, ranging from the technological level (e.g., optimization of the device mounting and layout) to the electrical level (e.g., investigation on how temperature affects the device electrical performance). A preliminary condition is, in all cases, the availability of reliable data for the thermal conductivity and the specific heat of the relevant materials, from semiconductors (sometimes exotic and recently developed like GaN or with a large variety of crystal polytypes with different thermal properties, like SiC) to metals and dielectrics [1]. All material parameters should, in principle, include nonlinearities, i.e., the temperature dependency. A crucial role in assessing the validity of model parameters and of simulation techniques is of course played by the availability of device-level thermal measurements, including both averaged parameters (such as the thermal resistance) and detailed surface temperature maps. Under this framework, the proposed measurement techniques vary from standard pulse characterization [2] to optical photoluminescence mapping [3], [4], Raman spectroscopy [5], and photocurrent measurements [6].

The purpose of thermal modeling (and therefore the simulation approaches and tools of choice), also depends on the kind of analysis required. At the technological level, thermal modeling focuses on evaluating the device temperature distribution (see the "Heat Equation and Temperature Distribution" section), while at the electrical level (where such detailed information typically is redundant and scarcely manageable), an averaged thermal resistance, to be coupled to an electrical (temperature-dependent) device model in order to yield a self-consistent electrothermal model, is a more convenient choice (see the "Thermal Resistance Models" section). However, in modern RF and microwave communication systems exploiting broadband modulation schemes, the slow dynamic device response, also related to thermal effects, plays a significant role, making the issue of accurate dynamic thermal modeling (yielding the time-dependent temperature distribution, and the device thermal impedance) very hot today, as discussed in the "Dynamic Thermal Modeling" section.

Heat Equation and Temperature Distribution

Ideally, the detailed temperature distribution inside the device volume yields a complete picture of the thermal behavior. Under this respect, numerical simulation is able to resolve temperature variations at a submicron scale, and can therefore perform better than most experimental techniques, which are typically limited in spatial resolution and, often, to the surface of the active device [7], [8]. However, an exact device-level thermal model calls for the solution of the heat equation over the device volume in the presence of a self-consistent dissipated power distribution, which in turn has to be identified from the solution of a physics-based transport model, like for instance the drift-diffusion or hydrodynamic (see, e.g., [9]).

The resulting self-consistent physics-based coupled electrothermal model is, unfortunately, extremely demanding from a computational standpoint, particularly when the device structure under consideration is described through a realistic 3-D approach, including large-scale features such as the thermal mounting. In fact, this turns out to be a typical multiphysics problem, where the spatial scale of the relevant physical effects can vary over several orders of magnitude, from the submicron scale typical of the electric operation (electric field and carrier density distributions), to the millimeter or even centimeter scale typical of heat flow.

Furthermore, while electrical simulations are often restricted to the 2-D case, because of the nature of electron transport inside the device, accurate heat flow simulations do require a fully 3-D approach. As a result, the discretization mesh should be very dense over the active device area and coarser over all the other parts of the device that play little or no role in electrical simulations but need to be included for the thermal analysis (e.g., substrate, metal layers, heat sinks, etc.). Numerical multigrid approaches (in which the thermal and electrical models are discretized on different meshes, individually optimized for the respective spatial scale), or a simplified treatment of the heat boundary conditions [10] have been proposed in the past to make the approach viable; however, although most of the available physics-based simulation tools today allow for the self-consistent coupling of the heat equation to the transport equation, the practical use is confined to idealized 2-D cases, and the accurate modeling of realistic, 3-D large-scale devices still is beyond reach.

Even neglecting consistency in a physics-based, microscopic sense, the 3-D solution of the heat equation with an assigned dissipated power density is the kind

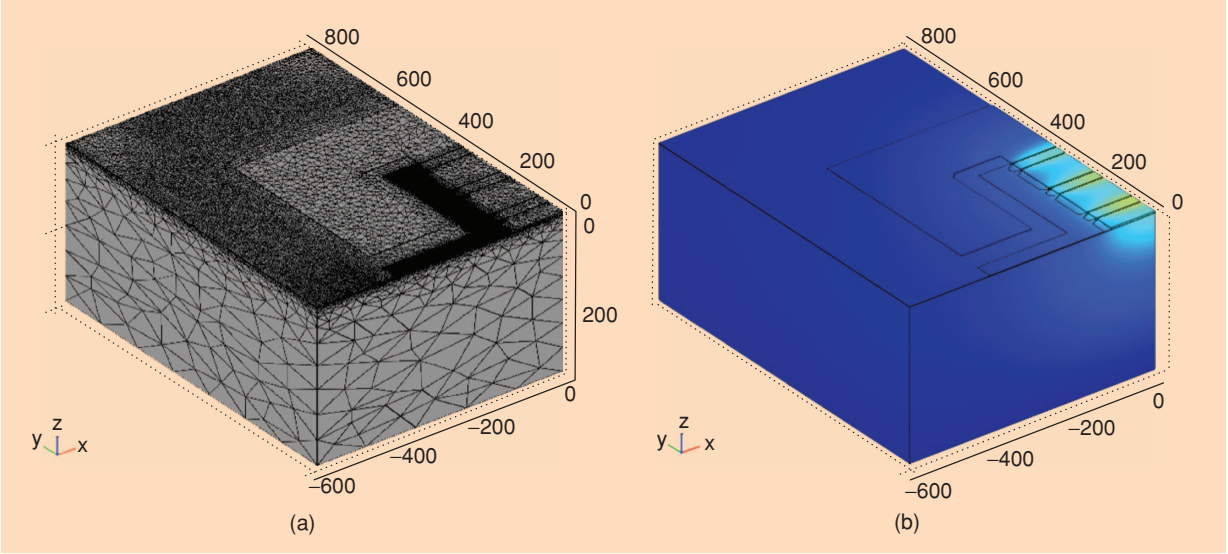

Figure 1. *Example of layout exploited in FEM thermal simulations. One quarter of a 12 × 200 μm AlGaN/GaN multifinger HEMT. (a) Nonuniform mesh discretizazion and (b) temperature profile. The dissipated power is 4 W, and the maximum temperature in the gates is, in this case, roughly 80 °C.*

of analysis typically required, e.g., for the device technological evaluation. Well known heuristic rules, derived from physics-based simulations or measurements, enable to approximate the dissipated power distribution with a number of constant or Gaussian sources (whose exact position and shape is however in some cases somewhat controversial, see [11]). The knowledge of the detailed temperature distribution on the device surface allows optimizing the layout (e.g., by decreasing the thermal coupling between parallel gates or emitters), estimating the device reliability (from the maximum channel or junction temperature), and optimizing the substrate thickness or the mounting.

Nonetheless, the very solution of the heat equation (also in the simplest case of stationary, dc conditions) poses various problems. The 3-D meshing problem is alleviated but not completely solved, since the volume where the dissipated power is significant is very small (micron scale) compared to the device dimensions (mm scale at least); this requires again the cumbersome management of extremely nonuniform grids, with possible inaccuracies in the problem solution and, in general, a worse numerical conditioning, which often becomes critical when the material parameters are temperature-dependent, and therefore the heat equation nonlinear. Nonlinearity is a difficult problem in itself, in most cases alleviated by the application of the Kirchhoff transformation [12], which allows to linearize the heat equation. Unfortunately, its exact application is limited to the case of homogenous materials. This is rarely the case in practical devices, which typically exploit inhomogeneous multilayered or fully 3-D geometries. For the general case of T-dependent material parameters, an approximate application is only possible for piecewise nonhomogeneous structures [13], while for time-varying (dynamic) analysis approximations are mandatory [14], [15]. Furthermore, fast, quasi-3-D (or 2.5-dimensional) planar solution methods, based on Fourier series or Green's function approaches, fail to capture fully 3-D geometries (although in some cases they can be modified to include, e.g., the effects of the metallizations [16]), and are not generally applicable to multilayered media with temperature-dependent parameters. In conclusion, the only safe approach to the large-scale temperature simulation of realistic device structures appears to be today the direct, three-dimensional solution of the heat equation via finite element (FEM) discretization; to this aim, several commercial software tools are available on the market, e.g., [17]–[19]. It should be stressed that, although such tools are today quite efficient and flexible from a computational standpoint, most of the problem solving time is spent in training the program user (sometimes a not overly enthusiastic Ph.D. student), and in setting up a working 3-D mesh.

As an example of a finite element heat equation solution for a multifinger HEMT (12 × 200 μm) on AlGaN/GaN, Figure 1 shows the discretization of one quarter of the device with the resulting grid, and the temperature profile for a dissipated power of 4 W.

Thermal Resistance Models

FEM-based 3-D models are too computationally intensive to be exploited within a circuit design environment, e.g., in connection with a self-consistent electrothermal circuit-level compact model; moreover, the information they provide is redundant. Electrical-level thermal modeling should yield an average or maximum value of the device active region temperature as a function of the total dissipated power, possibly derived

in turn as a time-varying unknown that depends on the device instantaneous working point. To this purpose, the so called thermal resistance approach is widely exploited, in which the temperature increase in the device active volume is related to the dissipated electrical power by a lumped model (often associated to an equivalent thermal network, as shown in Figure 2). This allows evaluating the temperature of a limited number of thermal nodes, corresponding to physical hot spots on the device layout.

In the simplest case, the active area of an electrical device reduces to a single thermal node and the thermal network to a simple resistance. However, this simplification may be inaccurate in multifinger (multigate or multiemitter) devices where the active region temperature undergoes significant variations along (and across) the device fingers; in this case, each finger should be divided into subsections, each one having its own temperature and its equivalent thermal network also accounting for thermal coupling between subsections. In the static (dc) case, and neglecting the material nonlinearity, the thermal network can be derived by repeatedly solving the heat equation through FEM 3-D approaches and by surface temperature averaging and/or heat source lumping; in this case, the output nodal temperature of the thermal network can be identified as a linear combination, through self and mutual thermal resistances, of the powers dissipated in each output node. More advanced methods rely on system-oriented approaches like model reduction identification tools [20], [21]. In many cases, the Green's function approach seems to be especially useful for the extraction of the thermal network, as it aims at the evaluation of the temperature distribution only for selected observation or injection points in the device [16], [21], but it is inherently restricted to the linear case, and therefore can be exploited only when the Kirchhoff transformation is applicable. The identification of a nonlinear equivalent thermal network still is a problem, although simplified system-level approaches can be applied to this purpose.

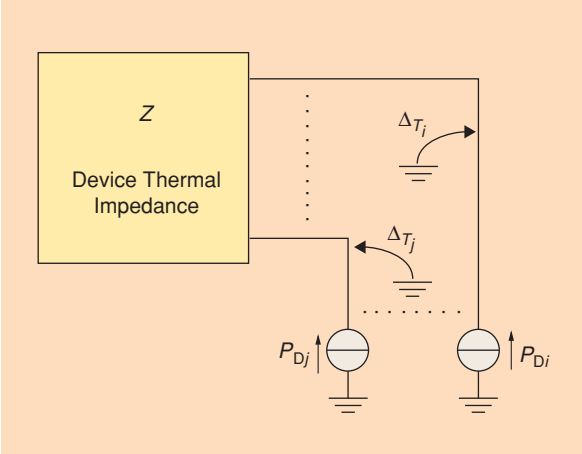

Figure 2. *Electron device equivalent thermal network [26]. Indices i and j refer to different thermal nodes (i.e., either different devices or fingers of the same device); ΔT is the corresponding temperature raise with respect to reference (ambient) temperature.*

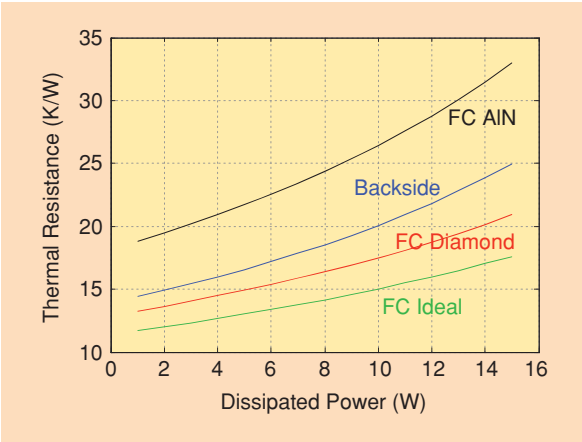

Figure 3. *Thermal resistance of an AlGaN/GaN device as a function of the dissipated power and for different mountings setups [25].*

Thermal resistance models can be effectively exploited as technology level models since they yield a concise picture of the device thermal behavior, which can help in assessing the available technology choices. As an example, Figure 3 shows the impact of substrate and mounting choices [backside on SiC or flip-chip (FC) on AlN, diamond or ideal heat sink] and nonlinearity on the thermal resistance of a 1 mm (10×100 μm) AlGaN/GaN HEMT on SiC from SELEX SI.

Dynamic Thermal Modeling

Although temperature does not respond as fast as the electrical variables to an external excitation, because of the slower dynamics involved in the very nature of heat flow, exact dynamic thermal models for electron devices are still difficult to be identified or extracted. Very often, simplified models are exploited, based on the bare idea that a cut-off frequency must exist, above which temperature is in some sense frozen and cannot vary following the fast variation of the instantaneous dissipated electrical power. According to this approach, thermal models are often limited to the identification of a reasonable time constant to set the limit of the thermal dynamics. The device is then modeled by a first-order dynamic system, i.e., a simple resistance-capacitance (*RC*) thermal circuit, where *R* is given by the thermal resistance, while the capacitance is set in order to yield the cut-off frequency corresponding to the chosen time constant.

Unfortunately, as will be shown further on, results based on more accurate thermal models point out that the thermal impedance of an electron device (i.e., the temperature response to a small-signal time-varying sinusoidal dissipated power) has a frequency behavior often far from being comparable to a simple, single-pole *RC* circuit low-pass response,

showing on the contrary a much slower decay at low frequency. This is not unexpected, since the dynamic heat equation corresponds to a distributed system, with nonrational frequency-domain transfer functions, whose detailed behavior is directly related to the large-scale multilayered inhomogeneous structure (semiconductor substrate, composite mounting, heat sink); higher-order *RC* circuits are therefore required to approximate, on a given frequency bandwidth, the complex frequency behavior of the thermal impedance.

Dynamic thermal models can be divided into two main categories: numerical models based on the space (FEM) and time discretization of the heat equation, including material nonlinearities in the thermal conductivity and specific heat [12]; semianalytical models, where the dynamic heat equation is given a closed-form solution in simple geometries and for ideal point, or line, heat sources. More complex geometries and/or heat sources can be recovered using source superposition (see, e.g., [22], where the thermal impedance is evaluated by approximating the heat source through a superposition of spherical sources) and enforcing the boundary conditions with the images technique, or by partitioning the device volume into subdomains coupled through proper interface conditions, in order to build an equivalent thermal network of the complex device. Examples may be found in [23] and [24]; in particular the latter presents such a methodology applied to the analysis of IGBTs. These approaches exploit the Kirchhoff transformation, therefore the remarks given above apply.

The use of simple time-domain excitations (pulses or steps) enables to reproduce the dynamic response over all relevant time scales, although they require accurate, adaptive time discretization to manage widely different time scales, and proper techniques to transform the response into frequency domain (more suited to CAD simulators). As an example of the kind of results that can be achieved with the FEM heat equation solution, Figure 4 reports the normalized frequency dependency of the device thermal impedance of the AlGaN/GaN HEMT device with different mounting setups, whose static (thermal resistance) behavior was shown in Figure 3 [25]. Figure 5 shows instead the frequency-domain elements of the (2×2) thermal impedance matrix for a two-finger heterojunction bipolar transistor (HBT) device from SELEX SI, directly extracted through the spherical sources approach [22], complemented by the application of the images method to impose adiabatic boundary conditions at the substrate periphery, see [26], [27].

Let us finally address the material nonlinearity issue in the dynamic case. Due to the *T*-dependency of the material parameters and to the dispersive nature of the thermal problem, the thermal impedance will change as a function of the dissipated power level. In such case a

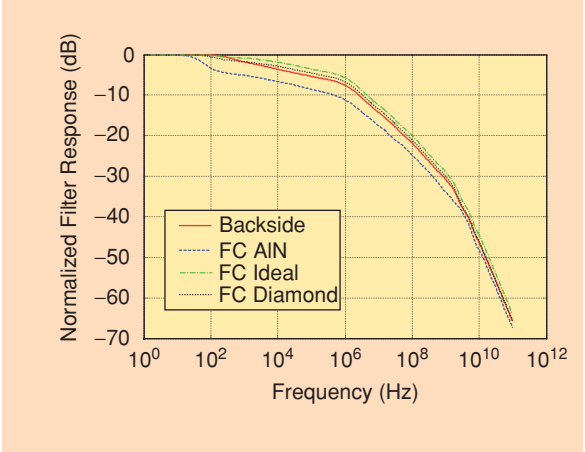

Figure 4. *Normalized frequency response of the thermal impedance of the 1 mm AlGaN/GaN HEMT device from SELEX SI.*

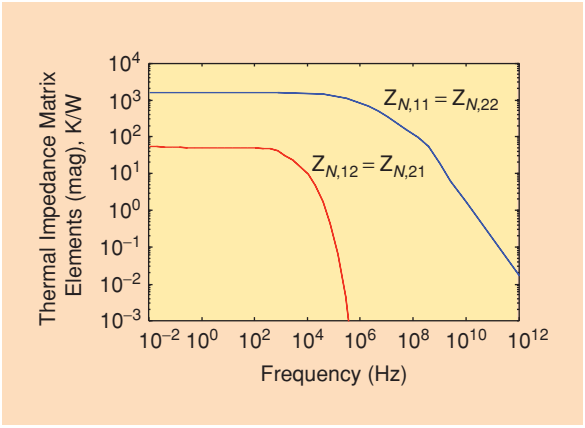

Figure 5. *Magnitude of the thermal impedance matrix elements of a two-finger AlGaAs/GaAs HBT device from SELEX SI [25].*

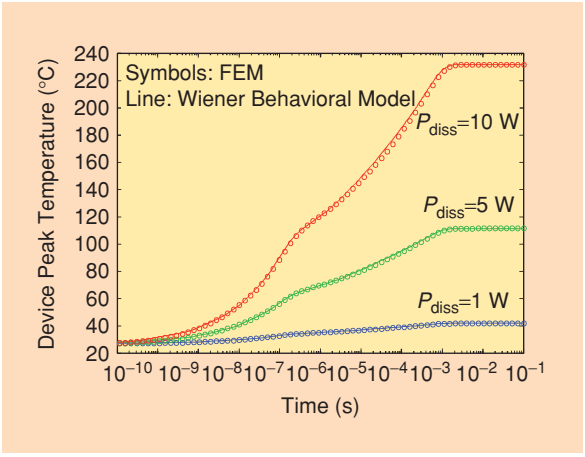

Figure 6. *Comparison between FEM reference solution and Wiener behavioral model. Peak temperature in the 1 mm AlGaN/GaN HEMT device from SELEX SI versus power dissipation level.*

system-level description, valid for any input dissipated power excitation waveform, can be derived by applying well-established results from the modeling of nonlinear dynamic systems. Approximate device thermal models can be based on Wiener, Hammerstein or Wiener-Hammerstein approaches [28]; as an example, Figure 6 shows that a Wiener-like model, based on the cascade of a memory-less nonlinear block followed by a linear filter, is able to accurately reproduce the transient thermal response of the AlGaN/GaN HEMT of Figure 4 at different power levels. Besides being suited to be implemented into CAD tools, the Wiener approach is readily identified with a two step procedure: first the static simulation, at various power dissipation levels, is performed in order to identify the nonlinear block; then only one time-domain simulation at a low level of dissipation (linear case) suffices to identify the frequency dependent block.

Electrothermal Simulation and Results

Coupling a suitable thermal model to an electrical model allows for the investigation of some critical effects occurring in power RF and microwave devices. The examples discussed here concern the stability of III-V based HBTs and the dispersion in the intermodulation behavior of power GaN HEMTs due to the thermal feedback. All these phenomena cannot be addressed unless a fully coupled electrothermal model is adopted. Similar issues and modeling solution techniques are also present in Si RF LDMOS devices, see, e.g., [29]–[31].

A preliminary remark refers to the fact that power microwave devices such as HBTs, heterostructure field effect transistors (HFETs), HEMTs (but also Si LDMOS) most often are characterized by a multifinger layout to decrease parasitic resistances and to allow for scalability. In this case, it is sometimes necessary to extract an electrical model for each device finger in order to couple it to a multiport thermal model allowing for the description of the thermal cross-coupling among the various device parts.

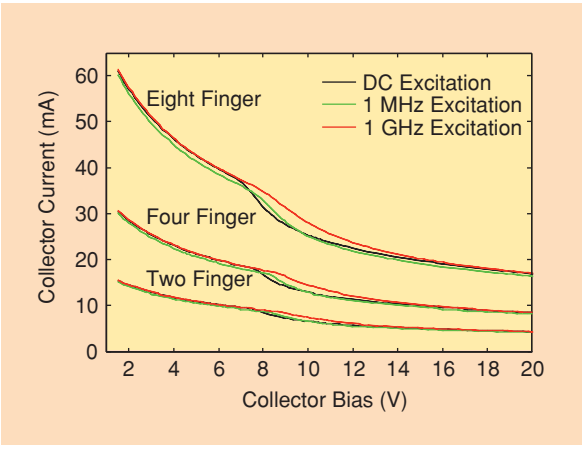

Figure 7. *Current collapse of three multifinger AlGaAs/GaAs HBT devices in static and dynamic conditions.* $I_{B,DC} = 0.4; 0.8; 1.6\ \mu A$. $I_{B,AC} = 0.3; 0.6; 1.2\ \mu A$.

This is particularly relevant for multifinger HBTs, where current collapse instabilities are due to thermal effects combined with device asymmetries. Current collapse is a sudden decrease in the output collector current as a function of the collector-emitter voltage. Collapse causes degradation in the HBT power capabilities, and possibly device failure due to thermal runaway. Self-consistent electrothermal HBT simulations have been widely used to assess the current collapse onset mechanism; this is due to current hogging induced by small layout or finger-to-finger electrical asymmetries whereby, with increasing temperature, a single device finger drains all the device current while the other fingers are ultimately turned off. HBT thermal simulations allow for the identification of the thermal coupling, and allow for layout optimization or the design of suitable emitter or base resistive ballasting.

In the vast majority of cases, HBT stability analysis through electrothermal simulations is based on the thermal resistance approach [32], where all thermal dynamic effects are neglected. However, experimental evidence shows that the onset of the instability occurs at different collector voltages in static or pulsed dc measurements, thus suggesting that the phenomenon is affected by significant dispersion, to be also attributed to thermal dynamical effects.

A full investigation of a SELEX SI HBT technology on GaAs has been carried out, including the thermal analysis of various devices (two, four, and eight fingers; some results have been already presented in the previous sections) and the estimation of dc and small-signal electrical performances [26], [27]. Each HBT finger has been modeled in dc through a modified Gummel-Poon model, where temperature impacts the current gain β and the reverse collector saturation current through the T-dependency of the material energy gap [27]. The dynamic electrical part is added by means of a standard set of capacitances (one between the intrinsic collector and base, the second between the intrinsic and the extrinsic emitter contacts), that in some modeling approaches are also made temperature dependent [33]. A time delay between emitter and collector currents was also added. The electrical model has been coupled with the thermal impedance evaluated as explained in the previous section (see Figure 5 for the two-finger HBT). The self-consistent electrothermal simulations have been carried out at the circuit level by an in-house simulator implementing the Harmonic Balance, frequency domain approach. Figure 7 shows the dc output characteristics, including the current collapse. Moving to dynamic conditions, the dc component of the collector current is shown as an answer to a single tone base current excitation superimposed to the dc value at two different frequencies. Significant dispersion is observed on the value of collector voltage where current collapse takes place. Notice that in this particular simulation no trap effects are present, so that all memory (dispersive)

phenomena are to be ascribed entirely to thermal dynamics. It is important to remark that thermal effects play a significant role also at very high frequencies (as can be seen by the marked difference between the 1 MHz and the 1 GHz excitation), showing that the frequency decrease of the thermal impedance is extremely slow compared to a single pole *RC* network.

Instability problems are peculiar of bipolar devices: in fact, the active device self-heating may act as an effective positive feedback, basically due to the current increase with temperature. FET based devices, instead, are intrinsically immune to this kind of problems, since in this case device self-heating mainly degrades carrier mobility in the semiconductor, thus decreasing the drain current. The net effect is equivalent to a negative feedback that, equalizing the several current contributions, makes multifinger devices inherently stable.

In the field of power amplifier design, great interest is given to the impact of thermal effects on device nonlinearities. In fact, accurate modeling of intermodulation products (IMPs) allows for the design of high linearity modules, either by exploiting the device sweet spots or by using proper predistorting stages. This issue is particularly important for advanced wireless communication systems, characterized by nonconstant envelope broadband modulation schemes: in this case the design of power modules, especially for base stations, is extremely demanding in term of linearity and efficiency. Broadband signals are influenced not only by the instantaneous nonlinearities typical of the electrical device operation at high frequencies, but also by long-term memory effects such as the ones related to thermal dynamics. Under this respect, the design of advanced power amplifiers requires the device to be described by a full electrothermal model, including nonstatic thermal effects. The latter, as explained in the previous section, can be described by a thermal impedance network in the linear case, or, more generally, by a system-oriented nonlinear dynamic model.

Among high-power compound semiconductor microwave and RF technologies, AlGaN/GaN HEMT have shown appealing performances in terms of power density. A fully self consistent electrothermal model extracted and tailored on the coplanar power GaN HEMTs on SiC already seen in the previous sections will be used to gain insight into the impact of thermal effects on amplifier linearity. The FET model is a cubic Curtice [34], with temperature dependent parameters (see [25] for additional details). The thermal behavior is modeled by the Wiener approach (see Figure 6), with the dynamic part approximated through a multipole rational transfer function, cascaded with a nonlinear memoryless block. The electrothermal model has been implemented within the Agilent advanced design system (ADS) environment exploiting a symbolic defined device (SDD) block. The model has been fitted on dc, pulsed dc, multibias scattering parameters, time-domain RF waveforms and the P_{in}-P_{out} output characteristics [25].

Interesting results on nonlinear performance is observed in the simulated P_{in}-P_{out} curves on optimum loads, shown in Figure 8. To highlight the effect of different thermal dynamic modeling choices, the full-dynamic model is compared to an isothermal one, in which the temperature is kept constant at the value assumed at quiescent bias point, i.e., the two models are equivalent for zero output power. The results relative to single tone excitation, at class A, AB, and B bias, are reported in Figure 8. For class A operation, the simplified model does not take into account the cooling effect on the device of the RF for large signal input power, thus underestimating the actual output power (more than 1 dBm in the present case) in a sort of worst case approach. On the contrary, the same strategy applied in class B (at dc the device is in this case cold and becomes hotter and hotter for increasing RF power), leads to overestimating the output power. In the present case, class AB operation is a sort of average condition in which the results from both models are comparable.

To highlight the role of thermal feedback on device nonlinearity, let us consider the third order IMPs (IMP3s), a typical figure of merit (FOM) to assess the amplifier linearity in case of a two-tone excitation. Thermal memory effects lead to IMP3 asymmetries (in the upper and lower products) and dependence on tone spacing [35], in a way similar to the one introduced by bias networks used for dc feed and RF blocks. IMP3 asymmetry and dispersion make the design of predistortion circuits extremely critical, and should be possibly reduced, or at least accurately modeled.

Figure 9 compares the IMP3 for the same GaN power amplifier of Figure 8, evaluated through different electrothermal models, as a function of the tone spacing. In addition to results from the full dynamic, and to the already mentioned isothermal model, Figure 9 also shows, for the sake of comparison, the IMP3s from two additional models a single-pole low-pass

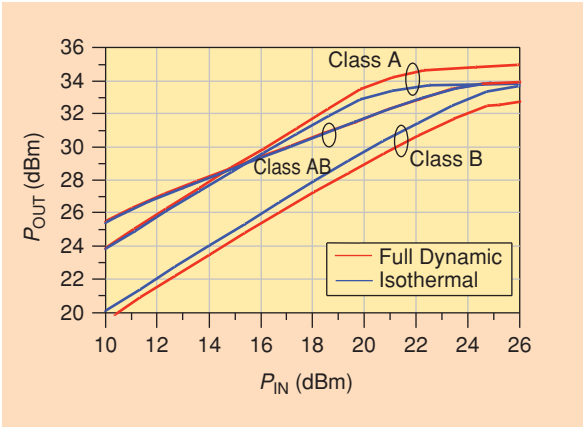

Figure 8. *Single tone simulation of the 1 mm GaN based HEMT device from SELEX SI. All simulated results are at 4 GHz on optimum load for power.*

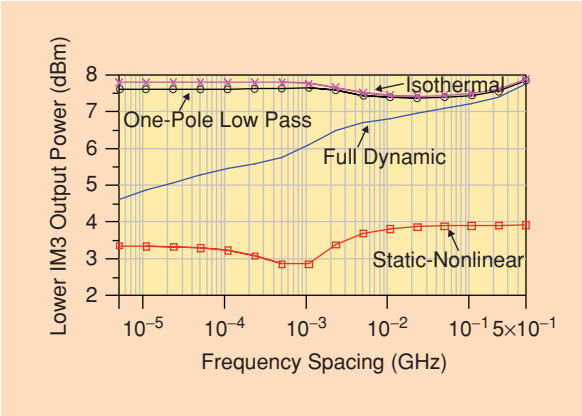

Figure 9. *Two-tone simulation of the 1 mm GaN based HEMT device from SELEX SI as the function of the tone spacing. The center frequency is 4 GHz and the power level is 25 dBm total input power, corresponding to 36 dBm output power.*

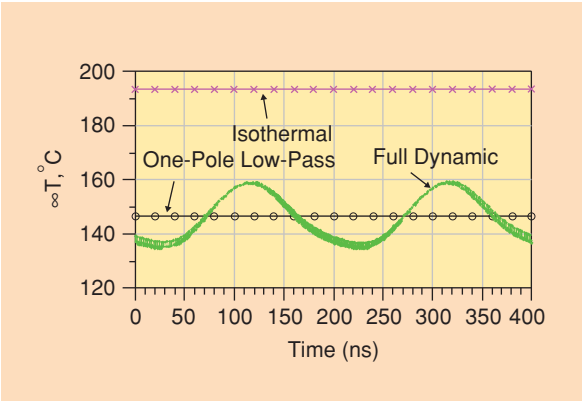

Figure 10. *Temperature variation as a function of time in a class A two tone simulation with 4 GHz center frequency, 5 MHz tone spacing and 24 dBm input power.*

model (cut-off frequency of 10 Hz) and a nonlinear memoryless model (implying that the temperature variation instantaneously follows the RF modulation). The different behavior of the predicted IMP3s is evident, and can be related to the amount of electrothermal feedback, that is null for the isothermal model, and increasingly significant for the single pole low-pass, the full-dynamic, and finally the nonlinear memoryless model. In fact, IMP3s can be approximately interpreted as originating from two contributions: the device electrical nonlinearity (electrical IMP3s), and the mixing between second-order IMPs in the device temperature (caused by second-order IMPs in the instantaneous dissipated power) and the signal fundamental frequency (thermal IMP3s). Such a mixing occurs because of the temperature sensitivity of the device current. In the case considered, the two contributions cancel each other; moreover, the magnitude of second-order IMPs in the instantaneous temperature greatly depends on the dynamic low-frequency response of the thermal model. For the static nonlinear model, the magnitude of thermal IMP3s is not influenced by tone spacing, and cancellation is most effective. In the low-pass or isothermal model, thermal IMP3s are virtually suppressed, and the total IMP3s are maximized. In the full-dynamic, more accurate model, the amount of cancellation depends on the tone spacing, since the IMP2s in the instantaneous temperature are stronger for small tone spacing, weaker for large tone spacing, leading to the result in Figure 9. Cleary, the detailed IMP3 behavior critically depends on the thermal impedance frequency response at low frequency. For a more detailed discussion, see [25] and [35].

When the dynamic part of the thermal impedance is a very slow decreasing function of frequency, electrothermal simulations may also be critical from the standpoint of CAD implementation and numerical solution. The two-tone simulation can serve as an example to highlight this point. For such a simulation, and more in general for modulated narrowband signals, the harmonic balance solution technique is often replaced by the widely known envelope strategy, presenting relevant advantages on the computational standpoint. However, the envelope algorithm relies upon the assumption of negligible overlapping between the envelopes spectra centered on the carrier harmonics. With slow decaying thermal impedances, this assumption may become inaccurate because of the envelope broadening emphasized by the presence of the thermal nonlinearities. To point out the insufficient accuracy of the envelope approach for the present case, the harmonic balance have been used as a reference solution for two-tone analysis carried out on the same GaN HEMT previously described. The results are shown in Figure 10: the temperature waveform from the self-consistent electrothermal model clearly shows the presence of residual high frequency components, overlapped to the low frequency envelope of the two tone modulation. Such a ripple is not compatible with the assumption underlying the envelope simulation, which, in this case, would have led to inaccurate results.

Conclusions

Compound semiconductor power RF and microwave device modeling requires, in many cases, the use of self-consistent electrothermal equivalent circuits. The slow thermal dynamics and the thermal nonlinearity should be accurately included in the model; otherwise, some response features subtly related to the detailed frequency behavior of the slow thermal dynamics would be inaccurately reproduced or completely distorted. Two examples have been shown, concerning current collapse in HBTs and modeling of IMPs in GaN HEMTs. Accurate thermal modeling can be made compatible with circuit-oriented CAD tools through a proper choice of system-level approximation; in the discussion presented we have exploited a

Wiener approach, but of course the strategy should be tailored to the specific problem under consideration.

Acknowledgments

The work has been supported by the Italian PRIN GANFET and FPFET projects and by the EUROFINDER KORRIGAN project.

References

[1] F. Cappelluti, M. Furno, A. Angelini, F. Bonani, M. Pirola, and G. Ghione, "On the substrate thermal optimization in SiC-based backside-mounted high-power GaN FETs," *IEEE Trans. Electron. Devices*, vol. 54, no. 7, pp. 1744–1752, July 2007.

[2] S. Augaudy, R. Quere, J.P. Teyssier, M.A. Di Forte-Poisson, S. Cassette, B. Dessertenne, and S.L. Delage, "Pulse characterization of trapping and thermal effects of microwave GaN power FETs," 2001, in *IEEE MTT-S Int. Microwave Symp. Dig.*, May 2001, vol. 1, pp. 427–430.

[3] P. Landesman, E. Martin, and P. Braun, "Temperature distribution in power GaAs field effect transistors using spatially resolved photoluminescence mapping," in *Proc. 7th Int. Physical and Failure Analysis of Integrated Circuits*, 1999, pp. 185–190.

[4] N. Shigekawa, K. Shiojima, and T. Suemitsu, "Optical study of high biased AlGaN/GaN high-electron mobility transistors," *J. Appl. Phys.*, vol. 92, pp. 531–535, 2002.

[5] M. Kuball, J.M. Hayes, M.J. Uren, T. Martin, J.C. Birbeck, R.S. Balmer, and B.T. Hughes, "Measurement of temperature in active high power AlGaN/GaN HFETs using Raman spectroscopy," *IEEE Electron Device Lett.*, vol. 23, no. 1, pp. 7–9, Jan. 2002.

[6] P. Regoliosi, A. Reale, A. Di Carlo, P. Romanini, M. Peroni, C. Lanzieri, A. Angelini, M. Pirola, and G. Ghione, "Experimental validation of GaN HEMTs thermal management by using photocurrent measurements," *IEEE Trans. Electron Devices*, vol. 53, no. 2, pp. 182–188, Feb. 2006.

[7] R. Aubry, J.-C. Jacquet, J. Weaver, O. Durand, P. Dobson, G. Mills, M.-A. di Forte-Poisson, S. Cassette, and S.-L. Delage, "SThM temperature mapping and nonlinear thermal resistance evolution with bias on AlGaN/GaN HEMT devices," *IEEE Trans. Electron Devices*, vol. ED-54, no. 3, pp. 385–390, Mar. 2007.

[8] A. Reale, A. Di Carlo, M. Peroni, C. Lanzieri, and S. Lavagna. "Thermal maps of GaAs P-HEMT: A novel system based on the photocurrent spectral analysis," *IEEE Trans. Electron Devices*, vol. 54, no. 4, pp. 879–882, Apr. 2007.

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

[10] G. Ghione and C.U. Naldi, "High-resolution self-consistent thermal modeling of multi-gate power GaAs MESFETs," in *Proc. IEDM*, Dec. 1989, pp. 147–150.

[11] A. Reale, A. Di Carlo, A. Angelini, F. Cappelluti, F. Bonani, and G. Ghione "Channel T-distribution of high-power GaN HEMTs from spatially resolved photoconductance measurements and physical interpretation from thermal simulations," in *Proc. 11th Int. Symp. Microwave and Optical Technology (ISMOT-2007)*, Monte Porzio Catone, Italy, 17–21 Dec. 2007, pp 173–179.

[12] H.S. Carlslaw and J.C. Jaeger, *Conduction of Heat in Solids*, 2nd ed. London, U.K.: Oxford Univ. Press, 1959.

[13] F. Bonani and G. Ghione, "On the application of the Kirchhoff transformation to the steady-state thermal analysis of semiconductor devices with temperature-dependent and piecewise inhomogeneous thermal conductivity," *Solid-State Electron.*, vol. 38, no. 7, pp. 1409–1412, July 1995.

[14] W. Batty and C.M. Snowden, "Electro-thermal device and circuit simulation with thermal nonlinearity due to temperature dependent diffusivity," *Electron. Lett.*, vol. 36, no. 23, pp. 1966–1968, Nov. 2000.

[15] W. Batty and C.M. Snowden, "Reply to comment on "Electro-thermal device and circuit simulation with thermal nonlinearity due to temperature dependent diffusivity," *Electron. Lett.*, vol. 37, no. 24, pp. 1482–1483, Nov. 2001.

[16] F. Bonani, G. Ghione, M. Pirola, and C.U. Naldi, "A new, efficient approach to the large-scale thermal modeling of III–V devices and integrated circuits," in *Proc. IEEE Int. Electron Devices Meeting*, Dec. 1993, pp. 101–104.

[17] ANSYS Thermal Analysis System, ANSYS, Inc. [Online]. Available: http://www.ansys.com

[18] COMSOL Multiphysics Heat Transfer Module, COMSOL AB. [Online]. Available: http://www.comsol.com

[19] MD PATRAN, MSC Software Corp. [Online]. Available: http://www.mscsoftware.com

[20] B. Shapiro, "Creating compact models of complex electronic systems: An overview and suggested use of existing model reduction and experimental system identification tools," *IEEE Trans. Comp. Packag. Technol.*, vol. 26, no. 1, pp. 165–172, Mar. 2003.

[21] J. Mazeau, R. Sommet, D. Caban-Chastas, E. Gatard, R. Quéré, and Y. Mancuso, "Behavioral thermal modeling for microwave power amplifier design," *IEEE Trans. Microwave Theory Tech.*, vol. 55, no. 11, pp. 2290–2297, Nov. 2007.

[22] T. Veijola, "Model for thermal spreading impedance in GaAs MESFETs," in *ICECS Tech. Dig.*, 1996, p. 872.

[23] W. Batty, C.E. Christoffersen, A.J. Panks, S. David, C.M. Snowden, and M.B. Steer, "Electrothermal CAD of power devices and circuits with fully physical time-dependent compact thermal modeling of complex nonlinear 3-D systems," *IEEE Trans. Comp. Packag. Technol.*, vol. 24, no. 4, pp. 566–590, Dec. 2001.

[24] A.R. Hefner, "A dynamic electro-thermal model for the IGBT," *IEEE Trans. Ind. Applicat.*, vol. 30, no. 2, pp. 394–405, Mar.–Apr. 1994.

[25] V. Camarchia, F. Cappelluti, M. Pirola, S. Donati Guerrieri, and G. Ghione, "Self-consistent electrothermal modeling of class A, AB, and B Power GaN HEMTs under modulated RF excitation," *IEEE Trans. Microwave Theory Tech.*, vol. 55, no. 9, pp. 1824–1831, Sept. 2007.

[26] F. Cappelluti, F. Bonani, S. Donati Guerrieri, G. Ghione, M. Peroni, A. Cetronio, and R. Graffitti, "A new dynamic, self-consistent electro-thermal model of power HBTs and a novel interpretation of thermal collapse loci in multi-finger devices," in *Proc. Custom Integrated Circuits Conf.*, San Diego, 6–9 May 2001, pp. 397–400.

[27] F. Cappelluti, F. Bonani, S. Donati Guerrieri, G. Ghione, C.U. Naldi, and M. Pirola, "Dynamic, self consistent electro-thermal simulation of power microwave devices including the effect of surface metallizations," in *Proc. GAAS 2002*, Milan, Italy, 23–24 Sept. 2002, pp. 209–212.

[28] M. Schetzen, "Nonlinear system modeling based on the Wiener theory," *Proc. IEEE*, vol. 69, no. 12, pp. 1557–1773, Dec. 1981.

[29] W.R. Curtice, J.A. Pla, D. Bridges, T. Liang, and E.E. Shumate, "A new dynamic electro-thermal nonlinear model for silicon RF LDMOS FETs," in *1999 IEEE MTT-S Int. Microwave Symp. Dig.*, June 1999, vol. 2, pp. 419–422.

[30] Wenhua Dai, P. Roblin, and M. Frei, "Distributed and multiple time-constant electro-thermal modeling and its impact on ACPR in RF predistortion," in *Proc. ARFTG Microwave Measurements Conf.*, 2003, 4–5 Dec. 2003, pp. 89–98.

[31] M. Guyonnet, R. Sommet, R. Quere, and G. Bouisse, "Non-linear electro thermal model of LDMOS power transistor coupled to 3D thermal model in a circuit simulator," *Proc. of 34th European Microwave Conference*, vol. 2, 13 Oct. 2004, pp. 573–576.

[32] W. Liu, S. Nelson, D.G. Hill, and A. Khatibzadeh, "Current gain collapse in microwave multifinger heterojunction bipolar transistors operated at very high power densities," *IEEE Trans. Electron. Devices*, vol. 40, no. 11, pp. 1917–1927, Nov. 1993.

[33] University of California San Diego HBT model. [Online]. Available: http://hbt.ucsd.edu/

[34] W.R. Curtice and M. Ettenberg, "A nonlinear GaAs FET model for use in the design of output circuits for power amplifiers," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-33, no. 12, pp. 1383–1394, Dec. 1985.

[35] J.H.K. Vuolevi, T. Rahkonen, and J.P.A. Manninen, "Measurement technique for characterizing memory effects in RF power amplifiers," *IEEE Trans. Microwave Theory Tech.*, vol. 49, no. 8, pp. 1383–1389, Aug. 2001.