TESI DI DOTTORATO

UNIVERSITÀ DEGLI STUDI DI NAPOLI "FEDERICO II"

DIPARTIMENTO DI INGEGNERIA ELETTRONICA
E DELLE TELECOMUNICAZIONI

DOTTORATO DI RICERCA IN
INGEGNERIA ELETTRONICA E DELLE TELECOMUNICAZIONI

MEASUREMENT AND SIMULATION OF ELECTROTHERMAL EFFECTS IN SOLID-STATE DEVICES FOR RF APPLICATIONS

SALVATORE RUSSO

Il Coordinatore del Corso di Dottorato
Ch.mo Prof. Niccolò RINALDI

Il Tutore
Ch.mo Prof. Niccolò RINALDI

A. A. 2010–2011

”To my grandfather
Salvatore.”

Contents

List of Figures

Chapter 1

Introduction

The invention of the first transistor in 1947 by John Bardeen and Walter Houser Brattain and the subsequent invention of the integrated circuits by Jack Kilby in 1958 were destined to change not only the electronics but all fields of human activities. It was 1965 when Gordon Moore wrote a paper [1] that guided the evolution of the semiconductor industry in the forthcoming years. In the middle of 60's he was able to predict with an incredible accuracy the past 45 years of the electronics [1].

> The future of integrated electronics is the future of electronics itself. The advantages of integration will bring about a proliferation of electronics, pushing this science into many new areas. Integrated circuits will lead to such wonders as home computers or at least terminals connected to a central computer, automatic controls for automobiles, and personal portable communications equipment.

Moore formulated a law prefiguring the doubling of the number of integrated devices per chip every year, lately [2] in 1975 he corrected this law by considering that the device number doubles every two years as reported in Fig. 1.1. The most astonishing thing is that this law has been pulling semiconductor industry ahead till today.
The industry ability to follow Moore's law has been the core of a virtuous cycle

- Transistor scaling allows a better performance to cost ratio of products.
- This induces an exponential growth of the semiconductor market.

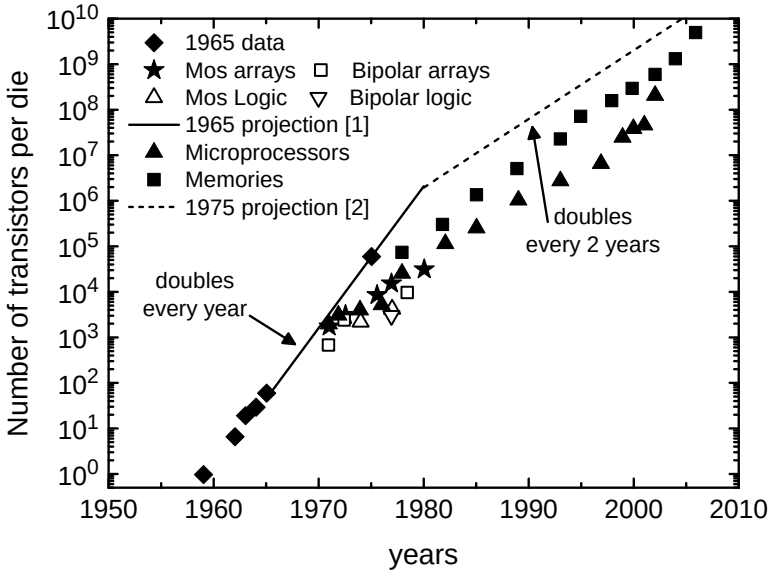


Figure 1.1: Number of transistors per die as a function of the year of production. Projection data (solid line) corresponding to [1] and (dashed line) [2] are reported as well. Data source: [3].

- Moreover, it allows further investments in new technologies that can enable further scaling.

However, the scientific community agrees that the law validity will cease in about 15 years. A new (and more fundamental) barrier must be faced in the next years because materials are made of atoms and the technology is now approaching the atomic dimensions [3]. Such barriers can be overcome, e.g. by exploiting new materials and using 3-D gate structures with multiple gates [3]. Furthermore, the dissipation of energy in the form of heat has long been recognized as a potential issue that may limit information processing [1]. This is more evident in Fig. 1.2 that reports the power density increase versus the production year. Nowadays devices show a power density comparable with the one of a nuclear reactor, this makes thermal management an utmost priority.

The demand for reduced parasitics necessary to guarantee improved (with respect to previous generation devices) performances has forced researchers to adopt increasingly aggressive insulation solutions.
Insulation schemes relying on oxide or even air-filled deep trenches have

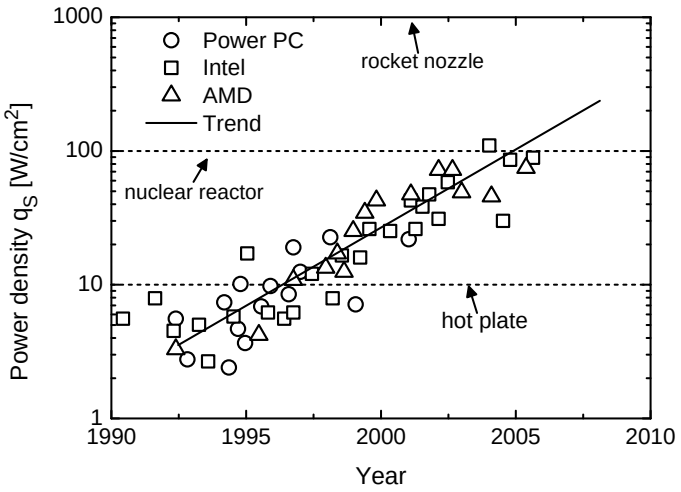


Figure 1.2: Trends of on-chip power density over the past 10+ years. The solid line marks an exponential trend. Horizontal lines refer to the cases of the power density of a hot plate (in the order 10 W/cm^2) and that of a nuclear reactor (around 100 W/cm^2), while the surface of the sun is about 7000 W/cm^2. Data compiled by F. Labonte, Stanford [4].

been constantly introduced to minimize parasitics. Nowadays the silicon-on-insulator (SOI) technology is widely adopted to (electrically) insulate the active area of devices from the silicon substrate through a buried silicon dioxide layer, thus favoring a remarkable gain in circuit performance. However, the major drawback of aggressive insulation schemes is connected to the fact that most of the good electrical insulators exhibit a low thermal conductivity, thus acting also as good thermal insulators. Therefore the employment of such materials can deteriorate the thermal path and may impose a limit on the current density for high speed devices [4, 5]. An example of a very aggressive insulation scheme is the silicon-on-glass (SOG) technology where the lossy silicon substrate is completely removed and substituted by an insulating glass thus enhancing thermal issues [6].
As noticed in [7], the "keyterm" of the market, today, is the diversification of functions (electronic, mechanical, etc...) that can be integrated, all together, in a single product but not necessarily in the same chip. This represents the outline of the System-in-Package (SiP) that can be considered "more than Moore" while, on the other hand, System-on-Chip (SoC) is "more Moore". As evident from Fig. 1.3 the objective of "More-than-Moore" is to extend the use of the silicon-based technology developed in the microelectronics industry to provide new, non-digital functionalities. It often inherits the scaling capabilities of the "More Moore" developments to incorporate digital and non-digital functionalities into compact systems [8]. Integration of many different systems in the same package resorts to 3-D technologies that can enhance (apart integration and diversification) communication between ICs (larger bandwidth, lower latency, and lower energy per bit) by stacking multiple chips into the same package. Aside from issues relating to manufacturing, power delivery and cooling of a stack of logic chips become more complex as compared to the single chip case.
Additionally the electronic scene has seen the introduction of novel semiconductor materials like III-V compounds in order to achieve higher performances than the ones possible with silicon. As an example, it was recently demonstrated [9] the operation of InP heterojunction bipolar transistor (HBT) at cut-off frequencies as high as 1 THz. On the other hand many of these semiconductor materials can present unfavorable thermal properties if compared to silicon (GaAs presents a thermal conductivity that is a third of the Si value).
Therefore it is of the utmost importance to perform an accurate thermal and electrothermal analysis of all modern state-of-the-art circuits/devices (not only power devices). This is particularly true for very high frequency operating

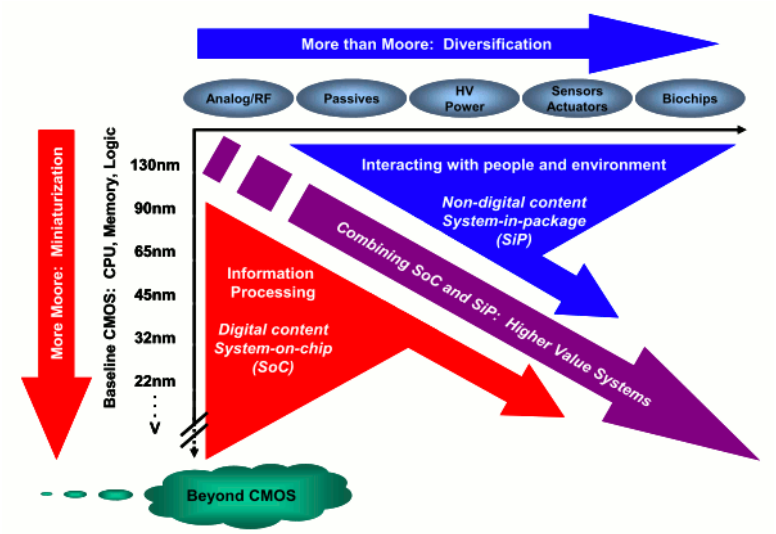


Figure 1.3: Functional diversification refers to the incorporation into devices of functionalities that do not necessarily scale according to "Moore's Law" but provide additional value in different ways. The "More-than-Moore" approach typically allows for the non-digital functionalities to migrate from the system board level into a particular package-level (SiP) or chip-level (SoC) implementation.

Figure 1.4: A 3-finger InGaP/GaAs transistor destroyed by electrothermal effects. The place of each device is taken by a “crate”.

transistors where the annihilation of electrical parasitics (necessary to achieve good electrical performances) making use of aggressive insulation schemes coupled with scaled geometries can truly enhance electrothermal issues so as to hamper performances or even lead to the destruction of the devices (an example is reported in Fig. 1.4).

1.1 Thermal resistance

The temperature influences the device/circuit behavior, it is related to the heat generated because of the Joule effect thus being dependent on the device power. As a consequence, it is more convenient to define a quantity that can describe the steady-state thermal behavior independently of dissipated power P_D. The thermal resistance for solid-state devices is defined as the temperature increase ΔT over ambient normalized to dissipated power P_D

$$R_{TH} = \frac{T_j - T_B}{P_D} = \frac{\Delta T}{P_D}, \tag{1.1}$$

where T_j is the junction temperature[1] and T_B is the ambient (baseplate) temperature. R_{TH}, also called self-heating thermal resistance, is a parameter that depends both on the material properties and the geometry of the structure. In analogy with the electrical resistance it represents the device "inability" to heat conduction.

A more complex situation exists when dealing with multiple devices integrated on the same chip. Considering, as an example, a pair of bipolar transistors, the operating temperature of each elementary device depends on its self-heating thermal resistance and also on the thermal coupling with the neighboring one. This thermal coupling can be modeled by a "mutual" thermal resistance[2]. Let us suppose that the device 1 is not dissipating power, then it is possible to define mutual thermal resistance as its temperature increase over ambient ΔT_1 normalized to the power P_{D_2} dissipated by device 2

$$R_{TH_{12}} = \frac{\Delta T_1}{P_{D_2}} = R_{TH_{21}} = R_M.$$

The procedure can be effortlessly extended to the case of n elementary devices

$$R_{TH_{ij}} = \left.\frac{\Delta T_i}{P_{D_j}}\right|_{P_{D_k}=0 \quad k \neq j}. \tag{1.2}$$

Thus is possible to write, by superposition

$$\Delta T_i = R_{TH_{i1}} \cdot P_{D_1} + \cdots + R_{TH_{ii}} \cdot P_{D_i} + \cdots + R_{TH_{in}} \cdot P_{D_n}, \tag{1.3}$$

[1] T_j, in bipolar transistors, refers to base-emitter junction temperature which is the one influencing the electrical behavior of the device.

[2] While self-heating thermal resistance represents the inability to heat conduction, the mutual thermal resistance quantifies the thermal coupling between elementary devices; it is called mutual thermal resistance because it has the same dimensions of self-heating thermal resistance K/W. Some authors denote it as a mutual thermal coefficient to avoid misunderstandings.

that is a system of linear equations that can be written in matricial form as

$$\begin{bmatrix} \Delta T_1 \\ \vdots \\ \Delta T_n \end{bmatrix} = \begin{bmatrix} R_{TH_{11}} & \cdots & R_{TH_{1n}} \\ \vdots & \ddots & \vdots \\ R_{TH_{n1}} & \cdots & R_{TH_{nn}} \end{bmatrix} \begin{bmatrix} P_1 \\ \vdots \\ P_n \end{bmatrix}.$$

It is also useful to define the normalized temperature increase

$$\theta(x, y, z) = \frac{T(x, y, z) - T_B}{P_D}, \tag{1.4}$$

that represents a sort of "distribution" of thermal resistance and can give thermal information independent of dissipated power (only dependent on materials and geometries) in every point (x, y, z) of the domain of interest.

1.2 Thermal impedance

The quantity that can help to model the dynamic behavior of a device is the thermal impedance. In time domain, it is defined as the thermal transient response to the unitary power step. Let us suppose that at time instant $t = 0$ the dissipated power steps instantaneously from 0 to P_D; the thermal impedance is defined as

$$Z_{TH}(t) = \frac{T_j(t) - T_B}{P_D} = \frac{\Delta T(t)}{P_D}, \tag{1.5}$$

where $T_j(t)$ is the junction temperature evolution during power step and T_B is the ambient or baseplate temperature. Obviously the self heating thermal resistance corresponds to the steady-state value of $Z_{TH}(t)$

$$Z_{TH}(+\infty) = R_{TH}. \tag{1.6}$$

Another two important parameters of interest, very useful to have information on the dynamic thermal behavior, are the rise time t_R and the thermal cut-off frequency f_{TH}. The rise time t_R is the time interval separating the points where the transient response reaches the 10% and 90% of the steady-state value (R_{TH}), respectively. This parameter is an indicator of the time needed to reach the steady-state in time domain.
The thermal cut-off frequency f_{TH} is defined as the frequency at which the magnitude of the thermal impedance reduces to the value of $R_{TH}/\sqrt{2}$.

1.3 Thermal conductivity dependence on temperature

Thermal conductivity is in general reviewed as a constant property for a material. However it is not completely true for many semiconductor materials that show a temperature dependence of thermal conductivity. When this dependence cannot be neglected it is possible to account for it through[3] [10]

$$k(T) = k_0 \cdot \left(\frac{T}{T_0}\right)^{-m}, \tag{1.7}$$

where k_0 is the thermal conductivity for a reference temperature $T_0 = 300$ K, T is the temperature and m an adimensional positive parameter. It is worth noting that for range of interest in semiconductors $m > 1$, that is, thermal conductivity reduces when temperature increases and this reflects into an increase of thermal resistance (being inversely proportional to thermal conductivity).

1.4 Electrothermal effects in solid-state devices

For semiconductor devices it is necessary to define the voltage and current working points in which the device can operate without self-damaging. This is done by defining a safe operating area (SOA) the border of which represents the boundary for the normal working conditions of the devices. Fig. 1.5 depicts the SOA of a bipolar transistor; the limits for safe operation of devices are identified by

- A max current I_{MAX} for which the current gain has still a reasonable value (i.e., 10).
- A maximum voltage V_{MAX} determined by avalanche multiplication.
- The maximum power line $P_{MAX} = const$ is related to the maximum junction temperature $T_{j_{MAX}}$ by recalling (1.1).
- A line with increased slope due to electrothermal effects [11, 12] and/or impact ionization (II) [13].

The SOA of solid-state devices can be limited by a mechanism that is often referred to as the “snapback” or “flyback” effect [11]. It is caused by the

[3] This dependence cannot be neglected for semiconductors of interest in microelectronics like, e.g., Si and GaAs.

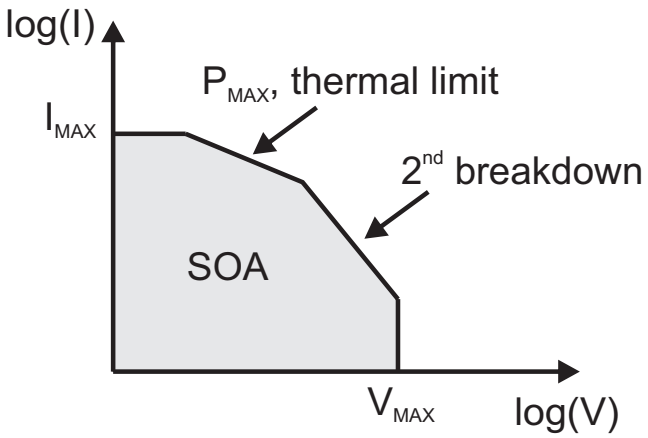


Figure 1.5: DC safe operating area for a bipolar transistor.

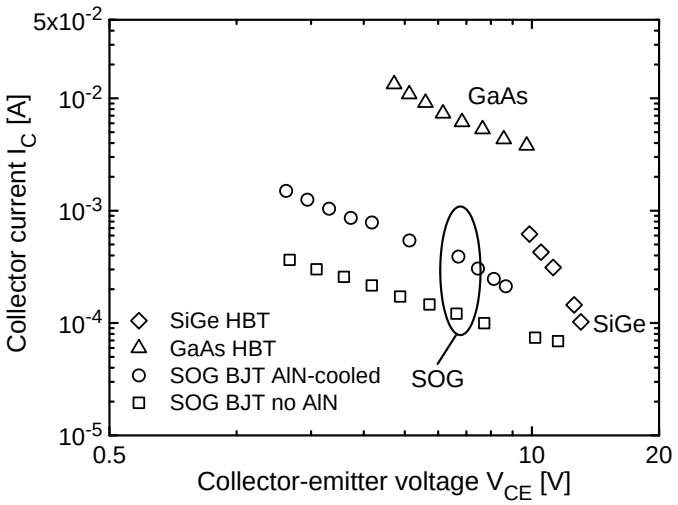


Figure 1.6: Measured flyback loci (i.e., SOA boundaries) for various single-finger devices. (Rhombi) SiGe HBT , (triangles) GaAs HBT, (circles) SOG AlN-cooled BJT and (squares) SOG BJT. All devices have an emitter area of 20 μm^2 with the exception of GaAs HBT that has a horse-shoe shaped emitter with an area of 60 μm^2.

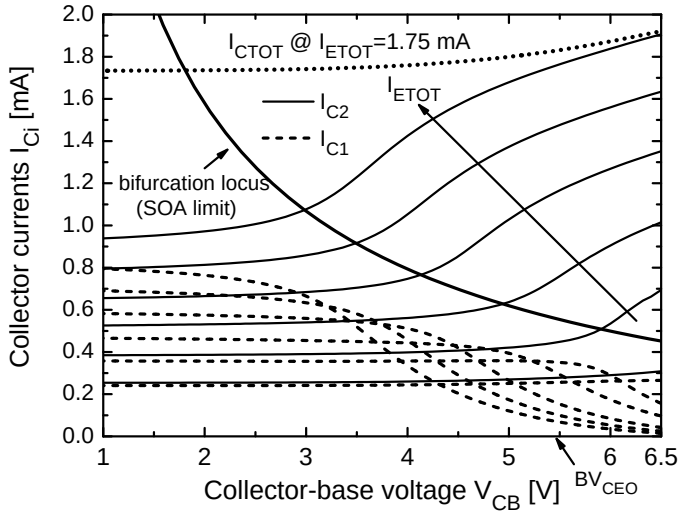


Figure 1.7: Experimental common-base output characteristics of an AlN-cooled two-finger SOG device biased with a constant emitter current I_{ETOT} ranging from 0.5 to 1.75 mA in steps of 0.25 mA. Also shown are voltage BV_{CEO} , (dotted line) the total collector current I_{CTOT} corresponding to $I_{ETOT} = 1.75$ mA, and (solid thick line) the bifurcation locus (i.e., the SOA limit).

positive feedback due to self-heating, and results in the occurrence of a negative differential resistance in the output characteristics. The instability locus, i.e., the collection of points in the output plane which defines the onset of the negative differential behavior, is usually taken as the SOA 2nd breakdown boundary. Impact-ionization can introduce a positive feedback which may lead to a snapback phenomenon. Fig. 1.6 shows the experimental SOA boundaries corresponding to various transistor families. It can be observed that the addition of a 4-µm-thick AlN layer (AlN-cooled SOG) can considerably enlarge the SOA of a SOG *npn* bipolar junction transistor (BJT) due to the strong reduction in self-heating thermal resistance [14].

Another instability effect is a sudden current focusing due to a nonuniform temperature distribution in the device [12]. This effect is often referred to as current hogging/focusing and has been studied in greater detail in multifinger BJTs. Fig. 1.7 shows the currents flowing in the fingers of a two-finger AlN-cooled (i.e., a 4-µm-thick layer of AlN is used as heatspreader) silicon-on-glass

BJT under constant emitter current driving conditions. Although each finger is expected to carry an equal amount of the output current (about one half of the overall emitter current), it can be seen that when a critical condition is reached called "bifurcation condition" one of the fingers suddenly absorbs most of the current, while the other one runs dry. This non-uniform current distribution among the fingers is undesirable as it leads to uneven device area utilization and a higher junction temperature. A theoretical analysis has shown that this effect can be attributed to the existence of multiple solution branches introduced by thermal and/or impact-ionization phenomena [12, 13]. It is worth noting that device stability can be improved by a stronger thermal coupling (i.e., increasing mutual thermal resistance by reducing finger spacing).

1.5 Thesis outline

After this brief introductory chapter, containing some definitions and a quick overview of thermal instabilities, the following 4 chapters represent the core of the work.

Chapter 2 is dedicated to the modeling of the thermal behavior of solid-state devices. It describes the use of thermal equivalent networks to model the thermal behavior of solid-state devices. The *in house* software conceived to perform the network identification (starting from thermal transient) is detailed in this chapter as well. Moreover, a procedure is used to evaluate the thermal properties of the structure and its defects by making use of thermal transient details. A routine has been written in order to extract the necessary parameters from the thermal response to the power step. Furthermore, scalable thermal models of thermal capacitance and resistance for bipolar transistors are described. The analytical formulation of compact thermal models (CTMs) for simulation of complex domain is discussed and an approach resorting to *shape functions* is proposed.

The development of a software platform based on a commercial simulator for the numerical analysis is discussed in **Chapter 3**. This platform is adopted to study several families of heterogeneous devices, that is, both single- and multifinger GaAs HBTs, GaN HEMTs, SOG BJTs, and SiGe HBTs to attain the guidelines to follow during the design phase. The numerous analyses allowed also the extraction of thermal networks suitable to (thermally) describe the devices. Moreover the novel ultra-thin chip stacking

technology (UTCS) has been deeply investigated.

In **Chapter 4** a critical comparison of the thermal resistance measurement techniques is carried out in order to supply the designer with extensive information on their accuracy and limitations. After this study a novel measurement method for the thermal resistance is presented, allowing for the accurate evaluation of the thermal resistance in bipolar transistors significantly affected by Early effect. All the analysis is supported by experimental data and electrothermal circuital simulations. Thermal impedance measurements techniques are reported as well and the "on-the-fly" cooling curve technique is adopted to measure thermal transient of silicon-on-glass bipolar transistors.

An electrothermal circuital simulation environment developed *in house* is presented in **Chapter 5**. This tool is based on the enrichment of commercial simulators and is used to perform several simulations both steady-state and transient whose results are reported as well.

Finally, **Chapter 6** provides the main conclusions and recommendations for the future work.

Chapter 2

Modeling the thermal behavior of solid-state devices

Solid-state device operation is always associated with heat dissipation and the consequent self-heating effect. When thermal effects become relevant, the device performances can be hampered and a reliable design should be verified with electrothermal (ET) analysis. Several modeling techniques are available in the literature to accurately evaluate thermal behavior of solid state devices/circuits starting from analytical methods [15] for simplified geometrical structures and arriving to complex 3-D FEM simulations. However, when it is necessary to perform electrothermal simulations by coupling the electrical and thermal model it is paramount, to reduce the computational effort, to resort to simplified, yet accurate, thermal models. The common effective approach to electrothermal problem solution was originally proposed by Székely and Tarnay [16] and consists in the use of a thermal network.

2.1 Modeling the transient thermal behavior with equivalent networks

While in principle thermal networks represent a simple approach, the implementation of an accurate thermal network is not a trivial task. In fact, the analysis of heat conduction can be performed by discretizing the heat diffusion equation by resorting to a large number of thermal nodes where a temperature variable is defined. An equivalent *RC* network where, in analogy to the electrical problems, temperature corresponds to electrical voltage and heat flux to

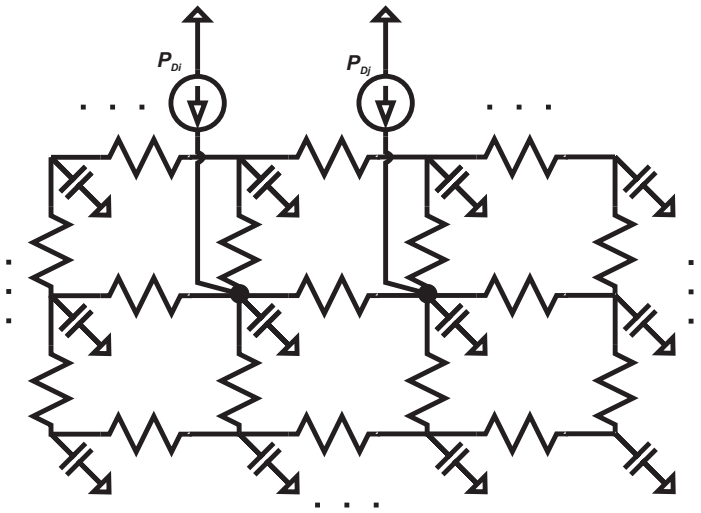


Figure 2.1: Possible schematic representation of a complex *RC* network equivalent to a typical heat conduction problem in solid-state devices.

electrical current can be obtained by using various discretization techniques based, for example, on finite differences or finite elements. As evident from Fig. 2.1 that represents a thermal *RC* network obtained from discretization, in general, a large number of nodes is necessary to model the behavior of a real thermal system with an acceptable accuracy.

2.1.1 Thermal network identification: accuracy vs complexity

It is clear that the accuracy of a model is highly desired, but at circuit level the thermal impact on electrical behavior of solid-state devices is taken into account by using a single temperature referred to as junction temperature T_j. This parameter is someway connected to the thermal distribution inside the dissipating device whose electrical behavior is modeled at the external contacts (the global voltages and currents are taken into account). Several definitions for T_j have been proposed in literature; some of these rely on a particular temperature value assumed in a domain within the device (i.e., temperature at the center of heat source, maximum temperature value, etc...), others on an averaged value within a region (i.e., value averaged over the base-emitter junction, value averaged over heat source, etc...). By observation of the Fig. 2.1 it is

clear that the fact that each node has an assigned temperature means that resorting to a complex representation (i.e., obtained by discretizing the layout of the device) a temperature distribution can be obtained. This, someway, results to be in contrast with the needs of electrical circuital simulators. In fact, the circuital simulator can provide the global power (for a single device) to thermal network and needs the junction temperature disregarding all the node temperatures. For these reasons it is clear that the use of a complex thermal network for electrothermal simulations if from one side can give extensive information on the temperature field inside the device, on the other hand complicates the scenario introducing not needed (by electrothermal simulator) variables that are necessary to be connected to the global (i.e., dissipated power and junction temperature) values. Once a connection between distributed values and global values is found it is possible to use a complex thermal network to perform electrothermal simulations. For example it is possible to average nodal temperatures of the complex circuit to obtain T_j, and to consider the single power sources (in the complex model) replaced by a parallel connection to the global power source (i.e., the one provided for the single device by electrothermal simulator).

Unfortunately, thermal networks obtained by directly discretizing the layout of a device tend to be very large, dramatically increasing simulation time and possibly leading to numerical problems [17]. From this it can be evinced that the increase in complexity of the network, as stated before, not always corresponds to an increase in accuracy of electrothermal simulation, and, in general, the validity of a thermal model is strictly connected to the scope it is conceived for. Nowadays the most used simulation approach in commercial electrothermal circuit simulators is to employ monodimensional networks, in particular a single *RC* network to model thermal behavior considering that the junction temperature and the dissipated (global) power are the only quantities of interest for coupled simulations. Obviously, the use of a single *RC* cell is an oversimplification of the problem owing to the fact that the heat crosses different layers in the device before reaching the heat-sink, each characterized by different thermal properties that can affect the global behavior. However, although the heat conduction in solid-state devices is intrinsically a 3-D phenomenon, monodimensional networks can be successfully employed to correctly take into account thermal behavior of solid-state devices in electrothermal circuit simulations.

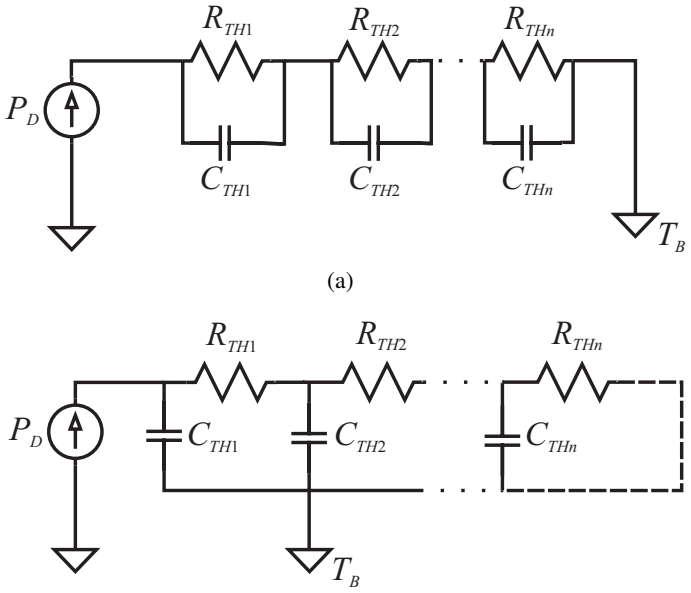


Figure 2.2: Foster (a) and Cauer (b) network models.

2.1.2 Modeling with different network topologies: Foster and Cauer

While the thermal response of a system is an intrinsically distributed phenomenon, every practical network model has a finite number of elements and for this it has to be considered of lumped nature [18]. Two different equivalent representations for monodimensional thermal networks exist, namely, the series network also known as Foster network (depicted in Fig. 2.2(a)) and the cascade network also known as the Cauer network (depicted in Fig. 2.2(b)). Both these network models can be used to describe the thermal transient behavior of a system but each representation has its own advantages and disadvantages.

The Foster Network has a very simple mathematical definition both in time and frequency domain. In fact a Foster network model is simply described, in time domain, by

$$Z_{TH}(t) = \sum_{i=1}^{n} R_{TH_i} \cdot \left[1 - exp\left(-\frac{t}{\tau_i}\right)\right], \tag{2.1}$$

and in frequency domain by

$$Z_{TH}(f) = \sum_{i=1}^{n} \frac{R_{TH_i}}{1 + j \cdot 2\pi \cdot f \cdot \tau_i}, \tag{2.2}$$

where R_{TH_i} and C_{TH_i} are the *i*-th thermal resistance and capacitance (showed in Fig. 2.2(a)) while $\tau_i = R_{TH_i} \cdot C_{TH_i}$ is the *i*-th time constant. The thermal capacitance C_{TH}, in general, can be viewed as the ability of a body to store heat and is measured in units of J/K.
Using the Foster model network, thermal resistances and capacitances can be obtained by fitting or by using several direct techniques proposed in literature [19–21]. Moreover, a direct frequency representation of thermal response is obtained so as to allow the discussion of results in frequency domain. For this reason all data sheets report R_{TH_i} and C_{TH_i} parameters for a Foster network to describe the thermal impedance of a module. Nevertheless, this model, while representing correctly the thermal behavior of the complete system, is purely formal [18, 22] due to the fact that internal nodes do not have a physical meaning. This can be inferred, intuitively, by the fact that changing the order of *RC* elementary networks in a Foster model does not affect the overall response; this means that there is no direct relation between the device physical structure and an elementary *RC* port. Moreover, thermal capacitances represent the energy storage capability of a volume inside the structure and, obviously, increase with it; that is not happening in Foster network model; to allow this all capacitances should be referenced to the same ground. For the same reason, it is not possible to connect directly two (or more) Foster networks to obtain the network representing the combination of subsystems; the result would be an unphysical faster transient[1].

The Cauer network depicted in Fig. 2.2(b) is more suitable than Foster network to describe the thermal transient of a system in a more "natural" way. In fact, it is noteworthy that the thermal capacitances are all connected to thermal references, differently from Foster network, and each cell is physically representative of heat-conducting structure layers, thereby describing correctly the internal temperature distribution

[1] An addition of a layer in a system causes an overall increase in thermal capacitance from a physical point of view. On the contrary, adding a cell to the Foster network results in a series connection between capacitances with an unphysical decrease of the overall thermal capacitance and a faster thermal response.

(i.e., the temperature rises at the internal nodes) [18, 22]. In this case, each cell is indicative of the contribution given to thermal impedance by a portion of the structure where the resistance can be determined on the basis of the thermal conductivity and the capacitance depends on the mass and specific heat of the delimited domain. Moreover, the Cauer model is a two-port network that, being representative of a subsystem, can be connected to other circuits (other two-port networks) to produce the thermal model of the complete system[2].
Unfortunately, there is no direct correspondence between time constants τ_i in (2.1) or (2.2) and thermal parameters (R_{TH_i} and C_{TH_i} for Cauer network) depicted in Fig. 2.2(b); in Cauer network, in fact, each time constant depends on all the thermal parameters (that is, all thermal resistances and capacitances) leading to a very complex mathematical representation and resulting in a severe difficulty in the identification of the network. In Laplace domain, the response of a Cauer network can be written indeed as

$$Z_{TH}(s) = Z_{TH_1}(s) + \cfrac{1}{Y_{TH_1}(s) + \cfrac{1}{Z_{TH_2}(s) + \cfrac{1}{Y_{TH_2}(s) + {}^1\!/\!_{\ldots}}}}, \quad (2.3)$$

where $Z_{TH_i}(s)$ and $Y_{TH_i}(s)$ are the thermal impedance and admittance of the *i*-th cell in the Cauer model.

A possible strategy to overcome the difficulties of Cauer network identification is to extract from thermal transient the equivalent Foster network and then transform, by using the formal identity of the two networks, the parameters obtained for one network in the Cauer model.
An advanced software platform has been developed in order to accomplish these tasks; this software is able to:

- Analyze thermal transient data obtained by measurements or simulations
- Obtain optimized thermal resistances and capacitances for a Foster network

[2] A typical example in solid-state devices is the thermal model of a die molded on its package. The overall thermal model can be obtained by simply connecting, in series, the two cascade networks relative to the chip and the package, respectively.

- Evaluate the minimum number of cells to warrant a prescribed accuracy
- Evaluate the continuous spectrum of time constants
- Transform the Foster network into the equivalent Cauer network.

The flow-chart is depicted in Fig. 2.3, it is worth noting that the software basically get thermal impedance versus time data as input along with the maximum allowed error and then the procedure evolves to find the best values (within the prescribed accuracy) for *RC* cells (the depicted flow-chart does not show the necessary steps for Foster to Cauer transformation).
More precisely, the software has three different modes of operation, namely

Maximum error. With this mode the system tries to optimize the parameters in order to return a result with an error less or equal to the value required with the minimum number of cells (i.e., *RC* pairs).

Fixed number of cells. When the software operates in this modality it optimizes the result in order to obtain the minimum possible error only for the prescribed number of networks (given as input by the user).

Full spectrum. In this operative mode the routine applies the method of *identification by deconvolution* proposed in [19,20] to gain the time constants spectrum (as it will explained later in this chapter). In this particular case there is no limitation on the number of elementary cells obtained, the limitation is only connected to numerical resolution of the thermal transient.

Apart from an optional filtering of the data that can be enabled, the first step, before optimization, is to evaluate the initial conditions, i.e. the initial values of thermal parameters (with the exception of the software working in full spectrum mode). This operation is not trivial and can affect dramatically the optimization in terms of (i) computational effort (time needed to retrieve network parameters) and (ii) results accuracy (solutions that are local minimums). For this reason several methods are used to obtain initial conditions and then the best possible solution is chosen to perform the optimization[3]. The optimization is then performed by varying the R_{TH_i} and C_{TH_i} values simultaneously

[3]The user can force the system to ignore the default behavior so as to optimize initial conditions obtained with the five methods described in the following section (with a bigger computational effort) and then choose the best fit. It is also possible to force the system to use a particular set of initial conditions.

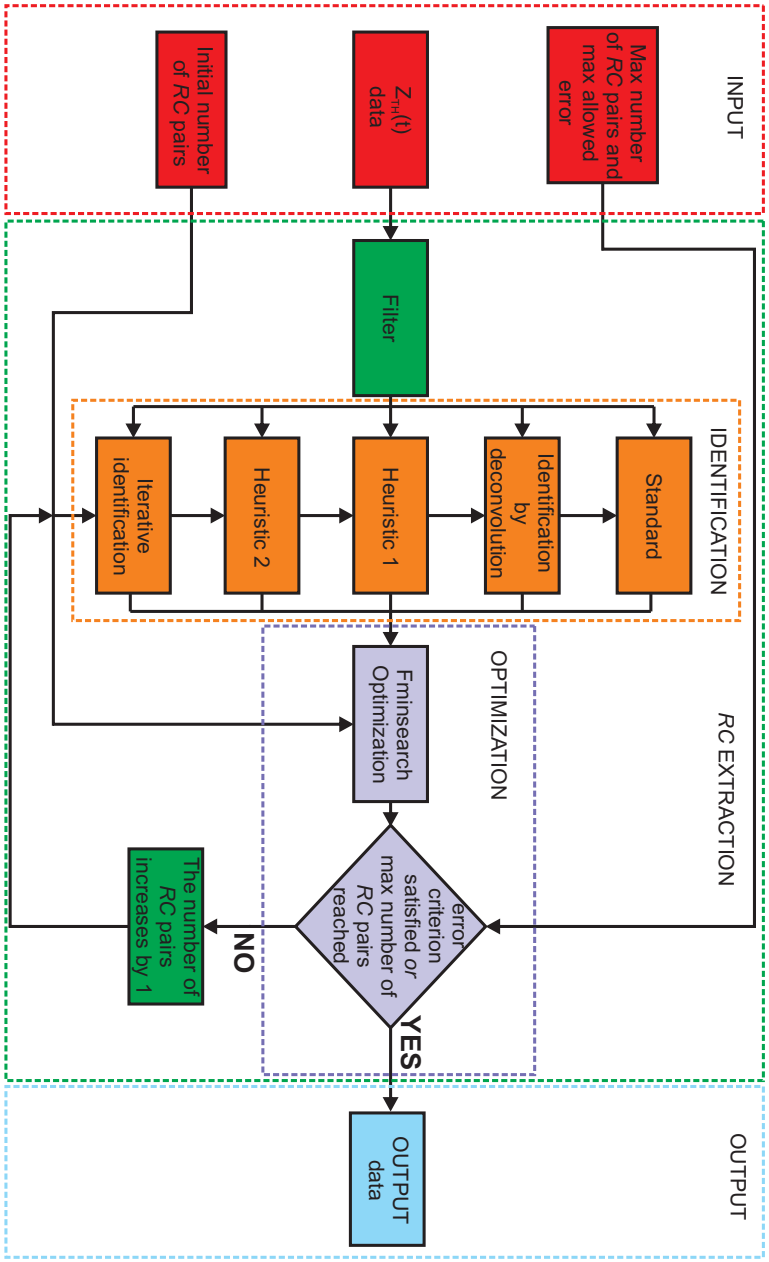


Figure 2.3: Flow-chart of the identification software.

accordingly to the Nelder-Mead simplex algorithm [23] until the sum of the squares of the differences between the (2.1) and the *target* curve (i.e., the input thermal transient) is minimized. When the *maximum error mode* is used, the software increases the number of degrees of freedom by adding another *RC* cell, if the resulting error exceeds a maximum value (preliminarily assigned by the user). The whole process is re-iterated until the maximum error criterion is satisfied. It is also possible to perform optimization only for a chosen number of elementary *RC* cells when setting the software in *fixed number of cells mode*. The methods used to evaluate the initial conditions are based on some direct identification methods proposed in literature [19–21] and on some *heuristic* techniques based on simplifications of (2.1).
All the methods to evaluate initial values for *RC* pairs (showed in the flowchart depicted in Fig. 2.3) are explained in the following.

Standard. This method is based on some simplifications of (2.1). In particular, it considers that the thermal transient is originated by a number n of equal time constants and equal thermal resistances. This means that all elementary cells share the same R^*_{TH} and C^*_{TH} values, hence it is possible to simplify the (2.1) in the case of n elementary ports as

$$Z_{TH}(t) = n \cdot R^*_{TH} \cdot \left[1 - exp\left(-\frac{t}{\tau^*}\right)\right], \tag{2.4}$$

where $R^*_{TH} = {}^{R_{TH}}/_{n}$ and $\tau^* = R^*_{TH} \cdot C^*_{TH}$ can be obtained by evaluating the rise time of thermal transient $t_R = t_2 - t_1$ and using the (2.4). It can be easily obtained that

$$t_1 = -\tau^* \cdot ln(0.9),$$

$$t_2 = -\tau^* \cdot ln(0.1),$$

$$\tau^* = R^*_{TH} \cdot C^*_{TH} = \frac{t_R}{ln(9)},$$

whence

$$C^*_{TH} = \frac{t_R \cdot n}{R_{TH} \cdot ln(9)}. \tag{2.5}$$

Obviously, this method is a rough approximation and gives acceptable results only if time constants and thermal resistances assume more or less the same values.

Heuristic 1 and Heuristic 2. These methods are very similar since they start from the same hypotheses: the time constants are equally (i) spaced and (ii) distant. With these assumptions it is possible to deduce that exponential terms in (2.1) don't influence each other. Once the time domain is equally (logarithmically) divided into a number n of intervals ($[t_0, t_1], [t_1, t_2], \cdots, [t_{n-1}, t_n]$), it is possible to use the $Z_{TH}(t_i)$ value, corresponding to the superior bound of each interval, to evaluate

$$R_{TH_i} = Z_{TH_i}(t_i) - \sum_{k=1}^{i-1} R_{TH_k}. \tag{2.6}$$

These two methods differs for the evaluation of $\tau_i = R_{TH_i} \cdot C_{TH_i}$. In the first method (*Heuristic 1*) the time constant is evaluated from the slope of the linear regression of the logarithm of the curve (the input thermal transient) in the *i*-th interval $[t_{i-1}, t_i]$ that is equal to $1/\tau_i$. In the second method (*Heuristic 2*) τ_i is evaluated on the basis of the time instant t_i^* defined as the time at which

$$Z_{TH}(t_i^*) = \frac{R_{TH_i}}{2} + Z_{TH_i}(t_{i-1}),$$

whence it is possible to approximately evaluate

$$C_{TH_i} = \frac{t_i^*}{R_{TH_i} \cdot ln(2)}. \tag{2.7}$$

Obviously, the two methods can provide a good approximation of thermal impedance only if the spacing between thermal constants exceeds one decade.

Identification by deconvolution. One of the methods proposed in literature [19, 20] is based on the substitution of R_{TH_i} in (2.1) with a continuous time constants spectrum; in this case, the summation should be replaced by integration

$$Z_{TH}(t) = \int_{-\infty}^{+\infty} R_{TH}(\zeta) \left[1 - exp\left(-\frac{t}{\zeta}\right)\right] d\zeta, \tag{2.8}$$

where $R_{TH}(\zeta)$ is the *time constant spectrum* and $\zeta = ln(\tau)$. Using the variable substitution $z = exp(t)$ and deriving $Z_{TH}(z)$ with respect to z it is possible to obtain

$$\frac{dZ_{TH}(z)}{dz} = R(z) \otimes W(z), \tag{2.9}$$

where $W(z) = exp(z - exp(z))$. By transforming (2.9) in the Fourier domain

$$Z'_{TH}(\Phi) = R_{TH}(\Phi) \cdot W(\Phi),$$

and

$$R_{TH}(\Phi) = \frac{Z'_{TH}(\Phi)}{W(\Phi)}, \tag{2.10}$$

where Φ is the transformed variable.
From (2.10) it is possible to obtain the continuous time spectrum by resorting to the inverse Fourier transformation. The main practical limit is that the higher Φ frequency components of $W(\Phi)$ are very small. This, obviously, enhances $R_{TH}(\Phi)$ values at very high Φ frequencies, that is, noise is enhanced as well [19,20]. For this reason an accurate filtering is required to reduce the high frequency noise introduced. As an alternative some other deconvolution methods, like Bayesian deconvolution, can be employed to annihilate it [19,20]. The final result of *network identification by deconvolution* (NID) is a continuous time spectrum whose integral is the self-heating thermal resistance R_{TH}. By sampling the spectrum into n logarithmically-spaced time constants it is possible to obtain an approximated Foster network. It is worth noting that this method can provide a very high number (if each time constant of the spectrum is considered) of elementary cells (limited only by numerical resolution) that do not need any further optimization as it will explained in Section 2.1.3. However, in this case the NID method is employed just to find a low number n of elementary cells to be further optimized in order to minimize the error defined by the user (by keeping the number of elementary ports as low as possible).

Iterative identification. This method also known as the *Peeling method* is based on the recursive elimination of time constants as showed in [21]. From (2.1) it can be directly deduced that

$$\sum_{i=1}^{n} R_{TH_i} - Z_{TH}(t) = \sum_{i=1}^{n} R_{TH_i} \cdot exp\left(-\frac{t}{\tau_i}\right),$$

considering that $Z_{TH}(+\infty) = \sum_{i=1}^{n} R_{TH_i}$, it can be demonstrated [21] that function $F_1(t) = [Z_{TH}(+\infty) - Z_{TH}(t)]$ has an asymptote for $t \to +\infty$

$$a_1(t) = -\frac{t}{\tau_1} + ln(R_{TH_1}).$$

The evaluation of the asymptote makes it possible to calculate R_{TH_1} and τ_1, where τ_1 is the biggest time constant (**identification**). Removing the influence of the calculated time constant (**elimination**) it is possible to obtain

$$F_2(t) = \left[Z_{TH}(+\infty) - Z_{TH}(t) - R_{TH_1} \cdot exp\left(-\frac{t}{\tau_1}\right)\right],$$

whose asymptote for $t \to +\infty$ is

$$a_2(t) = -\frac{t}{\tau_2} + ln(R_{TH_2}).$$

The knowledge of $a_2(t)$ permits to evaluate R_{TH_2} and $\tau_2 < \tau_1$, by iterating this procedure it is possible to obtain

$$F_i(t) = \left[Z_{TH}(+\infty) - Z_{TH}(t) - \sum_{k=1}^{i-1} R_{TH_k} \cdot exp\left(-\frac{t}{\tau_k}\right)\right], \quad (2.11)$$

$$a_i(t) = -\frac{t}{\tau_i} + ln(R_{TH_i}). \quad (2.12)$$

All other time constants are, thus, evaluated by gradually *identifying* and *eliminating*. The main difficulty of this procedure is connected to a correct numerical evaluation of asymptotes (carried out via a linear regression on a variable number of samples) and the minimum value of R_{TH_i} to take into account for constant identification (thermal resistances with values less than a part of $Z_{TH}(+\infty)$, i.e. 1%, are ignored).
This method returns a number of time constants that can be either higher or lower than n; for this reason there are different possible behaviors to obtain the wanted number of time constants to optimize.

1. When the number of time constants obtained by this identification technique equals that required, there is no further necessary action to perform.
2. If the routine determines a number of time constants that is less than n, the dominant time constant τ_d (the one with $R_{TH_d} = \max_i R_{TH_i}$) is selected and *split* so as to obtain the missing number of time constants. These time constants have the same numerical value of τ_d and a thermal resistance of $R_{TH_d}/n - n_1$, where $n_1 < n$ is the number of identified time constants.

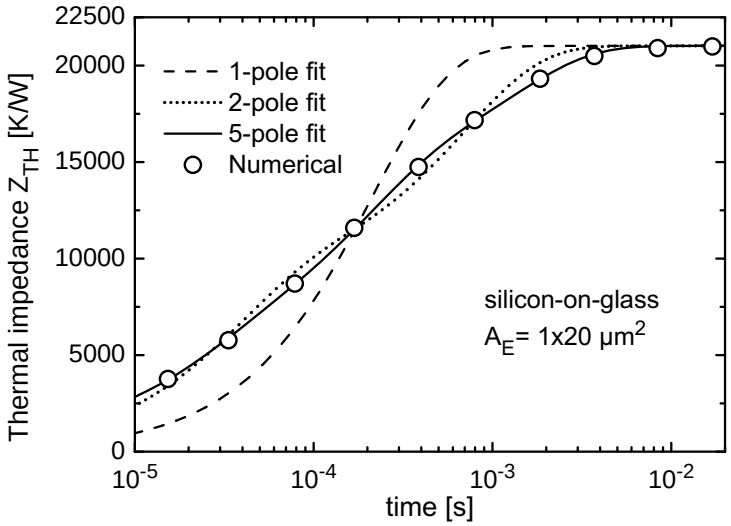


Figure 2.4: Thermal network identification for a silicon-on-glass bipolar transistor having an emitter stripe of 1×20 µm^2.

3. In the case that the software identifies a number $n_1 > n$ of elementary cells, the less influent time constants (the ones with the lower values of R_{TH_i}) are ignored and the values of the remaining n thermal resistances are rescaled proportionally in order to assure

$$Z_{TH}(+\infty) = \sum_{i=1}^{n} R_{TH_i}.$$

Owing to the fact that the final objective is the evaluation of the minimal number of *RC* elementary networks it is paramount to optimize the result. A typical example of application of the software is depicted in Fig. 2.4 where the thermal transient of a silicon-on-glass device is reported. It is straightforward to observe that a network built with a single cell is very inaccurate if compared to the 5-pole network (i.e., five elementary *RC* networks).
As formerly stated, the transformation of the Foster optimized model in the Cauer network can be very useful. It is possible, as pointed out in [24], to carry out the transformation by simply following the steps reported below.

- The transfer function of the Foster network is evaluated as a rational function in the form of $p_n(s)/q_n(s)$.

- The inverse of such function, i.e., $q_n(s)/p_n(s)$, is decomposed through the Euclid's algorithm into a first-order polynomial in s and a strict rational function referred to as remainder and denoted as $rem_n(s)$ so that

$$\frac{q_n(s)}{p_n(s)} = C_{TH_n} \cdot s + k_n + \frac{rem_n(s)}{p_n(s)},$$

  where C_{TH_n} is the value of the first thermal capacitance of Cauer network and $R_{TH_n} = 1/k_n$ is the value of the first thermal resistance. With this decomposition the thermal capacitance and resistance from heat source to ambient temperature are obtained (the first low-pass filter is the one directly connected to the power source).

- Subsequently, $p_{n-1}(s)$ and $q_{n-1}(s)$ are obtained as functions of $rem_n(s)$ and $p_n(s)$ as reported in [24]

$$q_{n-1}(s) = k_n \cdot p_n(s) + rem_n(s),$$

$$p_{n-1}(s) = -\frac{rem_n(s)}{k_n}.$$

- The procedure is iteratively applied to obtain all resistance/capacitance values until the remainder is equal to 0.

2.1.3 Evaluating thermal parameters and detecting structure defects by using thermal transient

As stated in the former section, the Cauer network model is the only one that correlates the equivalent lumped *RC* elements to "real" heat-conducting structure layers, thereby physically describing the internal temperature distribution (i.e., the temperature rises at the internal nodes). In particular once the Cauer network is evaluated it is possible to introduce the *cumulative structure* or *Protonotarios-Wing function* by the names of the authors of [25]. This function connects the cumulative thermal resistance R_Σ to the cumulative thermal capacitance C_Σ defined in the case of a Cauer network made of n elementary cells, by

$$R_{\Sigma_i} = \sum_{j=1}^{i} R_{TH_j}, \tag{2.13}$$

$$C_{\Sigma_i} = \sum_{j=1}^{i} C_{TH_j}, \tag{2.14}$$

respectively, while the *Protonotarios-Wing function* is

$$C_\Sigma = \sigma(R_\Sigma). \tag{2.15}$$

It is straightforward to demonstrate that (2.15) corresponds to a step function when a lumped *RC* network is considered [18]. If the *cumulative structure function* is known, it is possible to retrieve directly an equivalent "reduced" network approximated by a step-wise function. In fact, this monotonously increasing function has some peculiarities [26–29]; in particular, referring to structures where the cross sectional area is constant (that is most likely to occur in sandwich structures, where all layers have the same area):

- A change in the slope of the curve indicates a transition from one material to another.
- A plateau indicates a material with a small thermal capacitance and a large thermal resistance, that is, an insulator.
- The linearly increasing portions of the curve represent the fact that the heat flow is passing through higher (thermally) conductive materials.

Several algorithms can generate an approximation of the Cauer network, e.g. by finding the subsequent inflection points in the cumulative structure function. Fig. 2.5 depicts the structure function of an ultra-thin chip stacked device

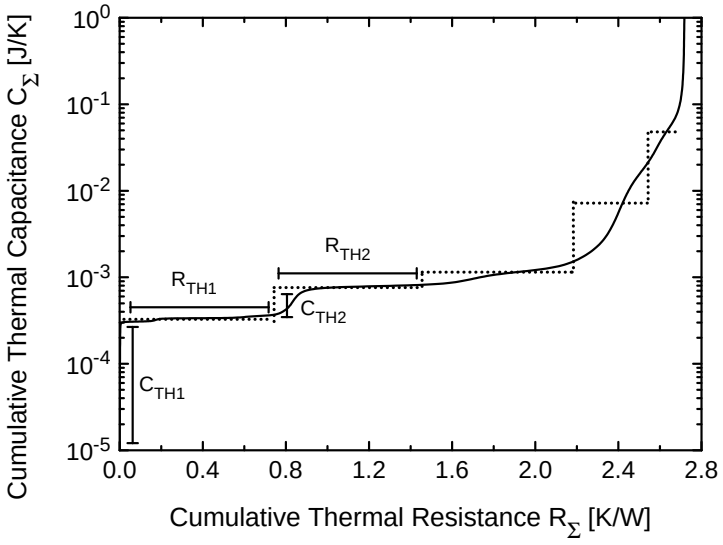


Figure 2.5: Cumulative structure function for a device fabricated with 3 stacked chips interleaved by BCB. The solid line represents the function while the dotted line is a step approximation to graphically deduce a Cauer reduced model. In figure are reported the parameters of the first two low-pass filters.

composed of three chips separated by a benzocyclobutene (BCB) layer (as detailed in Section 3.1.5).

The first vertical step, clearly visible in the cumulative structure function, represents the thermal capacitance of the top die (indicated as C_{TH_1} in the picture). The next element is the BCB (insulator) layer positioned just under the top die whose high thermal resistance is emphasized by the plateau in the figure (the thermal resistance associated with this layer is referred to as R_{TH_1}). In a similar fashion, thermal resistances and capacitances for remaining domains, that is, silicon chips and BCB layers as well as the silicon host and, finally, the heat sink encountered in the heat-flow path, can be evaluated by graphical means. On the basis of these considerations it is clear that this methodology can be used to evaluate thermal parameters, once geometrical parameters are known with good accuracy, or to identify inhomogeneities in the structure, like the presence of voids in the die attach.

To simplify this task another structure function was introduced by Szèkely and

Van Bien in [30] that is called *differential structure function* and is defined as the derivative of cumulative capacitance versus cumulative resistance

$$K(R_\Sigma) = \frac{dC_\Sigma}{dR_\Sigma}. \tag{2.16}$$

The demonstration of the direct connection between thermal properties and the *differential structure function* can be easily deduced by considering a slice of a homogeneous material (a parallelepiped, for example) having a cross section area of A and width of dl. Considering c_p the volumetric heat capacitance of the material and k its thermal conductivity, it is possible to write, for the infinitesimal element considered, that

$$dR_\Sigma = \frac{dl}{k \cdot A},$$

and

$$dC_\Sigma = c_p \cdot A \cdot dl,$$

whence

$$K(R_\Sigma) = c_p \cdot k \cdot A^2. \tag{2.17}$$

Hence, the value of $K(R_\Sigma)$ is directly connected to the structure of the system being proportional to its thermal properties and cross sectional area. Once cross sectional area is known it is possible to evaluate directly the thermal parameters. This method could be applied to evaluate the unknown thermal properties of a layer (geometrically defined) inside the structure of the system. The *differential structure function* has the following properties:

- The local peaks (a change of slope in *cumulative structure function*) indicate that new materials have been reached in the heat flow path.
- The distance between the local peaks along the horizontal axis provides the partial thermal resistance between the surfaces (as a result it is possible to evaluate thermal conductivity if geometrical parameters are known).
- The peaks usually arise at the middle of every material domain where the cross-sectional areas and the physical properties are uniform.
- Once the thermal conductivity is known, it is possible to evaluate the volumetric heat capacitance by using (2.17) for a peak (i.e., for the material that it represents).

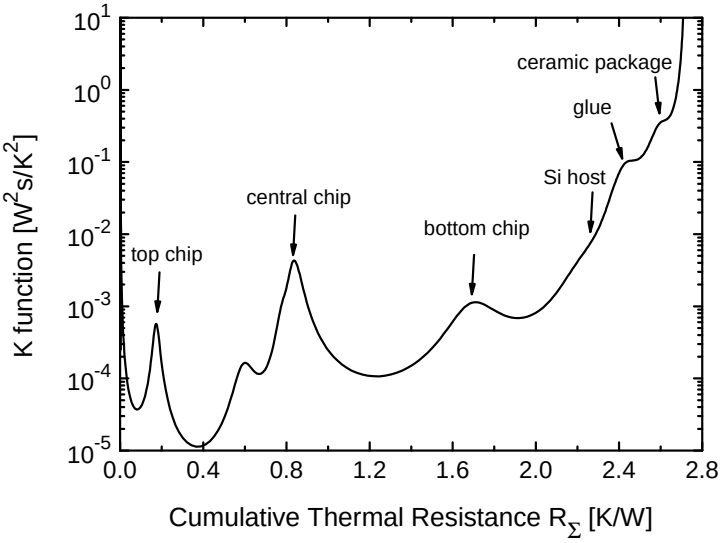


Figure 2.6: Differential structure function K for a device fabricated with 3 stacked chips interleaved by BCB. Peaks represent the reaching of the middle of the new material by the heat-flow. Materials reached are indicated as well.

Obviously this direct connection between heat-flow path and structure by means of differentials structure functions holds only if the heat-flow from source to heat sink is mostly unidirectional. However, this simplification is not very restrictive being verified in the most part of practical cases of interest. In particular, this function is very useful to individuate failures in the structure like voids or delaminations that can be observed like shifts of the peaks [26–29]. This is due to the increase in thermal resistance caused by the defects in the structure. The peaks difference (the difference between the peak of a good structure and a defected one) represents the thermal resistance increase. Fig. 2.6 depicts the case of a three stacked chips separated by BCB; it is worth nothing that a bad attach, because of voids in the glue would result in a shift of the glue peak (indicated in the picture) toward the right side (caused by an increase in partial thermal resistance due to difficulties in heat flowing through that material). The *full spectrum* mode of the programmed software is used to evaluate the complete spectrum of time constants. Once it is evaluated, each line (not equal to zero) represents a time constant and its amplitude the

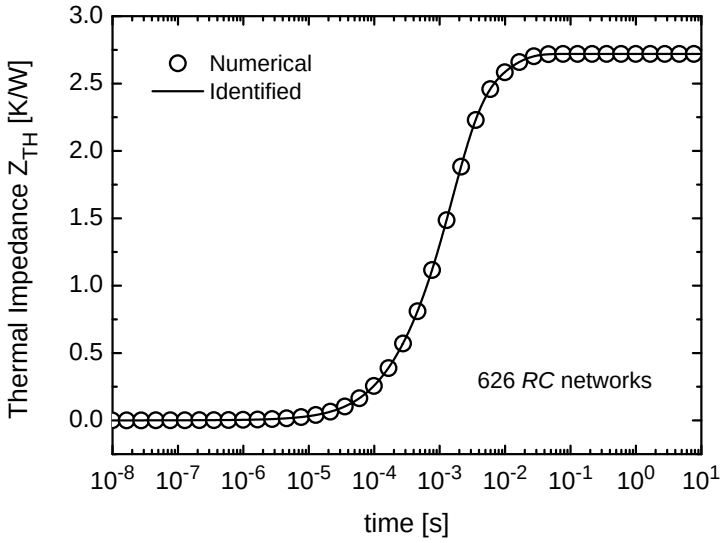


Figure 2.7: Numerical (symbols) and identified (solid line) thermal transient for a device fabricated with 3 stacked chips interleaved by BCB.

thermal resistance. Figs. 2.7 and 2.8 represent the transient data identified and the spectrum of time constants, respectively, evaluated by the software when operating in *full spectrum mode*.
In general, a large number of time constants is obtained by this method (the identification reported in Figs. 2.7 and 2.8 is obtained with more than 600 time constants) and then the Foster model must be transformed in Cauer model. This step can give rise to numerical problems, in fact, the evaluation of rational polynomial transfer function of Foster type

$$Z_{TH}(s) = \sum_{i=1}^{n} \frac{R_{TH_i}}{1 + s \cdot \tau_i}, \tag{2.18}$$

prescribe the multiplication of n monomials. Owing to the small values of τ_i as the number of monomials exceeds 20-30 the *double* precision is not sufficient to correctly evaluate all the polynomial terms. In order to overcome these limitations, the software for Foster to Cauer transformation has been extended

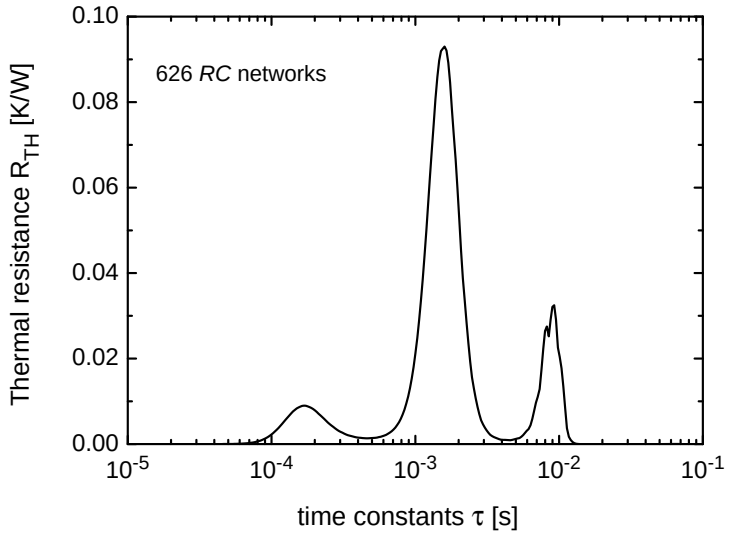


Figure 2.8: Time constants spectrum for a device fabricated with 3 stacked chips interleaved by BCB.

using an arbitrary precision mathematical library[4] so as to evaluate the Cauer model (with a very high accuracy) and, subsequently, the structure functions needed. Because of the extended mathematical precision the final results are written over files that can be read with standard methods and results can be re-converted (if necessary) to normal precision.

[4]The software makes use of the ARPREC (C++/Fortran-90 arbitrary precision package) library released under the BSD-LBNL license.

2.2 Compact models to enhance numerical analysis

The analysis of complex electronic systems can be simplified by resorting to compact thermal models (CTMs) as explained in [31–33]. The main approach for the development of CTMs is to partition the whole system into several sub-systems (parts), and then obtain a CTM for each single part. The connection between sub-systems is maintained through nodes, and the potential, defined for each node, is shared by all objects connected.
Obviously, under the action of the potentials there are flows (subject to conservation law) inward and outward the nodes connecting the objects. To complete the system description, a constitutive law per object relating the potentials of the object with the resulting flows is needed. These relations are directly connected to the first and second laws of thermodynamics used to derive constraints on compact models of each domain. Differently from detailed models where the potentials (temperatures) and fluxes (heat fluxes) are defined over each point of the boundary, in compact models the whole surface element S_i is associated to a the thermal node.
In general, it is possible to formulate linear relations between the (average) temperature and heat flux at the thermal nodes. These relations, typically, involve a certain number of parameters that can be simply obtained by means of numerical simulations or experimental characterization of each single part. As a result, it is possible to combine all the CTMs to study the behavior of the whole system.
The common approach in literature is to assume a uniform temperature distribution on the boundary surface elements S_i [31]. Nevertheless, the temperature field is rarely even and to obtain a good accuracy a fine discretization of the boundary surface (a high number of elementary S_i surfaces) is required. With a proper discretization it is possible to reproduce complex temperature fields or heat flux distributions on the boundary interface independently of the boundary condition. This means that changing a part of the system (i.e., adding a different package) or environmental condition does not affect CTMs accuracy. As suggested in [32, 33] this approach can be further improved by assuming a variable temperature or heat flux distribution over the boundary interface elements S_i. In particular, for example, assuming the temperature distribution at the thermal nodes to have a prescribed behavior governed by *shape functions* can effectively improve the accuracy of CTMs. The first step to obtain constitutive equations for CTMs is to find the solution of the heat

conduction equation

$$\nabla T^2 + \frac{g(r)}{k} = 0, \tag{2.19}$$

where k is the thermal conductivity (in general position dependent) and g is the power density [W/μm^3].
For the sake of simplicity, the following analysis considers a single heat source that is dissipating power; obviously the results obtained can be extended to the case of multiple power sources by superposition starting from the single source case. In order to develop a compact thermal model it is necessary to approximate the power density as

$$g(r) = P\frac{\sigma_g(r)}{V_g}, \tag{2.20}$$

where P is the dissipated power [W], V_g is the volume of the domain where the power is dissipated [μm^3], and $\sigma_g(r)$ is a suitable *shape function* describing the position dependence of the power density. This $\sigma_g(r)$ is defined to assume the value of 0 outside the active region of volume V_g. Owing to the fact that the volumetric integral of the power density must be equal to the total dissipated power P it is straightforward to obtain the relation that $\sigma_g(r)$ must satisfy

$$\int_{V_g} \sigma_g(r)dv = V_g. \tag{2.21}$$

It is assumed that this shape function is bias independent, that is, $\sigma_g(r)$ does not depend on the current and voltage (or dissipated power).
In some thermal studies, e.g. [31], a constant shape function is assumed (i.e., a uniform power density is considered). In this case, the shape function is simply equal to unity in the active region, and zero outside obviously verifying that $\int_{V_g} dv = V_g$.
Apart from power generation, it is compulsory to model the heat exchange between each thermal node (that is, each partitioned surface S_i) and the external environment. For example, the heat exchange between two connected sub-systems A and B occurs through the N surface elements S_i. From a mathematical point of view it is possible to define the temperature rise over ambient in a position r_i of the *i*-th partition of the surface S (indicated as S_i) that connects the subdomains

$$T|_{r_i \in S_i} = \phi_i(r_i), \tag{2.22}$$

where $\phi_i(r_i)$ is the prescribed temperature rise on S_i part. It is possible to assume that the temperature distribution on S_i can be expressed by a suitable

shape function $\sigma_i(r_i)$

$$T|_{r_i \in S_i} = T_i \cdot \sigma_i(r_i). \tag{2.23}$$

These shape functions must satisfy the condition

$$\int_{S_i} \sigma_i(r) ds = S_i, \tag{2.24}$$

whence it is possible to consider T_i as the *average temperature* on S_i as also evident by evaluating such a value

$$\frac{1}{S_i} \int_{S_i} T(r) ds = \frac{T_i}{S_i} \int_{S_i} \sigma_i(r) ds = T_i. \tag{2.25}$$

Equations for a sub-system are obtained by solving the heat equation

$$\nabla T^2 + \frac{P \cdot \sigma_g(r)}{V_g \cdot k} = 0, \tag{2.26}$$

subject to the boundary conditions described by (2.23) for all boundary surfaces S_i where each shape function σ_i is subject to the constraint (2.24). By using the Green's function method [34], the solution of the problem is given as

$$\begin{aligned} T(r) &= T_g(r) + \sum_{i=1}^{N} T_{BC_i}(r) = \\ &= \frac{P}{V_g \cdot k} \int_{V_g} \sigma_g(r') G(r, r') dv' \\ &- \sum_{i=1}^{N} T_i \int_{S_i} \sigma_i(r') \frac{\partial G(r, r')}{\partial n_i'} ds_i', \end{aligned} \tag{2.27}$$

where $G(r, r')$ is the steady Green's function (GF) for the problem at hand [34], and n_i is the outwardly pointing unit normal. The term T_g identifies the solution of the (2.26) accounting for power generation under homogeneous boundary conditions ($P \neq 0$ and $T_i = 0$). The term T_{BC_i} is the solution with no dissipated power ($P = 0$), a nonzero excitation at the *i*-th boundary surface S_i ($T_i \neq 0$), and zero excitation at all remaining boundaries ($T_j = 0$ $\forall j \neq i$). Therefore, all terms T_g and T_{BC_i} can be easily determined from $N + 1$ numerical simulations or experiments, where only one forcing term is activated; conveniently, sometimes symmetry can be exploited to reduce the number of simulations/experiments.

To summarize, it can be stated that a compact thermal model provides two sets of equations:

- An equation that relates the *junction* temperature, indicated by T_0 in the following part of this work, and the forcing terms P and T_i. This *junction* temperature is the average temperature at some location of interest (i.e, the base-emitter junction for bipolar transistors).
- The equations at the nodes of CTM relating the power flowing through the boundary surfaces S_i and the forcing terms P and T_i.

Moreover, some constraints deriving directly from the heat flux conservation (as formerly stated) or by applying the properties of the Green's functions can be obtained. The equations and constraints are derived below.

Junction temperature equation. The final goal of the thermal compact model is, undoubtedly, to allow the evaluation of the temperature in a specific region of interest indicated as V_0. This temperature, as already pointed out, is sometimes referred to as the *junction* temperature and can be defined as the average temperature calculated in a volume V_0

$$T_0 = \frac{1}{V_0} \int_{V_0} T(r) dv. \tag{2.28}$$

Substituting the solution (2.27) into (2.28) it is possible to obtain

$$T_0 = R_{TH_0} \cdot P + \sum_{i=1}^{N} a_i \cdot T_i, \tag{2.29}$$

where R_{TH_0} and coefficients a_i can be expressed, according to (2.27), in terms of Green's functions. It is worth nothing that R_{TH_0} represents the thermal resistance related to the sole heat generation.

Boundary equations. It is possible to express the heat flux flowing out from the *j*-th node (that is, the boundary surface S_j) as

$$P_j = -k \int_{S_j} \nabla T(r_j) \cdot n_j ds_j, \tag{2.30}$$

where n_j is the outwardly pointing unit normal and k the thermal conductivity. By using (2.27), new relations are obtained

$$P_j = q_j \cdot P + \sum_{i=1}^{N} \frac{T_i}{R_{ji}}, \tag{2.31}$$

for $j = 1, \ldots, N$. These relations include N non-dimensional parameters q_j and N^2 parameters R_{ji} representing the thermal resistances between the thermal nodes. These parameters can be expressed, according to (2.27), in terms of Green's functions as well.

Constraint relations. Coefficients in (2.31) are subject to some constraints imposed by heat conservation. The conservation of heat flux implies that the power generated by the heat source equals the power leaving the body. This yields, according to [31]

$$\sum_{j=1}^{N} q_j = 1 \quad \sum_{i=1}^{N} \frac{1}{R_{ji}} = 0. \tag{2.32}$$

It is possible to find other constraints for parameters reported in (2.29) and (2.31) assuming uniform boundary conditions [31], that is, $\sigma_i(r_i) = 1$,

$$\sum_{i=1}^{N} a_i = 1 \quad R_{ji} = R_{ij} \tag{2.33}$$

, for $j = 1, \ldots, N$. By applying these relations, the number of simulations/experiments necessary to generate the parameters can be drastically reduced. However, when dealing with non-uniform shape functions, relations (2.33) do not hold any longer.
This CTM formulation is used to simplify the case of the simulation of ultra-thin chip stacked devices discussed in Section 3.1.5.

2.3 Model for thermal resistance and capacitance scaling

Most of the existing formulations that describe the dependence of the thermal resistance upon the effective emitter dimensions (for bipolar transistors) are markedly inaccurate when applied to state-of-the-art bipolar transistors provided with isolation schemes. Nevertheless, (approximate) scalable models for thermal resistance and capacitance could be of extreme importance for the design of next generation devices/circuits.
Many advanced transistor models (like Mextram-504) include scaling rules for the electrical quantities that can allow the prediction (with a reasonable accuracy) of the device/circuits behavior due to scaling by means of simulations. Due to the fact that temperature strongly influences electrical characteristics of devices scalable models for thermal parameters (i.e., thermal resistance and capacitance) can be highly desired as well.
It was found that a scalable model recently proposed by Wu and coworkers [35] can be successfully employed for modern transistors if a careful, yet simple, parameter optimization is performed.
The model relies on the knowledge of the thermal resistance of a "reference" device $R_{TH_{ref}}$ with an emitter area $A_{E_{ref}} = W_{E_{ref}} \times L_{E_{ref}}$ and makes use of 3 non-dimensional fitting ("geometry") parameters connected to emitter area R_{TH_A}, emitter width R_{TH_W} and emitter length R_{TH_L} variation. The thermal resistance model is defined, according to [35], as

$$R_{TH}(L_E, W_E, A_E) = R_{TH_{ref}} \left[1 + R_{TH_A} \cdot \left(\frac{A_E}{A_{E_{ref}}} - 1 \right) + R_{TH_W} \cdot \left(\frac{W_E}{W_{E_{ref}}} - 1 \right) + R_{TH_L} \cdot \left(\frac{L_E}{L_{E_{ref}}} - 1 \right) \right]^{-1} , \quad (2.34)$$

where L_E, W_E and $A_E = W_E \times L_E$ are the length, width and area, respectively, of the emitter stripe.
In principle, apart from the value of a reference thermal resistance $R_{TH_{ref}}$, three sets of thermal resistance data are required to evaluate all parameters. In the latter case, from 2.34 it is possible to write a system of three equations in three unknowns and solving the system obtain R_{TH_A}, R_{TH_W} and R_{TH_L} values. However, a better procedure relies on the use of some additional sets of measurements to be used with an optimization procedure to gain a higher

accuracy.
On the basis of the thermal resistance model proposed in [35] a simple scalable model for thermal capacitance has been obtained as well

$$C_{TH}(L_E, W_E, A_E) = C_{TH_{ref}} \left[1 + C_{TH_A} \cdot \left(\frac{A_E}{A_{E_{ref}}} - 1 \right) + C_{TH_W} \cdot \left(\frac{W_E}{W_{E_{ref}}} - 1 \right) + C_{TH_L} \cdot \left(\frac{L_E}{L_{E_{ref}}} - 1 \right) \right] \quad , \tag{2.35}$$

where C_{TH_A}, C_{TH_W} and C_{TH_L} are non-dimensional scaling parameters connected to area, width and length of emitter, respectively, while $C_{TH_{ref}}$ is the reference thermal capacitance.
A software routine that performs the automatic optimization to evaluate fitting parameters has been developed as well.

Chapter 3

Numerical analysis of the thermal behavior of solid-state devices

The past decades have seen an unmatched advancement of semiconductor industry. In fact, the constant need to improve the performances of solid-state devices has driven the research toward the study and fabrication of faster devices/circuits. Nevertheless, the demand for reduced parasitics necessary to guarantee improved (with respect to previous generation devices) performances has forced researchers to adopt increasingly aggressive isolation schemes (i.e., silicon-on-glass technology [6]). As stated before, the major drawback of aggressive isolation schemes is connected to the fact that most of the good electrical insulators have the same insulating behavior from a thermal point of view that can deteriorate the thermal path and may impose a limit on the current density for high speed devices [4, 5].
Thermal problems can also arise because of the particular semiconductor material used. In fact in the past years there was a rapid ascent in the use of innovative materials like III-V compound semiconductors (like GaAs, InP, GaN etc...). These materials have been introduced mostly as a replacement for silicon because of the better performances from an electrical point of view. However, they often present worse thermal properties; GaAs, for example, has a thermal conductivity equal to a third of the silicon one.
Moreover, the continuous request of integrating even more features in a single chip and the consequent miniaturization of the single device (an aggressive scaling) contribute to the power density increase [4]. Nowadays, the power

density of a commercial processor is comparable to the one of a nuclear reactor as clearly evidenced in Fig. 1.2.
Hence, it is clear that the analysis of the thermal behavior of electronic devices and circuits is a necessary task to optimize performances and avoid undesirable effects. As explained in the following, 3-D finite elements simulations have been extensively used to analyze the thermal behavior of solid state devices and to suggest optimal design rules to mitigate thermal problems.

3.1 3-D FEM thermal simulations

Numerical thermal simulation makes it possible to evaluate the temperature distribution both under steady-state and transient conditions, identify critical heat flow paths, determine the effect of technology and layout parameters, and optimize the thermal architecture. In this section a detailed 3-D thermal simulation study based on the commercial finite-element-method (FEM) Comsol software package is presented. The impact of various technological parameters and material properties is investigated, as e.g., heat source area, thicknesses of silicon die, as well as the impacts of thermal physical parameters for various family of solid-state devices/circuits.

3.1.1 Common problems for the thermal numerical analysis of electronic devices

3-D FEM simulations are very useful to deeply investigate thermal behavior of solid state circuits and devices. Nevertheless, the attainment of an accurate simulation is not a trivial task because of the increase of CPU/memory requirements with degrees of freedom (influencing the accuracy of solution). For this reason it is necessary to optimize the design of the device structure within the simulation environment in order to attain an accurate solution within a reasonable time. Although a unique approach to simulate structures is not possible due to the large differences in technology used to fabricate circuits, some common methods can be employed, which are reported in the following.

Exploit symmetries. It is possible to reduce the number of elements used within a simulation, simply working on symmetry. In fact, the application of an adiabatic condition (i.e., zero heat flux) over the symmetry planes can help to reduce the number of elements as evident from Fig. 3.1(a) showing the top view of a SOG device with its symmetry

planes. In some cases it is possible to simulate only a quarter of the structure.

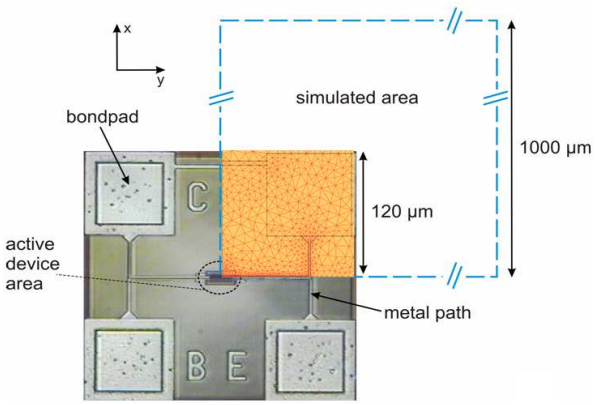
Reduce lateral dimensions. Active area in devices is very small if compared to the chip size or to a whole wafer. It is possible, after a preliminary study, to reduce lateral dimensions in order to further lower the number of elements containing the error within a ratio of 1%. In Fig. 3.1(a) it is shown that the area of 1000×1000 μm^2 is used for simulation and is enough to guarantee satisfactory results avoiding the simulation of the whole wafer having a diameter of approximately 20 cm as shown in Fig. 3.1(b).

Selective optimization of mesh. Even if the accuracy of solution is directly proportional to the number of elements used to mesh the structure it is almost impossible to use a very fine grid for all practical cases. In fact, the time needed to solve the problem grows exponentially with the number of mesh elements and the limited amount of memory can limit, even more, the possibility to obtain a solution in a reasonable time. An effective approach is to create a mesh as fine as possible in the zone where the higher gradient of temperature is expected and use an increasingly coarser mesh when moving away from the *hot zone* (i.e., the zone where the heat is generated). Moreover, the typical *sandwich structure* of many devices and the large differences between layers vertical and horizontal dimensions contributes to jeopardize this scenario[1].

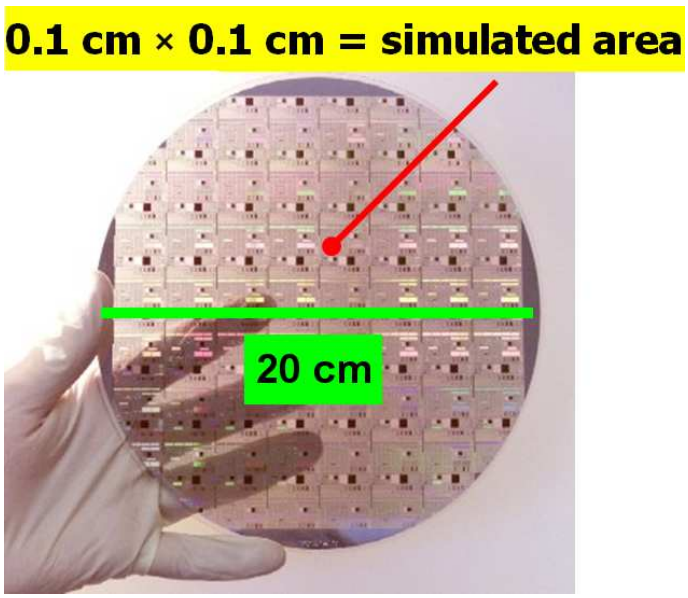
Accurate evaluation of physical parameters. The accuracy of solution is strictly connected to the goodness of thermal parameters. Unfortunately, novel device technologies introduced new materials where thermal parameters are pretty unknown due to fact that they are influenced by a multitude of factors (e.g., layer thickness, details of deposition technique, impurities, quality and shape of the crystal, and temperature). For this reason it could be necessary to perform a calibration of thermal parameters before running accurate simulations.

Simulations follow the general approach defined in the former section and Fig. 3.2 reports the schematic flow-chart of the semi-automatic Comsol-based software programmed *in house* to perform accurate simulations. The software, called *Autosolver* in the rest of this work, receives (as evident from Fig. 3.2)

[1]Layer thickness can be of the order of μm or nm while length and width can also be of the order of mm (see also Fig. 3.1(a)).



(a)



(b)

Figure 3.1: Top view (a) of a typical silicon-on-glass bipolar transistor with the simulated part compared to (b) a whole wafer. The portion of simulated device is enclosed by dashed blue lines. It is worth nothing that the active portion is 120×120 μm^2 whilst the simulated area is 1000×1000 μm^2.

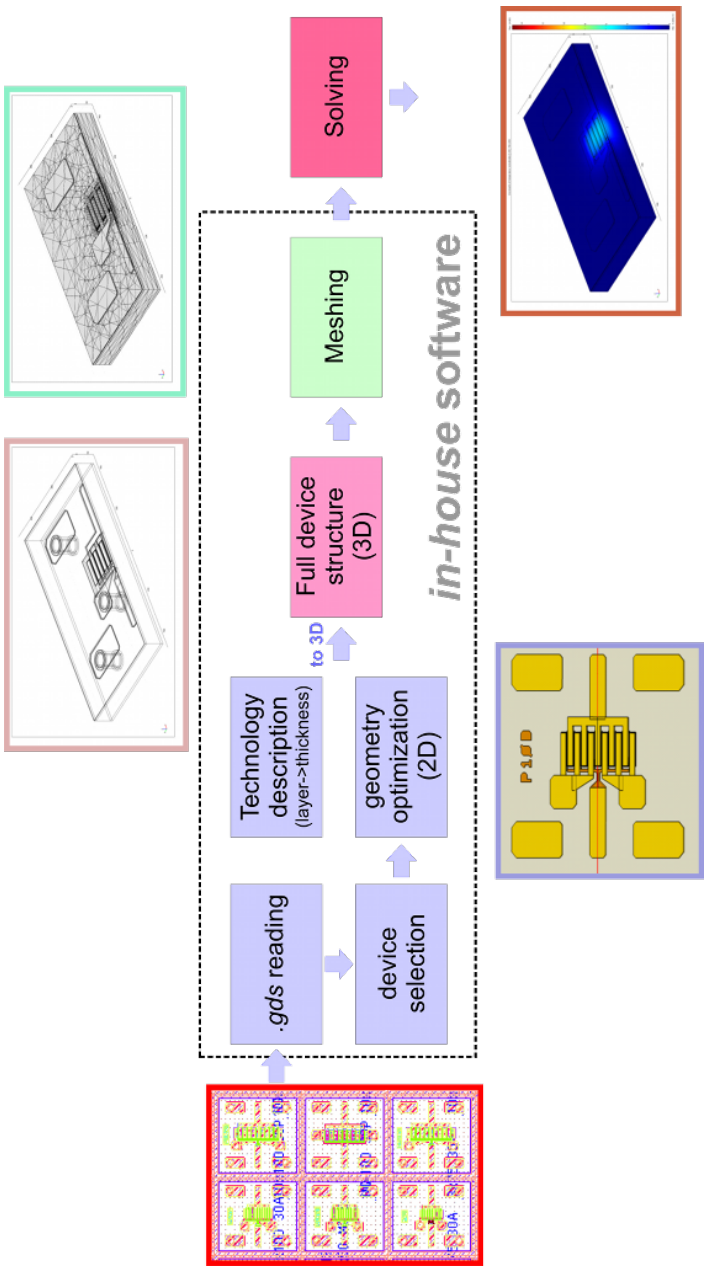


Figure 3.2: Flow-chart of the *Autosolver* software used to perform simulations for various technologies throughout this work. Steps performed by the software to simulate a GaN HEMT are reported as an example.

the layout of the device[2], a description of material layers and thermal parameters as inputs. By providing these inputs to *Autosolver* the software evolves as follows

- A 2-D structure on the basis of data contained in the layout is generated.
- The generated 2-D structure is further optimized by correcting geometrical errors (i.e., lines not connected etc...) and exploiting the possible symmetries (in order to reduce structure complexity).
- After these steps the software is able to automatically build a 3-D representation of the structure making use of the technology information (i.e., thickness of the layers, displacement etc...) and using 3-D transformation techniques (extrusion).
- Physical properties are applied to the problem to solve.
- *Autosolver* tries to generate an optimized mesh that can be further modified by the user (i.e., to enhance some details).
- The problem can, finally, be processed by the solver (a commercial FEM [36]) to obtain solution data that can be further postprocessed to obtain quantities (i.e., thermal resistance and impedance) and graphs (i.e., temperature distributions) of interest.

When some thermal parameters are unknown or not very accurate, an *in house* routine can be employed (its flow-chart is reported in Fig. 3.3) to accomplish the optimization of physical quantities starting from thermal resistance or impedance (i.e., obtained from measurements) for a reference structure[3].

[2]The software is able to automatically read a standard GDS file format and to extract geometrical data.

[3]Starting values of unknown thermal parameters can be obtained, in a first approximation, by using the literature values or applying the method described in Section 2.1.3.

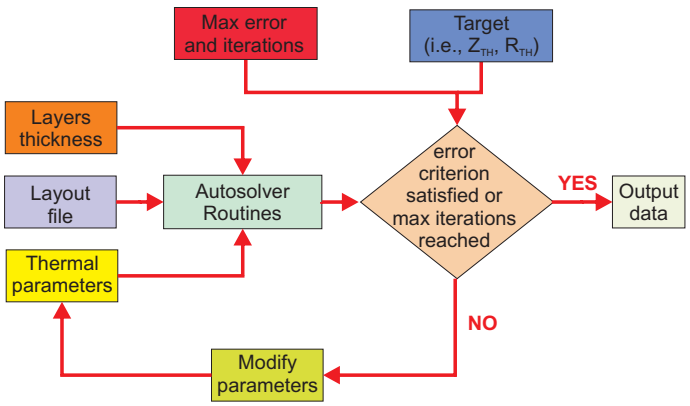


Figure 3.3: Calibration strategy used together with *Autosolver* routines in order to optimize thermal unknown parameters.

3.1.2 Silicon-on-glass technology

Silicon-on-glass technology can be considered the ultimate solution for reducing electrical parasitics [6]. In fact the lossy silicon substrate is replaced by a dielectric (i.e., glass) as evident from Fig. 3.4 depicting process steps.

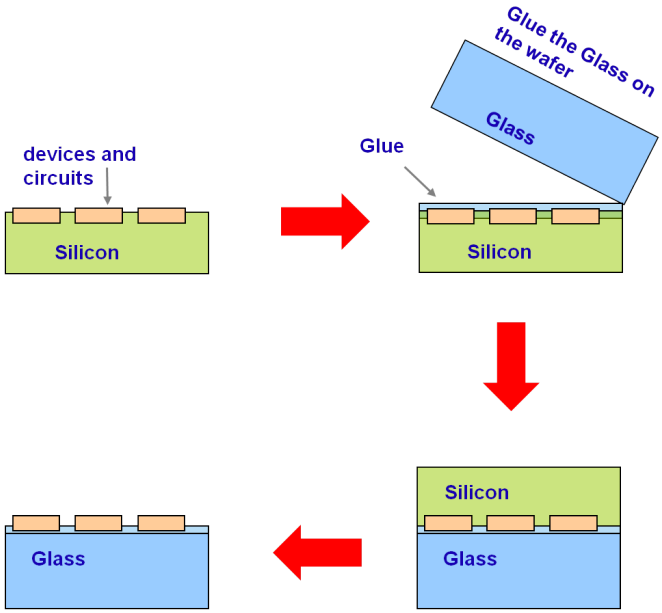
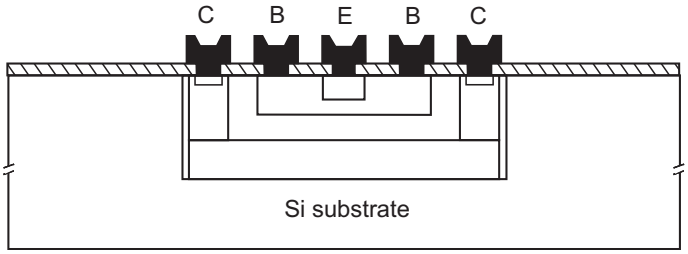


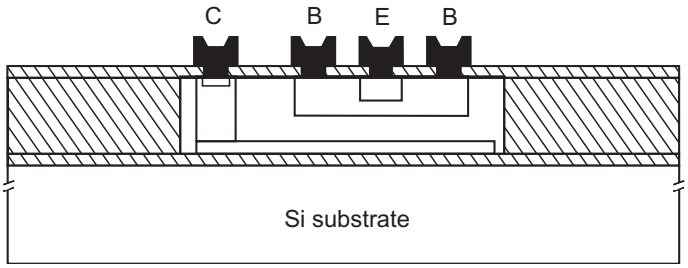
Figure 3.4: Schematic representation of main steps for the fabrication of SOG devices.

- The starting material to fabricate SOG devices is a SOI wafer, where devices are fabricated during the first step.
- After devices are fabricated, the glass substrate is glued onto the SOI substrate.
- The wafer is flipped and prepared to finally remove the excess silicon.
- Silicon is removed using the buried oxide (BOX) as etch stop.
- All devices are completely surrounded by insulating materials.

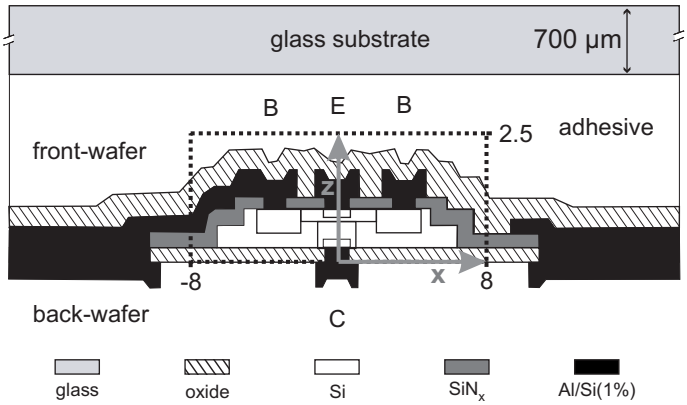
In Fig. 3.4 the double side contacting process steps are not reported in order to simplify the technology description. Nevertheless, a typical cross-section of a



(a) Bulk BJT $R_{TH} \approx 400\ K/W$



(b) SOI BJT $R_{TH} \approx 1200\ K/W$



(c) SOG BJT $R_{TH} \approx 16000\ K/W$

Figure 3.5: Schematic cross-sections of bipolar transistors with various isolation schemes: (a) conventional bulk-silicon technology, (b) trench-isolated device on a SOI substrate, and (c) SOG BJT. The numerically-evaluated thermal resistances refer to devices having the same emitter area of 1×20 µm^2.

SOG bipolar transistor is reported in Fig. 3.5(c) where the complete isolation of device is even more apparent. Fig. 3.5 shows various isolation schemes for silicon bipolar transistors, from bulk to silicon-on-glass. It is worth noting that in the latter situation (Fig. 3.5(c)) the conflict between electrical and thermal performance is pushed to the extreme. In this case bipolar transistors are characterized by a thermal resistance that can be two orders of magnitude higher than the one of comparable bulk-silicon transistors (Fig. 3.5(a)).
For this reason an accurate thermal analysis must be carried out in order to evaluate the impact of main design parameters on the static and dynamic thermal behavior of devices.

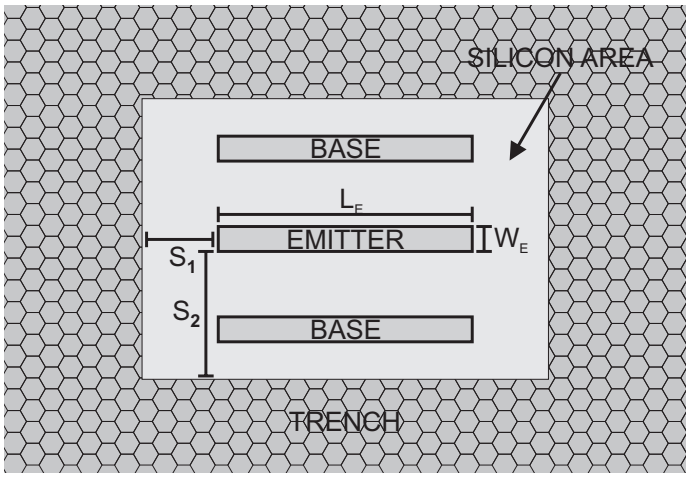


Figure 3.6: Representation of main layout parameters involved in the project of SOG bipolar transistors.

The main design parameters for a fully isolated bipolar transistor (single device) fabricated in this technology are reported in Fig. 3.6. They are the emitter length L_E and width W_E, as well as the distance of the emitter from the trenches that defines, someway, the silicon island surface. Such distances are identified as S_1 and S_2 in Fig. 3.6; other quantities of interest are the emitter area $A_E = L_E \times W_E$ and the aspect ratio $AR = {}^{L_E}/_{W_E}$.
Another important factor in the evaluation of the behavior of these devices is the possibility of the addition of aluminum nitride (AlN) layers that can be used to “beat the heat”. As depicted in Fig. 3.7, AlN layers can be integrated

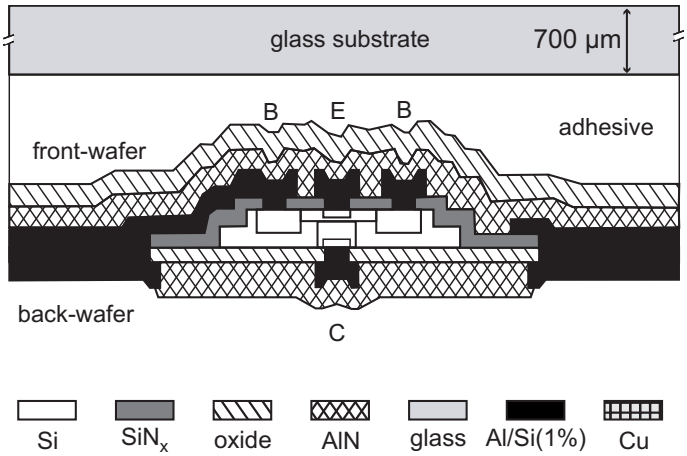


Figure 3.7: Cross section of a typical SOG bipolar transistors with AlN heatspreaders integrated during both front- and back-wafer processing.

during both the front- and back-wafer processing, and such films [14] can be an effective way of reducing the thermal issues of devices. More in general, thin films of high thermally conductive materials can be integrated in different technologies in order to reduce thermal issues. The deposition process of AlN µm-thick layers to be used as heatspreaders must overcome some difficulties due to the piezoelectric effects of the material and the tensile stress. Full details of AlN layers deposition are reported in [14, 37] but it is worth noting that AlN is fully compatible with silicon technology, it is neither contaminating nor poisonous (in contrast to other materials like, e.g., beryllia), and it can be deposited and etched by means already available in conventional silicon technology. Geometrical details of the simulated single-finger devices (corresponding to fabricated test structures) are reported in Table 3.1.

Calibration procedure. Not all thermal parameters for the materials involved in the structure simulation are known. For this reason the calibration procedure (as reported in the former section) was performed in order to evaluate the unknown thermal parameters. First, a reference structure without heatspreaders was numerically simulated by tuning (as reported in Fig. 3.3) the specific heats of silicon nitride, glass, and glue so as to guarantee the best fit with the experimental thermal impedance over the whole time range consid-

Table 3.1: Single finger device geometry details.

Device	**Emitter area** $A_E = W_E \times L_E$ **[µm × µm]**	**Aspect ratio** $AR = L_E/W_E$ **[µm / µm]**	**Emitter-to-trench distances**	
			S_1 **[µm]**	S_2 **[µm]**
devA	1×2.5, 2×5, 3×7.5	2.5	2.5	5.5
	1×5, 2×10, 3×15	5		
	1×10, 2×20, 3×30	10		
	1×20, 2×40, 3×60	20		
	1×30, 2×60, 3×90	30		
	3×6.6	2.2		
	4×5	1.25		
devB	1×2.5, 2×5, 3×7.5	2.5	2.5	10
	1×5, 2×10, 3×15	5		
	1×10, 2×20, 3×30	10		
	1×20, 2×40, 3×60	20		
	1×30, 2×60, 3×90	30		
	3×6.6	2.2		
	4×5	1.25		
devC	1×2.5, 2×5, 3×7.5	2.5	4	14
	1×5, 2×10, 3×15	5		
	1×10, 2×20, 3×30	10		
	1×20, 2×40, 3×60	20		
	1×30, 2×60, 3×90	30		
	3×6.6	2.2		
	4×5	1.25		
devD	1×2.5, 2×5, 3×7.5	2.5	6	14
	1×5, 2×10, 3×15	5		
	1×10, 2×20, 3×30	10		
	1×20, 2×40, 3×60	20		
	1×30, 2×60, 3×90	30		
	3×6.6	2.2		
	4×5	1.25		

ered. By converse, all other parameters are taken from the literature. In particular, the thermal conductivities of SiO_2 and SiN_x are set as in [38] and [39], respectively, and the mass density of the glass category adopted for the substrate transfer was taken from the datasheet. After these parameters are exctracted with a good approximation, a SOG device provided with a 2-µm-thick polycrystalline AlN layer on the front-wafer was simulated, and the thermal conductivity and specific heat of AlN were calibrated so as to favorably match the thermal impedance measured on the same AlN-cooled BJT. As far as the AlN thermal conductivity is concerned, the best fitting for the thermal resistance was achieved with a value of 17 W/mK; all optimized parameters are reported in Table 3.2.

Table 3.2: Thermal parameters for SOG technology.

Thermal Material	**Specific Conductivity [W/mK]**	**Mass Heat [J/gK]**	**Density [g/cm^3]**
Silicon (Si)	140	0.71	2.33
Silicon Nitride (SiN_x)	0.65	4	3.1
Silicon Oxide (SiO_2)	0.65	1.4	2.3
Aluminum (Al)	200	0.9	2.7
Aluminum Nitride (AlN)	17	0.63	3.3
Glue	0.55	0.85	1.5
Glass	0.55	0.01	2.43
Copper (Cu)	400	0.385	8.9

The reference structure used to accomplish the simulation tasks is reported in the Fig. 3.8 where the inherent symmetries of the device are exploited in order to reduce the problem (as also depicted in Fig. 3.1(a), only a quarter of the device is simulated). As an enhancement the base and emitter lines – which lie on the front-wafer in the actual (i.e., fabricated) devices – were more accurately defined within the environment of the simulator as clearly evidenced in the zoom-in of Fig. 3.8. A metal bondpad common to both emitter and collector metal tracks is considered not to jeopardize the structure symmetry; such an approximation was suitably found not to lead to loss of accuracy from the thermal standpoint. Even if such an enhancement, in the structure to be simulated, was shown not to significantly contribute to the evaluation of the (steady-state)

thermal resistance, it was revealed to be indispensable to correctly predict the thermal transient behavior in devices with an AlN layer deposited on the front-wafer. In this case, indeed, the emitter and base metal tracks represent a fundamental path for the heat to be quickly transferred from the active region to the AlN heatspreaders (by-passing the insulating SiN_x layer, see also Figs. 3.7 and 3.8).

Owing to the fact that the bondpads, during the performed measurements, are connected to measurement equipment and, to some extent, the heat can flow through the probes another boundary condition is fixed. In order to emulate this situation, a boundary condition (b.c.) of the 3rd kind (also known as Robin's or convective condition) was accounted for over the surface of the bondpads when numerically simulating the domains under analysis. The Robin's condition relates the outgoing heat flux to the temperature distribution over the bondpad through a parameter called *heat transfer coefficient*, which is defined as [34]

$$h(x,y,z)|_S = \frac{f(x,y,z)|_S}{T(x,y,z)|_S - T_{AMB}}, \tag{3.1}$$

where $f(x,y,z)|_S$ is the heat flux, $T(x,y,z)|_S$ is the temperature over a given surface S, and T_{AMB} is the ambient temperature. For SOG transient simulations, this parameter was set at a constant value of 5 W/K·cm^2. Moreover, nonlinear thermal effects (i.e., the dependence on temperature of the thermal conductivities) were disregarded for SOG devices without appreciable loss of accuracy in simulations. As can be clearly seen in Fig. 3.9, the simulated and measured values of thermal impedance versus time exhibit a good accuracy for structures with and without a 2-μm-thick AlN heatspreading layer. Further details on experimental characterization of silicon-on-glass bipolar transistors are provided in Chapter 4.

Impact of design parameters on thermal behavior. Figs. 3.10 and 3.11 show the influence of layout parameters on the thermal response of the simulated transistors. These figures report also the points (black dots) where the thermal impedance reaches 90% of the steady-state value (that is, the thermal resistance R_{TH}) being a good estimation of rise time t_R (see Chapter 1 for its definition) since the time instant needed to reach 10% of the steady-state value is two orders of magnitude lower. The thermal resistance decreases with increasing emitter area, whereas the rise time increases with A_E for devices either with or without AlN heatspreaders (as clearly shown in Fig. 3.10). A reduction of 35% is obtained, for layout devA in situations with or without

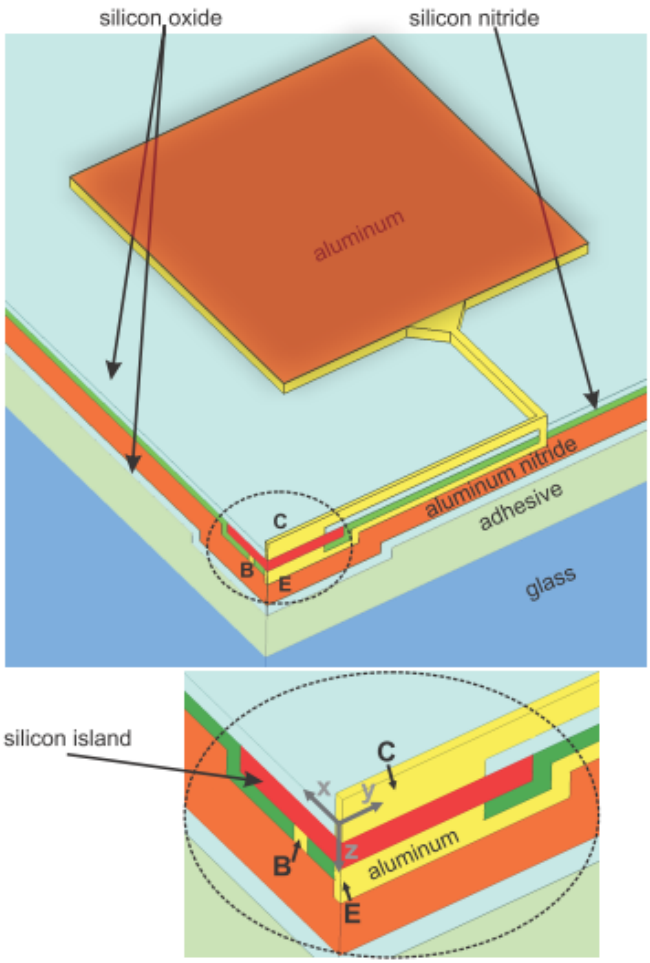


Figure 3.8: 3-D illustration (not to scale) of the structure of an SOG bipolar transistor used for thermal simulations and calibrations. A zoom-in of the silicon island evidencing the double side contact is shown as well.

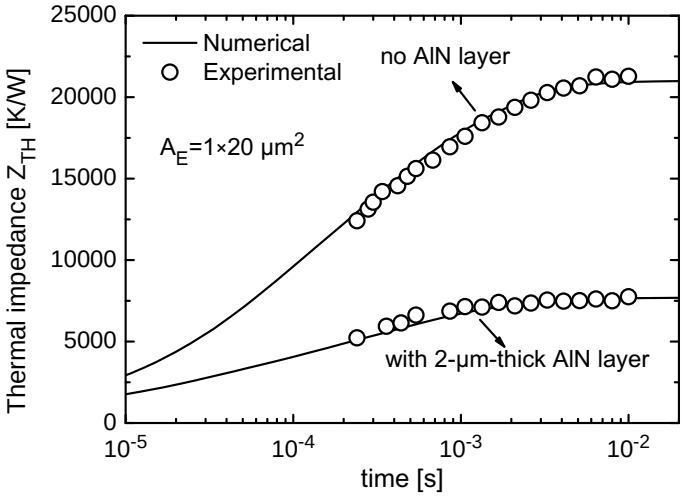


Figure 3.9: Measurements (symbols) and simulations (solid lines) both for devices with and without AlN heatspreaders. All data refer to geometry devA with an emitter area of 1×20 μm^2.

AlN heatspreaders, when comparing a structure with $A_E = 3 \times 60$ μm^2 to one with $A_E = 1 \times 20$ μm^2 (meaning that a device with a silicon area of 910 μm^2 is compared to one of 300 μm^2, respectively).

On the other hand, the rise time for the device with larger silicon island increases by 27%. Fig. 3.11 depicts the impact of aspect ratio variation for devices with and without a 2-μm-thick AlN layer on the thermal impedance. Like in the previous case the increase in AR leads to a growth of t_R and a decrease in the thermal resistance. A variation of AR from 1.25 to 20 produces a reduction in thermal resistance of 16% and a consequent increase of the rise time of 17% for devices both with and without AlN heatspreaders. In conclusion, if an increase in silicon area (obtained by either increasing the emitter area or the aspect ratio by keeping fixed the distance from the trenches) leads to lower R_{TH}, on the other hand, it causes an increment of t_R. This can also be observed when increasing the distance from the trenches, as evident from Fig. 3.12, where passing from layout devA to devD (increasing silicon area from 300 to 928 μm^2, respectively) there is a decrease in R_{TH} of almost 24% and an increase of 20% in rise time.

Thermal resistance decrease and t_R increase obtained by increasing the silicon island volume (by either increasing emitter area, aspect ratio or distance

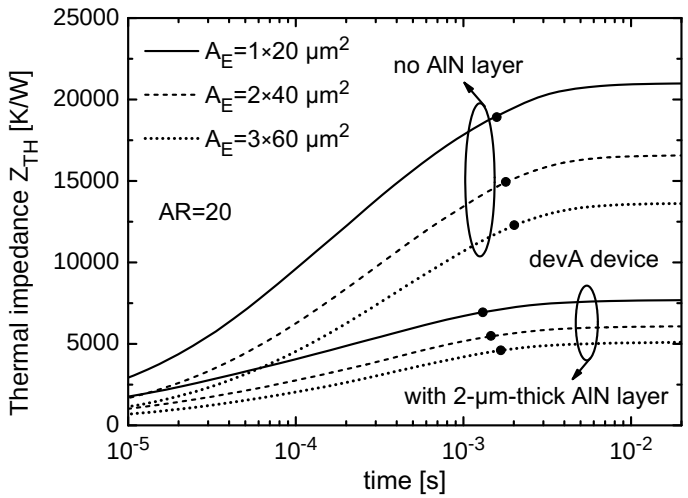


Figure 3.10: Simulated thermal impedance versus time for a devA layout by varying emitter area and keeping constant aspect ratio for devices without and with a 2-μm-thick AlN layer. The points where the thermal impedance reaches the 90% of the steady-state values are marked by black dots.

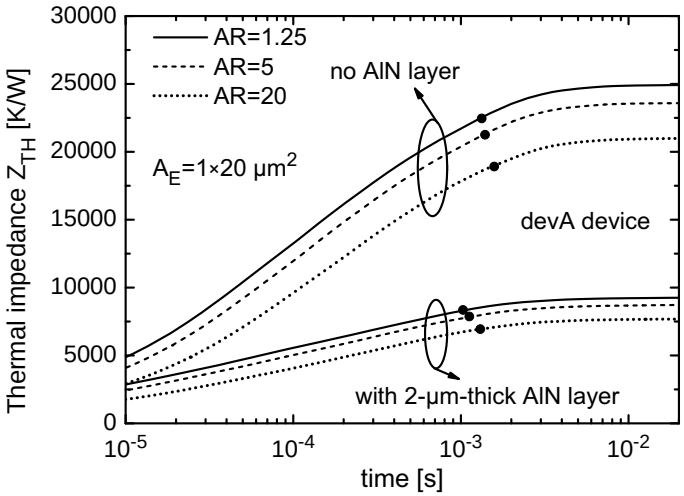


Figure 3.11: Simulated thermal impedance versus time for a devA layout by varying aspect ratio and keeping constant emitter area for devices without and with a 2-μm-thick AlN layer. The points where the thermal impedance reaches the 90% of the steady-state values are marked by black dots.

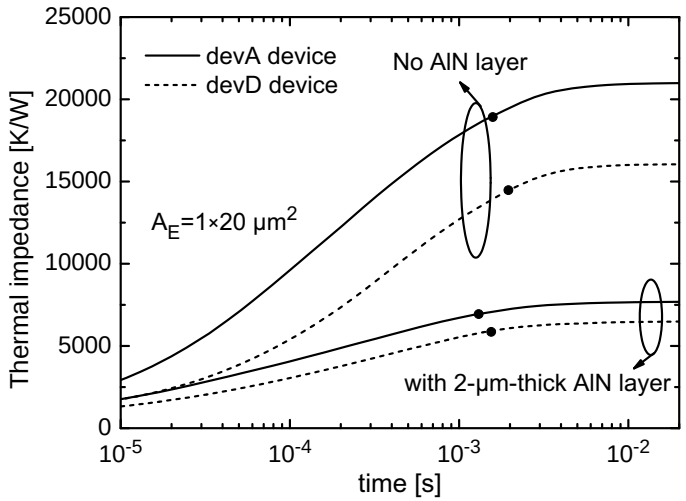


Figure 3.12: Simulated thermal impedance versus time for devices having a fixed emitter geometry by varying emitter-to-trench distance for devices without and with a 2-μm-thick AlN layer. The points where the thermal impedance reaches the 90% of the steady-state values are marked by black dots.

from the trenches independently from the addition of AlN) can be physically explained as follows. It is possible, to perform a first-order analysis, to assume that the dynamic electrothermal feedback can be described with a single-pole *RC* network. Due to the geometrical features of the typical SOG transistor, the thermal capacitance is much more sensible than the thermal resistance to variations of the silicon island size.

This is even more apparent in the Fig. 3.13 where an extensive numerical

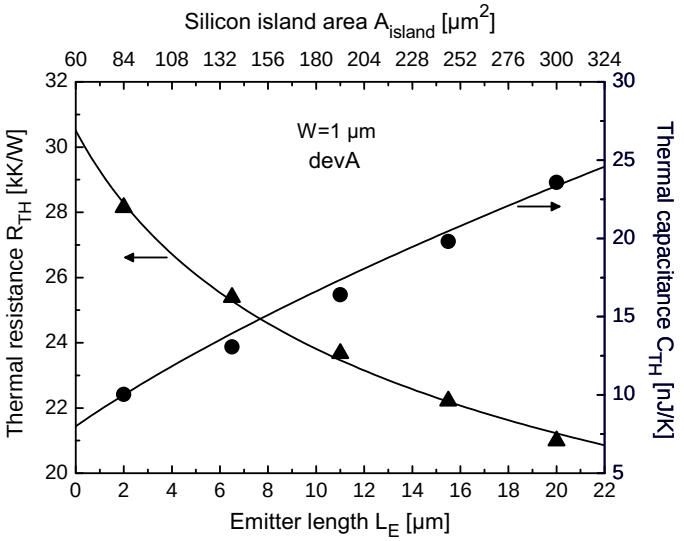


Figure 3.13: Thermal resistance and capacitance obtained by first order approximation versus emitter length for devices having a fixed width and devA layout. Simulations data (symbols) are reported together with (solid line) fitting curve. Silicon area values are reported as well.

analysis performed on devA-layout BJTs (not equipped with AlN layers) with $W_E = 1$ µm and by varying the emitter length L_E (keeping unchanged the distances between emitter edges and trench sidewalls) allows determining that $R_{TH} \propto A_{island}^{-0.23}$ whereas $C_{TH} \propto A_{island}^{+0.67}$. It should be underlined that the "sole" thermal capacitance was conventionally calculated as the C_{TH} value such as the quantity $(2\pi R_{TH} \cdot C_{TH})^{-1}$ equates the "real" thermal cut-off frequency resulting by the optimized multi-pole Foster network obtained by the software described in Chapter 2.

Fig. 3.14 shows that employing aluminum nitride layers beneficially influences the thermal behavior under both steady-state and transient conditions; the cooling action of AlN layers yields both a thermal resistance and rise time lowering. A 6-µm-thick AlN layer integrated on the front-wafer produces reductions amounting to 80% and 40% in R_{TH} and t_R, respectively. A physical expla-

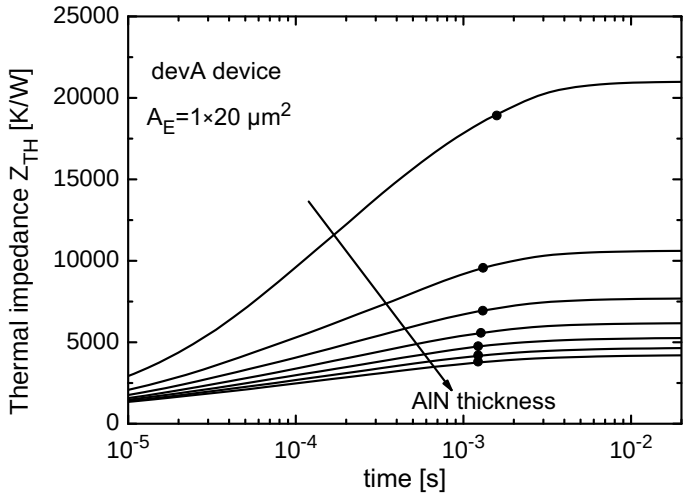


Figure 3.14: Simulated thermal impedance versus time for devices having a fixed emitter geometry and varying AlN layer thickness ranging from 0 to 6 µm. The points where the thermal impedance reaches 90% of the steady-state value are identified with filled symbols.

nation can descent directly from the inspection of the parameter values in the Table 3.2. This allows clarifying that the specific heat of the AlN film is lower than those corresponding with the other insulating materials surrounding the silicon island, while the thermal conductivity is much (proportionally) higher; as a result, the heat redistribution within the SOG structures is faster compared to AlN-free devices. An extensive number of calibrated 3-D thermal-only FEM simulations allowed to "monitor" that the AlN films effectively contribute – due to their relatively high thermal conductivity – to spread the heat from the power dissipation region toward both the pads and glued glass substrate, thereby reducing the base-emitter junction temperature as clear from Fig. 3.15.
The numerical tool can be also effectively exploited to detect temperature maps over specific device regions. This can be particularly helpful to

- Effortlessly explore the influence of alternative materials on the thermal device behavior.

- Quantify the thermal coupling between elementary transistors integrated in the same chip.

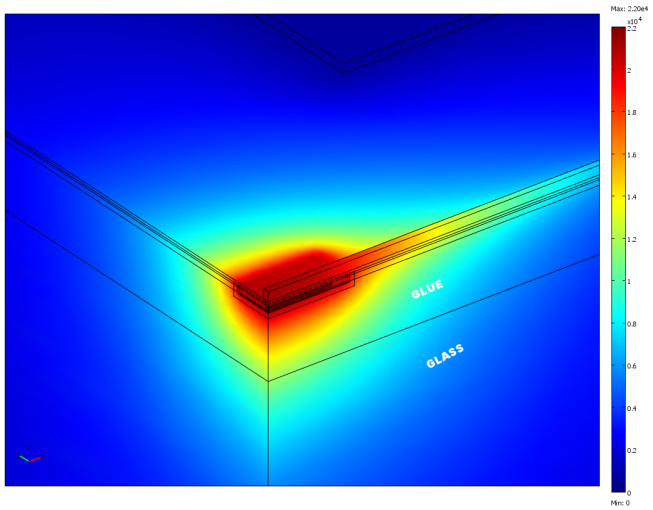
As an example, Fig. 3.16 reports the normalized temperature distributions within two SOG bipolar transistors when the steady-state condition is reached. It shows the simulated temperature increase over ambient normalized to dissipated power over the cross-section of a BJT without heatspreaders (enclosed within the dotted line in Fig. 3.5(c)) and with a 2-μm-thick AlN layer (the both with a 1×20 μm^2 emitter area and devA layout). An inspection of the pictures reveals that the temperature field

- Is fairly uniform within the silicon island (a kind of plateau can be discerned).

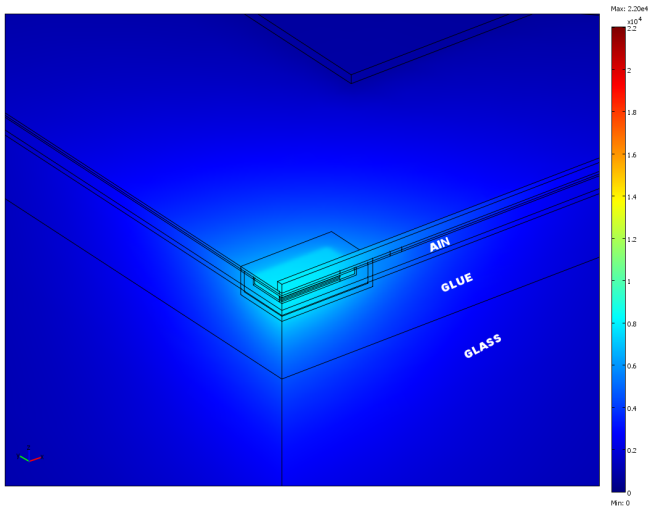
- Clearly benefits from the cooling action of emitter and base metal tracks (placed on the front-wafer) on the device without heatspreaders.

- Evidences the role of AlN in redistributing the heat generated within the thin island.

Moreover thanks to optimization software it is possible (for more details see Chapter 2) to report simulation results in frequency domain as showed in Fig. 3.17 where black dots identify the value of thermal impedance magnitude at the thermal cut-off frequency f_{TH}. It can be seen that f_{TH} decreases by enlarging the silicon island while increasing with AlN thickness. This can be also inferred from the rise time behavior since f_{TH} is inversely proportional to the rise time of the unitary step response. Comparing a device without AlN heatspreaders with one having a 2-μm-thick AlN layer, it could be possible to interpret the increase of f_{TH} as a disadvantage for small signal operation [40–42]. However, the very small value of thermal cut-off frequencies evaluated (around 1KHz) does not pose a critical threat for small signal operation of these devices; it should be considered indeed that the electrical cut-off frequency f_T of a SOG bipolar transistor with same geometrical parameters is several orders of magnitude (tens of GHz) higher than the thermal cut-off frequency (few KHz) as can be seen also in Section 4.2.2 containing experimental details.

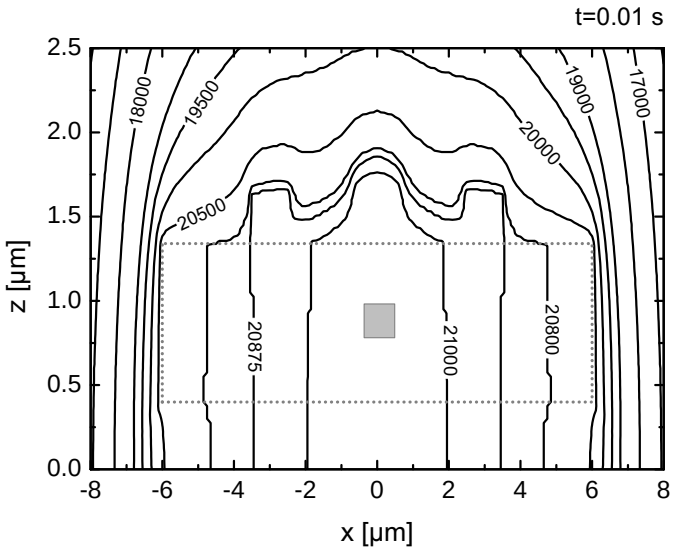


(a)

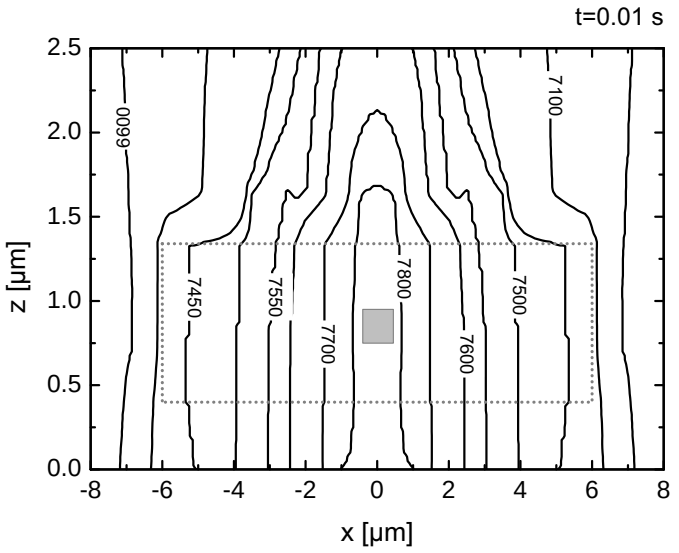


(b)

Figure 3.15: 3-D thermal simulations of bipolar transistors without (a) AlN heatspreader layer and with (b) a 2-µm-thick AlN layer for a SOG BJT having devA layout and $A_E = 1 \times 20$ µm^2.



(a)



(b)

Figure 3.16: Simulated contour plots of the temperature increase above ambient normalized to dissipated power for device devA with $A_E = 1 \times 20$ μm^2 (a) without and (b) with an AlN layer at one time instant, namely, 0.01 s, as evaluated over a section crossing the emitter stripe center. The dashed zone identifies the 12(×25)×0.94 silicon island whilst the small rectangle coincides with the heat source.

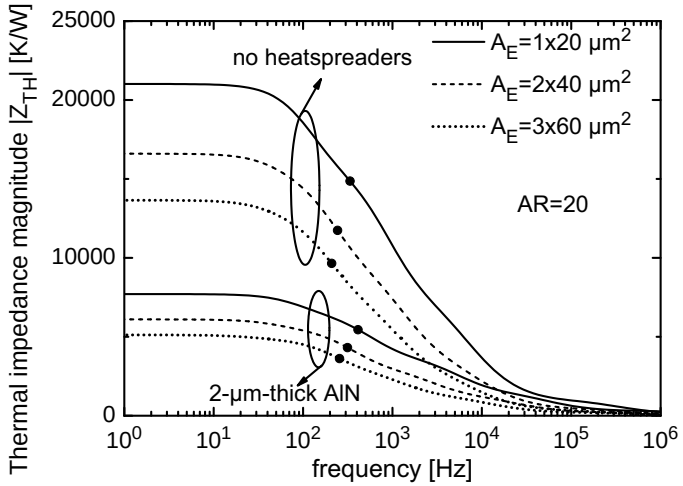


Figure 3.17: Thermal impedance magnitude as function of frequency for various emitter areas. All simulations refer to layout devA with and without a 2-µm-thick AlN layer on the front-wafer. Also identified are the magnitudes corresponding to the cut-off frequencies (black dots).

Multi-finger bipolar transistors. The calibrated FEM simulator was extensively used to evaluate the thermal behavior and thermal coupling of different layouts of multi-finger transistors. Moreover, it was employed to study the influence of emitter segmentation on the thermal behavior of silicon-on-glass transistors. In particular, several three- and four-finger layouts (depicted in Fig. 3.18 and Fig. 3.19, respectively) were simulated in order to provide a deep comprehension of the layout on the thermal coupling of devices. The values of self-heating ($R_{TH_{ii}}$) and mutual ($R_{TH_{ji}}$ with $j \neq i$) thermal resistances were extracted for all tested layouts. These values were used to perform accurate electrothermal simulations by means of electrothermal tools as explained in Chapter 5.

It is worth noting that in trench isolated 3-finger SOG transistors there is a non-uniform temperature distribution mostly due to the interconnection scheme. In fact it is clear in Table 3.4 that SOG devices of layout T32-HOR and T15-VER present a finger with a higher self-heating thermal resistance (#3 and #1, respectively). The addition of a 2-µm-thick AlN layer is beneficial for thermal resistance reduction and for an even temperature distribution. Dis-

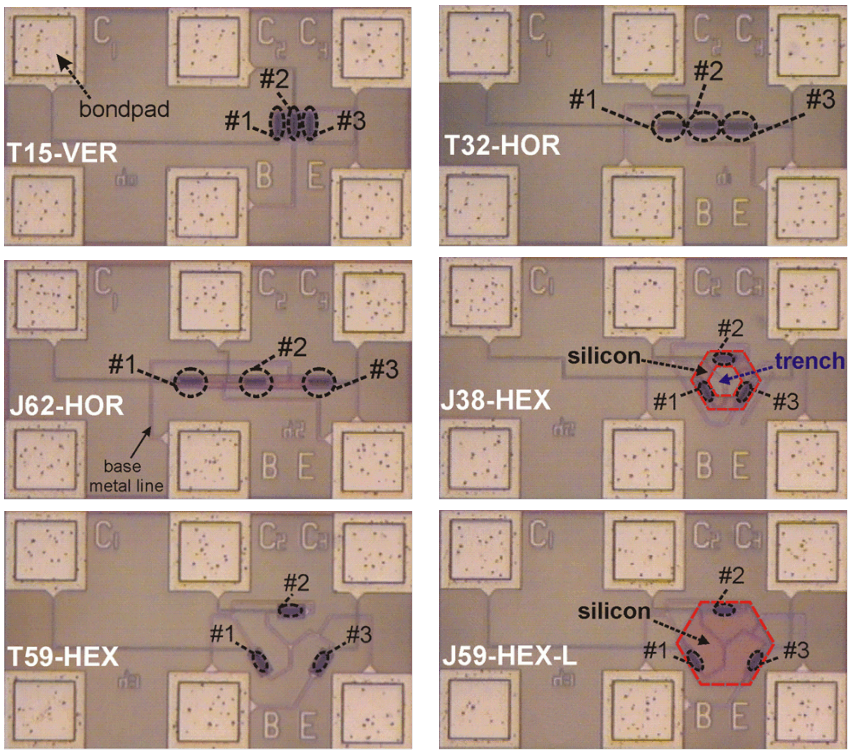


Figure 3.18: Microscope image of three-finger silicon-on-glass bipolar transistors having different topology and isolation schemes. Trench (left) isolation vertical (T15-VER) and trench (right) isolation horizontal (T32-HOR) three-finger bipolar transistors are shown in the two top images. Junction (left) isolation vertical (J62-VER) and junction (right) isolation hexagonal (J38-HEX) three-finger bipolar transistors are shown in the two middle images. Trench (left) isolation hexagonal (T59-HEX) and junction (right) hexagonal isolation (J59-HEX-L) three-finger bipolar transistors are shown in the two bottom images.

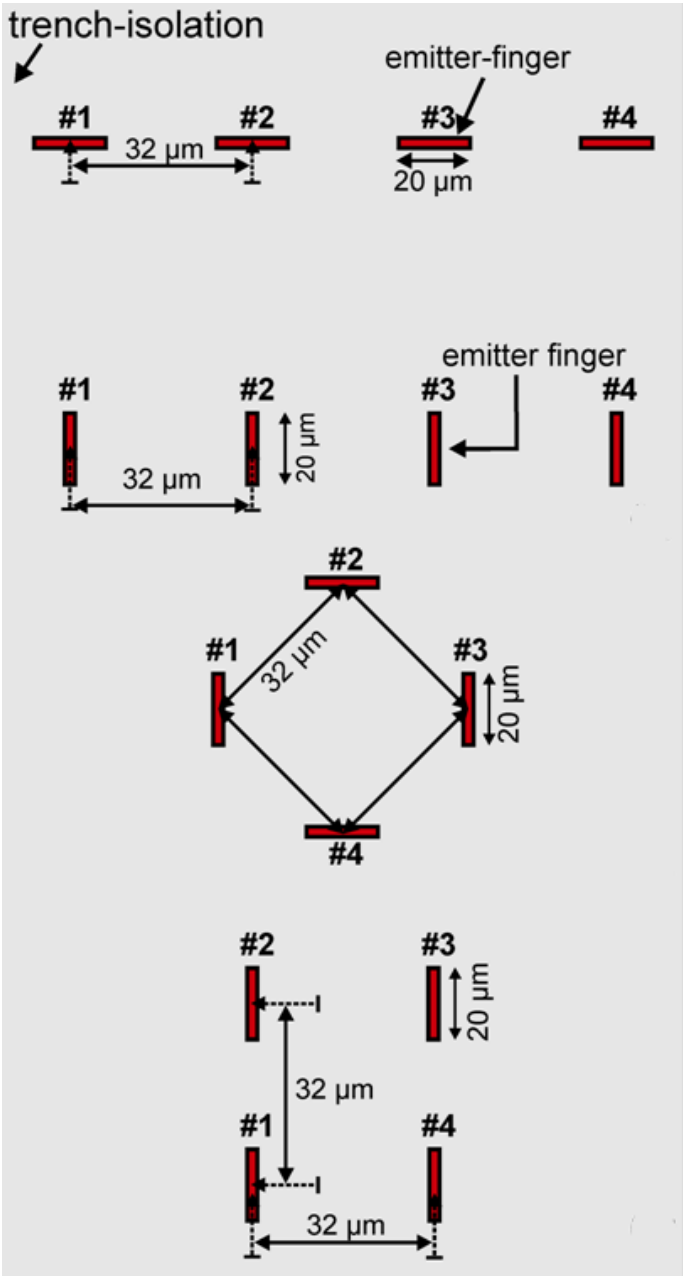


Figure 3.19: Schematic emitter layout of a four-finger silicon-on-glass device with different topologies, from top to bottom: horizontal, vertical, rhombus, and square configuration.

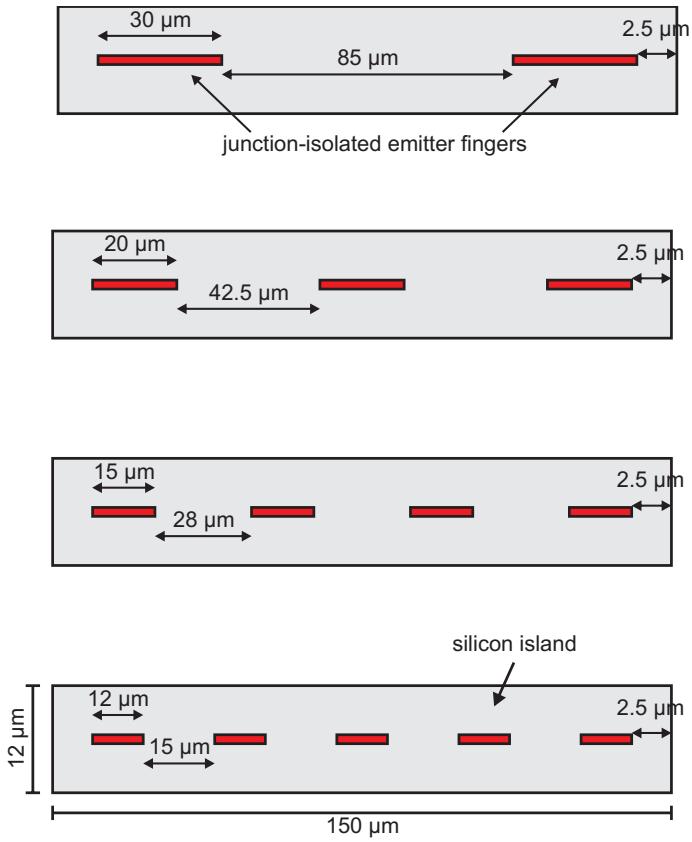


Figure 3.20: Schematic layout of emitter segmentation for a device having a total emitter stripe area $A_E = 1 \times 60$ µm^2. From top to bottom, respectively, are reported: two-, three-, four- and five-part segmented emitter.

Table 3.3: Device description of three-finger SOG bipolar transistors.

Device	Isolation	Si area [µm^2]	Distance between fingers [µm]	Topology	AlN heatspreader thickness [µm]
T32-HOR	Trench	3×(12×25)	32	Horizontal	0
T15-VER	Trench	3×(12×25)	15	Vertical	0
T32-HOR-2AlN	Trench	3×(12×25)	32	Horizontal	2
T15-VER-2AlN	Trench	3×(12×25)	15	Vertical	2
J62-HOR	Junction	1800	62	Horizontal	0
J38-HEX	Junction	1855	38	Hexagonal	0
T59-HEX	Trench	3×(12×25)	59	Hexagonal	0
J59-HEX	Junction	2873	59	Hexagonal	0
J59-HEX-L	Junction	5618	59	Hexagonal	0
T71-HEX	Trench	3×(12×25)	71	Hexagonal	0
J71-HEX	Junction	3209	71	Hexagonal	0
J71-HEX-L	Junction	7576	71	Hexagonal	0
T59-HEX-2AlN	Trench	3×(12×25)	59	Hexagonal	2
J59-HEX-2AlN	Junction	2873	59	Hexagonal	2
J59-HEX-L-2AlN	Junction	5618	59	Hexagonal	2
T71-HEX-2AlN	Trench	3×(12×25)	71	Hexagonal	2
J71-HEX-2AlN	Junction	3209	71	Hexagonal	2
J71-HEX-L-2AlN	Junction	7576	71	Hexagonal	2

Table 3.4: Thermal resistances for three-finger SOG bipolar transistors.

Device	$R_{TH_{11}}$	$R_{TH_{22}}$	$R_{TH_{33}}$	$R_{TH_{12}}$ $R_{TH_{21}}$	$R_{TH_{13}}$ $R_{TH_{31}}$	$R_{TH_{23}}$ $R_{TH_{32}}$
T32-HOR	20496	19332	21315	6744	3054	5619
T15-VER	18305	16532	17857	7902	4724	7716
T32-HOR-2AlN	6698	6701	6895	3382	2215	3363
T15-VER-2AlN	6593	6340	6482	4123	3172	4113
J62-HOR	15380	12697	15712	5642	2945	5784
J38-HEX	13563	13464	13635	8362	8444	8376
T59-HEX	22114	22882	22077	3317	2837	3057
J59-HEX	12656	13183	12782	5082	4937	5073
J59-HEX-L	9705	9855	9789	6024	5970	6038
T71-HEX	21867	22203	21490	2654	2416	2575
J71-HEX	12534	12533	12550	4061	4070	4096
J71-HEX-L	9319	9336	9327	5269	5253	5265
T59-HEX-2AlN	6804	7059	6822	2210	2116	2184
J59-HEX-2AlN	5737	5917	5753	2324	2260	2314
J59-HEX-L-2AlN	4999	5089	5014	2580	2538	2575
T71-HEX-2AlN	6753	6910	6738	1956	1896	1941
J71-HEX-2AlN	5885	5986	5891	2083	2049	2081
J71-HEX-L-2AlN	4968	5017	4956	2293	2268	2289

crepancies between elementary fingers still exist but the differences are very small so that they do not pose a threat for an even temperature distribution. A more symmetric design, like the hexagonal one, can warrant a more even temperature distribution as can be noticed by the self-heating and mutual thermal resistance values in Table 3.4. The junction-isolated devices present a lower self-heating thermal resistances and an higher mutual thermal resistance than the trench isolated counterparts (see for example T59-HEX and J59-HEX). The self-heating thermal resistance reduction can be easily explained by the increase of silicon volume while the mutual thermal resistance increase can be explained by the fact that elementary devices share the same silicon island (differently from trench-insulated devices) and the heat can favorably spread within the silicon island having an higher thermal conductivity with respect to neighboring materials. Large discrepancies in self-heating thermal resistance can trigger thermal instabilities to the point that the device with the larger R_{TH} is the one which takes the biggest amount of current as it will be clarified by electrothermal simulations in Chapter 5.

Very similarly, in the case of 4-finger layouts is possible to see how the horizonthal design presents the higher self-heating thermal resistance (in particular finger #4) by inspecting Table 3.5. All simulations evidence that a symmetric design leads to a more homogeneous distribution of temperature.

A possible way to enlarge the safe operating area of devices is to resort to emitter segmentation [43, 44], namely, the division of emitter stripe in several pieces as reported in picture Fig. 3.20. When adopting this strategy, the total current that can be driven remains unchanged while the thermal ruggedness is enhanced. In fact the temperature, owing to the increase of parts, becomes more uniform by increasing the number of devices since they are approaching (the center-to-center distance of elementary part is reducing), thereby leading to an increase in thermal resistance as can be seen in Table 3.6.

Table 3.5: Thermal resistances for four-finger SOG bipolar transistors with and without AlN heatspreaders.

Device	Topology	$R_{TH_{11}}$	$R_{TH_{22}}$	$R_{TH_{33}}$	$R_{TH_{44}}$
FOURA	Horizontal	21694	18619	19681	22237
FOURB	Vertical	21218	21550	21497	21185
FOURC	Rhombus	20782	20472	20770	20581
FOURD	Square	21102	21040	21029	21032
FOURA-4AlN	Horizontal	4878	4788	4869	4974
FOURB-4AlN	Vertical	5088	5108	5111	5087
FOURC-4AlN	Rhombus	5003	4982	5009	5003
FOURD-4AlN	Square	5066	5066	5067	5055

Device	Topology	$R_{TH_{12}}$ $R_{TH_{21}}$	$R_{TH_{13}}$ $R_{TH_{31}}$	$R_{TH_{14}}$ $R_{TH_{41}}$	$R_{TH_{23}}$ $R_{TH_{32}}$	$R_{TH_{24}}$ $R_{TH_{42}}$	$R_{TH_{34}}$ $R_{TH_{43}}$
FOURA	Horizontal	6847	3599	2059	6389	2910	5702
FOURB	Vertical	5196	2646	1792	5024	2621	5168
FOURC	Rhombus	5838	3784	5543	5589	3829	5779
FOURD	Square	4655	3539	5913	5908	3583	4669
FOURA-4AlN	Horizontal	2413	1734	1359	2407	1694	2418
FOURB-4AlN	Vertical	2469	1795	1439	2469	1796	2472
FOURC-4AlN	Rhombus	2526	2143	2516	2516	2144	2527
FOURD-4AlN	Square	2422	2134	2565	2565	2134	2422

Table 3.6: Thermal resistances for segmented emitter SOG bipolar transistors.

2 parts $\mathbf{R_{TH_{ij}}}$	**1**	**2**
1	4200	1250
2	1250	4200

3 parts $\mathbf{R_{TH_{ij}}}$	**1**	**2**	**3**
1	4700	1800	1200
2	1800	4100	1800
3	1200	1800	4700

4 parts $\mathbf{R_{TH_{ij}}}$	**1**	**2**	**3**	**4**
1	4950	2200	1500	1150
2	2200	4750	2100	1500
3	1500	2100	4750	2200
4	1150	1500	2200	4950

5 parts $\mathbf{R_{TH_{ij}}}$	**1**	**2**	**3**	**4**	**5**
1	5220	2470	1700	1350	1150
2	2470	4850	2300	1650	1350
3	1700	2300	4650	2300	1700
4	1350	1650	2300	4850	2470
5	1150	1350	1700	2470	5220

3.1.3 Simulation of GaAs HBTs

Heterojunction bipolar transistors (HBTs), differently from homojunction bipolar transistors, are fabricated by employing different semiconductor materials for emitter and base. When referring, e.g., to InGaP/GaAs HBTs the emitter is constituted by InGaP and the base (as well as collector, with different dopings, obviously) by gallium arsenide. GaAs HBTs introduce some advantages over BJTs

- High electron mobility with a consequent decrease in electron transit time in the base region and a growth of the cut-off frequency for transistor (f_T).
- Lower leakage currents if compared to silicon fabricated BJTs being directly proportional to intrinsic carrier concentration (GaAs has a 4 orders of magnitude lower intrinsic concentration with respect to Si).
- Low noise.

These advantages make GaAs very suitable to power microwave and military applications. Nowadays, GaAs HBTs are very diffused in power amplifiers for mobile communications. Unfortunately, GaAs HBTs present, apart the advantages exposed, a fundamental limit due to thermal issues. In fact, thermal conductivity of GaAs is one third of the one corresponding to silicon thus exacerbating the thermal issues so that the thermal management of devices must be accurately analyzed. A typical example of thermal issues in GaAs HBTs is the *collapse* of current gain. This phenomenon can be evinced [45–49] by the rapid decrease of forward gain while increasing V_{CE} as clearly evidenced in Fig. 3.21 (reporting the collapse of current gain for a three-finger GaAs HBT) posing a serious limitation for power amplifiers.

InGaAs/GaAs fabricated by Skyworks Inc. CA are numerically analyzed in order to evaluate the thermal behavior and to extract the thermal matrix (i.e., self-heating and mutual thermal resistances, see Section 1.1 for definitions) suitable to be used in electrothermal simulations. Table 3.7 reports the values of some electrical parameters for a "horse-shoe" shaped emitter having an emitter area of about 60 μm^2; each value is reported with the corresponding measurement condition. These devices feature a collector-base mesa and several thin layers forming the emitter mesa as evidenced in the SEM image reported in Fig. 3.22. It is worth noting the typical contact scheme "CEBEC" is used to reduce the base-collector capacitance.

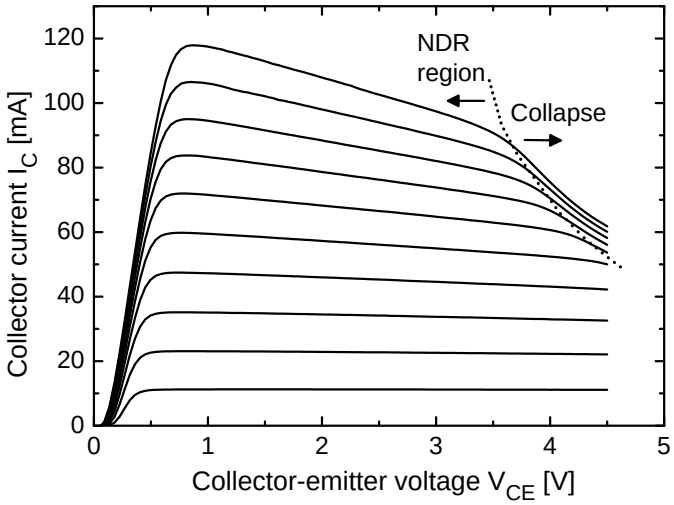


Figure 3.21: Collapse of current gain of a three-finger InGaP/GaAs. Also evidenced is the "stable" negative differential resistance (NDR) region.

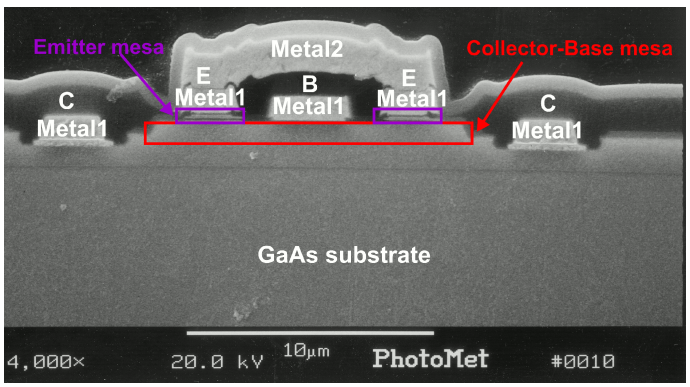


Figure 3.22: SEM image of a horse-shoe shaped InGaP/GaAs having an area of about 60 μm^2. Emitter and collector-base mesa are evidenced as well. Image courtesy of Skyworks CA Inc.

Table 3.7: Main electrical parameters of a "horse-shoe" shaped InGaP/GaAs, courtesy of Skyworks Inc. CA.

Parameter	**Measurement conditions**	**Value**
Current Gain (β)	$I_B = 0.1$ mA/μm^2	115
Base sheet resistance		225 Ω/□
BV_{EBO}	$J_E = 100$ nA/mm^2	9.3 V
BV_{CBO}	$I_C = 200$ nA	27 V
BV_{CEO}	$J_C = 0.5$ mA/mm^2	14.8 V
f_T	$I_B = 0.1$ mA/mm^2, $V_{CE} = 1.5$ V $V_{CE} = 1.5$ V	31 GHz
	$I_B = 0.25$ mA/mm^2, $V_{CE} = 1.5$ V	46.9 GHz
MAG	$I_B = 0.1$ mA/mm^2, $V_{CE} = 3.5$ V, $f = 0.9$ GHz	30.4 dB
	$I_B = 0.1$ mA/mm^2, $V_{CE} = 3.5$ V, $f = 1.9$ GHz	27.1 dB

Thermal simulation of InGaP/GaAs HBTs. As evident from Fig. 3.22 these HBTs have a complex 3-D structure, mainly due to the particular emitter shapes. Fig. 3.23 shows a 3-D illustration of the horse-shoe shaped emitter device without 2nd metal level to appreciate the complexity of the device; the emitter stack is composed by several thin stacked layer as evident by schematic cross section in Fig. 3.24.

The *Autosolver* software was successfully employed to simulate complex HBT structures both single- and three-finger devices with different geometries. Geometric details of single-finger devices are given in Table 3.8 while details of thermal conductivities are reported in Table 3.9.

Table 3.8: GaAs HBTs geometry.

Name	Emitter shape	Area [μm^2]
Q117	Rectangular	120
Q117R	Horse-shoe	120
Q56	Rectangular	59
Q56R	Horse-shoe	59
Q130	Rectangular	130
Q160	Rectangular	160

Table 3.9: Thermal conductivities for InGaP/GaAs HBTs.

Material	Thermal conductivity [W/mK]
GaAs	44.6
AlGaAs	11.0
InGaP	5.2
InGaAs	4.9
Ti	21.9
Au	270.0

It is possible to appreciate the level of simulation detail in Fig. 3.25 that shows both the simulated pads-equipped Q56R device and a microscope image of the same transistor; it is evident the small size of active device area (delimited by a green rectangle in microscopic image) compared to the complete struc-

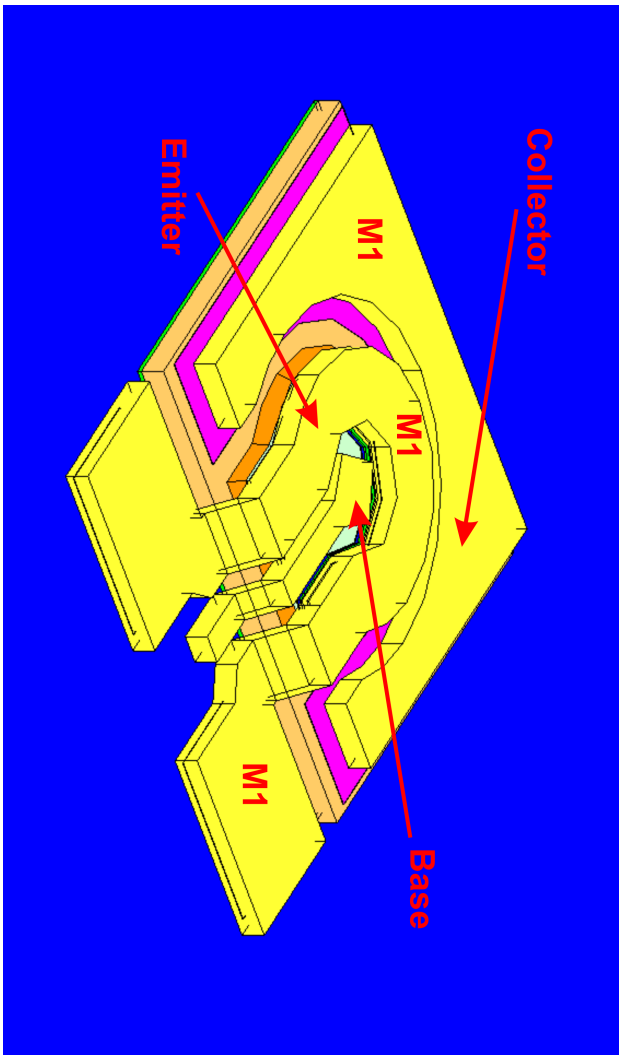


Figure 3.23: 3-D illustration of a horse-shaped InGaP/GaAs having an area of about 60 μm^2.

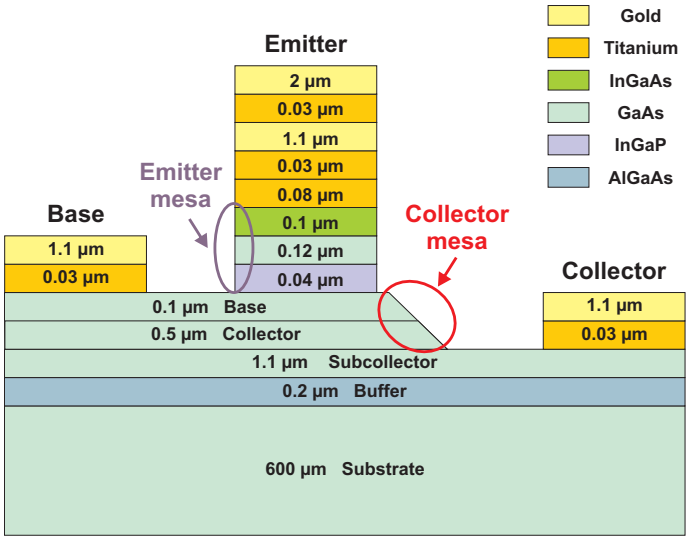


Figure 3.24: Illustration (not to scale) of the cross section of a InGaP/GaAs HBT. Emitter and collector-base mesa are evidenced as well.

ture. Thermal non-linearities are taken into account due to the fact that GaAs thermal conductivity presents a power law dependence on temperature (as previously shown, see Section 1.3 for further details). This results into an increase of thermal resistance due to temperature, and, hence, to dissipated power as reported in Fig. 3.26. When the power increases up to 20 mW, R_{TH} increases of 20% from the value obtained at low dissipated power (i.e., by deactivating thermal non-linearities). Fig. 3.27 depicts the thermal resistance dependence on emitter area. It is worth noting that round emitter shaped devices present a slower thermal resistance decrease with A_E if compared to rectangular emitter shaped ones. This can be obviously explained with the fact the the structure geometry strongly affects thermal behavior; it can be evinced that a doubling of emitter area (i.e., from about 60 µm^2 to 120 µm^2) turns in a decrease of 25% and 18% for square and horse-shoe shaped emitters, respectively. Thermal resistances extracted for both single-finger devices and three-finger structures are reported in Table 3.10 and Table 3.11 (only values related to one active finger are reported for sake of brevity), respectively.

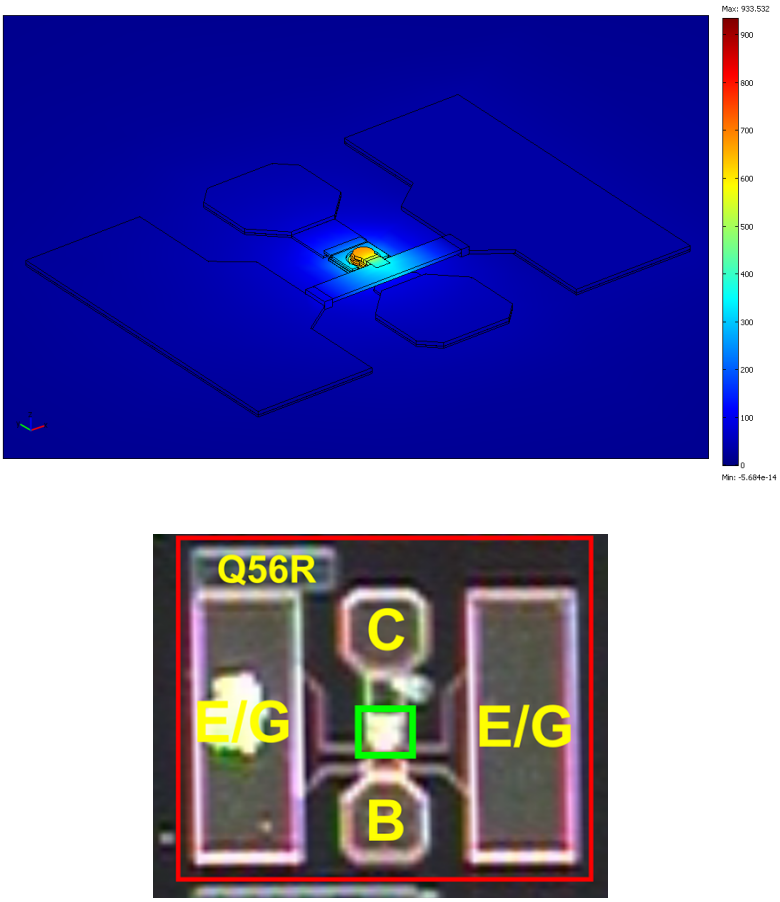


Figure 3.25: 3-D thermal simulation (top image) of a horse-shoe shaped InGaP/GaAs having an area of about 60 μm^2. A microscope image of the device is reported (bottom picture) as well. The small size of the active part of the device is evidenced.

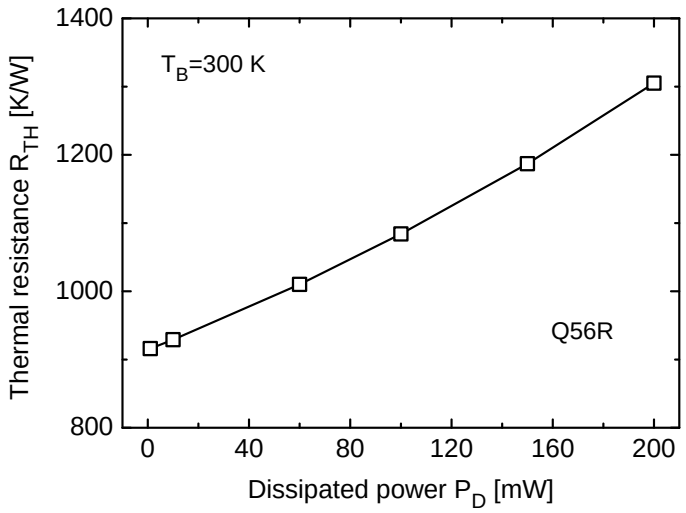


Figure 3.26: Thermal resistance dependence on dissipated power of a horse-shoe shaped InGaP/GaAs having an area of about 60 μm^2.

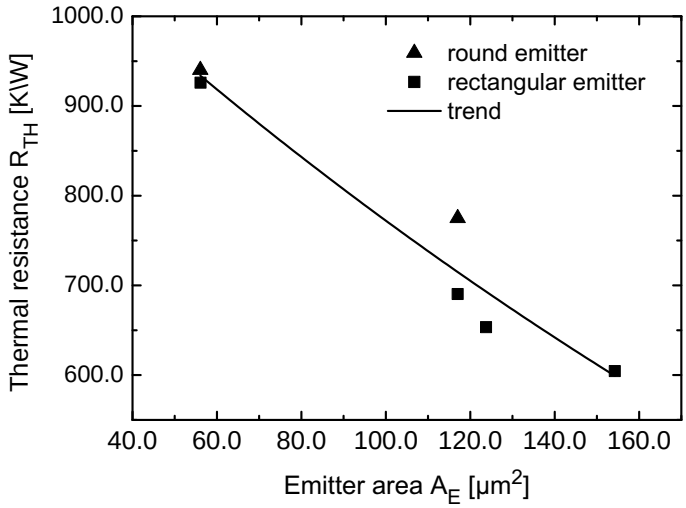


Figure 3.27: Thermal resistance dependence on emitter area A_E for different design of emitter for InGaP/GaAs devices. Horse-shoe (triangles) and rectangular (squares) shaped emitters data are showed.

Table 3.10: Self-heating thermal resistance extracted for various single InGaP/GaAs HBTs.

Device	**Thermal resistance [K/W]**
Q56	926
Q56R	940.2
Q117	690.4
Q117R	774.9
Q130	653.4
Q160	604.4

Table 3.11: Thermal resistance extracted for various three-finger InGaP/GaAs HBTs.

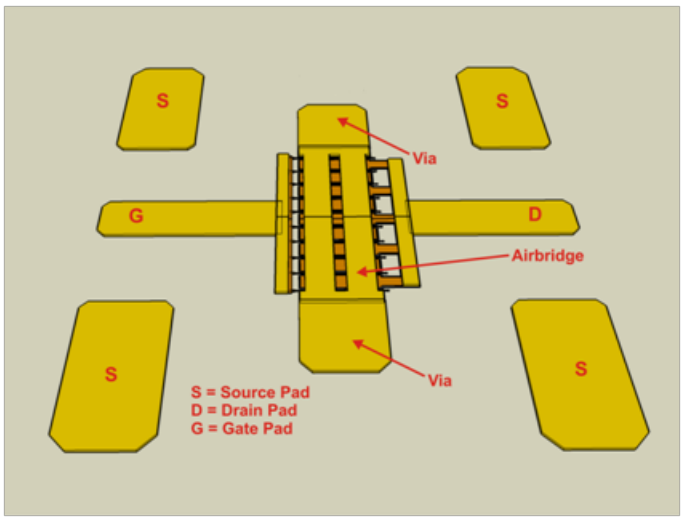
Geometry	**Number of devices**	**Thermal resistance [K/W]**	**Distance between centers [µm]**		
			24	**48**	**72**
Q56R	3	$R_{TH_{11}}$	983	1005.1	1020.7
		$R_{TH_{12}}$	161.2	76	49.5
		$R_{TH_{13}}$	78.7	37	24.4
Q117R	3	$R_{TH_{11}}$	769.6	787.6	783.8
		$R_{TH_{12}}$	160.8	75	49.9
		$R_{TH_{13}}$	79	37.2	24.4

3.1.4 Simulation of GaN HEMTs

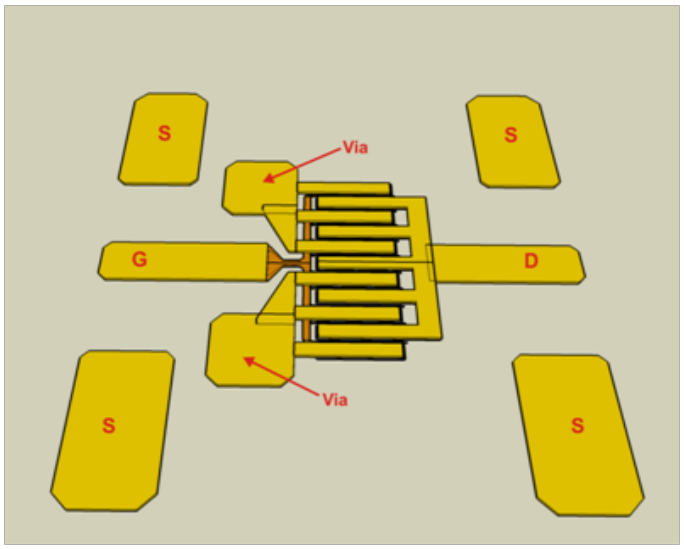
High electron mobility transistors (HEMT) are unipolar field effect transistors in which the channel is represented by the junction between two materials with different bandgaps (differently from MOSFET); they are adopted for applications at high frequencies where both high gain and low noise are required (radar, microwave communications, radio astronomy). In the past years, the power density handled by HEMTs is constantly increased; as a consequence the thermal management has become essential in order to minimize the temperature-induced degradation of performances [50]. AlGaN/GaN HEMTs analyzed in this section are fabricated by Selex SI.

10-finger HEMTs. Devices analyzed in this section are called A10D and P10D; an illustration of the both is reported in Fig. 3.28. The difference between the two topologies is due to the fact that in A10D design, source terminals are connected all together with a gold airbridge on the frontwafer; such an airbridge is connected to the "global" source terminal (gold layer on the backwafer) by using 2 gold vias (see also Fig. 3.28(a)); on-wafer measurement pads are connected to global terminal with other vias. Concerning P10D, sources are connected to 2 vias assuring the connection with the backwafer gold layer ("global" source) where the pads are connected through vias. The two structures share the same die surface, i.e., 880×980 μm^2. The cross section of simulated HEMTs is depicted in Fig. 3.29 while the indicated dimensions are summarized in the Table 3.12.

Many material parameters show a temperature dependence as detailed by (1.7). All values for GaN HEMTs are reported in Table 3.13. Conversely, the thermal conductivity of $Al_xGa_{1-x}N$ can be reasonably considered as temperature-insensitive and equal to 50 W/mK for $T > 300$ K within the range of aluminum molar fractions x of interest [51]. Concerning the GaN layer, it is possible to find large ranges of variation for both k_0 (50-130 W/mK) and m (0.43-1.4) as reported in [52, 53] depending on the doping (hydrogen), impurities and defects. The substrate of devices is made of 6H-SiC showing a strong dependence of thermal conductivity on temperature as reported in [54] and can change with doping and purity of samples. The features of the transition layer are not known; for this reason a simple parametric analysis changing both thermal conductivity and thickness has been carried out. This analysis was devoted to investigate the impact of transition layer properties on the thermal behavior of the device. In particular, thermal conductivity was varied between 10 e 100 W/mK while thickness between 10 and 100 nm.



(a)



(b)

Figure 3.28: Illustration of the simulated 10-finger AlGaN/GaN HEMTs with airbridge (a) connecting individual sources (A10D) and without airbridge (b). In the latter case (P10D), sources are connected to backwafer contact with two more vias.

Table 3.12: Dimensions of the main parameters of interest in HEMT structures.

Parameter	**Dimensions [µm]**
L_{DS}	4 (from layout)
L_G	0.4 (from layout)
L_{GS}	1.3 (from layout)
L_{GD}	2.3 (from layout)
L_S	20 (from layout)
L_D	20 (from layout)
$t_{S0} = t_{D0}$	0.220
$t_{S1} = t_{D1}$	0.6
$t_{S2} = t_{D2}$	4
t_{G1}	0.04
t_{G2}	0.3
t_{Trl}	unknown
t_{Sub}	70
t_{Bck}	10

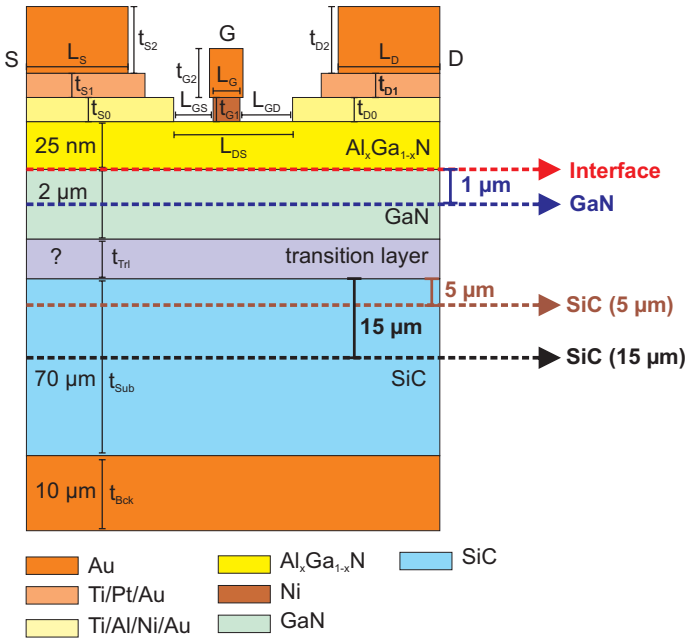


Figure 3.29: Illustration of the cross section of an individual finger in a HEMT device. Planes at a fixed z are depicted as well (namely, interface, GaN and SiC).

Figs. 3.30(a) and 3.30(b) show the normalized temperature increase for de-

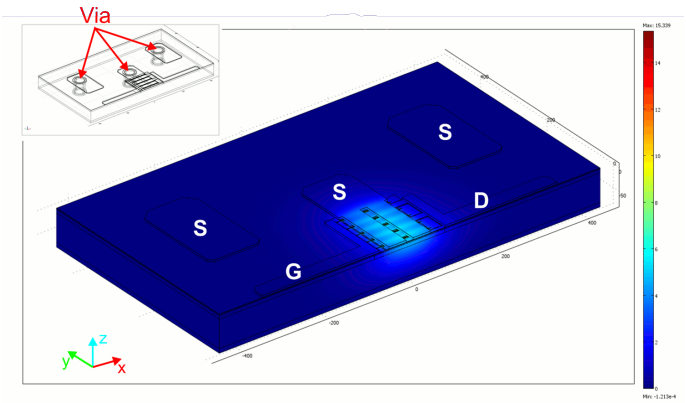
Table 3.13: Thermal parameters employed in HEMT simulations.

Material	Thermal conductivity (k_0) [W/mK]	m
6H-SiC	370	1.49
GaN	130	1.4
AlGaN	50	0
Transition layer	10-100	0
Ni	90	0
Ti/Al/Ni/Au	203	0
Ti/Pt/Au	131.2	0
Au	300	0

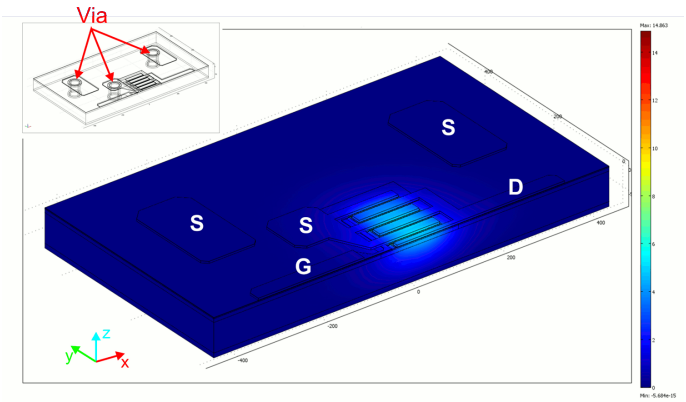
vices A10D and P10D, respectively with a transition layer having a thickness of 50 μm and a thermal conductivity of 50 W/mK. It is worth noting that in both cases the *Autosolver* software exploited the intrinsic symmetry of devices to reduce the number of degrees of freedom (horizonthal symmetry plane). All the analyses have been performed by taking into account non-linear thermal effects and considering all the power sources "activated" (i.e., all the 10 elementary devices in conduction). In Figs. 3.31 and 3.32 thermal resistance dependence on transition layer thickness and thermal conductivity are reported, respectively. Fig. 3.31 depicts the thermal resistance of both devices by varying the thickness t_{Trl} with a fixed power of $P_D = 10$ W, a thermal conductivity of $k_{Trl} = 50$ W/mK, and a baseplate temperature $T_B = 300\ K$. The increase of thermal resistance with thickness of the transition layer amounts to about 1.3% for device with airbridge (namely, A10D) and 3.9% for P10D. Thermal resistance of A10D exceeds the one of P10D by 6.3% considering $t_{Trl} = 10$ nm and 3.7% for $t_{Trl} = 90$ nm. The reason of this discrepancy is twofold

- In the "colder" transistor P10D, gold vias (contacting source pads to back wafer gold) are nearer to heat dissipating regions than in A10D , thus enhancing the heat removal effect.

- To a minor extent, it has to be considered that P10D gates are longer (116 μm) than the ones of A10D (112 μm), thereby reducing the power density (for the same dissipated power).



(a)



(b)

Figure 3.30: 3-D view of the normalized temperature increase for HEMTs (a) A10D and (b) P10D. The inset shows the vias connecting individual sources to the "global" source contact (i.e., the gold layer on the backwafer).

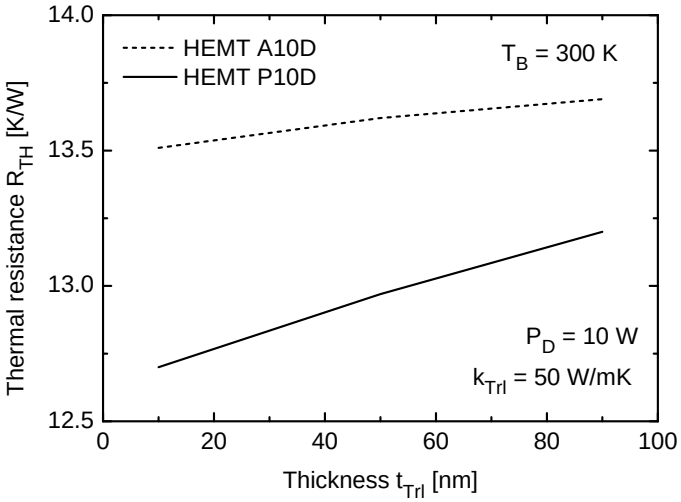


Figure 3.31: Thermal resistance dependence on transition layer thickness both for A10D and P10D HEMTs. Non-linear thermal effects are accounted for.

Fig. 3.32 shows the thermal resistance of A10D and P10D HEMTs as a function of thermal conductivity (from 10 to 100 W/mK) of the transition layer for various thicknesses (i.e., 10, 50, and 90 nm). For $t_{Trl} = 50$ nm, it is possible to obtain a R_{TH} reduction of 3.1% by increasing the thermal conductivity in the range considered for both devices.

Due to thermal non-linearities of some materials used to fabricate HEMTs, it follows that the thermal resistance depends on the dissipated power. This dependence is clarified by Fig. 3.33 where the dissipated power is varied in the range between 1 and 25 W while keeping constant the geometrical and physical parameters of the transition layer (in particular, $k_{Trl} = 50$ W/mK e $t_{Trl} = 50$ nm). The thermal resistance shows a power dependence on P_D for both A10D and P10D HEMTs, namely,

$$R_{TH} = R_{TH_0} + A \cdot (P_D)^n$$

where R_{TH_0} is the thermal resistance for very low dissipated power (i.e., when disregarding non-linear thermal effects), while A and n are fitting parameters. For both curves in Fig. 3.33 it was possible to evaluate $n = 1.49$ (somehow recalling the m parameter of SiC thermal conductivity). In Fig. 3.29 are depicted the planes for a fixed z, namely,

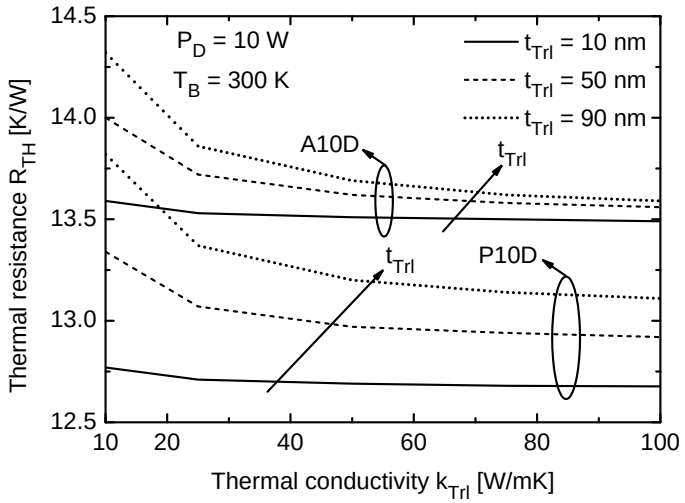


Figure 3.32: Thermal resistance dependence on transition layer thermal resistance both for A10D and P10D HEMTs. Curves are reported for various values of thickness (i.e., 10, 50, and 90 nm). Non-linear thermal effects are accounted for.

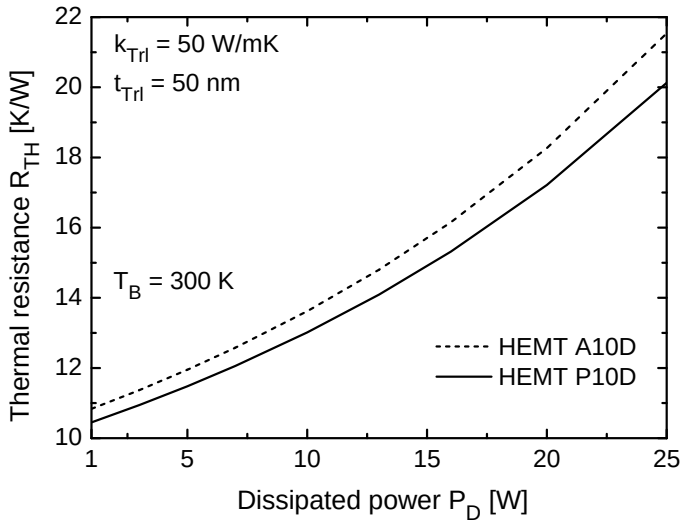


Figure 3.33: Thermal resistance dependence on dissipated power both for A10D and P10D HEMTs.

Interface is the plane that divides the AlGaN and GaN layers.

GaN is placed in the middle of GaN layer (i.e., 1 µm under the interface layer).

SiC (5 µm) is the plane positioned 5 µm under the interface between transition layer ans SiC substrate.

SiC (15 µm) is placed 10 µm under the SiC (5 µm) plane.

A 2-D normalized temperature increase distribution on the above planes for both the A10D and the P10D is reported in Fig. 3.34; the total dissipated power P_D amounts to 10 W. Internal fingers are more prone to self-heating with respect to lateral ones because of the higher thermal coupling. Another interesting result is the reduction of the maximum value moving through the SiC to reach the heat-sink; in fact, the maximum value of normalized temperature reduces of almost $1/4$ with respect to the one evaluated on the interface plane.

Gate dimensions impact on thermal behavior. The investigation of the impact of gate dimensions on the steady-state thermal behavior was conduced on simplified structures where the metalization layers are removed in order to simplify the analysis. The effect of channel width W and the number of fingers for a fixed distance between gates L_{GG} (i.e., center-to-center distance between fingers). Let us suppose that each finger is not thermally coupled with the neighboring ones thus the thermal resistance for a single finger can be written like

$$R_{TH} = \frac{R_{TH_W}}{W}, \tag{3.2}$$

where R_{TH_W} is the thermal resistance for a millimeter of gate width W [mm·K/W]. Hence, by ignoring the thermal coupling between n fingers it is possible to obtain, being all thermal resistances in parallel

$$R_{TH} = \frac{R_{TH_W}}{n \cdot W}. \tag{3.3}$$

Thermal resistance variation with periphery, defined as $n \cdot W$ (obtained by changing either W or n), is summarized in Fig. 3.35. From there, it is possible to evaluate $R_{TH_W} = 11$ mm·K/W. This first order approximation is, obviously, inaccurate when the thermal coupling starts to become relevant (i.e., fingers very close). However, the maximum error contemplates an underestimation of thermal resistance of 19% in the cases considered in Fig. 3.35 (i.e., for the

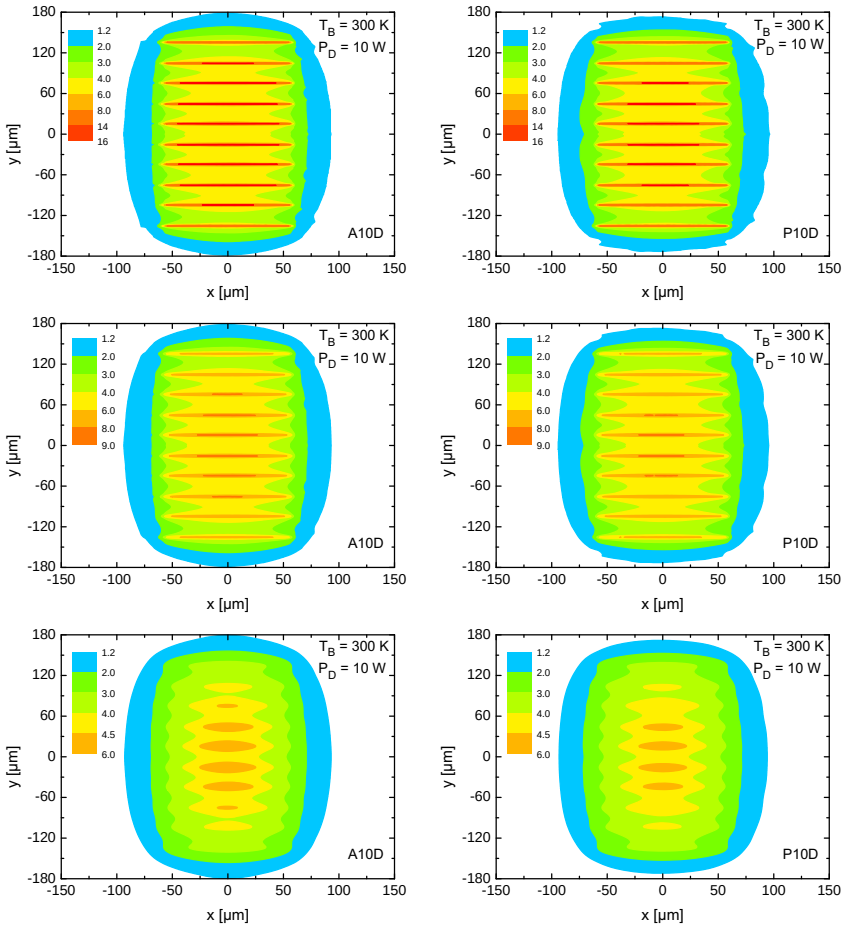


Figure 3.34: Normalized temperature increase over the planes of interest for both (left) A10D and (right) P10D. Pictures from top to bottom refer to *interface*, *GaN* and *SiC (5 µm)* planes, respectively.

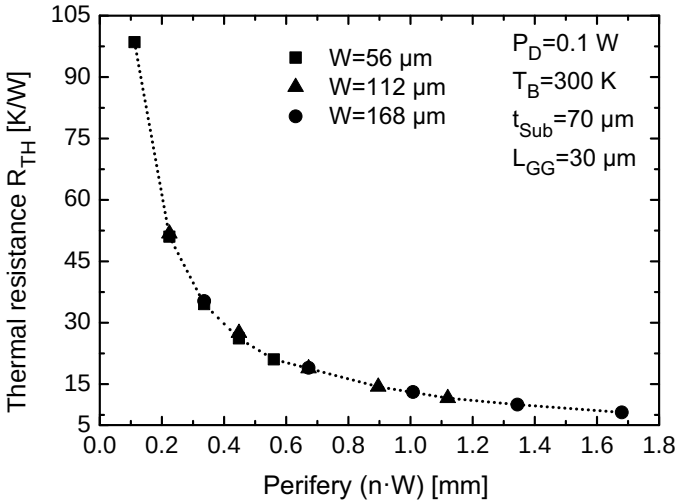


Figure 3.35: Thermal resistance dependence on the periphery. Data relate to various emitter widths and number of elementary fingers. The dotted curve represents the trend of R_{TH}.

largest periphery as expected). A better approximation can be obtained with a very small error by

$$R_{TH} = \frac{R_{TH_{Wa}}}{(n \cdot W)^p}, \tag{3.4}$$

where $R_{TH_{Wa}} = 12.5$ mm·K/W and $p = 0.94$ are fitting parameters that takes into account the thermal coupling. With these data the maximum error obtained amounts to 4%.

Dynamic behavior of P10D. By resorting to the approach proposed in Section 2.1.2 it was possible to optimize an equivalent thermal compact network suitable to describe the dynamic thermal behavior of HEMTs. Fig. 3.36 shows the results of the optimization in time domain for a GaN P10D HEMT. It is worth noting the inaccuracy of single pole approximation if compared to the excellent fit obtained with 5 *RC* cells. Foster network is then transformed to Cauer network so as to receive information on the dynamical heat propagation inside the structure of the device; the *RC* elementary cells can be associated to the various different material layers. The gradual vertical heat flux can be appreciated in Fig 3.37 where the dynamic temperature increases normalized

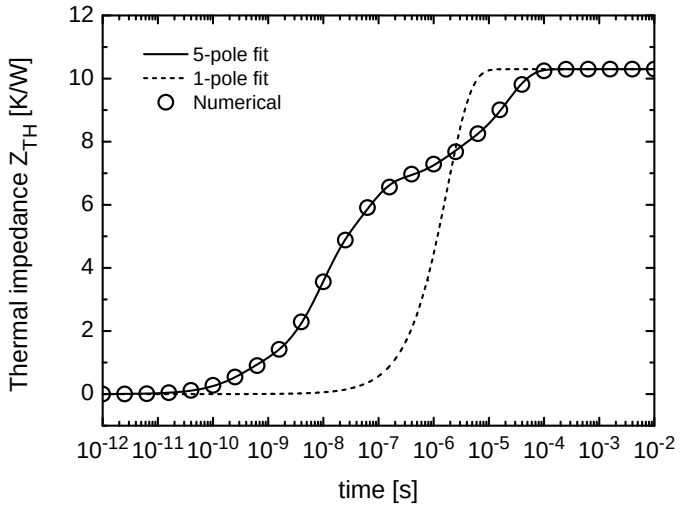


Figure 3.36: Thermal impedance versus time for a P10D HEMT. Simulation (symbols) data are reported together with (dashed line) single-pole and (solid line) five-pole fitting.

to dissipated power are shown for the Cauer network nodes.

Thermal networks optimized parameters are reported in Table 3.14 both for Foster and Cauer models.

The *full spectrum mode* (see Section 2.1.3 for detailed description) has been employed in order to verify the possibility of individuating attach failures (i.e., the die is not well fixed to the package). To study such a possibility, three different structures where the device is attached to an heatsink (see Fig. 3.38 for a cross section) with an epoxy resin[4] were conceived; this resin is considered homogeneously distributed in the ideal case while, in the other two cases it presents voids disposed following various geometrical schemes (namely, a big central void and a "checkers" distribution), to model a non ideal attachment. Fig. 3.39 depicts the simulated curves and the curves obtained by employing the *time constant spectrum* for all cases of interest (i.e., the ideal case with no void, and the two cases with different geometry of voids). The software evaluated 349 *RC* cells in the ideal case; on the contrary when voids

[4]The epoxy resin is Diemat 6030HK, characterized by a thickness of 120 μm, mass density $\rho = 4.5$ g/cm^3, thermal conductivity $k = 60$ W/mK.

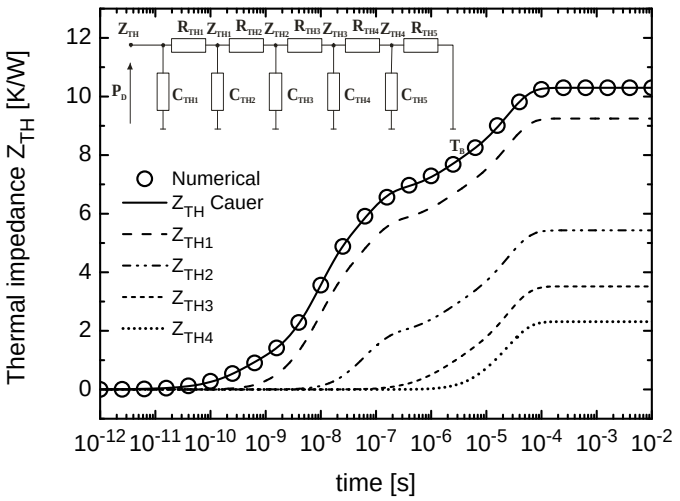


Figure 3.37: Thermal impedance versus time for a P10D HEMT at the various nodes of Cauer network model. (Symbols) Simulation data are reported together with the (solid line) "global" thermal impedance evaluated by Cauer model.

Table 3.14: Foster and Cauer optimized network parameters for HEMT P10D.

Device	$\mathbf{R_{TH_1}, R_{TH_2}, R_{TH_3}, R_{TH_4}, R_{TH_5}}$ **[K/W]**		$\mathbf{C_{TH_1}, C_{TH_2}, C_{TH_3}, C_{TH_4}, C_{TH_5}}$ **[nJ/K]**	
	Foster	***Cauer***	***Foster***	***Cauer***
P10D	0.745847	1.04877	0.405807	0.341103
	3.41908	3.81694	2.35841	1.95424
	2.51465	1.91374	23.4375	28.8121
	0.916398	1.20311	1583.9	1407.73
	2.70016	2.31357	8243.35	8060.49

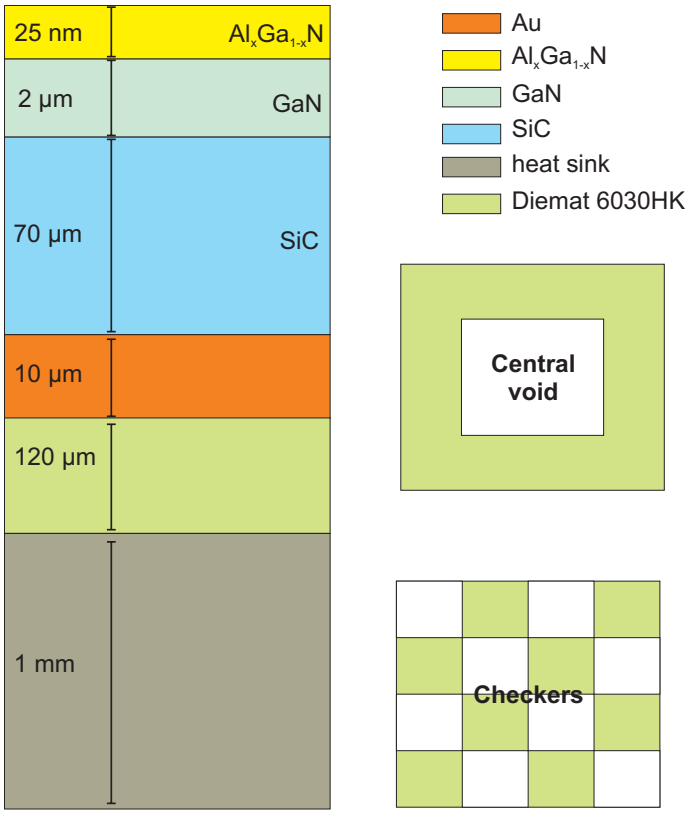


Figure 3.38: Illustrative cross section of a GaN HEMT fixed on a heat-sink with an epoxy resin. The top views of the two void models used in simulations to emulate a bad attach are reported as well.

are present the number of elementary networks jumps to 747 and 757 for a central void and a "checkers" void, respectively. It is worth noting that

- The curves obtained from equivalent networks coincide with the numerical ones.
- There is no influence of the voids on the thermal behavior up to the time instant 5×10^{-5} s since the heat has not reached epoxy resin yet.

Once the *time constants spectrum* is known, it is possible to employ the enhanced transformation routine to obtain a Cauer model (with a big number of elementary *RC* cells) so as to evaluate the differential structure function $K(R_\Sigma)$ as explained in Section 2.1.3. The K function is reported versus R_Σ in Fig. 3.40, the peaks indicated by the arrows identify the epoxy resin layer. The shift of the peaks is connected to the increase of resin defects, representing a thermal resistance growth (the presence of voids counteracts the heat flow). This method, for this reason, can be a valid, non destructive, solution to indirectly estimate the effectiveness of attach process.

To solve the numerical issues connected to transformation method from Foster to Cauer model, the software arbitrary numerical precision[5] was set to 80 digits.

[5] See Section 2.1.3 for major details on Foster to Cauer transformation and the need of arbitrary mathematical precision.

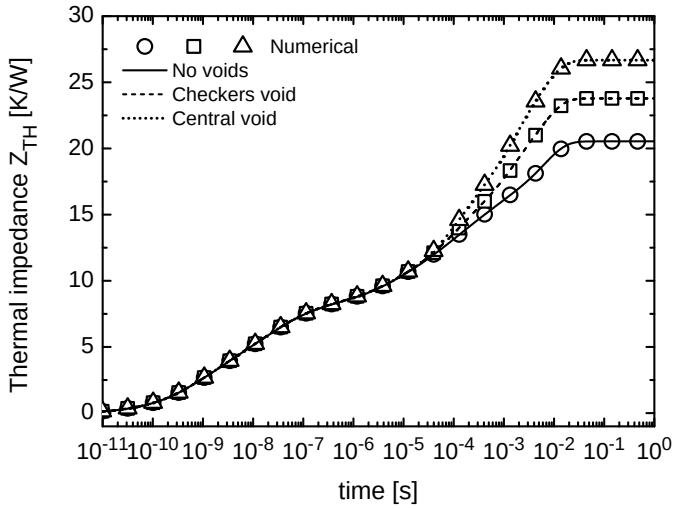


Figure 3.39: Thermal impedance versus time of a GaN HEMT for different die attach situations (ideal, central and checkers void); (symbols) numerical curves are compared to (lines) modeled characteristics (see Chapter 2 for details).

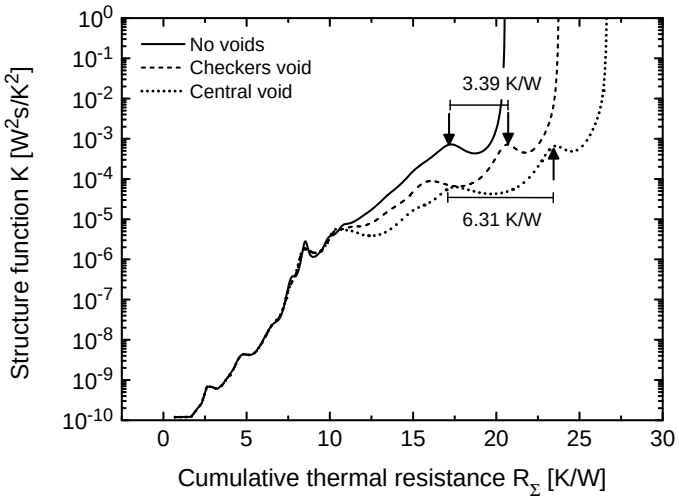


Figure 3.40: *Differential structure function* versus cumulative thermal resistance for a GaN HEMT with various die attachment conditions. Peaks relative to the resin are indicated by the arrows.

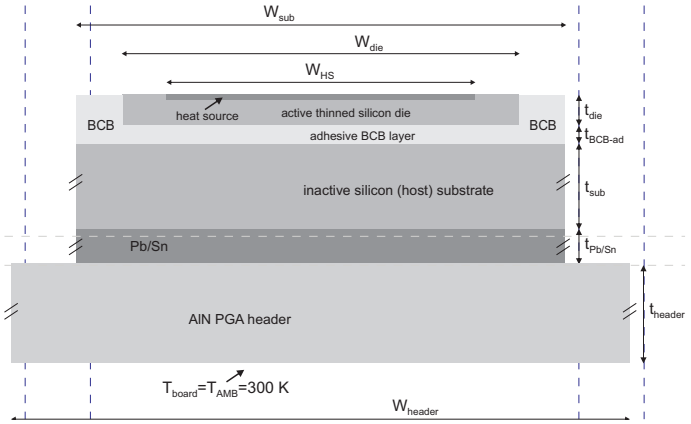


Figure 3.41: Illustration of the schematic cross-section of the single-level UTCS module.

3.1.5 Simulation of UTCS devices

Semiconductor industry is investing resources into three-dimensional stacked-die architectures; these are devised to effectively increase the integration density of semiconductor systems, thereby leading to smaller, lighter, and cheaper products. Nevertheless, 3-D stacked architectures are particularly prone to thermal effects since the increase in dissipated power is not accompanied by a corresponding improvement in cooling efficiency. For this reason cooling strategies need to be investigated to increase the heat removal capability from the dissipating regions. Thermal effects can be even more severe in ultra-dense packaging technologies resulting from the recent advances in thinning and attachment techniques. *Ultra-thin chip stacking* (UTCS) technology is one of these advanced technologies suffering from thermal issues. This is mainly due to the low thermal conductivity of benzocyclobutene (BCB) layers adopted to electrically insulate silicon chips thinned up to 10 µm. This is necessary to vertically integrate chips in the structure on the same inactive host silicon substrate [55,56]. Numerical simulator can help in predicting the thermal behavior and find technology solutions to mitigate thermal issues.

Simulation of single-chip module. A two-dimensional cross section of the structure analyzed is illustrated in Fig. 3.41 where

- W_{die} , L_{die} , and t_{die} are the width, length, and thickness of the thinned

silicon die containing the dissipating circuitry, respectively.

- W_{HS} and L_{HS} denote the width and length of the superficial (indefinitely thin) heat source, located on the die top.
- W_{sub} , L_{sub} , and t_{sub} designate the width, length, and thickness of the inactive silicon substrate.
- t_{BCB-ad} is the thickness of the BCB layer between active die and substrate.
- W_{header} , L_{header} , and t_{header} are the width, length, and thickness of the AlN PGA header.
- $t_{Pb/Sn}$ is the thickness of the thermally conductive resin that attaches the module (i.e., the silicon host) to the package.

The geometrical parameters of a "reference" single-chip structure are reported in Table 3.15. A "reduced" structure where the resin and header layers are removed is considered as well. It was found that non-linear thermal effects do not influence simulation results and for this reason are disregarded[6]. Metallic interconnections were neglected since their small size does not sensibly favor the heat removal [55]. The thin heat source was accounted for by enabling a uniform heat flux entering the domain. Adiabatic boundary conditions were assumed for both top and lateral faces of the structure.

The AlN header bottom in contact with the board is considered isothermal at ambient temperature, while the top and lateral surfaces are assumed adiabatic. The inherent system symmetries are exploited to reduce the number of elements by simulating only one quarter of the structure[7]. The self-heating thermal resistance of the "reference" single-die module was evaluated to be 1.24 K/W, while the value corresponding to the "reduced" module was found to be 1.02 K/W. Hence it is possible to conclude that the contribution of the layers beneath the silicon substrate amounts to about 17.5%. The rather low R_{TH} values are due to the large dissipated areas. The adiabatic condition at top and lateral faces was justified by demonstrating that a natural convection with

[6] A maximum error amounting to less than 1% was determined at a dissipated power of 20 W.

[7] A typical mesh for UTCS structures comprises about 5×10^5 elements and 7×10^5 degrees of freedom. When disregarding nonlinear thermal effects, a steady-state solution is evaluated in about 300 seconds when using a workstation equipped with 2 hexacore 2.43 GHz CPUs and a 100 GB RAM

Table 3.15: Geometrical parameter values of the reference structure.

Parameter	Value [µm]
$W_{die} = L_{die}$	5620
$W_{HS} = L_{HS}$	4240
$W_{sub} = L_{sub}$	6200
$W_{header} = L_{header}$	15000
t_{die}	10
t_{BCB-ad}	3
t_{sub}	500
$t_{Pb/Sn}$	50
t_{header}	760

Table 3.16: Thermal conductivities employed for UTCS FEM simulations.

Material	Thermal conductivity [W/mK]
Si	148
BCB	0.18
Pb/Sn	36
AlN	150

a heat transfer coefficient h amounting to 2 W/m^{2}K applied to the top and lateral surfaces [56] does not influence appreciably the results. To better quantify the heating action of the BCB layer, a fictitious structure where all the BCB is substituted by silicon was simulated. It is worth noting that the resulting R_{TH} reduces to about 0.39 K/W, that is, a value more than three times lower than that of the BCB-equipped module. By considering the temperature rise over ambient normalized to dissipated power along a vertical line crossing the center of the dissipating region (see Fig 3.42), the influence of the 3-μm-thick BCB layer is apparent. Fig. 3.43 depicts the R_{TH} variation with heat source area $A_{HS} = W_{HS} \times L_{HS}$, which was varied by keeping all other parameters (including $A_{die} = W_{die} \times L_{die}$) equal to the reference values, for both the complete and reduced modules. The thermal resistance increases from 0.92 ($A_{HS} = 2.5 \times 10^7$ μm^2) to 537.4 K/W ($A_{HS} = 100$ μm^2) for the reference thickness of the die (10 μm). When increasing the die thickness to 20 μm the analogous values were found to be 0.94 K/W and 408.8 K/W. This means that the beneficial impact of the increased silicon die thickness, due to the larger volume where the heat can spread, is higher for low dissipating areas; in particular, a maximum thermal resistance decrease of about 38% was detected. On the contrary, the influence of the resin and PGA header was estimated to be negligible for small heat source dimensions, while increasing up to 19% larger value of heat source considered (i.e., $A_{HS} = 2.5 \times 10^7$ μm^2).

The influence of the adhesive BCB layer interposed between the thinned silicon die and the host substrate was examined by varying its thickness from 3 (the reference value) to 20 μm. Fig 3.44 groups the results that can be explained and summarized as follows

- The thermal resistance increases from 1.24 to 5.44 K/W by ranging the BCB thickness from 3 to 20 μm for a reference die ($t_{die} = 10$ μm).
- For a thicker silicon die ($t_{die} = 10$ μm) the cooling effect due to the spreading action is more perceptible at higher values of t_{BCB-ad}; in fact, the thermal resistance reduction amounts to 6.6% for $t_{BCB-ad} = 20$ μm while being only 2.7% for $t_{BCB-ad} = 3$ μm.

In Fig. 3.45, Z_{TH} is represented against time for the "complete" single-chip module, the "reduced" package-free structure, and the corresponding devices obtained by replacing the adhesive BCB layers with silicon. As evident from the figure, a longer transient behavior is obtained for the package-equipped modules, for which steady-state conditions are reached at about 5×10^{-2} s. On the contrary, "reduced" modules are characterized by a faster response, the

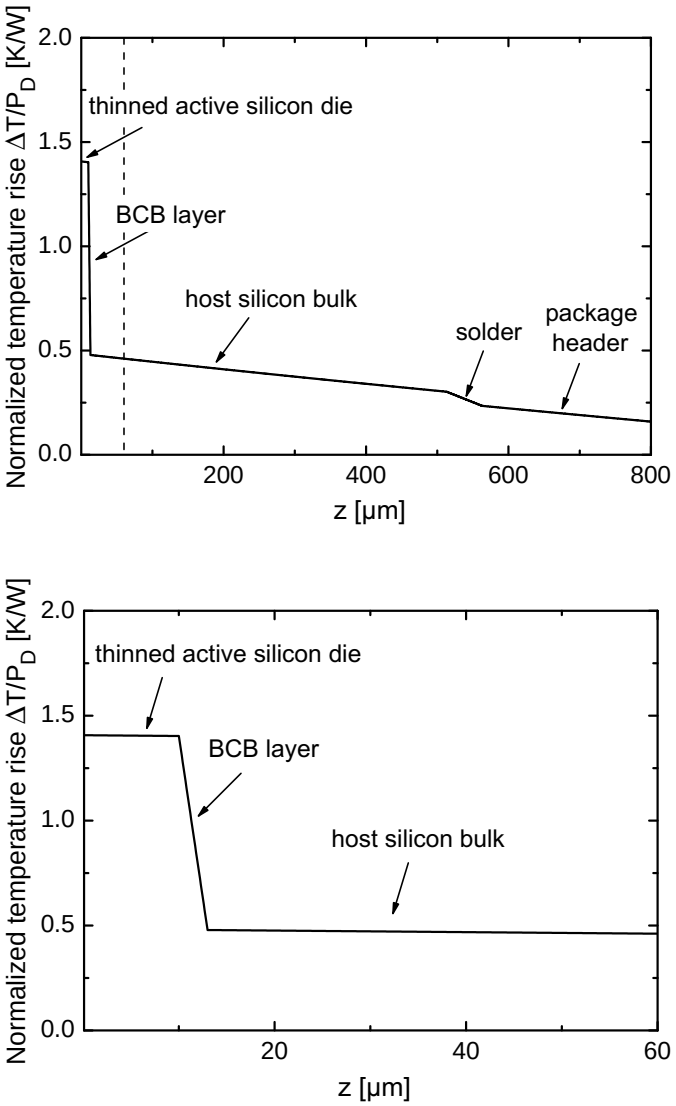


Figure 3.42: Temperature rise over ambient normalized to dissipated power versus z along a vertical cut crossing the heat source center; the bottom picture is a magnification of the normalized temperature behavior within the region delimited by the dashed line in the top figure.

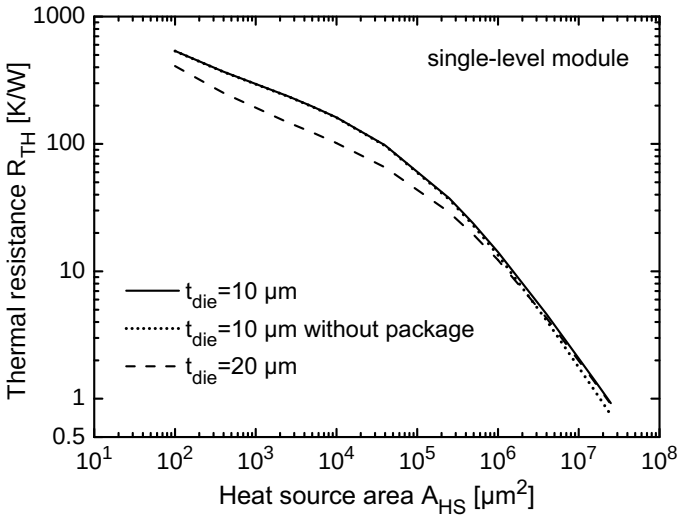


Figure 3.43: Self-heating thermal resistance R_{TH} as a function of heat source area A_{HS} for (solid line) $t_{die} = 10$ µm and (dashed) $t_{die} = 20$ µm; also shown is (dotted) the curve corresponding to the "reduced" domain.

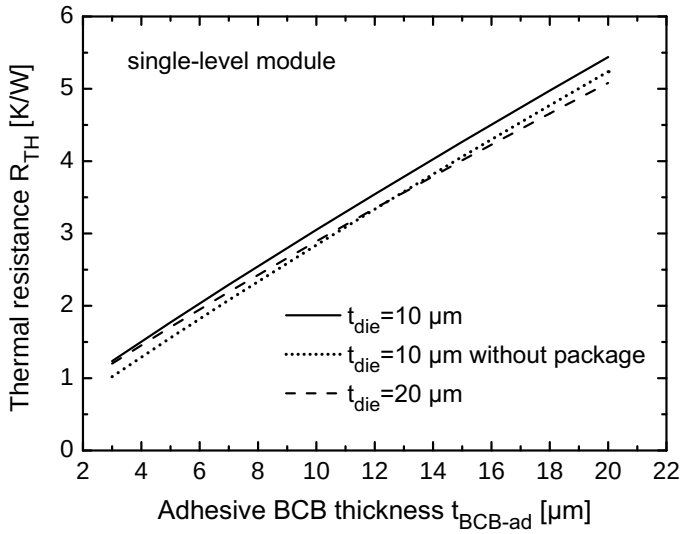


Figure 3.44: Self-heating thermal resistance R_{TH} as a function of BCB thickness t_{BCB-ad} for (solid line) $t_{die} = 10$ µm and (dashed) $t_{die} = 20$ µm; also shown is (dotted) the curve corresponding to the "reduced" domain.

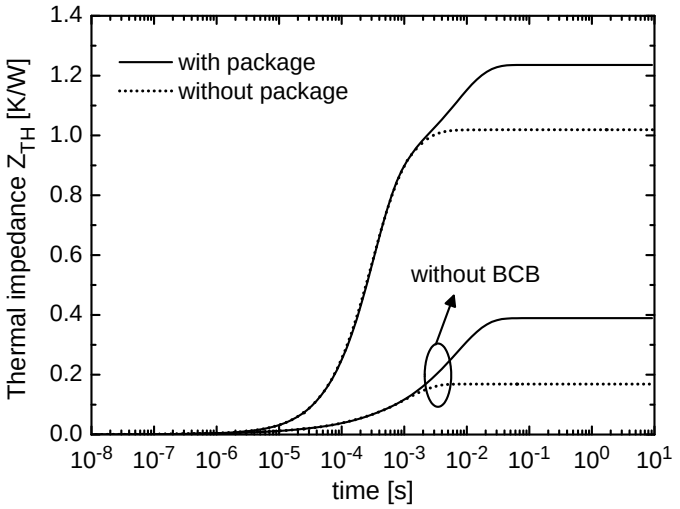


Figure 3.45: Thermal impedance Z_{TH} as a function of heat source area A_{HS} for (solid line) $t_{die} = 10$ µm and (dashed) $t_{die} = 20$ µm; also shown is (dotted) the curve corresponding to the "reduced" domain.

impedance growth ceases at 7×10^{-3} s. It is worth noting that for times shorter than 5×10^{-6} s the heat is still propagating along the 10-µm-thick active silicon die, while at $t = 1 \times 10^{-3}$ s it is approaching the bottom of the silicon substrate. The adverse impact of the adhesive BCB layer on the efficiency of heat spreading within the structure is apparent, regardless of the presence of the AlN header. The values of specific heat and mass density employed for the FEM simulations can be found in [55].

Simulation of three-chip module. The cross section of the "complete" three-chip structure is represented in Fig. 3.46. The analysis was performed by alternatively switching on the individual thinned dies, in order to quantify the values of the thermal resistance matrix (see Chapter 1 for definitions) and their sensitivity to the BCB interlayer thickness t_{BCB}. All the geometrical parameters were kept to the reference values with the exception of t_{BCB}. As a first step, the 1st-level (lower) chip was switched on, while deactivating the other two.

The thermal resistance variation versus the (common) thickness of the planarization BCB layers interposed between the thinned silicon dies is reported

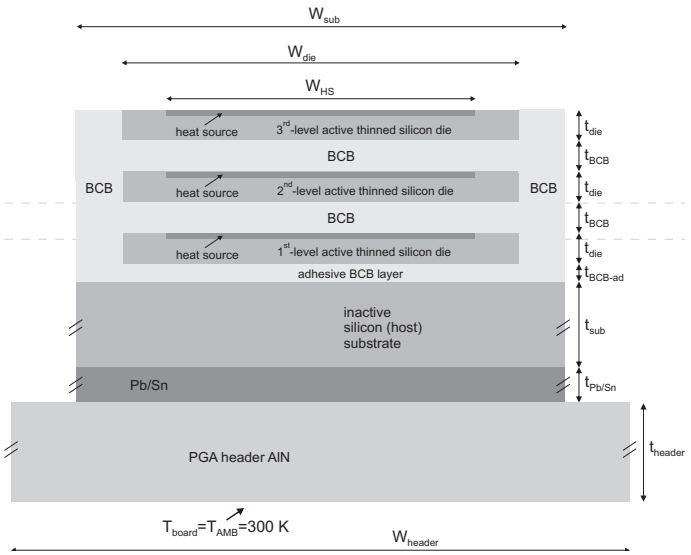


Figure 3.46: Illustration of the schematic cross-section of the three-level UTCS module.

in Fig. 3.47. It is shown that $R_{TH_{11}}$ slowly increases because of the slight decrease in the upward heat flow (while the downward remains unaltered). Both thermal resistances $R_{TH_{21}}$ and $R_{TH_{31}}$ decrease by 8.4% since the thermal coupling is growingly inhibited. $R_{TH_{31}}$ diminishes up to 12.1% due to the increase in thickness of BCB layers interposed between the first and the third chips. Subsequently, the 2$^{\text{nd}}$-level thinned die is activated, while the others are kept switched off. The behavior of the thermal resistances is reported in Fig. 3.48; $R_{TH_{22}}$ swiftly increases from 1.94 to 5.81 K/W, since the heat is progressively more constrained within the dissipating die by the thicker overhanging and overhung BCB layers. The mutual resistance $R_{TH_{32}}$ between the upper and the (dissipating) central chip linearly grows mostly because of the inhibition of the downward heat flux (due to the enlargement of BCB layer); on the contrary, $R_{TH_{12}}$ weakly decreases since the downward heat-flux from the dissipating region is increasingly weakened. Fig. 3.49 illustrates the behavior of the thermal resistances as a function of t_{BCB} when the 3$^{\text{rd}}$-level silicon die is switched on, while maintaining the others inactive. The main effect of the t_{BCB} increase is that the heat is mostly confined in the top device; this causes an increase both in R_{TH33} (spanning from 2.72 to 10.2 K/W) and R_{TH23} (ranging from 1.89 to 5.39 K/W); on the contrary, R_{TH13} reduces since

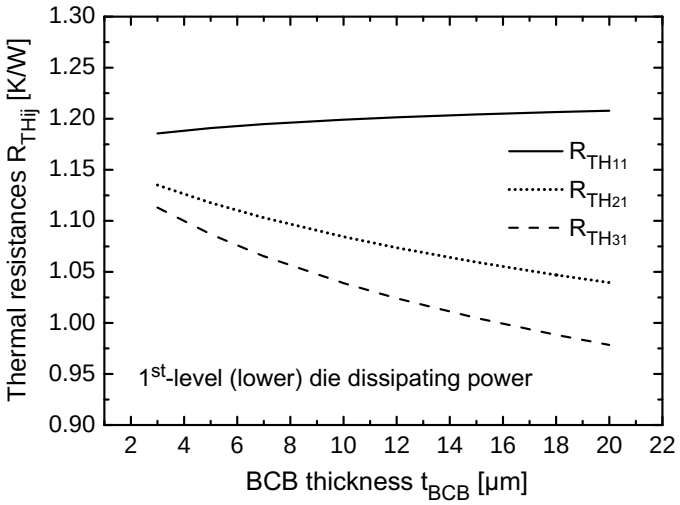


Figure 3.47: Self-heating (solid line) thermal resistance $R_{TH_{11}}$, and mutual thermal resistances (dotted) $R_{TH_{21}}$ and (dashed) $R_{TH_{31}}$ as a function of BCB thickness t_{BCB} in the stacked three-chip module in which only the lower thinned silicon die is activated.

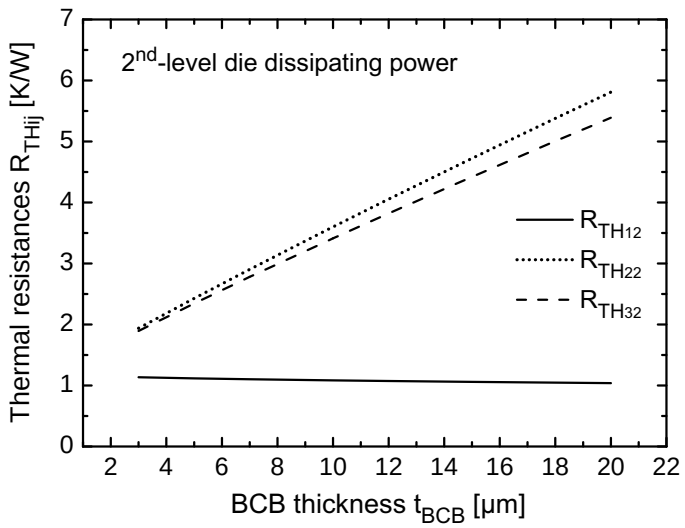


Figure 3.48: Self-heating (dotted line) thermal resistance $R_{TH_{22}}$, and mutual thermal resistances (solid) $R_{TH_{12}}$ and (dashed) $R_{TH_{32}}$ as a function of BCB thickness t_{BCB} in the stacked three-chip module in which only the central thinned silicon die is activated.

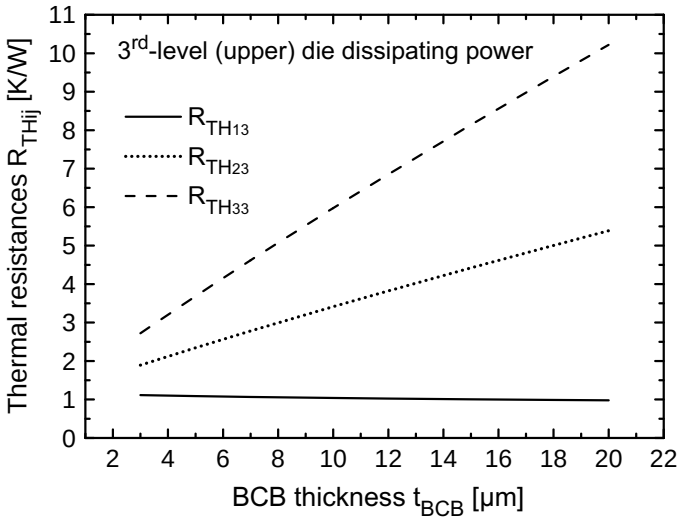


Figure 3.49: Self-heating (dashed line) thermal resistance $R_{TH_{33}}$, and mutual thermal resistances (solid) $R_{TH_{13}}$ and (dotted) $R_{TH_{23}}$ as a function of BCB thickness t_{BCB} in the stacked three-chip module in which only the bottom thinned silicon die is activated.

the bottom chip is increasingly far (divided by enlarged BCB layers) from the top chip. Considering a vertical cut crossing the center of the dissipating region ($t_{BCB} = 3$ μm), Fig. 3.50 clearly shows that the maximum normalized temperature rise is equal to 3.22 K/W while the self-heating thermal resistance R_{TH33} amounts to 2.72 K/W. Fig. 3.52 details the transient simulation data corresponding to the latter case (the 3rd-level die is turned on). It is shown that the temperature rise over ambient of the activated silicon die becomes perceptibly different from zero in correspondence of $t = 10^{-5}$ s, while that of the closest chip(s) starts growing at $t = 10^{-4}$ s, after the heat has propagated through the electrically (and thermally) insulating BCB layer. This feature is shared by the transient behavior of the different configurations as can be seen in Chapter 23 of [57] that collects the results of this work. The K (see Section 2.1.3 for further details) function for this configuration, showing the peaks corresponding to each part of the structure, is reported as an example in Fig. 2.6.

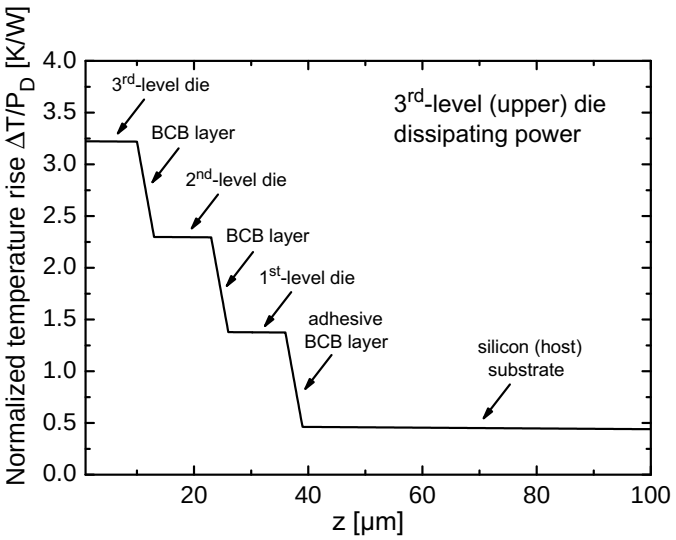


Figure 3.50: Temperature rise above ambient normalized to dissipated power along a vertical cut crossing the heat source center when the top chip is active.

Compact thermal models application to single-chip structure. A major issue in the numerical thermal analysis of UTCS modules is related to the complexity of the structure, in which heat flows through different regions showing a huge variation in size and material properties. While the investigation presented above could take advantage of the symmetry of the structure and dissipation pattern (so that only one quarter of the module could be considered), however, for non-symmetric structures this simplification is not possible. As extensively explained in Section 2.2, in this situation, it is possible to resort to compact thermal models (CTMs) [31–33]. The structure shown in Fig. 3.41 can be subdivided into two subsystems A and B, as illustrated in Fig. 3.53. Hence, it can be seen that the contact surface between parts A and B can be discretized into e.g., 2 elements. The thermal resistance of single-level modules characterized by PGA headers with different thermal conductivities was evaluated by using the CTM approaches relying on either a uniform or the position-dependent shape function (see also Chapter 2). In the latter case, the shape function used for the calculation of the CTM parameters was determined for the reference single-level module (i.e., that characterized by an AlN header with $k_{AlN} = 150$ W/mK). Fig. 3.54 reports the reference shape

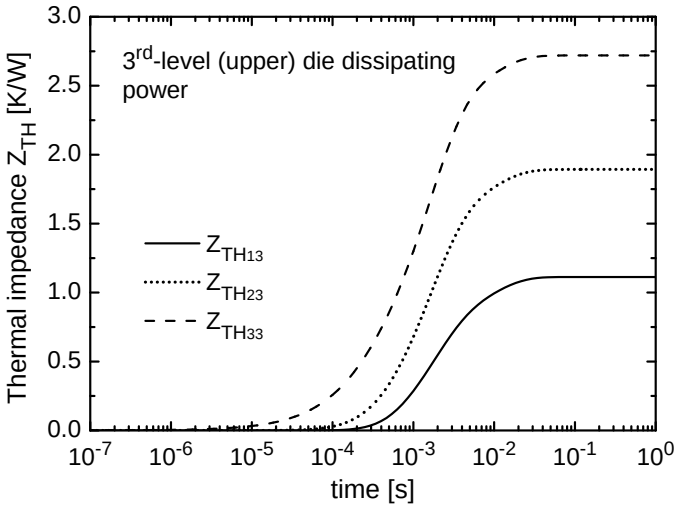


Figure 3.51: Thermal impedance for UTCS three-chip module in which only the 3rd-level silicon die is dissipating power: (dashed line) self-heating thermal impedance $Z_{TH_{33}}$, and mutual impedances (solid) $Z_{TH_{13}}$, and (dotted) $Z_{TH_{23}}$.

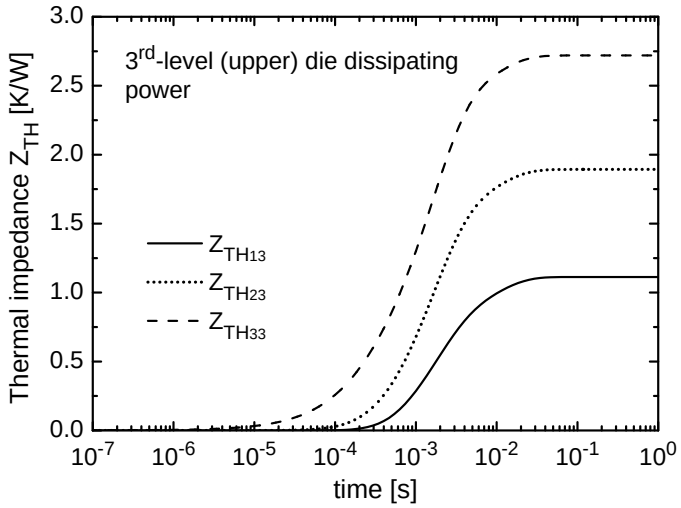


Figure 3.52: Thermal impedance for UTCS three-chip module in which only the 3rd-level silicon die is dissipating power: (dashed line) self-heating thermal impedance $Z_{TH_{33}}$, and mutual impedances (solid) $Z_{TH_{13}}$, and (dotted) $Z_{TH_{23}}$.

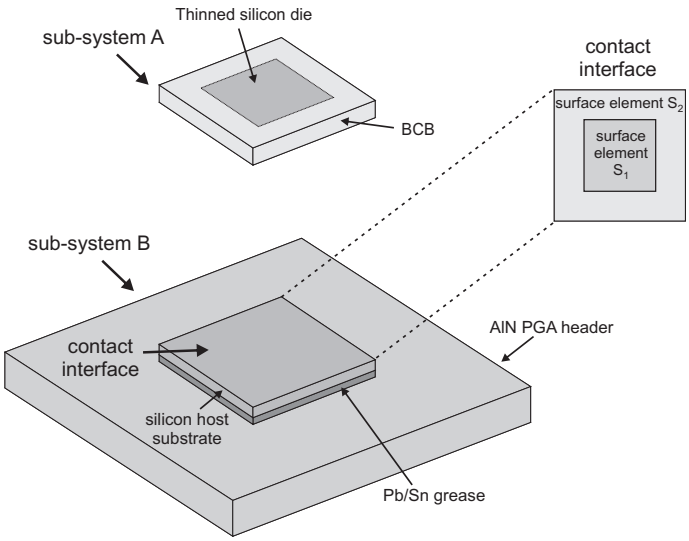


Figure 3.53: Illustration of the CTM formulation for the stacked single-level module depicted in Fig. 3.41. The system is partitioned into two sub-systems A and B. The contact interface between subsystems is discretized into 2 surface elements S_1 and S_2.

function used and a 2-D profile of the shape functions obtained by varying the thermal conductivity of PGA header. The calculation of the thermal resistance was repeated for $N = 1, 2, 3$ to quantify the accuracy improvement with increasing the discretization level. CTM-based results are compared to numerical data for each thermal conductivity of the header in Fig. 3.55. Considering a single thermal node ($N = 1$), the error is above 10% for the uniform case ($\sigma(r) = 1$), and decreases with increasing k_{header}. Obviously, this can be evinced by the fact that the temperature distribution at the contact surface becomes more uniform if k_{header} increases. In the case of the CTM based on a nonuniform shape function, the error is below 6% and obviously vanishes for the reference value of $k_{header} = k_{AlN}$. The error can be lowered by means of a finer discretization of the contact surface. Increasing N from 1 to 3 reduces the error in thermal resistance evaluation of about 50% both for uniform and nonuniform cases as far as the k_{header} values farthest from k_{AlN} (i.e., 50 and 250 W/mK) are concerned.

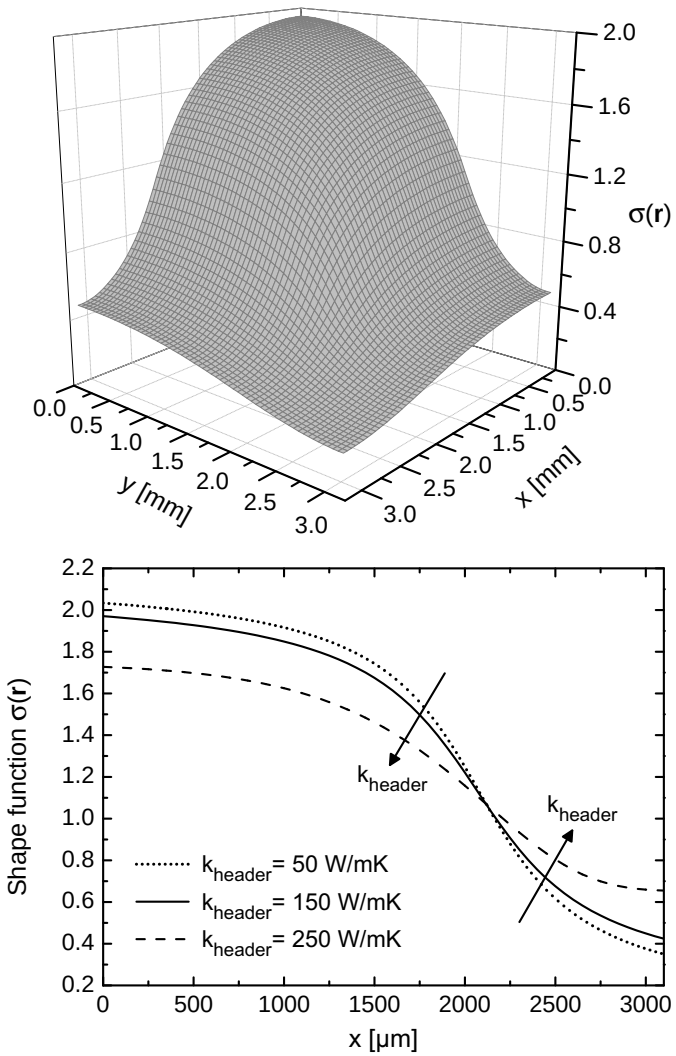


Figure 3.54: (Top) 3-D and (bottom) 2-D representations of the shape function $\sigma(r)$ over the contact interface, as determined through a 3-D thermal simulation of the single-level module illustrated in Fig. 3.41; the 2-D shape function shown is determined along a cut crossing the contact interface center for three values of the thermal conductivity of the PGA header, namely, (dotted line) 50, (solid) 150 (i.e., the reference value), and (dashed) 250 W/mK.

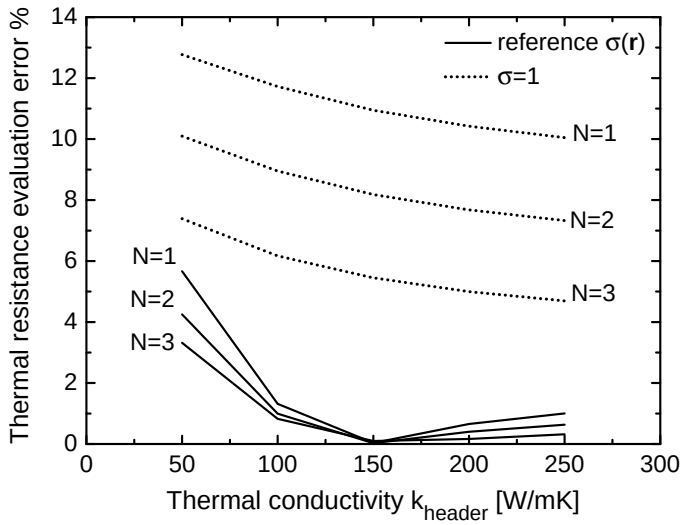


Figure 3.55: Percentage error (in absolute value) in evaluating the thermal resistances by CTMs in comparison to FEM 3-D results for the single-layer module shown in Fig. 3.41 as a function of thermal conductivity of the PGA header. The CTM-based R_{TH} are calculated by (dotted lines) a uniform shape function $\sigma(r) = 1$ [31] and (solid) the position-dependent $\sigma(r)$ determined for the reference domain with $k_{header} = k_{AlN} = 150$ W/mK illustrated in Fig. 3.54. Three discretization levels ($N = 1, 2, 3$) are considered for both cases.

3.1.6 Simulation of SiGe HBTs

Recent years have seen a constant interest of the electronic industry in the area of SiGe HBTs (Silicon-Germanium Heterojunction Bipolar Transistors), in particular due to the low cost (SiGe process integrates without high costs with silicon process) and the high performances. The European Project DOTFIVE that involves several academic (among which the University of Naples) and commercial partners (Infineon technologies, ST Microelectronics) was conceived to reach the frequency of 500 GHz at room temperature with SiGe HBTs. This kind of performance is usually thought to be only possible with III-V compound semiconductor technologies. Obviously, SiGe HBTs if compared to III-V compound semiconductor devices can combine high density and high frequency of operation with low-cost integration so to make them suitable for consumer applications. In this scenario, the scaling of devices and the use of isolation schemes can enhance thermal issues and for this reason a thoroughly thermal analysis is necessary to avoid performance degradations.

Thermal transient behavior and network identification. Infineon (IFX) and ST Microelectronics (STM) SiGe technologies were analyzed by employing the identification software in order to obtain (from 3-D FEM simulations) the minimum number of *RC* cells and study results in frequency domain. Moreover, it is possible, from frequency discussion, to obtain the sole equivalent thermal capacitance C_{TH} of the transistor in a way that the thermal cut-off frequency f_{TH} of the single-pole pair involving the self-heating thermal resistance R_{TH} (the sum of elementary resistances) and the mentioned capacitance C_{TH} equates the "multi-pole" network cut-off frequency

$$f_{TH} = \frac{1}{2\pi \cdot R_{TH} \cdot C_{TH}}.$$

Hence, it is possible to observe the cut-off frequency scaling with emitter dimensions together with thermal capacitance and resistance and obtain models to be included in CAD software both for IFX and STM transistors. W_E and L_E are the width and length of the emitter stripe (that are assumed equal to the width and length of the heat source), respectively, while $W_{E_{ref}}$ and $L_{E_{ref}}$ are the reference values. $W_{E_{ref}}$ amounts to 0.2 (IFX) and 0.13 (STM) μm while L_{Eref} amounts to 2.67 (IFX) and 9.88 (STM) μm, respectively.
A parametric analysis of the dynamic thermal behavior of both IFX and STM HBTs was performed by varying the emitter length L_E (for $W_E = W_{Eref}$). Fig. 3.56 shows the thermal impedance versus time as obtained through the

identification strategy and numerical simulations for various emitter stripe lengths for both technologies; as can be seen, excellent matching is achieved. As seen for other category of devices the thermal resistance (the steady-state value of transient thermal characteristics) reduces with emitter enlargement. Afterward, network optimized in the time domain (consisting of 8 *RC* pairs obtained with a maximum error of 3%; see also Section 2.1.2 for further details) was exploited to investigate the dynamic behavior in the frequency domain. Fig. 3.57 details the frequency response (i.e., the magnitude of Z_{TH} vs. frequency) for both IFX and STM devices characterized by various values of emitter length. The Z_{TH} values corresponding to the thermal cut-off frequencies f_{TH} are identified by symbols. It is found that the cut-off frequency reduces (for both technologies) by lengthening the emitter stripe (i.e., by increasing the length of the heat source). This means that the thermal capacitance growth prevails over the thermal resistance reduction due to the enlargement of the silicon volume, a result already found for fully-isolated silicon-on-glass BJTs in the previous section. The f_{TH} decrease with increasing L_E is apparent in Fig. 3.58, which shows that the reduction rate is higher for small L_E values, while f_{TH} tends to be insensitive to L_E as this grows higher for both technologies. From Fig. 3.58 it can be evinced that STM HBTs exhibit higher thermal cut-off frequency for a prescribed L_E; it should be however remarked that $W_E = 0.2$ μm for IFX HBTs while it is 0.13 μm for STM HBTs along the curves.

Details on the structures and technologies are confidential and can be found only in DOTFIVE technical reports.

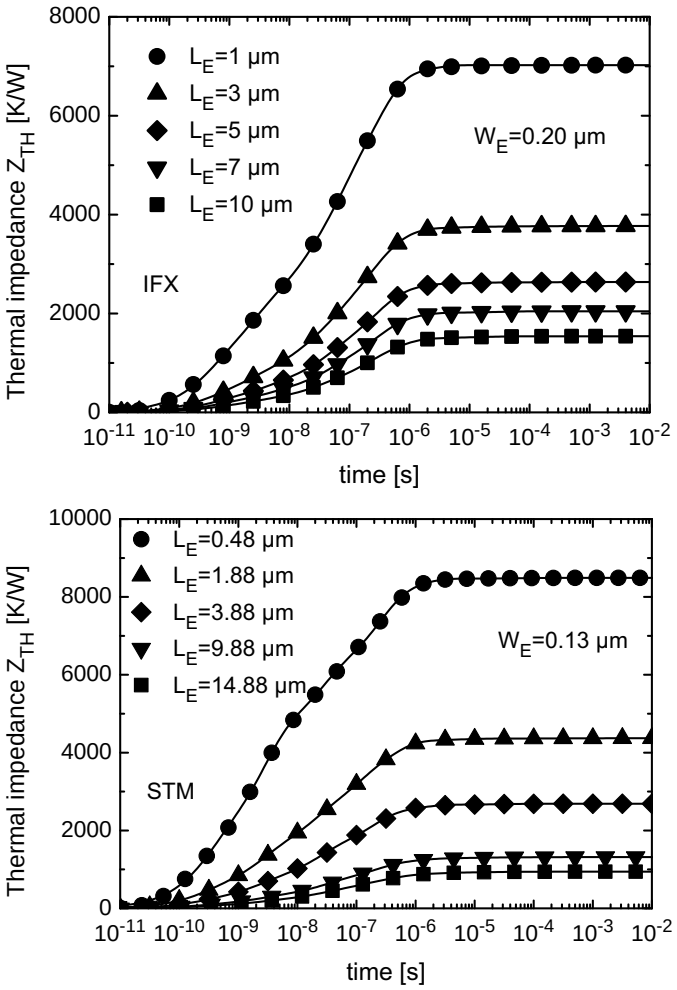


Figure 3.56: Thermal impedance versus time for various emitter lengths both for (top) IFX and (bottom) STM devices. In both situations, (symbols) numerical and (solid line) data from optimized networks are reported for $W_E = W_{Eref}$.

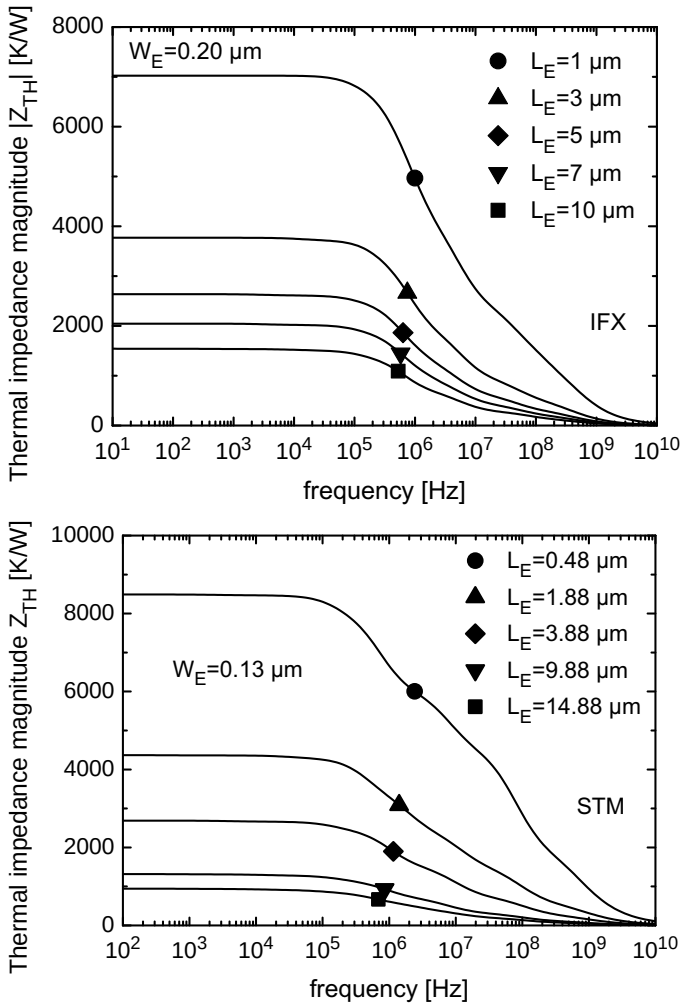


Figure 3.57: Thermal impedance versus frequency for various emitter lengths both for (top) IFX and (bottom figure) STM devices. In both situations, the values of the thermal impedance magnitude at the cut-off frequency are evidenced by symbols.

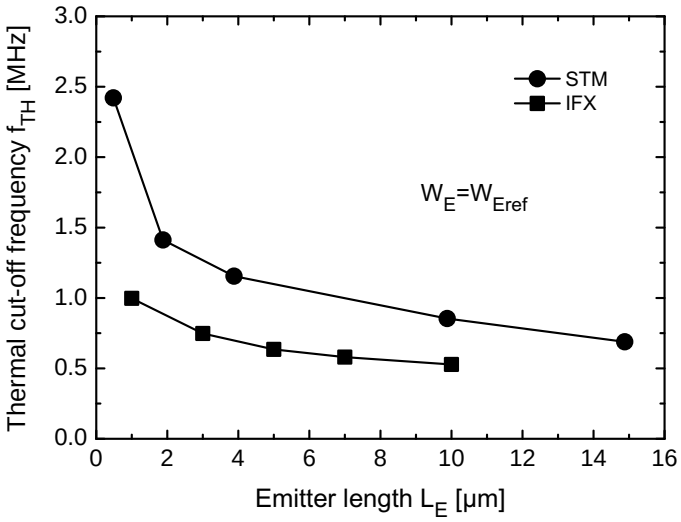


Figure 3.58: Thermal cut-off frequency for both technologies (namely, IFX and STM) versus emitter length for $W_E = W_{Eref}$.

Chapter 4

Measurement techniques for the thermal resistance and impedance

Thermal experimental characterization of solid-state devices is continuously gaining in importance in many technologies, also for low-power applications. Nowadays, the development of experimental methods allowing a reliable evaluation of the actual junction temperature T_j of the devices is highly desired both in steady-state (i.e., the self-heating thermal resistance R_{TH}) and transient (i.e., thermal impedance in time or frequency domain Z_{TH}) regimes. The last two decades have seen the proposal of many measurement techniques relying on different approaches (infrared microscopy, Raman spectrography, pulsed measurements of electrical parameters, etc..) each one having its advantages and disadvantages. In this scenario, the possibility of accurately evaluating thermal parameters of interest for solid-state devices with common (and cheap) laboratory equipment must be deeply investigated.

4.1 Measurement techniques for the thermal resistance

In the last two decades, a number of methods have been conceived and developed to indirectly determine R_{TH} through straightforward measurements of DC electrical characteristics [44, 58–64] for bipolar transistors. However, a systematic comparison of all these approaches is not found in literature until

now. Moreover, the applicability of these techniques to silicon BJTs, where the base-width modulation can hamper the thermal resistance evaluation, was never clearly discussed. The scope of this section is twofold

- Provide a "critical" review of thermal resistance methods based on DC measurements both for GaAs HBTs and silicon BJTs.
- Propose a novel technique more suitable for transistors strongly affected by Early effect.

4.1.1 A critical review of thermal resistance DC measurement techniques

To better understand the most important methods for R_{TH} evaluation, as well as their range of application and limitations is useful to detail (although not in deep details) the hypotheses from which their formulation is obtained. Some techniques are not described (i.e., [44, 59, 65]) because they are just a direct derivation of previously proposed techniques (i.e., [58]) and do not represent a major improvement (in terms of accuracy and simplicity) with respect to the original technique.

A. Dawson *et al.* technique [58]. The first technique to measure the thermal resistance was proposed in literature by Dawson and his coworkers in 1992 [58]. According to the authors, the method is suitable to evaluate the R_{TH} for HBTs, where the base width modulation (Early effect) is negligible. This technique is based on the observation that a transistor (HBT or BJT) changes its β as well as its V_{BE} according to the temperature rise due to device power dissipation. From the direct observation that these parameters depend almost linearly on the baseplate temperature (T_B) it is possible to deduce:

$$\beta = \beta_1 + \frac{\Delta\beta}{\Delta T} \cdot (T_j - T_{j_1}) \,, \tag{4.1}$$

or

$$V_{BE} = V_{BE_1} + \frac{\Delta V_{BE}}{\Delta T} \cdot (T_j - T_{j_1}) \,, \tag{4.2}$$

where β_1 and V_{BE_1} are the current gain and base-emitter voltage evaluated at the device junction temperature T_{j_1}, being T_j the (base-emitter)

junction temperature of device. $\Delta\beta/\Delta T$ and $\Delta V_{BE}/\Delta T$ are the slopes of current gain and V_{BE} versus junction temperature, respectively. By recalling that

$$T_j - T_B = R_{th} \cdot P_D, \tag{4.3}$$

where T_B is the baseplate temperature , P_D the power dissipated by the device and R_{TH} is the thermal resistance, it is possible, combining (4.1) and (4.2) with (4.3), to obtain:

$$\beta\left(T_B, P_D\right) = \beta_1 + \frac{\Delta\beta}{\Delta T} \cdot \left(T_B + R_{TH} \cdot P_D - T_{j1}\right),$$

and

$$V_{BE}\left(T_B, P_D\right) = V_{BE1} + \frac{\Delta V_{BE}}{\Delta T} \cdot \left(T_B + R_{TH} \cdot P_D - T_{j1}\right),$$

respectively. Hence, by simply selecting two different power dissipation points (P_{D1} and P_{D2}) at the same baseplate temperature (namely, T_{B1}) and a point at an higher baseplate temperature (T_{B2}) but the same power dissipation (P_{D1}) it is possible to evaluate the thermal resistance as

$$R_{TH} = \frac{\frac{V_{BE}(T_{B1},P_{D1}) - V_{BE}(T_{B1},P_{D2})}{P_{D1} - P_{D2}}}{\frac{V_{BE}(T_{B1},P_{D1}) - V_{BE}(T_{B2},P_{D1})}{T_{B1} - T_{B2}}}, \tag{4.4}$$

and

$$R_{TH} = \frac{\frac{\beta(T_{B1},P_{D1}) - \beta(T_{B1},P_{D2})}{P_{D1} - P_{D2}}}{\frac{\beta(T_{B1},P_{D1}) - \beta(T_{B2},P_{D1})}{T_{B1} - T_{B2}}}, \tag{4.5}$$

where (4.4) is considered more accurate than (4.5) by authors because V_{BE} has a smaller deviation from linearity with respect to T_B than β (see Fig. 4 compared to Fig. 2 in [58]). This technique is based on the following simplifications:

- The thermal resistance is threated as a constant so it is possible to use it just in cases where there are no dependencies of thermal conductivity on temperature, which in turn causes a R_{TH} dependence on both dissipated power and baseplate temperature.
- The technique is rigorously valid only if the linear dependence of β and V_{BE} on the baseplate temperature is valid in a large range, but this is strictly true if $T_j \approx T_B$, which implies that the technique can be used only for low dissipated powers.

For these reasons, this technique is suitable to accurately evaluate the thermal resistance only at low dissipated powers.

B. Reisch technique [61]. In order to adopt this technique, the device must be biased with a constant V_{BE} under common-emitter configuration and the I_B versus V_{CB} curve must be measured. The method assumes that transistors have a negligible thermal generation in the base-collector diode (i.e., $I_B(0, V_{CB}) = 0$) [61] and self heating can be neglected. The base current is considered to be independent on V_{CB} up to the onset of carrier multiplication. The linear increase in $\partial I_B/I_B$ is due to the increasing power dissipation in device proportional to V_{CB}. The extraction of R_{TH} is performed plotting $\partial I_B/I_B$ versus the change in dissipated power expressed as $P(V_{BE}, V_{CB}) - P(V_{BE}, 0)$ where $P(V_{BE}, V_{CB}) = I_B \cdot V_{BE} + I_C \cdot V_{CE}$, so it is possible to obtain:

$$\frac{\partial}{\partial P}\left(\frac{\Delta I_B}{I_B}\right) = \frac{\partial \ln(I_B)}{\partial P} = R_{TH}\frac{\partial \ln(I_B)}{\partial T},$$

whence

$$R_{TH} \approx \frac{k \cdot T_j^2}{E_g - q \cdot V_{BE}} \cdot \frac{\partial}{\partial P}\left(\frac{\Delta I_B}{I_B}\right). \tag{4.6}$$

It is noteworthy that T_j is the junction temperature (i.e., actually an unknown of the problem) but under the approximation of a low power dissipation $T_j = T_B$ the thermal resistance is directly evaluated. On the other hand, considering (4.6) it is possible to extend the technique and obtain an expression of R_{TH} dependent on both T_B and P_D; in fact, considering $C_1 = \frac{k}{E_g - q \cdot V_{BE}} \cdot \frac{\partial}{\partial P}\left(\frac{\Delta I_B}{I_B}\right)$ it is possible to obtain from (4.6):

$$R_{TH}^2 + \left(\frac{2 \cdot T_B \cdot C_1 \cdot P_D - 1}{P_D^2 \cdot C_1}\right) \cdot R_{TH} + \frac{T_B^2}{P_D^2} = 0, \tag{4.7}$$

and thus

$$\begin{aligned} R_{TH}(T_B, P_D) = &- \left(\frac{2 \cdot T_B \cdot C_1 \cdot P_D - 1}{2 \cdot P_D^2 \cdot C_1}\right) + \\ &- \sqrt{\left(\frac{2 \cdot T_B \cdot C_1 \cdot P_D - 1}{2 \cdot P_D^2 \cdot C_1}\right)^2 - \frac{T_B^2}{P_D^2}}. \end{aligned} \tag{4.8}$$

It is worth noting that (4.8), as an extension of the formulation proposed in [61], takes into account the dependence on T_B and P_D, but the range of dissipated power is limited to the submultiplication zone. On the contrary, by using (4.6), the dependence on dissipated power is totally neglected while the T_B dependence is taken into account.

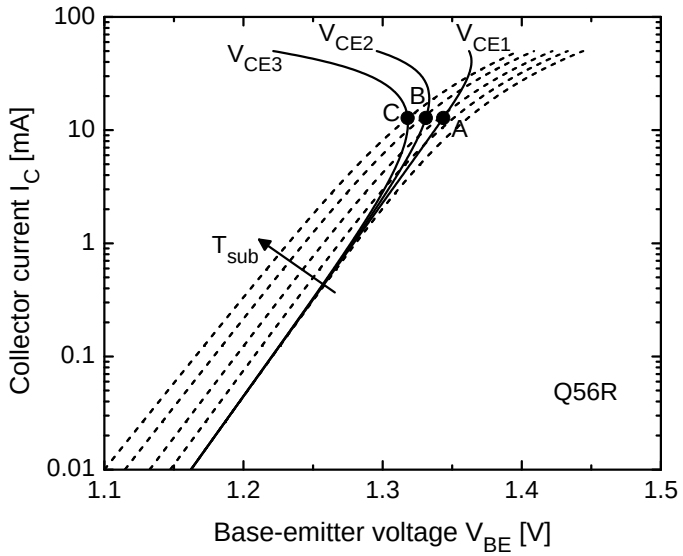


Figure 4.1: Collector current versus base-emitter voltage for (solid lines) various V_{CE} at a fixed T_B. Calibration (dashed lines) curves at various T_{sub} at a low V_{CE} are reported as well.

.

W. Liu and A. Yuksel technique [62]. The method proposed in 1995 by Liu and Yuksel [62] relies on the measurements of several I_C versus V_{BE} characteristics at a low V_{CE0} (i.e., low dissipated power) for various baseplate temperatures that are used as calibration curves. Then other measurements are performed at a fixed baseplate temperature T_B in order to obtain the flybacks for prescribed collector-emitter voltages (at values higher than V_{CE0}) as shown in Fig. 4.1. Considering the points of intersection between calibration curves and higher voltage curves (i.e., A, B and C in Fig. 4.1) at a fixed I_C it is possible to write, according to [62]

$$\begin{aligned} T_{jA} &= T_B + R_{TH_A} \cdot P_{DA} = \\ &= T_{subA} + R_{TH_0} \cdot P_{D0} \\ T_{jB} &= T_B + R_{TH_0} \cdot P_{DB} = \\ &= T_{subB} + R_{TH_0} \cdot P_{D0} \end{aligned}, \tag{4.9}$$

where T_{subA} and T_{subB} are the baseplate temperatures for calibration curves in points A and B, respectively; $P_{DA} = V_{CE1} \cdot I_C$, $P_{DB} =$

$V_{CE2} \cdot I_C$ and $P_{D0} = V_{CE0} \cdot I_C$ are the dissipated powers. Iterating the procedure from (4.9) it is possible to obtain

$$\begin{aligned} T_{j0} &= T_B + R_{TH_0} \cdot 0 = T_B = \\ &= T_{sub0} + R_{TH_0} \cdot P_{D0} \end{aligned}, \tag{4.10}$$

where T_{sub0} is evaluated by extrapolation (at $V_{CE} = 0$) from the curve obtained by plotting the calibration temperatures T_{sub} versus the V_{CE} values. Thermal resistance extraction with this method is not extremely accurate because of the approximation done in considering R_{TH} as a function of dissipated power and not of the baseplate temperature which, in turn, during the baseplate temperatures sweep at low V_{CE} cannot be neglected if there are thermal non-linearities (i.e., thermal conductivity dependent on temperature). In fact, in this case, increasing the baseplate temperature at a fixed I_C (see also Marsh technique and Fig. 4.3) R_{TH} increases almost linearly with temperature.

C. Bovolon *et al.* technique [60]. This technique is a direct extension of [58] with the difference that R_{TH} is considered dependent on dissipated power P_D, as well as on T_B. For this reason $\Delta P = P_{D2} - P_{D1}$ and $\Delta T_B = T_{B2} - T_{B1}$ have to be chosen small enough because the β linearization is considered a "local" property, with this hypothesis it's possible to apply the (4.5) to high powers and current densities to extract $R_{TH}(P_D, T_B)$. It is noteworthy that [58] and [60] techniques are strictly dependent on the difference between baseplate temperatures used in the chosen working points; this means that an error in chuck temperature control affects directly the measurement of thermal resistance. If ΔT_B is very small the percentage error in evaluation can be very high; just to fix the ideas, if temperature precision is ± 0.1 K (i.e., this is the case of a good thermal chuck) it is possible that in the two temperature points evaluation a maximum error (absolute value) of 0.2 K can be committed which implies an error of 10% in the R_{TH} evaluation in the case of a $\Delta T_B = 2$ K. This, obviously, makes it necessary to use a higher ΔT_B in R_{TH} evaluation but keeping in mind that an higher R_{TH}, on the other hand, voids the hypothesis of local linearization of β. In general a good compromise is to choose a value of 10-20 K for ΔT_B.

D. Marsh technique [63]. The technique proposed by Marsh is based on the measurements of I_C versus V_{CE} characteristics at three differ-

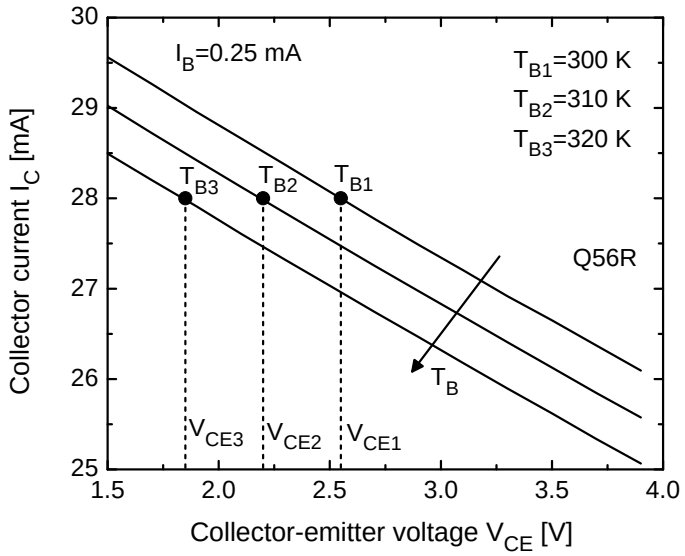


Figure 4.2: Collector current versus collector-emitter voltage at three temperatures (namely, 300, 310, and 320 K). Data refers to measurements of GaAs HBT Q56R with a fixed base current. Points at the same I_C for various T_B are indicated.

.

ent baseplate temperatures and a fixed base current I_B as reported in Fig. 4.2 showing the characteristics of a GaAs HBT (see Section 3.1.3 for description of device). Due to the slope of collector current (if base width modulation can be neglected) it is possible to select three points (one point for curve) having the same collector current (as reported in Fig. 4.2) meaning that in those points β is the same and therefore it is possible to affirm that the junction temperature T_j is the same for each working point. Moreover, this technique assumes that R_{TH} is not constant but can be considered dependent on T_B at each junction temperature T_j. This can be proved by means of 3-D FEM simulations as shown in Fig. 4.3

$$R_{TH}\left(T_j, T_B\right) = A + T_B \cdot B, \tag{4.11}$$

where A and B depend on junction temperature. Combining (4.3) and (4.11) for each single polarization point, and remembering that R_{TH} is

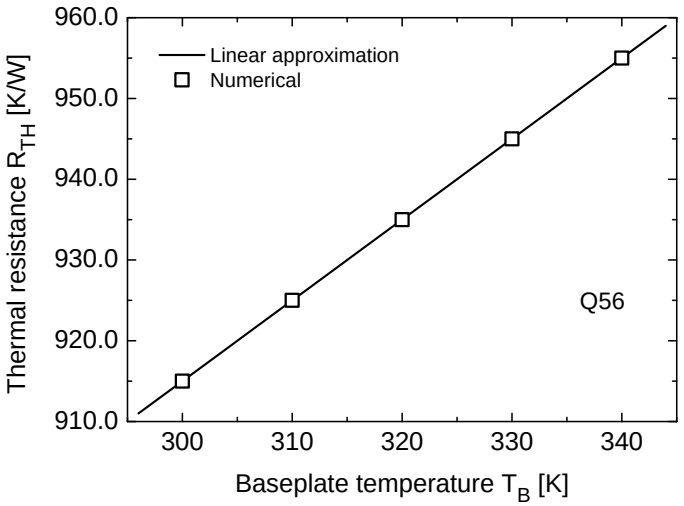


Figure 4.3: Thermal resistance dependence on baseplate temperature for a fixed T_j. Data refers to 3-D numerical simulations of GaAs HBT Q56.

.

different for each chosen point, it is possible to find

$$\begin{cases} T_j = T_{B1} + P_{D1} \cdot (A + T_{B1} \cdot B) \\ T_j = T_{B2} + P_{D2} \cdot (A + T_{B2} \cdot B) \\ T_j = T_{B2} + P_{D3} \cdot (A + T_{B3} \cdot B) \end{cases}. \tag{4.12}$$

With the simple resolution of the system (4.12) it is possible to find T_j and, consequently, R_{TH} for the particular biasing condition.

E. Menozzi *et al.* technique [64]. The technique proposed by Menozzi and his coworkers is based on the observation that when the Early effect is negligible (as for GaAs HBTs). Thus, it is possible to write that

$$I_C = I_{C00} \cdot [1 + k \cdot (T_j - T_{j00})], \tag{4.13}$$

where I_{C00} and T_{j00} are collector current and junction temperature, respectively, corresponding to a reference point in which $V_{CE} = V_{CE0}$

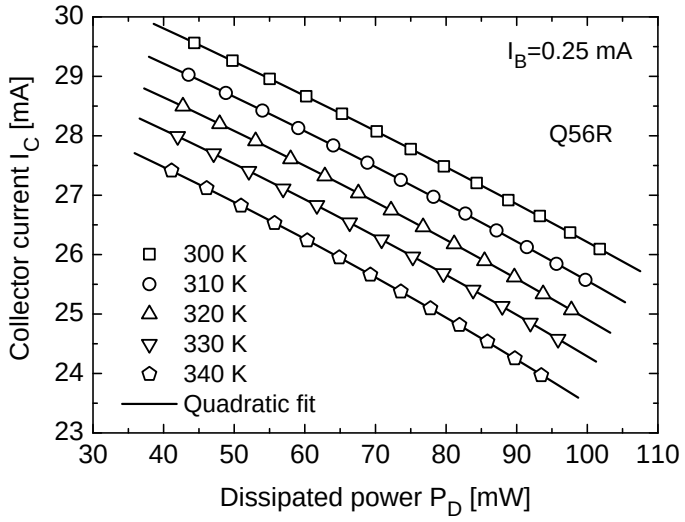


Figure 4.4: Collector current dependence on dissipated power for various T_B. (Symbols) measurements and (solid line) quadratic fit refer to an HBT GaAs Q56R.

.

and $T_B = T_{B0}$ so that $P_{D00} = V_{CE0} \cdot I_{C00}$ and k[1] is a constant. Considering (4.3) it is possible to obtain that

$$T_j - T_{j00} = T_B - T_{B0} + R_{TH} \cdot P_D - R_{TH00} \cdot P_{D00}, \qquad (4.14)$$

where R_{TH00} is the thermal resistance at the working point (I_{C00},V_{CE0}). In this technique, the thermal resistance is considered dependent on both dissipated power and baseplate temperature through the following relationship that has to be written at each T_B

$$R_{TH} = R_{TH0} + \frac{dR_{TH}}{dP_D} \cdot (P_D - P_{D0}), \qquad (4.15)$$

where $R_{TH0} = R_{TH}(T_B, P_{_{D0}})$ and $P_{D0} = V_{CE0} \cdot I_C(T_B, V_{CE0})$. By substituting (4.14) and (4.15) into (4.13) it can be seen that there is a quadratic dependence of I_C on P_D at each baseplate temperature as clearly visible in Fig. 4.4

$$I_C(T_B, P_D) = a_2 \cdot P_D^2 + a_1 \cdot P_D + a_0, \qquad (4.16)$$

[1] In the original technique k sign is minus, here the plus sign is used to maintain generality.

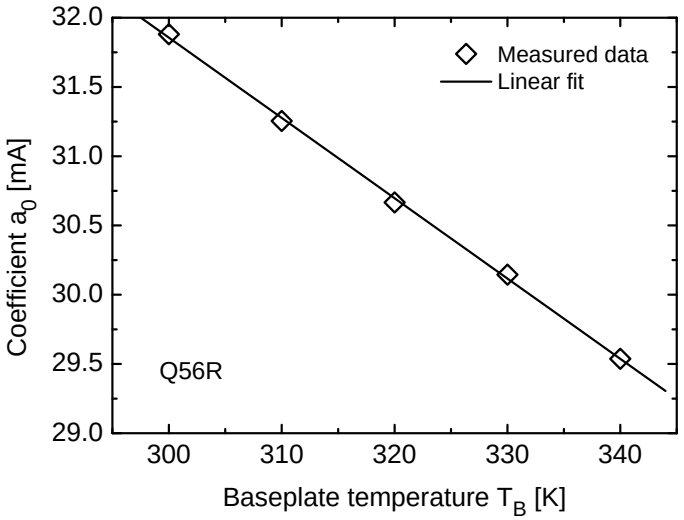


Figure 4.5: Coefficient a_0 versus T_B for an HBT GaAs Q56R. (Symbols) measured data and (solid line) linear fit are reported as well.

.

where

$$\begin{aligned} a_2 &= I_{C00} \cdot k \cdot \frac{dR_{TH}}{P_D} \\ a_1 &= I_{C00} \cdot k \cdot \left(R_{TH0} - \frac{dR_{TH}}{P_D} \cdot P_{D0} \right) \\ a_0 &= I_{C00} \cdot [1 + k \cdot (T_B - T_{B0} - R_{TH00} \cdot P_{D00})] \end{aligned} \quad , \tag{4.17}$$

Observing that a_0 depends only on T_B (see Fig. 4.5) by means of a linear fitting it is possible to obtain $I_{C00} \cdot k$ from a_0 and dR_{TH} from a_2 as well as R_{TH0} from a_1 at each T_B.

4.1.2 A novel approach to measure the thermal resistance of Si bipolar transistors subject to Early effect

Almost all methods proposed were devised for GaAs HBTs and disregard Early effect. Obviously this may lead to unacceptable errors when applied to Si BJTs owing to the fact that the Early effect is "interpreted" as an additional thermal effect (as it will clearly shown in the next section). For this reason a novel method was proposed in this work as an extension of the well-known [58] approach; this method is also referred to as *Voltage Plane* method. If the Early effect can be neglected, the slope of the V_{CB}-V_{BE} characteristics at a fixed I_E is reduced because of the sole self-heating. This thermal contribution γ_{SH} can be expressed as

$$\gamma_{SH} = \frac{-\phi\left(I_E\right) R_{TH} I_E}{1+\phi\left(I_E\right) R_{TH} I_E} \approx -\phi\left(I_E\right) R_{TH} I_E. \tag{4.18}$$

It is worth noting, that (4.18) coincides with the method described by Dawson [58] and $\phi(I_E)$ can be determined by the slope of V_{BE}-T_B curves with very low dissipated power at a fixed I_E ($T_j \approx T_B$). Nonetheless, if the Early effect is not negligible, (4.18) leads to an overestimation of R_{TH} values (i.e., in silicon BJTs, where Early effect may play an important role). More generally speaking, the slope reduction of the V_{CB}-V_{BE} (with V_{CB} sufficiently lower than the open-emitter breakdown voltage BV_{CBO}) is due both to the electrical phenomenon (Early effect) and self heating. Hence, it is possible to write

$$\gamma = \gamma_{EL} + \gamma_{SH}\left(I_E\right), \tag{4.19}$$

where γ is the slope of V_{CB}-V_{BE} characteristics at a fixed emitter current I_E. Considering that the sole electrical contribution γ_{EL} is constant in a wide emitter current range as stated in [66], it is possible to evaluate thermal resistance by two sets of measurements. In fact, as can be observed from (4.18), the term γ_{SH} is dependent upon I_E and considering two biasing conditions (two different emitter currents I_{E1} and I_{E2}) it is possible to free the thermal resistance evaluation from the electrical feedback (almost independent of the current level). From (4.19) it is possible to obtain

$$\begin{aligned} \gamma_1 &= \gamma_{EL} + \gamma_{SH}\left(I_{E1}\right) \\ \gamma_2 &= \gamma_{EL} + \gamma_{SH}\left(I_{E2}\right) \end{aligned}, \tag{4.20}$$

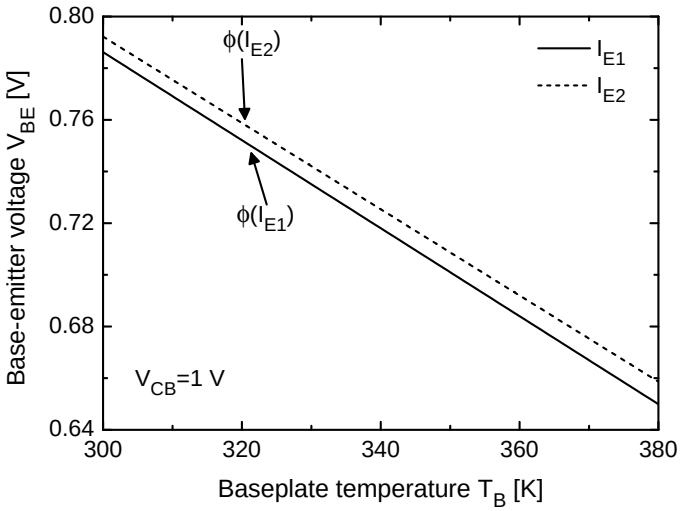


Figure 4.6: V_{BE}-T_B characteristics for a fixed $V_{CB} = 1$ V and two different emitter currents used to evaluate $\phi(I_{E1})$ and $\phi(I_{E2})$.

.

and hence

$$R_{TH} = \frac{\gamma_1 - \gamma_2}{\phi\left(I_{E2}\right) I_{E2} - \phi\left(I_{E1}\right) I_{E1}}, \tag{4.21}$$

where $\phi(I_{E1})$ and $\phi(I_{E2})$ can be obtained by the slopes of V_{BE}-T_B for the two biasing conditions I_{E1} and I_{E2}, respectively, as shown in Fig. 4.6. γ_1 and γ_2 can be evaluated as shown in Fig. 4.7. It should be considered that this approach does not affect the correct evaluation of thermal resistance also when Early effect is negligible (γ_{EL} can be neglected with respect to γ_{SH}).

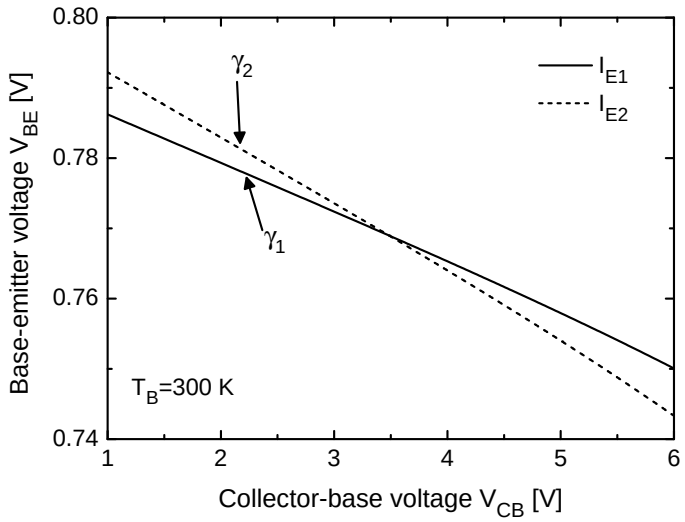


Figure 4.7: V_{BE}-T_B characteristics for a fixed $T_B = 300$ K and two different emitter currents used to evaluate γ_1 and γ_2.

.

4.1.3 Comparison of thermal resistance measurement techniques

The thermal resistance, in general, is an unknown of the problem, so, apparently, there is no possibility to compare different techniques based on DC characterization. The strategy conceived to evaluate and compare the accuracy of all the techniques is based on the use of "simulated experiments" where the thermal resistance in assigned by the user[2]. It is explained as follows.

1. Isothermal (i.e., pulsed) measurements were performed on GaAs HBTs and bulk-silicon BJTs with emitter area amounting to 60 and 20 μm^2, respectively, at various baseplate (i.e., thermochuck) temperatures. With this approach it was possible to extract all the standard and temperature-scaling parameters of detailed in-house DC electrothermal (ET) transistor models that include the following physical mechanisms:
   - Early and high-injection effects.
   - Series resistances.
   - Dependence on temperature of current gain and base-emitter voltage.
   - Impact ionization.
2. Accurate models were then employed to perform ET "simulated experiments" under any biasing conditions by adopting chosen values of self-heating thermal resistances (R_{TH}) in order to account for the ET feedback.
3. Realistic simulation results are obtained by activating nonlinear thermal effects (i.e., by including the thermal conductivity dependence on temperature) through the Kirchhoff transformation approach [65, 67] (see also Chapter 5 for implementation), thereby enabling the R_{TH} dependence on both dissipated power (P_D) and baseplate temperature (T_B).

By resorting to this strategy it is possible to "measure" (by calibrated simulations) all the DC characteristics needed by any of the measurement techniques to extract the thermal resistance. As formerly stated,

[2]The thermal resistance is obtained by using the DC "electrical" characteristics, that is, the R_{TH} input value is not available to the measurement techniques.

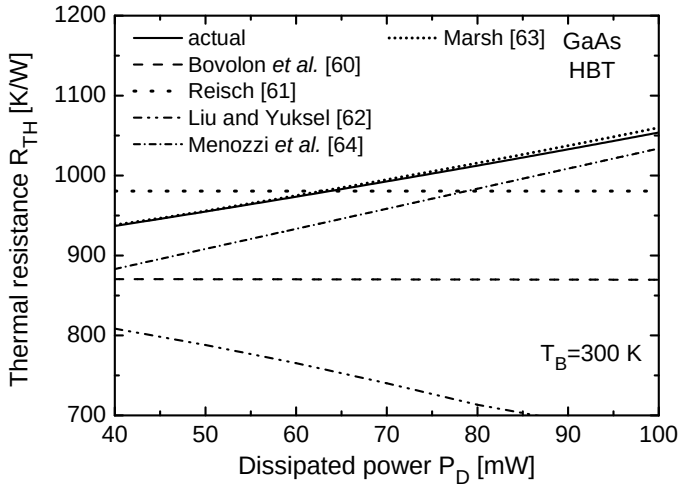


Figure 4.8: Self-heating thermal resistance of a GaAs HBT having an emitter area of 60 μm^2 as a function of dissipated power at $T_B = 300$ K: comparison between approaches [60–64] and the actual R_{TH}.

.

the key point of this approach is that, in contrast to real experiments, the self-heating thermal resistances (and their sensitivity on P_D and T_B) are known, and can be used to compare the accuracy of the extraction methods. It should be noted that R_{TH} values of 870 and 470 K/W corresponding to $T_B = 300$ K were adopted for GaAs HBTs and bulk-silicon BJTs, respectively, as suggested by calibrated 3-D thermal-only FEM simulations.

Comparison of techniques for GaAs HBTs. Fig. 4.8 reports the R_{TH} variation with dissipated power for a GaAs HBT having an area of 60 μm^2. A comparison is carried out between the results of the experimental procedures [60–64] and the actual R_{TH} at $T_B = 300$ K. As can be seen, the technique of Marsh [63] guarantees an excellent fit over the whole P_D range. Relatively good accuracy is also provided by Menozzi *et al.* [64], Bovolon *et al.* [60], and Reisch [61], although the latter two predict a R_{TH} that is insensitive to P_D. In [60], this is due to the approximations in the derivation of the approach; in particular, in this method the evaluation of R_{TH} is based on the use of biasing point at two base-

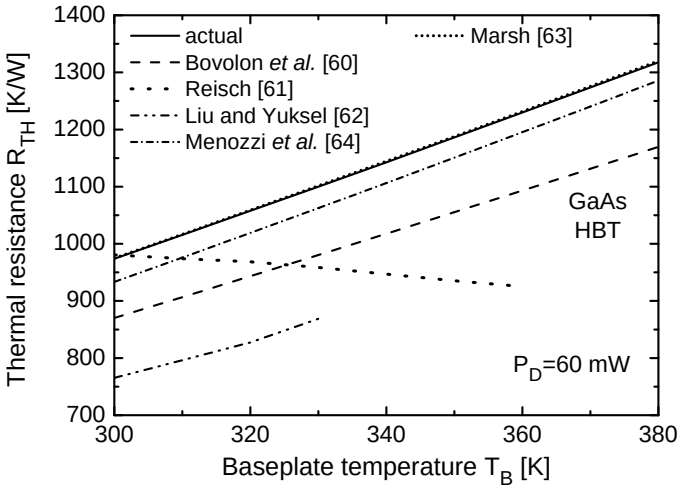


Figure 4.9: Self-heating thermal resistance of a GaAs HBT having an emitter area of 60 μm^2 as a function of baseplate temperature at $P_D = 60$ mW: comparison between approaches [60–64] and the actual R_{TH}.

.

plate temperatures. However, this is not strictly correct owing to the fact that the thermal resistance depends on baseplate temperature while in this approach such a dependence is totally neglected. In [61], the self-heating thermal resistance dependence on dissipated power is not accounted for. In contrast, the quite intricate technique proposed by Liu and Yuksel [62] leads to a relatively high error of about 14% at $P_D = 40$ mW and determines a physically meaningless R_{TH} reduction with P_D because of the simplifying assumptions (the thermal resistance dependence on baseplate temperature is neglected) on which the approach is based. The behavior of R_{TH} as a function of baseplate temperature at $P_D = 60$ mW is depicted in Fig. 4.9. Once again, the strategy proposed by Marsh [63] provides R_{TH} values virtually coinciding with the actual ones. The procedure of Menozzi and his coworkers [64] also guarantees a good accuracy: an average error of about 3% is found over the whole T_B range. The techniques proposed by Bovolon et al. [60] and Liu and Yuksel [62] correctly predict the R_{TH} increase with T_B, although they provide average errors of about 11% and 21%, respectively. Lastly, the technique of Reisch [62] accurately evaluates the R_{TH} value

at $T_B = 300$ K (as expected from the previous analysis), but improperly determines an "unphysical" reduction of the self-heating thermal resistance as the baseplate temperature increases, as a consequence of a strict assumption when deriving the formula on which the procedure relies. These results were confirmed by a large number of similar analyses. As a rule of the thumb, the method proposed by Marsh [63] always results in an admirable accuracy in evaluating the thermal resistance of a GaAs HBT over wide P_D and T_B ranges. Reliable results can also be obtained by applying the techniques developed by Menozzi *et al.* [64] and Bovolon *et al.* [60], whereas the other approaches are of little use in most applications since they are based on simplifying assumptions that may considerably affect the accuracy of the results.

Comparison of techniques for Si BJTs. In order to compare all the techniques under analysis in the case of silicon BJTs, the methods developed by Marsh [63] and Menozzi *et al.* [64] were extended to account for silicon BJTs, in which the common-emitter current gain exhibits a positive temperature coefficient. All other techniques, including the one proposed in this work, can be applied regardless of the bipolar transistor category. The bulk-silicon transistor analyzed is subject to a base width modulation with an Early voltage $V_A = 25$ V and characterized by an open-emitter breakdown voltage $BV_{CBO} = 16$ V. The thermal resistance is plotted as a function of dissipated power at $T_B = 300$ K in Fig. 4.10. As it was expected, the methods from Marsh [63], Menozzi *et al.* [64], and Bovolon *et al.* [60], lead to considerable evaluation errors (up to 200%) as a consequence of the Early effect. In fact in these approaches the Early effect is erroneously interpreted as an additional thermal feedback mechanism. A similar behavior can be observed in Fig. 4.11, which illustrates the thermal resistance as a function of baseplate temperature at $P_D = 20$ mW. It is noteworthy that the proposed approach, although not suitable for predicting the R_{TH} sensitivity with P_D, still provides a good accuracy within the whole P_D and T_B ranges (errors spanning from 0.5% up to 2.5% are found). As evident from Figs. 4.10 and 4.11 employing the approaches of Liu and Yuksel [62] and Reisch [61] can lead to acceptable results, although the latter noticeably fails when the baseplate temperature exceeds 350 K. It is important to underline that also for silicon BJTs with negligible Early effect (ideally with $V_A \to \infty$) the method proposed by Marsh [63] turns out to be

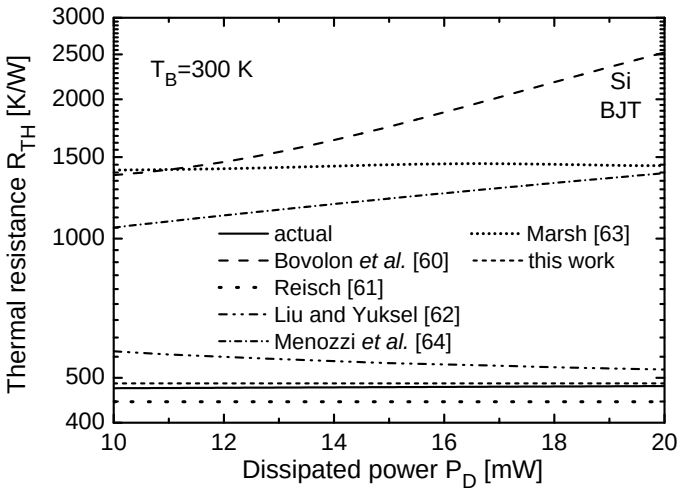


Figure 4.10: Self-heating thermal resistance of a Si BJT having an emitter area of 20 μm^2 as a function of dissipated power at $T_B = 300$ K: comparison between approaches [60–64], the one proposed, and the actual R_{TH}.

.

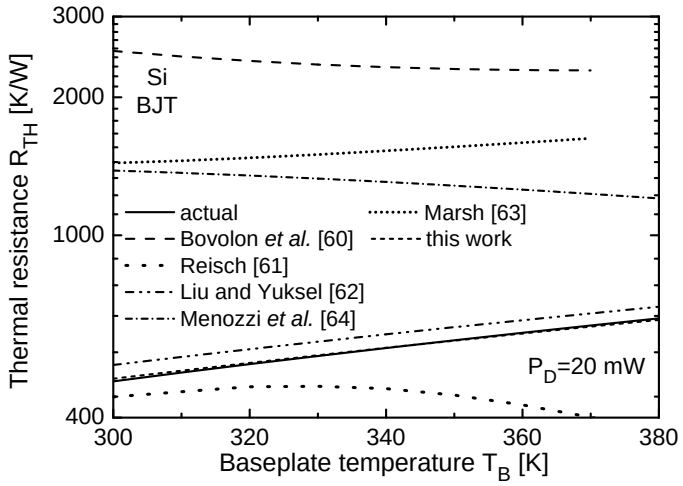


Figure 4.11: Self-heating thermal resistance of a Si BJT having an emitter area of 20 μm^2 as a function of baseplate temperature at $P_D = 20$ mW: comparison between approaches [60–64], the one proposed and the actual R_{TH}.

.

the most accurate. All the results of this study are summarized in [68].

Thermal resistance measurements of real devices. The techniques analyzed in the previous section have been successfully used to measure the thermal resistance of a number of different bipolar transistors. Fig. 4.12 depicts the experimentally determined self-heating thermal resistance for a GaAs HBT with emitter area of 60 μm^2 having a horseshoe shaped emitter (Q56R). The extraction was performed in the power range 40-90 mW by keeping the baseplate temperature constant at 300 K. From the results obtained in the previous analysis the curve obtained by applying the approach of Marsh [63] is expected to provide the most accurate results. In Fig. 4.13, an experimental analysis is presented to demonstrate the influence of emitter area (A_E) and aspect ratio (AR) of the emitter stripe on the thermal resistance of SOG BJTs. The investigation was conducted by keeping constant the spacing between the edges of the emitter and the island sidewalls. Two methods, namely, that presented in this work and the one of Bovolon et al. [60], were em-

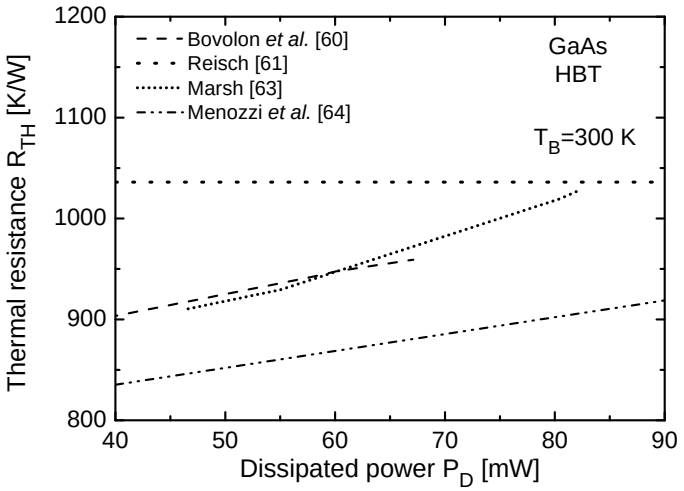


Figure 4.12: Measured self-heating thermal resistance of a GaAs HBT having an emitter area of 60 μm2 as a function of dissipated power at $T_B = 300$ K: comparison between approaches [60, 61, 63, 64].

.

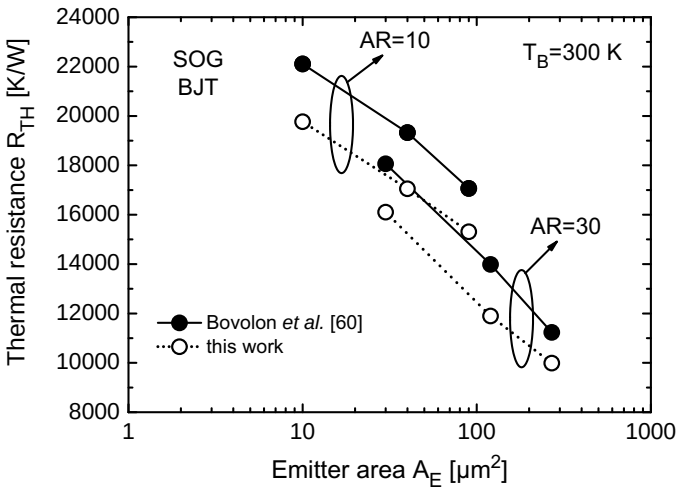


Figure 4.13: Measured self-heating thermal resistance of a SOG BJT versus emitter area at $T_B = 300$ K: comparison between approach [60] and the one proposed. Data for various aspect ratios are reported as well.

.

ployed to evaluate the thermal resistance. It is interesting to note that both strategies predict an almost linear lowering of thermal resistance by increasing A_E. Moreover it is worth noting that the method from Bovolon *et al.* [60] determines R_{TH} values 15-20% higher than those calculated by the proposed approach. This discrepancy can be explained by recalling that the results obtained with the Bovolon *et al.* [60] approach are most likely overestimated due to the Early effect (V_A was extracted to be about 35 V for the SOG devices under analysis), although the electrically-induced error is not as large as in silicon transistors with thermal resistances lower than a few thousand K/W. In fact in SOG devices the thermal feedback (thermal resistances as high as 20000 K/W) is predominant over the electrical feedback (Early effect).

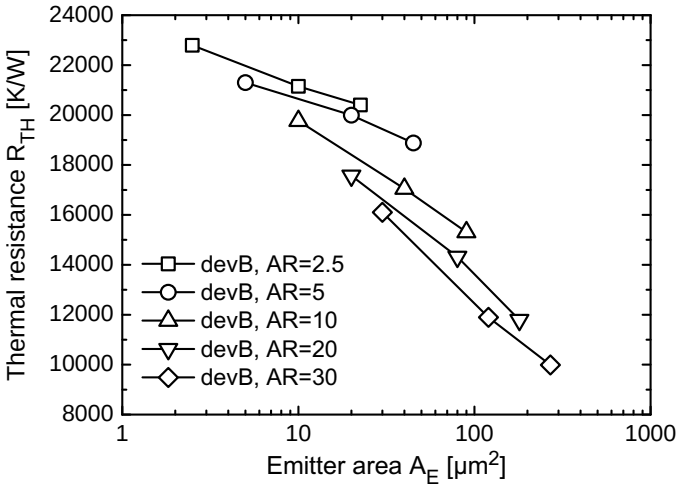


Figure 4.14: Experimental thermal resistance as a function of emitter area for various values of the aspect ratio of the emitter stripe. All data refer to geometry devB.

4.1.4 Experimental case studies

The described techniques were applied to a number of devices of different technologies in order to evaluate the thermal resistance. In particular, the method proposed in this work was used to experimentally evaluate the impact of scaling both on SOG bipolar transistors and SiGe HBTs.

Thermal resistance evaluation for SOG devices. By using the Voltage Plane technique it is possible to evaluate the impact of emitter area on thermal resistance for different aspect ratios of the emitter stripe (see Section 3.1.2 for the structure details). Fig. 4.14 depicts the almost linear increase of the thermal resistance as A_E reduces logarithmically for a fixed aspect ratio (as it was previously obtained in Section 3.1.2). It is noteworthy that the absolute value of the slope of each line increases with AR; in other words, the influence of the emitter area on the thermal behavior of the device is more pronounced for higher aspect ratios. This is even more evident in Fig. 4.15, where R_{TH} is plotted as a function of the emitter area for different distances of the emitter stripe from the trench sidewalls. As can be seen, if A_E and the spacing between emitter

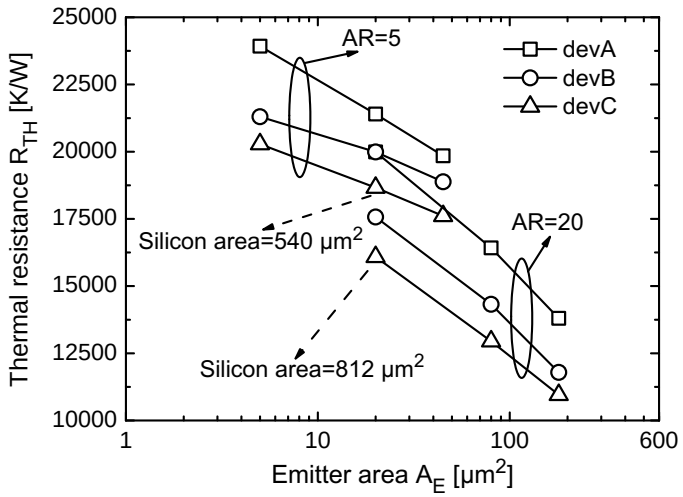


Figure 4.15: Experimental thermal resistance as a function of emitter area for geometries (squares) devA, (circles) devB, and (triangles) devC; two aspect ratio values (5 and 20) are considered for the emitter window. Values of the silicon island area are reported only for devices with $A_E = 20$ μm^2 and geometry devC.

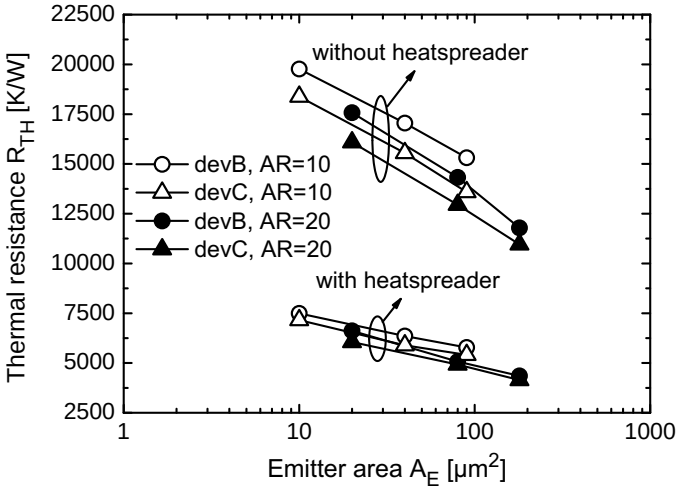


Figure 4.16: Experimental thermal resistance as a function of emitter area for devices with and without a 2-μm-thick AlN layer deposited on the emitter-base contact side of the devices. Geometries devB and devC are considered, each for two values of the aspect ratio (10 and 20) of the emitter stripe.

stripe and trench are kept constant, the thermal resistance increases by reducing the aspect ratio (as also seen in Section 3.1.2). This result was to be expected, since the volume of the fully-isolated silicon island decreases, thereby reducing the heat spreading from the dissipating region; besides, the heat removal is counteracted by the AR lowering itself. Reducing AR from 20 to 5 while keeping constant emitter area to 20 μm^2 and distances S_1 and S_2 from the trenches (devC layout), leads to a 34% decrease of the silicon island area (from 812 μm^2 to 540 μm^2). The addition of an AlN layer can help to counteract thermal issues as previously seen in Chapter 3. The beneficial effect of AlN is quantified in Fig. 4.16 that depicts the influence of a 2-μm-thick AlN layer, deposited onto the emitter-base side of the device, upon the thermal resistances of a series of SOG BJTs with different layouts, as well as with a variety of areas and aspect ratios of the emitter stripe. If AlN-covered devices are compared to their AlN-free counterparts it can be found that the introduction of the AlN layer leads to a reduction in R_{TH} higher than 60% for all the geometries considered in this figure, as a consequence of the

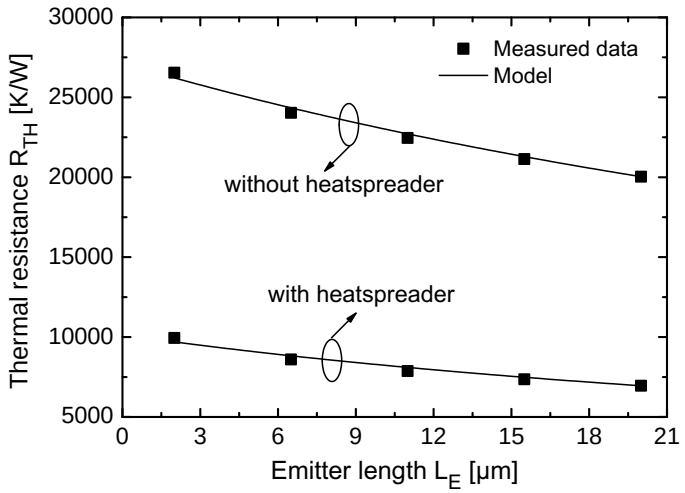


Figure 4.17: Thermal resistance as a function of emitter length L_E for SOG BJTs sharing devA layout and emitter width equal to 1 μm, with and without a 2-μm-thick AlN layer deposited onto the front-wafer; (symbols) numerical data are compared to (solid lines) model (2.34) with optimized parameters.

high-thermal-conductivity AlN path, which favors the heat transfer from the active area to both the pads and glass substrate.
Measured data were also used to optimize the parameters of the model, according to (2.34), that describe the thermal resistance dependence on the scaling of emitter geometry. The parameters R_{TH_A}, R_{TH_W}, and R_{TH_L} were extracted considering as reference devices two SOG BJTs with devA layout and emitter area amounting to 1×20 μm^2, without and with a 2-μm-thick AlN layer as a heatspreader, respectively. It was found that the attained values lead to errors lower than 2% (without AlN layer) and 3.5% (with the AlN layer). Fig. 4.17 illustrates the comparison between model and numerical data as a function of emitter length for both transistor categories. The model parameters are reported in Table 4.1 for both AlN-free and AlN-cooled devices. Further details can be found in [69].

Table 4.1: Empirical parameters of (2.34) optimized for the SOG BJTs under investigation.

Parameters	**AlN-free BJT**	**AlN-cooled BJT**
$R_{TH_{ref}}$ [K/W]	20034	6954
$A_{E_{ref}} = W_{E_{ref}} \cdot L_{E_{ref}}$ [μm^2]	1×20	1×20
R_{TH_A}	-0.0227	-0.0219
R_{TH_W}	0.0417	0.0423
R_{TH_L}	0.2848	0.3364

Thermal resistance evaluation for SiGe HBTs. As previously done for SOG BJTs, it is also possible to evaluate the impact of emitter geometry on the thermal resistance for SiGe HBTs. Although several devices both from IFX and STM were measured, for sake of brevity only results relative to STM devices are reported. Fig. 4.18 depicts the R_{TH} dependence upon L_E for various emitter widths W_E and two different technologies, referred to as B3T (presenting $f_{MAX} = 350$ GHz) and B4T (with an f_{MAX} of 400 GHz).
It is evident, from Fig. 4.18, that for a fixed emitter geometry, the thermal resistances of B4T devices are 20-27% higher than the B3T coun-

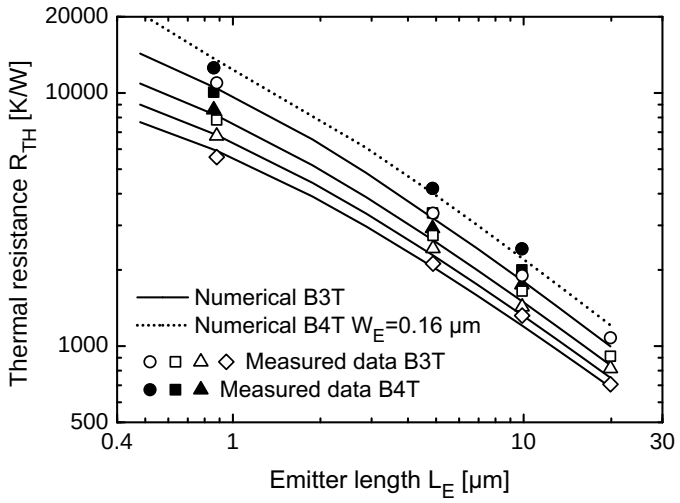


Figure 4.18: (Symbols) experimental and (lines) numerical self-heating thermal resistance as a function of emitter length L_E for various W_E. As concerns B3T test structures, W_E amounts to (open circles) 0.13, (open squares) 0.38, (open triangles) 0.63, and (open rhombi) 0.88 µm; for B4T HBTs, W_E is equal to (filled circles) 0.16, (filled squares) 0.36, (filled triangles) 0.61 µm. FEM simulations referring to B4T transistors are shown only for $W_E = 0.16$ µm (dotted line).

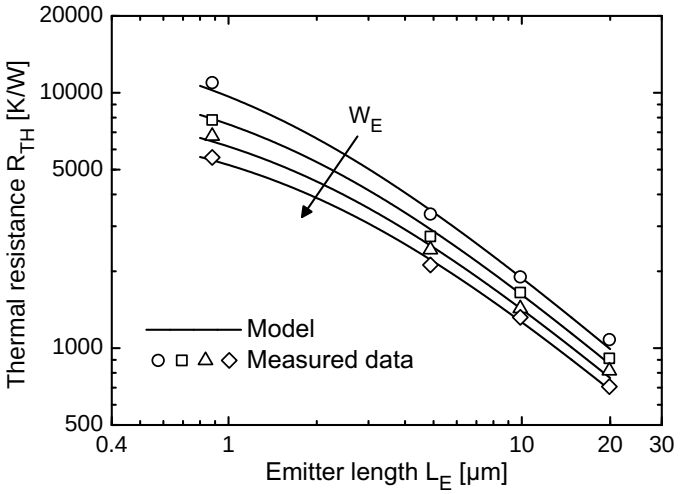


Figure 4.19: Self-heating thermal resistance as a function of emitter length L_E for W_E values of 0.13, 0.38, 0.63, and 0.88 μm for SiGe HBTs in B3T technology: comparison between (solid lines) the scalable model (2.34) and (symbols) experimental results.

terparts. This result was expected, since B4T HBTs are characterized by larger trench depths (e.g., 3.5 μm instead of 2.5 μm) and a smaller spacing between heat source and shallow trench. Moreover, the thermal resistance sensitivity to emitter length is markedly high for small emitter lengths while drastically lessening for longer devices regardless of the technology, and independently of emitter width.

Measured data were employed to optimize the parameters in (2.34) to describe the thermal resistance dependence on the scaling of emitter geometry for both technologies.

Data measured and modeled for B3T structures are reported in Fig. 4.19. As can be seen, a fairly good agreement is obtained. The average and peak discrepancies amount indeed to about 3.5% and 7% for both B3T and B4T HBTs. Model parameters are reported in Table 4.2.

Table 4.2: Empirical parameters of (2.34) optimized for the SiGe HBTs under investigation.

Parameters	**B3T**	**B4T**
$R_{TH_{ref}}$ [K/W]	1901	2423
$A_{E_{ref}} = W_{E_{ref}} \cdot L_{E_{ref}}$ [µm^2]	0.13×9.88	0.16×9.86
R_{TH_A}	0.06206	0.01550
R_{TH_W}	0.02278	0.01907
R_{TH_L}	0.83179	0.71351

4.2 Measurement techniques for the thermal impedance.

The transient thermal impedance represents the thermal response of the system to a unitary power step and is defined, as previously stated, like

$$Z_{TH}(t) = \frac{T_j(t) - T_B}{P_D}, \tag{4.22}$$

where T_j is the base-emitter junction temperature, T_B is the baseplate (i.e., thermochuck) temperature, and P_D is the power constantly dissipated by the device. The experimental evaluation of the thermal impedance versus time can be performed through the "classical" electrical methods described in the early Blackburn and Oettinger paper [70], namely, the *pulsed heating curve* and the *cooling curve* techniques. Both are based on the monitoring of a temperature-sensitive electrical parameter acting as a thermometer (e.g., the forward voltage drop across a diode). The temperature T_j is, in fact, generally evaluated by observing the behavior of an electrical temperature-sensitive parameter, like the base-emitter voltage V_{BE}, used as a thermometer

$$\phi = \left.\frac{\partial V_{BE}}{\partial T_j}\right|_{I_C}. \tag{4.23}$$

This thermometer can be obtained from isothermal (i.e., pulsed) measurements at various fixed I_C, or by using DC measurements at very low dissipated power (so that $T_j \approx T_B$) as depicted in Fig. 4.6. From (4.22) and (4.23) it follows that

$$Z_{TH}(t) = \frac{\Delta V_{BE}(t)}{\phi \cdot P_D} = \frac{V_{BE}(t=0) - V_{BE}(t)}{\phi \cdot P_D}, \tag{4.24}$$

where $\Delta V_{BE}(t)$ is the V_{BE} reduction due to self-heating effects starting from the instant at which the power step is applied (i.e., $t = 0$) to the (arbitrary) time t. When applying the pulsed *heating curve* method [70], the device under test is biased with a single power pulse, and the temperature is evaluated through the measurement of the voltage drop immediately after the pulse end, provided that a small sensing current (sufficiently low not to heat the device) is applied to allow the voltage monitoring. After letting the device completely cool to ambient temperature,

a power pulse of larger length is applied and the temperature is measured again as the pulse is switched off. The procedure is repeated by biasing the device with multiple pulses of increasing duration until steady-state conditions are reached (a typical number of recorded points can be in the order of 50/100). When the cooling curve approach is used, the device is first heated to steady-state conditions; afterward, the applied power is totally switched off and the voltage drop is measured during the cooling phase, provided that a small sensing current is supplied to the diode. The thermal impedance is then evaluated as [70]

$$Z_{TH}(t) = Z^{C}_{TH_{MAX}} - Z^{C}_{TH}(t), \tag{4.25}$$

where $Z^{C}_{TH}(t)$ (C stands for cooling) is the thermal impedance registered during the cooling phase and $Z^{C}_{TH_{MAX}}$ is its maximum (corresponding to the time instant $t = 0$). It should be remarked that, employing the traditional pulsed heating curve technique, it is possible "constructing" the thermal impedance behavior with a logarithmic time resolution, if pulses of logarithmically increasing duration are applied, whilst the cooling curve procedure does not allow acquiring the voltage drop for very short times. Yet, the cooling curve approach is generally preferred since it is simpler and faster. In both the versions of the electrical methods devised to detect the thermal impedance, the voltage drop V across the diode (and, therefore, the junction temperature T_j) is measured after turning off the power pulse(s). This is due to a twofold reason

- When the power excitation is supplied by applying a current pulse to the diode, measuring the voltage drop on a junction while heating the junction itself might be cumbersome.
- If the voltage drop monitoring is carried out after switching off the current pulse, one can calibrate the (voltage-junction temperature) thermometer without resorting to pulsed measurements, i.e., the value of temperature coefficient ϕ can be extracted under DC conditions by simply varying the baseplate temperature T_B, since $T_j \approx T_B$ when the current conducted by the diode is equal to a small sensing current (the dissipated power is low).

4.2.1 On-the-fly cooling curve technique

The "classical" cooling curve approach was extended in this work and an alternative version that can be referred to as *on-the-fly* since the voltage drop across the base-emitter junction of the BJT (V_{BE}) is monitored when a "cooling", yet different from zero, power is still applied to the device. The use of this method is possible since

- The power excitation is not provided by applying a current pulse to the base-emitter junction, but forcing a collector voltage pulse; as a consequence, the detection of the base-emitter voltage drop on the junction is not an issue.
- The thermometer was preliminarily calibrated through isothermal (i.e., pulsed) measurements performed on a wide current range.

This can be demonstrated analytically by assuming that (i) the dynamic behavior of the device can be represented by a single-pole thermal network, i.e., a parallel between a thermal resistance R_{TH} and a thermal capacitance C_{TH}, and (ii) nonlinear thermal effects can be disregarded. Moreover, it is worth noting that the power is reduced from a heating-up value P_{DH} to a "cooling" value P_{DC} (conversely, in the classical technique the power is totally switched off before performing the measurement, i.e., the voltage acquisition) so that $P_{DH} > P_{DC}$. First, P_{DH} is applied to the network until steady-state conditions are reached. The static temperature increase (above baseplate) of the base-emitter junction is denoted with $\Delta T_{jMAX} = T_{jMAX} - T_B$ and is equal (by definition) to $R_{TH} \cdot P_{DH}$. As a second step, the power is suddenly reduced to P_{DC} (at $t = 0$), during the cooling stage the junction temperature rise ΔT_j decreases from $\Delta T_{jMAX} = R_{TH} \cdot P_{DH}$ to the steady-state value $R_{TH} \cdot P_{DC}$ according to the equation

$$\begin{aligned} \Delta T_j^C(t) &= R_{TH}P_{DC} + \\ &+ (\Delta T_{j_{MAX}} - R_{TH}P_{DC}) \exp\left(-\frac{t}{R_{TH}C_{TH}}\right) = \\ &= R_{TH}P_{DC} + \\ &+ R_{TH}(P_{DH} - P_{DC}) \exp\left(-\frac{t}{R_{TH}C_{TH}}\right) \end{aligned} . \tag{4.26}$$

The cooling thermal impedance $Z_{TH}^C(t)$ is defined as the ratio between the temperature rise above baseplate during the cooling stage and the

difference between the heating-up (P_{DH}) and cooling (P_{DC}) powers, hence

$$\begin{aligned} Z_{TH}^{C}(t) &= \frac{T_j^C(t) - T_B}{P_{DH} - P_{DC}} = \frac{\Delta T_j^C(t)}{P_{DH} - P_{DC}} \\ Z_{TH_{MAX}}^{C} &= \frac{\Delta T_{jMAX}^C(t)}{P_{DH} - P_{DC}} = R_{TH}\frac{P_{DH}}{P_{DH} - P_{DC}} \end{aligned}, \tag{4.27}$$

By combining (4.26) with (4.27) it can be deduced

$$\begin{aligned} Z_{TH}^{C}(t) &= \frac{\Delta T_j^C(t)}{P_{DH} - P_{DC}} = \\ &= R_{TH}\frac{P_{DC}}{P_{DH} - P_{DC}} + R_{TH}\exp\left(-\frac{t}{R_{TH}C_{TH}}\right) \end{aligned},$$

and therefore

$$\begin{aligned} Z_{TH_{MAX}}^{C} - Z_{TH}^{C}(t) &= \\ &= R_{TH}\frac{P_{DH}}{P_{DH} - P_{DC}} - R_{TH}\frac{P_{DC}}{P_{DH} - P_{DC}} + \\ &\quad + R_{TH}\exp\left(-\frac{t}{R_{TH}C_{TH}}\right) = \\ &= R_{TH}\exp\left(-\frac{t}{R_{TH}C_{TH}}\right) = Z_{TH}(t) \end{aligned},$$

that proves the validity of the method. This demonstration can be extended to the case of a thermal circuit composed by a series of an arbitrary number of RC networks without any effort.

This technique is preferred with respect to the tradition cooling curve technique because the latter can be critical for the following reason. As previously mentioned, the supplied power cannot be ideally forced to zero during the cooling phase, since a sensing current is needed to allow measuring the V_{BE} evolution; as a consequence, a yet small power is necessarily dissipated by the device. However, it was found that the inaccuracy in the thermal impedance evaluation through (4.25) with $P_{DC} = 0$ (that is the traditional approach) is extremely sensitive to this small (sensing) power when the thermal resistance of the device under test is particularly high, which might lead to relatively high errors when dealing with BJTs characterized by full dielectric isolation (i.e., silicon-on-glass transistors). A simple on-the-fly version of the heating curve

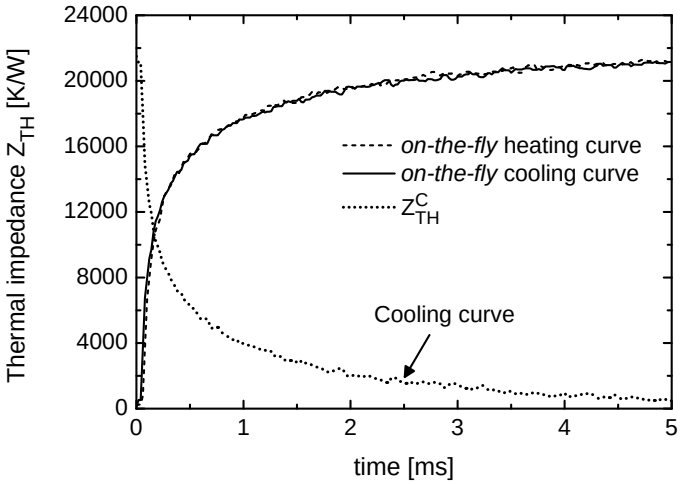


Figure 4.20: Measured thermal impedance versus time of a SOG BJT. Curves obtained both with on-the-fly cooling and heating techniques are represented. Z^C_{TH} is reported as well.

method can be exploited by applying only one V_{CB} pulse while keeping I_E constant; in this case, voltage V_{BE} is monitored during the power excitation. If the pulse is sufficiently long to reach steady-state conditions, one obviously looses the details of the thermal impedance over small times. As can be observed in the Fig. 4.20, results approximately coincide, thus demonstrating that the temperature dependence of the thermal parameters of the materials lead to an almost negligible influence on the thermal impedance behavior for the biasing conditions applied.

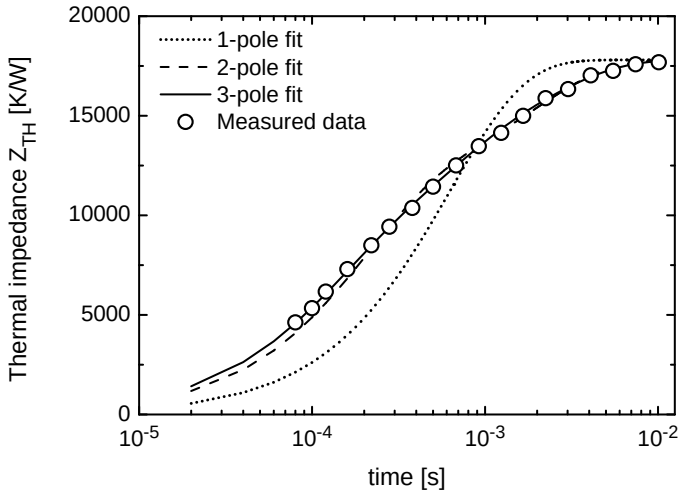


Figure 4.21: Thermal impedance as a function of time obtained by experiments and by (2.1) using different numbers of *RC* ports. All data refer to devB with an emitter area of 1×20 μm^2.

4.2.2 Experimental case studies

An extensive measurement campaign has been performed on SOG bipolar transistors to evaluate their transient response. For the first time thermal transient behavior of AlN-cooled devices was experimentally evaluated. Although the results of this study are thoroughly discussed in [71], it is interesting to report the main findings. The study was performed by using the *on-the-fly* cooling curve technique and resorting to identification software (see Chapter 2 for a description) for equivalent network modeling. It is worth noting (as evident from Fig. 4.21) that a three-pole fit guarantees a sufficient accuracy to correctly describe thermal transient of bipolar transistors.

Layout effects on thermal transient. Fig. 4.22 clarifies the effect of varying emitter area for a prescribed aspect ratio for devA transistors with and without a 2-μm-thick AlN layer acting as a heatspreader. The thermal resistance decreases while increasing emitter area whereas the rise time increases with A_E for devices either without or with an AlN heatspreader. In particular, changing A_E from 20 μm^2 to 180 μm^2 (i.e.,

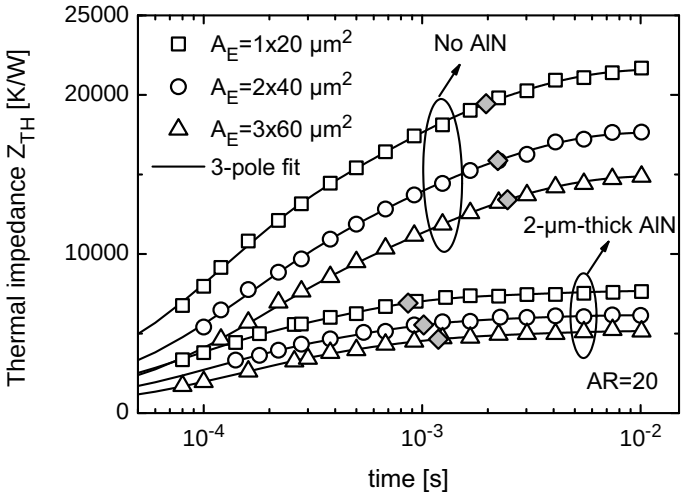


Figure 4.22: Transient thermal impedance of transistors sharing devA layout with and without a 2-μm-thick AlN heatspreader, for different emitter areas by keeping AR unchanged. Both (open symbols) measurements and (solid lines) the 3-pole model are presented. The points where the thermal impedance reaches 90% of the steady-state value are identified with filled symbols.

enlarging the silicon island area from 300 to 910 μm^2) leads to an increase of t_R of 25% and 37% for devices devA without and with an 2-μm-thick AlN layer, respectively. From Fig. 4.23, it can be observed that the increase in AR leads to a growth of t_R whilst the thermal resistance decreases (as also seen in the previous section); varying the AR from 1.25 to 20, a rise time increase of 14% and 18% arises for structures devA without and with an 2-μm-thick AlN layer, respectively. As also stated before (see Section 3.1.2), an increase in silicon area (e.g., obtained by increasing either the emitter area or the aspect ratio and keeping the distances between the emitter stripe and trench sidewalls unchanged) reduces the thermal resistance and increases the rise time. In order to give a qualitative interpretation of this behavior – which holds for both AlN-free and AlN-covered BJTs – let us assume, for the sake of simplicity, that the electrothermal feedback can be modeled with a single *RC* port. Under this hypothesis, it can be found that – due to the geometrical features of a fully-isolated BJT – the thermal capaci-

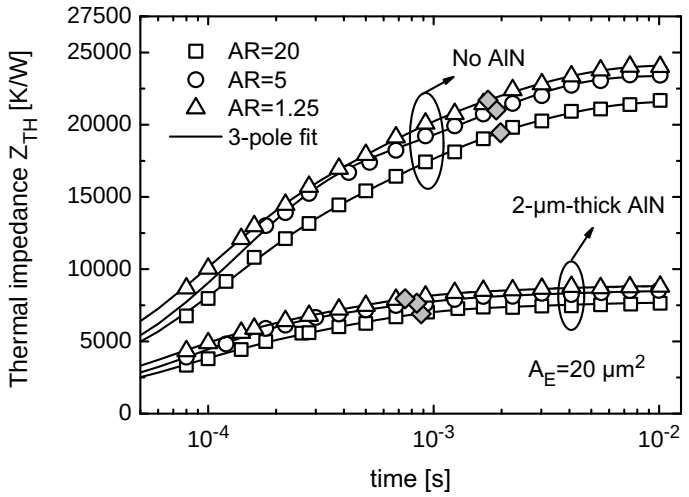


Figure 4.23: Transient thermal impedance of transistors sharing devA layout with and without a 2-µm-thick AlN heatspreader, for different aspect ratios by keeping $A_E = 20$ µm^2. Both (open symbols) measurements and (solid lines) the 3-pole model are presented. The points where the thermal impedance reaches 90% of the steady-state value are identified with filled symbols.

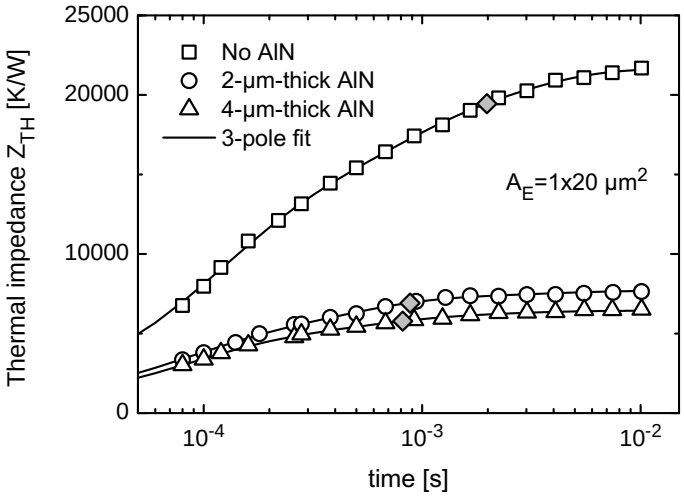


Figure 4.24: Transient thermal impedance variation with AlN thickness. All data refer to device geometry devA with $A_E = 1 \times 20$ µm^2. The points where the thermal impedance reaches 90% of the steady-state value are identified with filled symbols.

tance is much more sensitive to the silicon island size than the thermal resistance; consequently, the thermal capacitance increase due to the enlargement of the silicon island dominates over the thermal resistance reduction. As a conclusion, laterally extending the silicon island yields a slower redistribution of the heat in the structure. The integration of AlN during the front-wafer processing, as clearly evidenced in Fig. 4.24, influences both the steady-state and transient thermal behavior. First, a thermal resistance decrease is obtained due to the cooling action of such a layer, which makes it easier to remove heat from the dissipation region. Second, the heat redistribution within the structure is more effective, and the time needed to reach the steady-state condition is shortened. Such effects are enhanced by increasing the AlN thickness. In particular, it is found that the addition of a 4-µm-thick AlN layer produces a decrease in the rise time of 58% with respect to the same device without any heatspreader. By resorting to the single *RC* port description, it can be evinced that the thermal resistance lowering dominates over the small thermal capacitance growth; the AlN layer does indeed form a low-thermal-capacitance path for the heat flow due to a lower specific

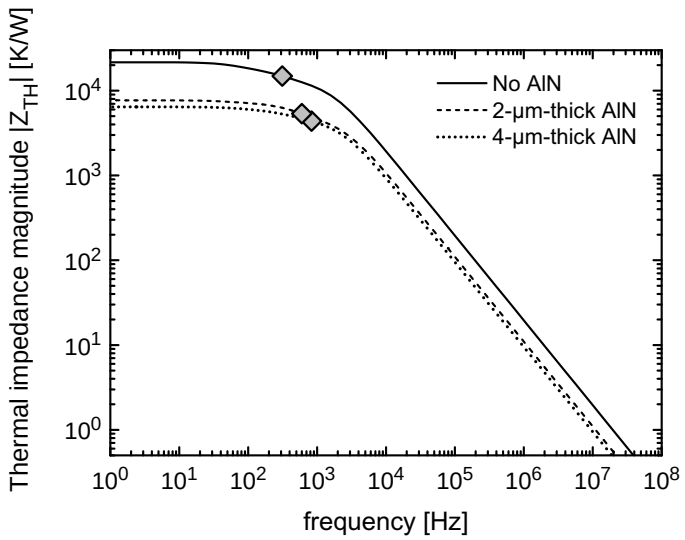


Figure 4.25: Thermal impedance versus frequency for SOG devices without and with AlN layers. For the latter case, two thicknesses, namely, 2 and 4 μm are considered. Also identified are (symbols) the magnitudes corresponding with the cut-off frequencies. All the devices share the geometry devA and $A_E = 1 \times 20$ μm^2.

heat than that of other materials surrounding the active device area, as e.g., silicon dioxide and nitride (see Chapter 3 for more details). These results were quantitatively confirmed by a thorough numerical analysis performed with the *Autosolver* software in Chapter 3. The thermal impedance can be expressed as a function of the frequency by means of (2.2); this allows the evaluation of thermal impedance magnitude versus frequency as reported in Fig. 4.25 that depicts the magnitude of thermal impedance versus frequency for three devices, i.e., one without any heatspreader and the others with AlN layers of either 2 or 4 μm. The analyzed BJTs share the same layout (devA) and emitter stripe geometry ($A_E = 1 \times 20$ μm^2). The symbols identify the thermal cut-off frequencies f_{TH}, at which the magnitude of the thermal response is reduced by a factor of $\sqrt{2}$ with respect to the maximum value. It can be observed that

- f_{TH} is confined to values well below 1 kHz, in agreement with the

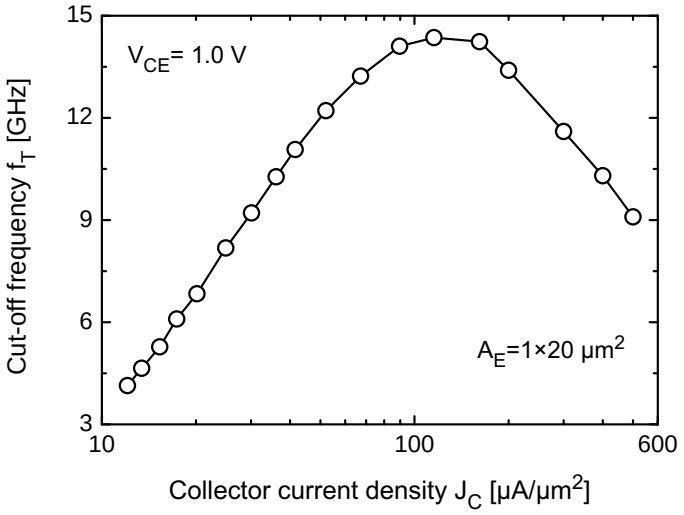


Figure 4.26: Measured cut-off frequency versus collector current density for a SOG BJT with $A_E = 1 \times 20$ μm^2. The device is biased with a collector-emitter voltage V_{CE} equal to 1 V.

results obtained in [72] through an AC measurement technique for comparable transistors.

– f_{TH} slightly increases with aluminum nitride thickness.

The latter result was expected from the t_R behavior since f_{TH} is inversely proportional to the rise time of the unitary step response. Such an increase could be in principle interpreted as a disadvantage for small-signal operation but it should be considered that the electrical cut-off frequency f_T of a typical SOG BJT is several orders of magnitude (tens of GHz, as shown in Fig. 4.26, which reports f_T as a function of the current density J_C) higher than the evaluated thermal cut-off frequency (hundreds of Hz). As a consequence, thermal issues should not pose a critical threat for small-signal operation of SOG BJTs. As extensively explained in Chapter 2, it is possible to transform Foster networks into Cauer model, results for measured devices sharing the same layout (devA) for the case of a device with and without a 2-μm-thick AlN layer are reported in Table 4.3.

Table 4.3: Evaluated RC values for both AlN-free and AlN-provided SOG BJTs with 1×20 μm^2 emitter and devA layout.

Device	$R_{TH_1}, R_{TH_2}, R_{TH_3}$ **[K/W]**		$C_{TH_1}, C_{TH_2}, C_{TH_3}$ **[nJ/K]**	
	Foster	**Cauer**	**Foster**	**Cauer**
No AlN	6080	13348	84	8.15
	4850	5751	470	95.25
	10700	2531	9.2	760
2-μm-thick AlN	3140	5027.5	120	14.84
	625	2212.7	4800	157.15
	3930	454.8	17	6399
4-μm-thick AlN	1760	4652	230	16.59
	4010	1415	18	290.37
	660	363	2800	4691.65

Chapter 5

Development of an electrothermal circuit simulation environment

In solid state devices the current (and hence the dissipated power) is a function of device temperature, which, in turn, is determined by the dissipated power. Therefore the determination of device current and temperature represents a coupled problem. To solve the circuit equations the device temperature must be known $I_i = f(T_i)$, but on the other hand, to calculate the temperature the power dissipated by the devices should be known $T_i = f(P_i)$. The importance of thermal effects is evident in Fig. 5.1 where isothermal (i.e., constant temperature for device) simulations are compared to measured data of the output characteristics of a Q56R InGaP/GaAs heterojunction bipolar transistor; the power increase affects device operation augmenting its temperature and causing the decrease (in the case of GaAs HBT) of current gain. Obviously, this can complicate the scenario of accurate circuital simulation. The solution is to resort to circuit simulation tools where the electrical and thermal problems are solved simultaneously as shown in the schematic representation of Fig. 5.2.

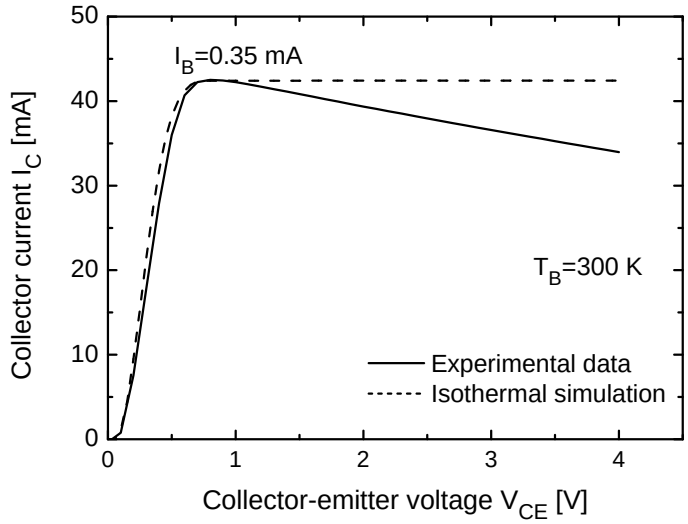


Figure 5.1: Isothermal simulation compared to real measured data of output characteristics of a Q56R.

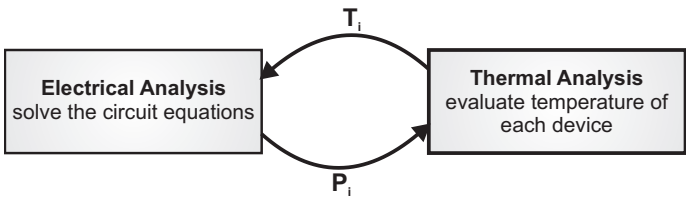


Figure 5.2: Illustration of the electrothermal coupled problem.

5.1 Advanced tools for electrothermal simulation

In general, there are two possible solutions to solve the coupled electrothermal problem

- Direct coupling between electrical and thermal solvers; this approach is based on the solution of the heat diffusion equation at each iteration point (i.e., by using a FEM solver) but also needs a huge computational effort.
- Use of reduced thermal models (e.g., R_{TH} matrix or equivalent thermal networks) whose parameters can be extracted in advance through measurements, numerical simulations, or analytical models.

The latter approach is the most convenient from a computational point of view, although details on the internal temperatures, different from T_j cannot be directly deduced. Moreover it is worth noting that

- Some commercial circuit simulators (PSPICE) does not present thermal models, so in theory the temperature is a constant parameter and does not change during the simulation process;
- Most advanced simulators (like ADS and Spectre) include advanced electrothermal models for devices; if enabled, the electrothermal feedback can be accounted for through a simple single-pole equivalent network; however:
  - Only self-heating effect is included, that is, thermal coupling between devices is not taken into account leading to markedly inaccurate results when the temperature of a transistor is affected by other (close enough) devices
  - Thermal parameters (R_{TH}, C_{TH}) have to be set manually in the thermal circuit
  - Single-pole thermal circuits have insufficient accuracy when describing transient evolution (as diffusely discussed in the former chapters).

The software tool was developed following two approaches, one based on the macromodeling technique (suitable for SPICE-like simulators) [73, 74] that is employed to account for the electrothermal feedback,

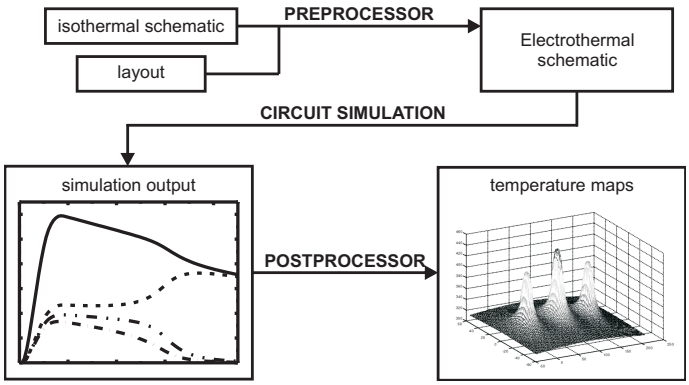


Figure 5.3: Flow-chart of the electrothermal tool.

temperature sensitivity of key electrical parameters, and unique phenomena like impact ionization. The other approach relies on the exploiting of the thermal node that most of the tools adopted in the CAD arena [75,76] incorporate in recent device models; this way it is possible to connect an electrothermal feedback block (projected by various approaches) to the temperature terminal. This way it is possible to evaluate temperature rise from the dissipated power and, hence, consider the evaluated temperature as an additional input that can influence temperature-sensitive parameters, thereby allowing the description of self-heating (SH) effects. According to the latter approach the electrothermal feedback block can be implemented by means of symbolically defined devices (SDD, that are multiport components where current on one port is defined as e.g., voltage on another port;), by Verilog-A (for simulators accepting this format) or equivalent networks (made of passive elements, can become very complicated, so it is unpractical for a large number of devices).

The software tool flow-chart is reported in Fig. 5.3; it is possible to identify the following phases

Preprocessing. This phase takes in charge, substantially, the generation of thermal feedback block (i.e., the thermal matrix). The tool offers several possibilities for the generation of the thermal resistance/impedance network. One possibility relies on the automatic evaluation of thermal feedback block through the closed-form analytical formulations proposed in [15, 77] for the steady-state and transient cases, respectively. In this particular operative mode the

software calculates steady-state or transient properties by using the layout informations (i.e., emitter window for bipolar transistors). Another possibility is to provide a thermal feedback block by using other tools (i.e., FEM simulator, measurements) through a thermal matrix file (i.e., a simple textual file containing information on the thermal feedback block). Once the thermal matrix has been generated, the ET feedback block has to be integrated in the "solely" electrical schematic in order to obtain an electrothermal network.

Circuit Simulation. It is worth noting that there are no differences between the approach of "macromodeling" and "thermal node" up to this point. When using the macromodeling approach (i.e., for SPICE-like simulators) an electrothermal macromodel must be created for the device to extend the simulator capabilities, which contains all the electrical laws governing the device behavior with temperature-dependent parameters and should be connected to the ET block (that is obtained in the previous phase). On the contrary, if the simulator offers the possibility of using a thermal node the simple electrical isothermal schematic is enriched (as shown in Fig. 5.4) with the ET feedback block (that can be realized in Verilog-A, SDD or equivalent networks). The ET feedback block is employed to calculate the corresponding junction temperature increases above ambient (node voltages of the ET feedback block) starting from the powers dissipated by all transistors (currents in the thermal equivalent network) during the electrothermal simulation. It is worth noting that internal thermal networks (namely, a single RC network) must be disabled in order to correctly perform simulation. The simulation of the whole ET network is performed by a commercial simulator (i.e., PSPICE or ADS). After the simulation is finished, results are passed back to the tool for post-processing phase.

Postprocessing. The post processing stage is partly charged to the simulator (i.e., generating IV characteristics etc..) but mostly is assigned to the external tool. The software imports the results from the commercial circuit tool and is capable of use them for further purposes like evaluating thermal maps (by means of analytical methods). It is possible to visualize thermal maps over a regular grid at different DC bias points (using layout information). Fig. 5.5 shows the thermal maps evaluated for a simulated three-

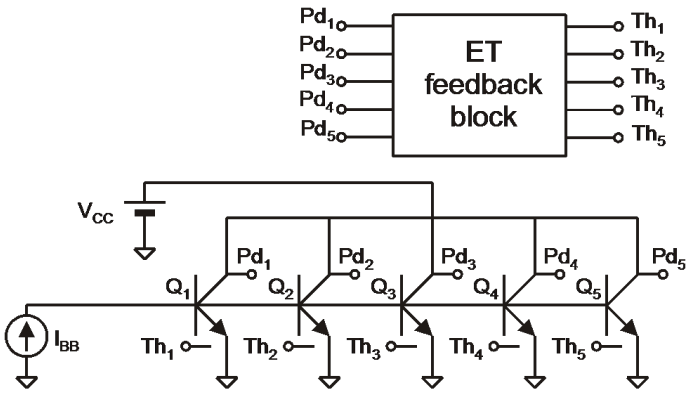


Figure 5.4: Illustration of the electrothermal schematic for a 5-finger simulation. The additional thermal and power nodes used to interconnect the pure electrical device to the thermal feedback block are evidenced.

finger GaAs device dissipating a (total) power of 300 mW.

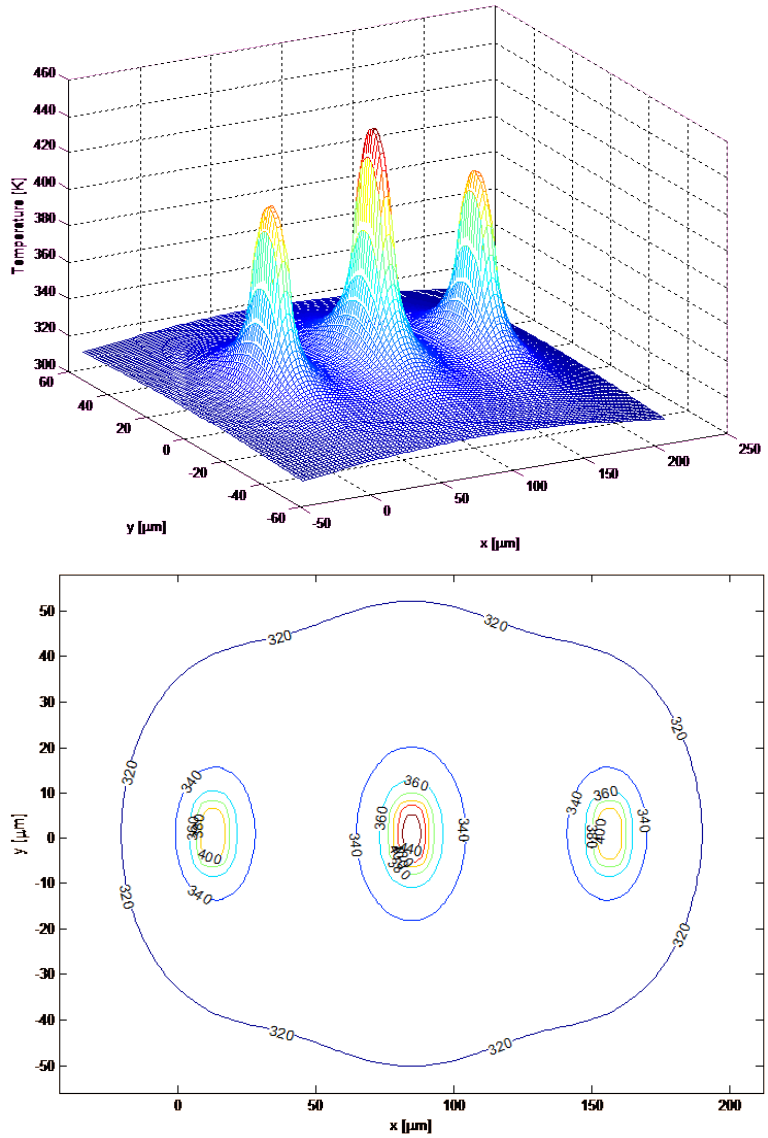


Figure 5.5: Thermal map (top) and contour (bottom) plot of a 3-finger InGaP/GaAs HBT dissipating 300 mW.

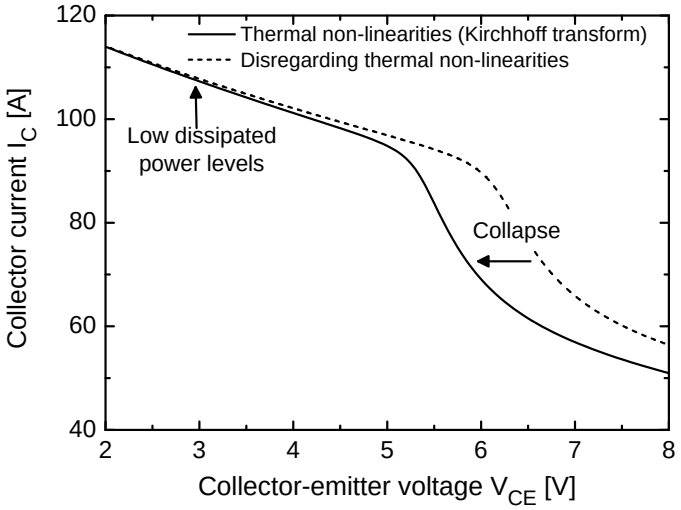


Figure 5.6: Simulated output characteristics of a 3-finger GaAs HBT with (solid line) and without (dashed line) thermal non linearities.

5.2 Non-linear thermal effects

Many semiconductor materials show non-linear thermal behavior because of temperature dependence of physical parameters (i.e., thermal conductivity (1.7)). The effect of thermal non-linearities is an exacerbation of the thermal issues as clearly depicted in Fig. 5.6 showing the (simulated) output characteristics of a 3-finger GaAs HBT with and without those effects. It is worth noting that at low dissipated power, the curves are identical, indeed at low dissipated power levels the temperature increase ($\Delta T = R_{TH} \cdot P_D$) is negligible and the thermal conductivity is nearby equal to k_0. On the contrary, the rise in dissipated power reduces the thermal conductivity, thus causing the reduction of device capability of heat conduction. This can be inferred by the leftward shift of the collapse of current gain, that is, the device thermal instability condition is reached before the predicted point. For this reason it is clear that thermal non-linearities have to be accurately discussed and properly included in simulations. Conveniently, if the thermal conductivity dependence on temperature can be expressed as a power law (see (1.7)) the non-linear thermal problem can be solved through a very effective ana-

lytical formulation, namely, the Kirchhoff transformation [67]. This is a linear transformation and it is possible to implement it in electrothermal circuit simulation tools by an additional block that modifies the linear increment of temperature in the non-linear counterpart. The procedure can be described as follows [40, 67, 78–82]

- The heat diffusion linear equation (where $k(T) = k_0$ is constant) is solved so as to obtain a temperature increase above T_0 (i.e., $= 300$ K) that can be indicated as $\Delta T_{lin} = T_{lin} - T_0$, where T_{lin} is also referred to as "pseudo-temperature" [78];
- In the electrothermal circuit simulator $\Delta T_{lin} = R_{TH} \cdot P_D$ is evaluated;
- It is possible to determine the non-linear problem solution ΔT by using [79]

$$\Delta T = T - T_0 = T_0 \cdot \left[1 - (m-1) \cdot \frac{\Delta T_{lin}}{T_0}\right]^{\frac{1}{1-m}} - T_0, \quad (5.1)$$

  where T is the "absolute" temperature corresponding to the non-linear problem.

Obviously, the non-linear temperature rise is different (in general higher[1] if $m > 1$) and hence, calculating thermal resistance ($\Delta T/P_D$) a new value is obtained as it was done in Section 4.1.1 to compare thermal resistance measurement techniques. This procedure is correct if the baseplate temperature is equal to the reference temperature ($T_B = T_0$); when $T_B \neq T_0$ the problem must be tackled in a slightly different way. Let us suppose $T_B > T_0$; in this case, (1.7) can be written as

$$k(T) = k(T_B) \cdot \left(\frac{T}{T_B}\right)^{-m}, \quad (5.2)$$

whence (5.1) can be rewritten as

$$\Delta T = T - T_B = T_B \cdot \left[1 - (m-1) \cdot \frac{\Delta T_{lin}(T_B)}{T_B}\right]^{\frac{1}{1-m}} - T_B, \quad (5.3)$$

where $\Delta T_{lin}(T_B)$ is the linear temperature increase evaluated when the thermal conductivity $k = k(T_B) = k_0 \cdot {T_B}/{T_0}^{-m}$. The thermal resistance

[1] For the temperature range of interest in semiconductors $m > 1$, thus thermal conductivity increases with temperature.

obviously modifies in comparison to the value at T_0; according to [65] it is possible to write, with a good approximation

$$R_{TH}(T_B) = R_{TH} \cdot \left(\frac{T_B}{T_0}\right)^m, \tag{5.4}$$

where R_{TH} is the thermal resistance evaluated at T_0, so that

$$\Delta T_{lin}(T_B) = R_{TH}(T_B) \cdot P_D. \tag{5.5}$$

It is worth noting that $R_{TH}(T_B)$ is to be evaluated through (5.4) only if another formulation is not provided (i.e., a linear dependence, values provided with 3-D FEM simulations, measured values). These results were thoroughly verified by means of 3-D FEM simulations.

5.3 Electrothermal circuit simulations

The software tool was extensively used with a number of different bipolar families in order to predict thermal instabilities like "flyback" and current "bifurcation" in single- and multifinger transistors (see Section 1.4 for more details). This section reports the main findings both for InGaP/GaAs HBTs (see Section 3.1.3 for structure details) and SOG BJTs (see Section 3.1.2 for structure details). It is worth noting that the use of an accurate thermal simulator allows the correct prediction of SOA and to access temperature details of the single devices of the circuit in order to counteract thermal instabilities at the design stage (i.e., by adopting emitter ballasting, or emitter segmentation). The models used to perform electrothermal simulations both for InGaP/GaAs HBTs and SOG BJTs are developed *in house* and includes all the relevant physical mechanisms [13] (i.e., high injection, impact ionization, etc...).

5.3.1 Steady-state circuit simulations

Steady state circuit simulations are very useful in order to predict the DC SOA of devices under test and to study possible solutions to counteract electrothermal issues and, eventually, enlarge safe operating area.

InGaP/GaAs HBTs simulations. The devices are detailed in Section 3.1.3 together with the thermal resistance matrix extracted for multifinger transistors. Fig. 5.7 shows the measured output characteristics for two Q56R 3-finger configurations that differ by the center-to-center distance between fingers. It is noteworthy that devices having a bigger distance between centers (namely, 72 µm) present the collapse behavior at a lower V_{CE} with respect to devices characterized by a distance of 24 µm. This is due to the fact (see also Table 3.11) that when the devices are close, mutual thermal resistances are higher (self-heating thermal resistance is lower) with respect to the situation of distant fingers, so that the temperature distribution is more even and the onset of the collapse (one finger starts to drain all collector current) is shifted rightward. Fig. 5.8 shows electrothermal simulation data, where the thermal matrix is obtained by means of *Autosolver*, compared to experimental data for both the cases. Excellent agreement is achieved between measurement and simulation data. The explanation of the collapse is evident

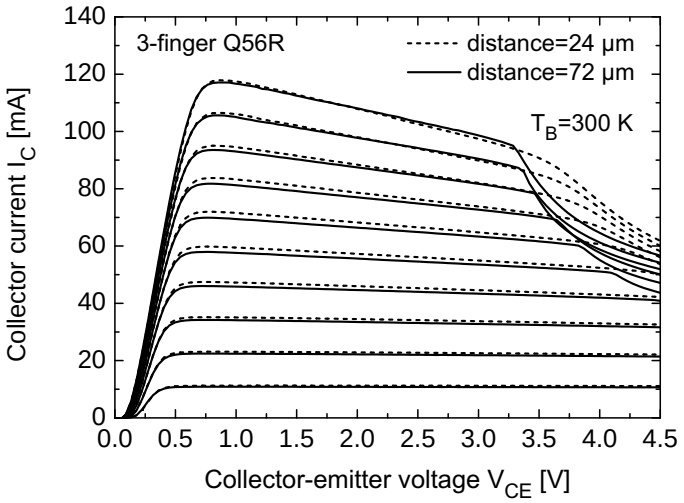


Figure 5.7: Output characteristics of two 3-finger GaAs HBT configuration having different center-to-center distances between elementary devices: (dashed line) 24 and (solid line) 72 µm. Elementary devices have 60 μm^2 horse-shoe shaped emitters.

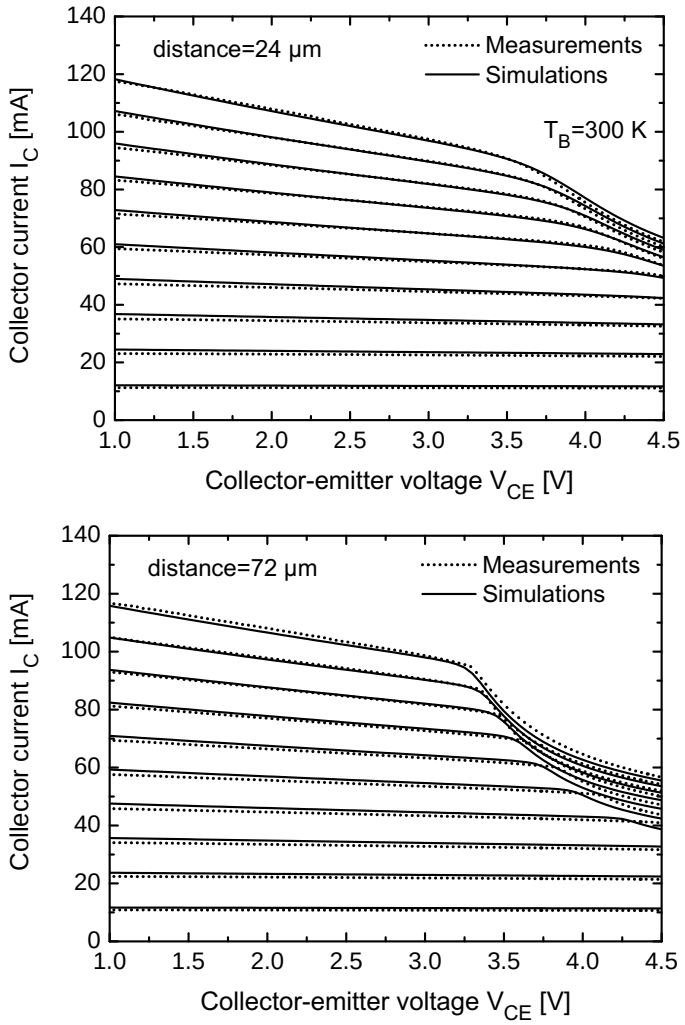


Figure 5.8: Measured (dotted line) versus simulated (solid line) output characteristics of two 3-finger GaAs HBT configuration having different center-to-center distances between elementary devices: (top) 24 and (bottom) 72 µm. Elementary devices have 60 μm^2 horse-shoe shaped emitters.

in Fig. 5.9 depicting the collector currents flowing in the individual fingers together with the temperature increases over ambient. It is worth noting that by means of the electrothermal simulation tool it is possible to study how the slight differences in device parameters, like (parasitic) internal emitter resistances, can affect the thermal instabilities as clearly shown in Fig. 5.10 depicting the output characteristics of a 3-finger device where each finger has a different emitter resistance. The device with the smaller emitter resistance is the one draining all the collector current on the onset of collapse behavior.

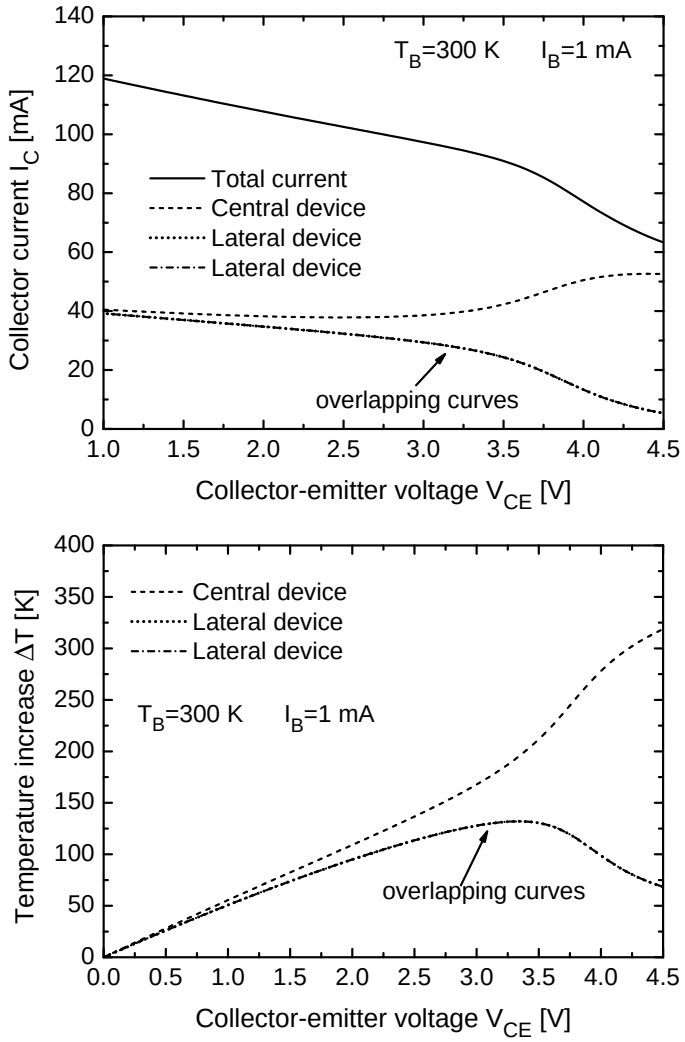


Figure 5.9: Simulated output (top) characteristics and temperature (bottom) increases of two 3-finger GaAs HBT configuration having a center-to-center distance of 24 μm. Elementary devices have 60 μm^2 horse-shoe shaped emitters. Characteristics relative to each single finger are reported as well.

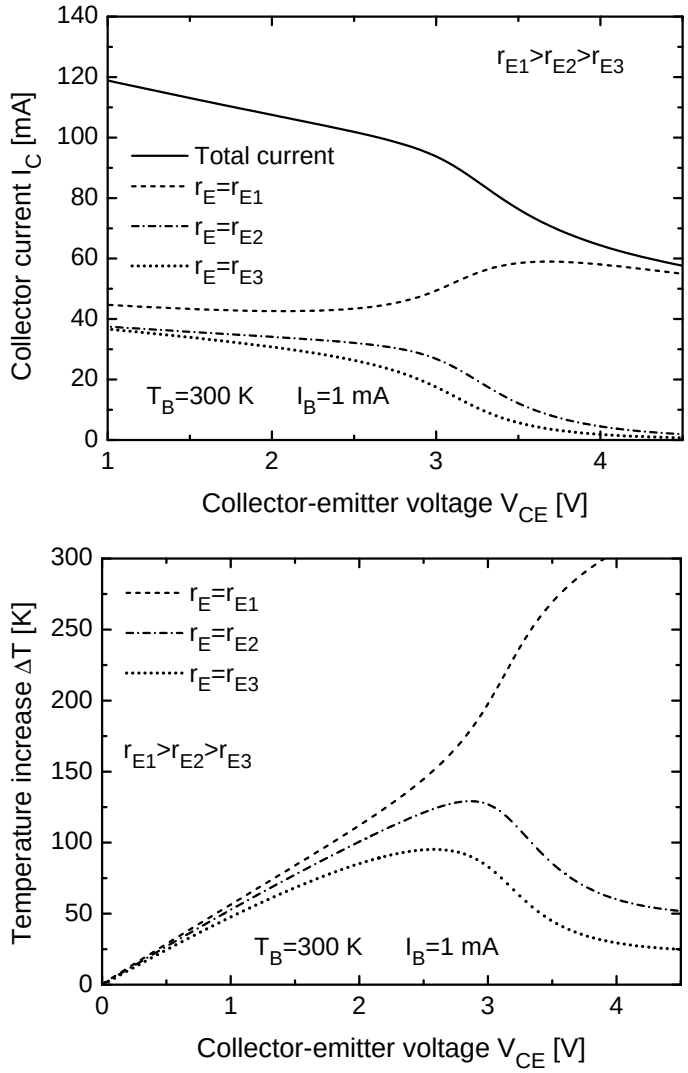


Figure 5.10: Simulated output (top) characteristics and temperature (bottom) increases of two 3-finger GaAs HBT configuration having a center-to-center distance of 24 μm. Elementary devices have 60 μm^2 horse-shoe shaped emitters. Characteristics corresponding to each single finger are reported when the parasitic (internal) emitter resistance is slightly different for each elementary device.

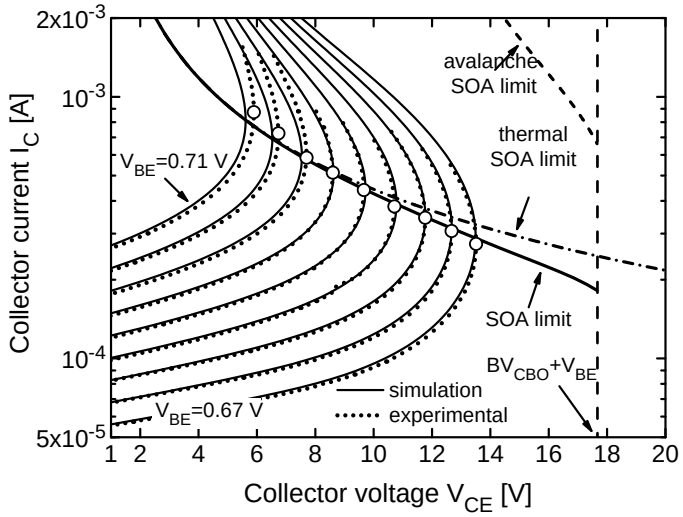


Figure 5.11: Common-emitter output characteristics of device devA with 2-µm-thick AlN layer. (Dotted lines) Experimental curves are compared with (solid) simulation results. Also shown are the (solid thick line) flyback locus including both ET and II effects as determined by [13] and those obtained by deactivating either (dot-dashed line) II or (dashed line) ET effects. The measured flyback points are identified by open circles.

Silicon-on-glass bipolar transistors. Electrothermal simulator has been widely used to study the influence of layout on thermal instabilities [83, 84] on single- and multifinger devices in order to understand the coupled effect of avalanche and thermal effects on SOA. It was also used to understand how innovative design patterns can influence the electrothermal behavior of multifinger transistors. For the SOG devices considered in this work, the analysis demonstrated that ET effects, due to the high thermal resistances, play a dominant role in defining the limit of safe operation at low/medium V_{CE} values. This can be evinced by the output plane shown in Fig. 5.11, individual and combined influence of ET and II effects are clarified. Both simulated and experimental common-emitter I_C-V_{CE} characteristics are shown for a SOG device having an area of 3×90 µm^2 and layout devA with a 2-µm-thick AlN layer. It can be observed that the SOA boundary, due to the

interdependence between ET and II, becomes increasingly lower than the ET-only limit as V_{CE} increases and the flyback point gets closer to the $BV_{CBO} + V_{BE}$ boundary (i.e., for low V_{BE} values). For low V_{CE}, the two coincide because avalanche effects are negligible. This example supports the general conclusion that the ET-II interplay will always restrict the SOA more than when solely ET or II effects play a role as also been reported in [13, 85] in contrast with the unjustified conclusion drawn in [86].

As previously seen, the segmentation of the emitter into more pieces would strongly reduce current crowding phenomena as compared to the one stripe emitter of equal total length. The thermal resistances obtained in Section 3.1.2 were used in the electrothermal simulation tool, that accounts for both ET and II effects, so that the behavior of the device could be analyzed for the cases where the emitter stripe was segmented in two, three, four or five pieces as reported in Fig. 3.20. Results of the simulation are reported in Fig. 5.12 where it is possible to observe that when the number of segments increases, the separation of the individual currents starts at higher V_{CB} values, as shown in Fig. 5.12. It should be noted that, in all cases, the total current remains the same. Fig. 5.12 also shows the temperature difference between the hottest and coldest segments in each of the designs shown in Fig. 3.20. All the elementary transistors operate at closer temperatures for a larger range of V_{CB} if the overall device is split into more segments, which means, in other words, that the temperature distribution has become more uniform across the silicon island. The characteristics of three different hexagonal layout (see Table 3.3 and Fig. 3.18 for more details) solutions are shown in Fig. 5.13. It is worth noting that, the heat propagates mostly through the silicon and metal lines; consequently, increasing the amount of silicon markedly improves the overall thermal stability (J59-HEX topographies are more stable than T59-HEX). As can be deduced directly from the thermal resistance matrix reported in Table 3.4 and Fig. 5.13 the second finger is the one conducing the highest individual current having the higher thermal resistance. Four different configurations of the emitter stripes of a four-finger bipolar transistor are depicted in Fig. 3.19 while the simulated thermal resistances are shown in Table 3.5. The experimental characteristics of device FOURA (horizonthal topography) are illustrated in Fig. 5.14, along with the electrothermal simulation results. Measurements were performed applying two different voltages to the

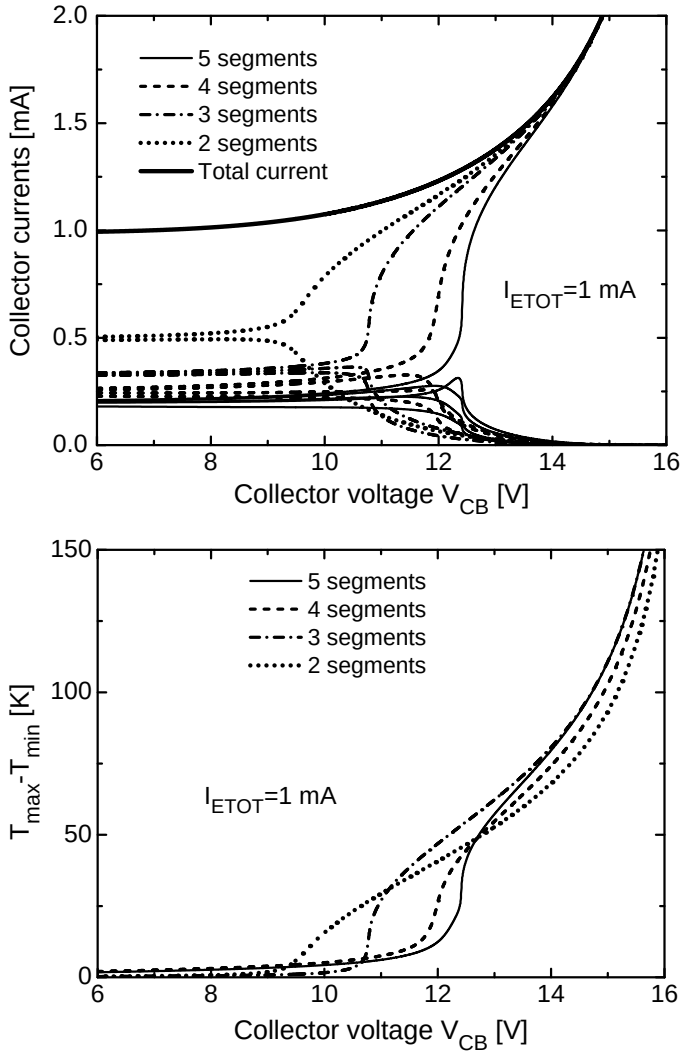


Figure 5.12: Simulated common-base characteristics of a device with emitter stripe partitioned into a different number of segments: (top figure) Individual collector currents, and (bottom figure) difference between the temperatures of the hottest and coldest emitter segment versus collector voltage V_{CB}.

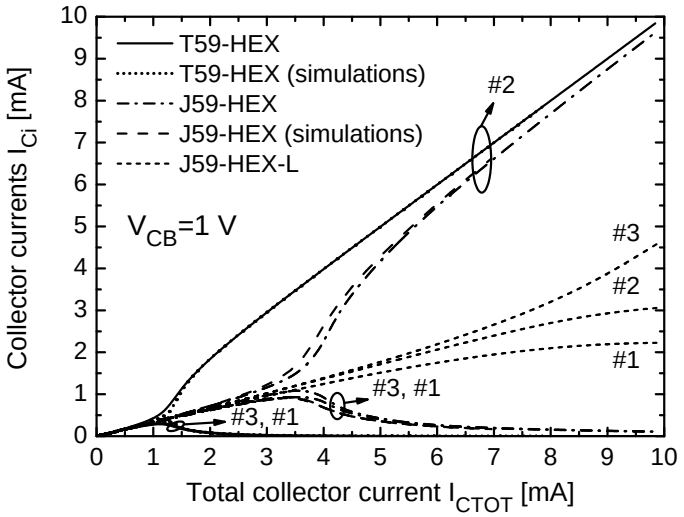


Figure 5.13: Measured and simulated individual collector currents as a function of the total collector current I_{CTOT}. Bipolar transistors with 59 μm center-to-center distance between two adjacent emitter stripes without any heat spreader. Details of topology are reported in Table 3.3.

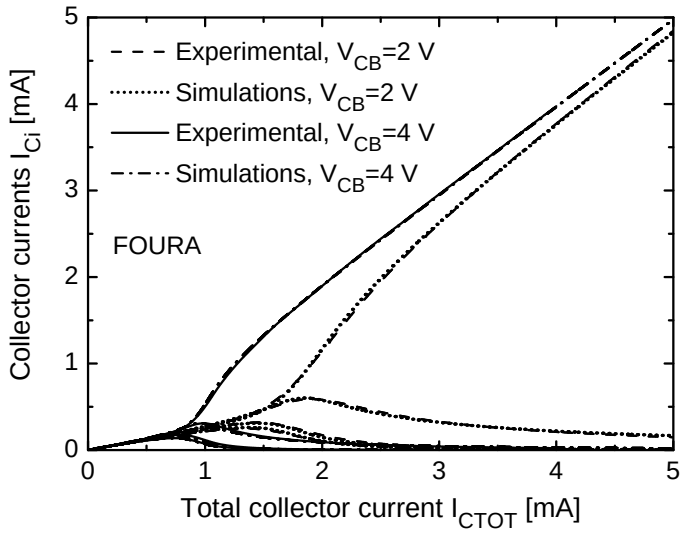
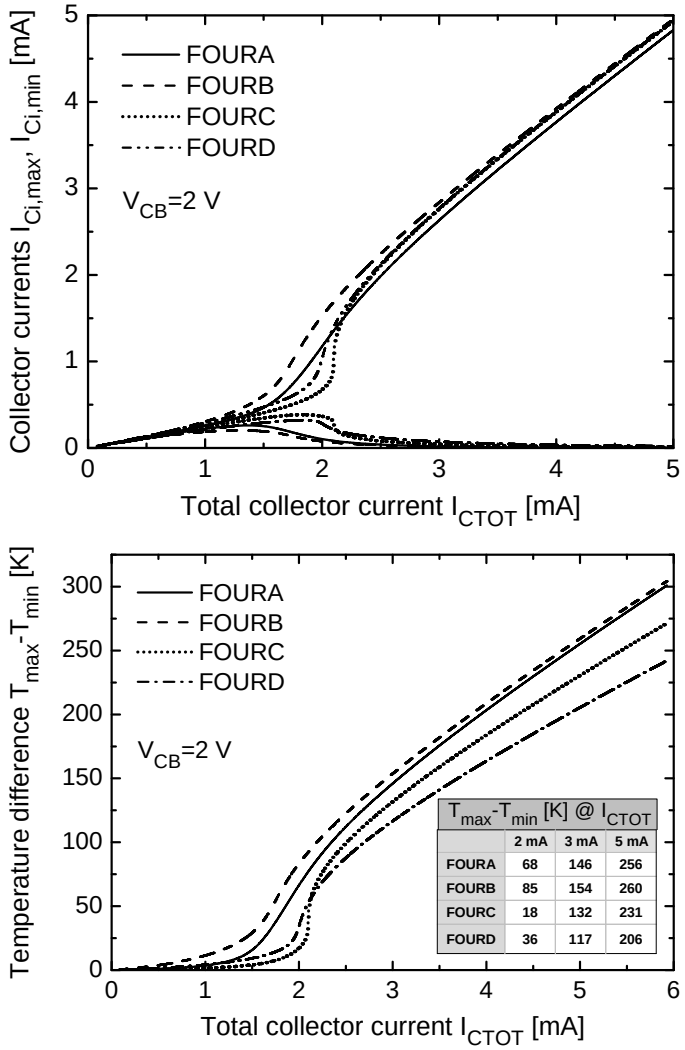


Figure 5.14: Measured and simulated individual collector currents as a function of I_{CTOT} for the trench-isolated four-finger bipolar transistor with emitter layout FOURA as in Fig. 3.19 at two V_{CB} values.

collector terminal; as expected, the separation of the individual currents is triggered at lower values of the total collector current for a higher collector voltage. The electrothermal behavior of the other three layouts was simulated, and the collector currents of the hottest and coldest fingers for each design are shown in Fig. 5.15 (the temperature difference between the hottest and coldest finger are reported as well). The ramification of the currents occurs at higher values of I_{CTOT} for devices FOURC and FOURD, which benefit from a more symmetrical geometry. In other words, for the given biasing conditions, the operating temperature of the fingers is more uniform in the rhombus and square layouts than in devices FOURA and FOURB. This has been quantified in Fig. 5.15. For instance, the maximum temperature gradient in the overall device, namely, the difference between the temperatures of the hottest and coldest individual transistors, is 68 and 85 K at $I_{CTOT} = 2$ mA for FOURA and FOURB, respectively, while being much lower for devices FOURC and FOURD, as can be read in the inset table. As previously seen for three-finger transistors, as well as in devices comprised of four elementary BJTs, the transition between the regions of symmetric and asymmetric current distribution is more sharply delimited when the differences between individual fingers are reduced. This is clear from the simulations in Fig. 5.15, in which the device with the most abrupt branching of the currents (i.e., device FOURC) is also the one with the most uniform values of individual self-heating thermal resistances (reported in Table 3.5).



T_{max}-T_{min} [K] @ I_{CTOT}

	2 mA	3 mA	5 mA
FOURA	68	146	256
FOURB	85	154	260
FOURC	18	132	231
FOURD	36	117	206

Figure 5.15: Simulated (top figure) individual collector currents of the hottest and the coldest finger and (bottom figure) temperature difference between the hottest and the coldest finger as a function of ICTOT for all the designs schematized in Fig. 3.19. The inset table (bottom figure) provides the maximum temperature discrepancies at various values of I_{CTOT}.

5.3.2 Transient circuit simulations

The electrothermal transient simulations allowed deepening the comprehension of the benefits achieved by using AlN layers in silicon-on-glass devices [71]. In Fig. 5.16, the simulated collector current as a function of time is shown for devices with different AlN thicknesses. It is worth noting that the thermal networks used as ET feedback block are reported in Table 4.3. The transistors are biased with a fixed value of collector-emitter voltage V_{CE} and by applying a voltage step in V_{BE} from 0 up to 0.75 V with a rise time of 50 ns. Points A, B, and C, at which the electrothermal transient has by far been extinguished for the three devices, are also identified on the steady-state I_C-V_{CE} characteristics reported in the inset. In Fig. 5.16 the simulated temperature increase ΔT_j as a function of time under the same biasing condition is reported as well. The inset illustrates the steady-state ΔT_j-V_{CE} curves. The response of devA to the V_{BE} voltage step is given in Fig. 5.17 for three V_{CE} values. It is worth noting that the biasing condition $V_{CE} = 2.8$ V does not correspond to any stable equilibrium point (i.e., the steady-state flyback behavior takes place at a collector voltage lower than 2.8 V, as shown in the insets of Fig. 5.17). As a consequence, the current and temperature tend to infinity, thus entailing a sudden device destruction as pointed out in [11]. From this analysis, it can be concluded that a calibrated electrothermal simulator can be a useful predictive tool for circuit designers.

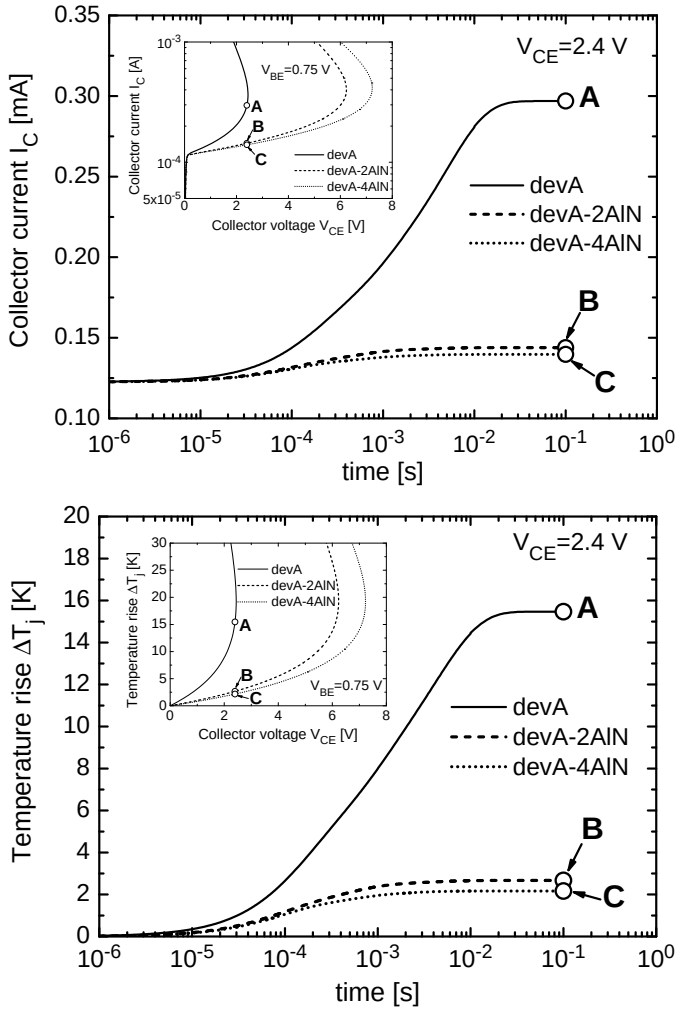


Figure 5.16: Simulated (top picture) collector current and (bottom picture) junction temperature rise over ambient as a function of time for SOG BJTs without and with AlN layers having a thickness of either 2 µm or 4 µm. The devices are all characterized by layout devA and area $A_E = 1 \times 20$ µm^2. Also shown in the insets are the simulated (top figure) collector current and (bottom figure) junction temperature over ambient as a function of V_{CE} under DC conditions. The steady-state equilibrium points A, B, and C corresponding to the analyzed devices are identified both at the end of the transient current/temperature evolutions and on the DC curves reported in the insets. The electrothermal transient simulations are performed for $V_{CE} = 2.4$ V and stepping V_{BE} from 0 V to 0.75 V.

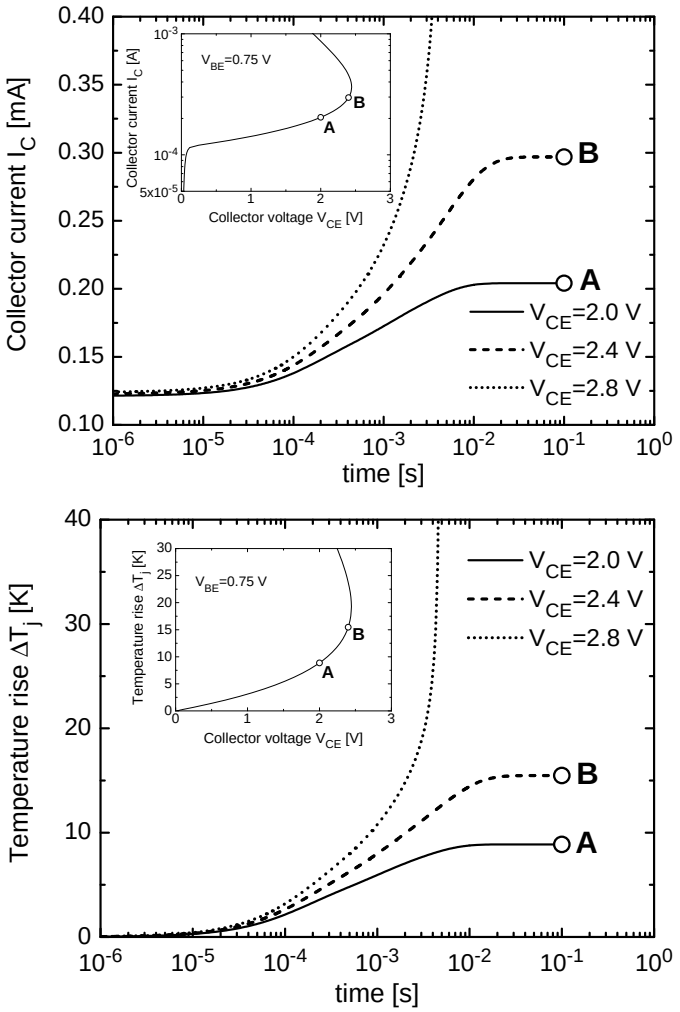


Figure 5.17: Simulated (top figure) collector current and (bottom figure) junction temperature rise over ambient as a function of time for a SOG BJT with layout devA and $AE = 1 \times 20$ μm^2 at three V_{CE} values. Also shown in the insets are the simulated (top figure) collector current and (bottom figure) junction temperature over ambient as a function of VCE under DC conditions. The (steady-state) equilibrium points A and B corresponding to $V_{CE} = 2.0$ V and 2.4 V, respectively, are identified both at the end of the transient current/temperature evolutions and on the DC curves reported in the insets. The electrothermal transient simulations are performed by stepping V_{BE} from 0 V to 0.75 V.

Chapter 6

Conclusions and recommendations

6.1 Conclusions

It has been shown that modern – even low-power – electronic devices/circuit can suffer from ET effects at the point that these can hamper the reliability and performances. In this thesis, an approach to the analysis of solid-state devices ranging from simulations to measurements has been proposed with the aim to optimize thermal behavior at device level. In particular,

- An advanced software *Autosolver* has been conceived to automatically run detailed 3-D FEM simulations using layout and technology data. This software can be extensively used to deepen the comprehension of thermal issues in solid-state devices and to understand the influence of the main technology and layout parameters on devices/circuits electrothermal behavior. In particular, is has been successfully applied to analyze a large variety of technologies, namely, SOG BJTs, GaAs HBTs, GaN HEMTs, SiGe HBTs, and UTCS technology devices , thus favoring the definition of useful guide-lines to improve their thermal management.
- An effective equivalent network identification tool has been developed to extract network data from thermal transients with the following twofold purpose:

Modeling. The tool allows obtaining accurate equivalent thermal network with a low number of elementary *RC* cells so as to remain into a prescribed accuracy. Those extracted thermal networks are suitable to be used in electrothermal simulators and to describe the thermal behavior of devices in frequency domain.

Analysis. The identification by deconvolution can be used to achieve the *full spectrum* of time constants and the structure function in order to gain information on the heat propagation through the device and to identify eventual problems (i.e., bad attach).

- A routine has been developed in order to transform the Foster network model in the Cauer model. Moreover, this routine can handle arbitrary precision numbers that are necessary when dealing with a large number of elementary *RC* networks.
- A novel CTM has been proposed in order to reduce the simulation complexity still maintaining a good level of accuracy.
- An extended overview of the DC thermal resistance measurement techniques presented in literature was carried out in order to identify the most appreciable methods and their capability to predict the thermal resistance behavior with changing baseplate temperature and dissipated power. From the study, it emerged that the technique proposed by Marsh is the most accurate when measuring devices with a negligible Early effect.
- A novel thermal resistance DC measurement technique has been proposed, which enables and accurate evaluation of R_{TH} regardless of the presence of Early effect. This technique has been adopted to verify the findings of simulations performed on SOG devices.
- An extension of the *cooling curve* technique has been successfully adopted to measure the thermal impedance of AlN-cooled SOG devices. As in simulations, it was found that the adoption of the AlN layer considerably decreases thermal resistance and favors a faster thermal transient.
- An electrothermal tool has been developed, which is based on the improvement of commercial circuit simulators. All the relevant physical effects, as well as self-heating and thermal coupling

mechanisms are included, so as to guarantee an accurate prediction of the SOA borders of the analyzed devices. The tool can be successfully employed to support the design stage so as to improve the thermal ruggedness of the transistors of interest.

6.2 Future work

This work presents a number of techniques for thermal characterization of solid-state devices from simulations to measurements; however there is still room for improvements and further research.

- In the last few years it is becoming more evident that the Fourier conduction equation is not suitable to describe the thermal behavior of very miniaturized (at the nanoscale) devices [87]. For this reason the Boltzmann's transport equation should be used. This could be useful to explain the 25-35% discrepancies between the measured and simulated thermal resistances of last generation SiGe HBTs fabricated by STM and IFX [88].
- Another interesting topic is the extension of the electrothermal simulator (see Chapter 5) to include a self-consistent model for the mutual thermal impedances without making use of Verilog-A.
- The possibility to identify thermal parameters as well as to diagnose problems in the structure by a transient measurement should be experimentally extended to better understand how the measurement noise can affect it.
- The analyses of many families of electronic devices posed the bases for the fabrication of an *in house* electronic pulser system useful for
  - The experimental isothermal characterization of devices to extract temperature-scaling parameters necessary for modeling purposes.
  - The measurement of thermal transient that can be used for diagnosis and modeling purposes.

  From the approaches shown in this thesis, it was possible to define the specifications of the system in order to adapt it to various families of devices, the Fig. 6.1 reports the maximum power allowed to perform an isothermal measurement (maximum temperature increase is less than 1 K) at a fixed pulse width.

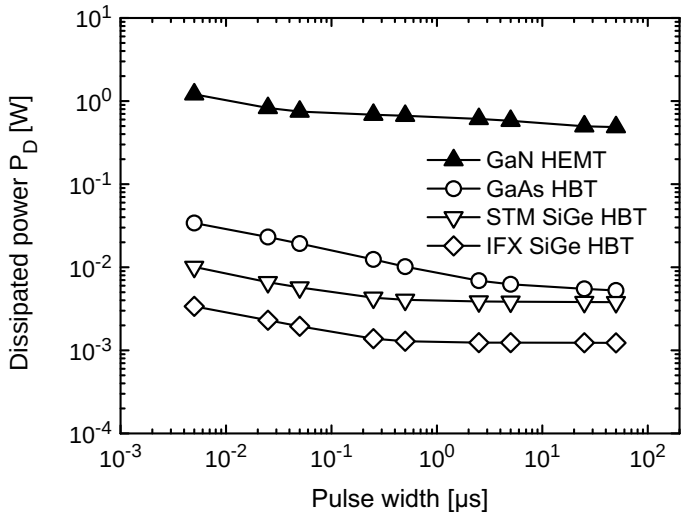


Figure 6.1: Maximum power allowed to perform isothermal measurements (the temperature increase is less than 1 K) as a function of pulse width for various device categories.

Bibliography

[1] G. E. Moore, "Cramming more components onto integrated circuits," *Electronics*, vol. 38, no. 8, pp. 114–117, 1965.

[2] G. E. Moore, "Progress in digital integrated electronics," in *Proc. IEEE International Electron Devices Meeting (IEDM)*, vol. 21, pp. 11–13, 1975.

[3] G. E. Moore, "No exponential is forever: but "forever" can be delayed!," in *Proc. IEEE International Solid-State Circuits Conference (ISSCC)*, vol. 1, pp. 20–23, 2003.

[4] L. E. Larson, "Silicon technology tradeoffs for radio-frequency/mixed-signal "systems-on-a-chip"," *IEEE Transactions on Electron Devices*, vol. 50, no. 3, pp. 683–699, 2003.

[5] N. Nenadović, L. K. Nanver, and J. W. Slotboom, "Electrothermal limitations on the current density of high-frequency bipolar transistors," *IEEE Transactions on Electron Devices*, vol. 51, no. 12, pp. 2175–2180, 2004.

[6] L. K. Nanver, N. Nenadović, V. d'Alessandro, H. Schellevis, H. W. van Zeijl, R. Dekker, D. B. de Mooij, V. Zieren, and J. W. Slotboom, "A back-wafer contacted silicon-on-glass integrated bipolar process. Part I. The conflict electrical versus thermal isolation," *IEEE Transactions on Electron Devices*, vol. 51, no. 1, pp. 42–50, 2004.

[7] A. van Roosmalen, "Roadmap challenges - there is more than Moore," in *Proc. IEEE International Conference on Thermal and Mechanical Simulation and Experiments in Microelectronics and Microsystems (EuroSimE)*, p. 3, 2004.

[8] "ITRS edition 2009," tech. rep., International Technology Roadmap for Semiconductors, 2009.

[9] F. Schwierz and J. J. Liou, "RF transistors: Recent developments and roadmap toward terahertz applications," *Solid-State Electronics*, vol. 51, no. 8, pp. 1079–1091, 2007.

[10] V. Palankovski and S. Selberherr, "Thermal models for semiconductor device simulation," in *Proc. IEEE European Conference on High Temperature Electronics (HiTEN)*, pp. 25–28, 1999.

[11] N. Rinaldi and V. d'Alessandro, "Theory of electrothermal behavior of bipolar transistors: Part I - single-finger devices," *IEEE Transactions on Electron Devices*, vol. 52, no. 9, pp. 2009–2021, 2005.

[12] N. Rinaldi and V. d'Alessandro, "Theory of electrothermal behavior of bipolar transistors: Part II - two-finger devices," *IEEE Transactions on Electron Devices*, vol. 52, no. 9, pp. 2022–2033, 2005.

[13] N. Rinaldi and V. d'Alessandro, "Theory of electrothermal behavior of bipolar transistors: Part III-impact ionization," *IEEE Transactions on Electron Devices*, vol. 53, no. 7, pp. 1683–1697, 2006.

[14] L. La Spina, E. Iborra, H. Schellevis, M. Clement, J. Olivares, and L. K. Nanver, "Aluminum nitride for heatspreading in RF IC's," *Solid-State Electronics*, vol. 52, no. 9, pp. 1359–1363, 2008.

[15] N. Rinaldi, "Thermal analysis of solid-state devices and circuits: an analytical approach," *Solid-State Electronics*, vol. 44, no. 10, pp. 1789–1798, 2000.

[16] V. Székely and K. Tarnay, "Accurate algorithm for temperature calculation of devices in nonlinear-circuit-analysis programs," *Electronics Letters*, vol. 8, no. 19, pp. 470–472, 1972.

[17] L. Codecasa, D. D'Amore, and P. Maffezzoni, "Compact modeling of electrical devices for electrothermal analysis," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 50, no. 4, pp. 465–476, 2003.

[18] F. Masana, "A straightforward analytical method for extraction of semiconductor device transient thermal parameters," *Microelectronics Reliability*, vol. 47, no. 12, pp. 2122–2128, 2007.

[19] V. Székely, "On the representation of infinite-length distributed RC one-ports," *IEEE Transactions on Circuits and Systems*, vol. 38, no. 7, pp. 711–719, 1991.

[20] V. Székely, "Identification of RC networks by deconvolution: chances and limits," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 45, no. 3, pp. 244–258, 1998.

[21] Z. Jakopović, Z. Benčić, and R. Končar, "Identification of thermal equivalent - circuit parameters for semiconductors," in *Proc. IEEE Workshop on Computers in Power Electronics*, pp. 251–260, 1990.

[22] P. Bagnoli, C. Casarosa, M. Ciampi, and E. Dallago, "Thermal resistance analysis by induced transient (TRAIT) method for power electronic devices thermal characterization. I. Fundamentals and theory," *IEEE Transactions on Power Electronics*, vol. 13, no. 6, pp. 1208–1219, 1998.

[23] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The Computer Journal*, vol. 7, no. 4, pp. 308–313, 1965.

[24] Y. C. Gerstenmaier, W. Kiffe, and G. Wachutka, "Combination of thermal subsystems by use of rapid circuit transformation and extended two-port theory," *Microelectronics Journal*, vol. 40, pp. 26–34, 2009.

[25] E. Protonotarios and O. Wing, "Theory of nonuniform RC lines, Part I: Analytic properties and realizability conditions in the frequency domain," *IEEE Transactions on Circuit Theory*, vol. 14, no. 1, pp. 2–12, 1967.

[26] V. Székely, M. Rencz, S. Török, and S. Ress, "Calculating effective board thermal parameters from transient measurements," *IEEE Transactions on Components and Packaging Technologies*, vol. 24, no. 4, pp. 605–610, 2001.

[27] M. Rencz, V. Székely, A. Morelli, and C. Villa, "Determining partial thermal resistances with transient measurements, and using the method to detect die attach discontinuities," in *Proc. IEEE Semiconductor Thermal Measurement and Management Symposium (Semi-Therm)*, pp. 15–20, 2002.

[28] M. Rencz and V. Székely, "Measuring partial thermal resistances in a heat-flow path," *IEEE Transactions on Components and Packaging Technologies*, vol. 25, no. 4, pp. 547–553, 2002.

[29] M. Rencz and V. Székely, "Structure function evaluation of stacked dies," in *Proc. IEEE Semiconductor Thermal Measurement and Management Symposium (Semi-Therm)*, pp. 50–54, 2004.

[30] V. Székely and T. V. Bien, "Fine structure of heat flow path in semiconductor devices: A measurement and identification method," *Solid-State Electronics*, vol. 31, no. 9, pp. 1363–1368, 1988.

[31] Y. C. Gerstenmaier, H. Pape, and G. Wachutka, "Rigorous model and network for static thermal problems," *Microelectronics Journal*, vol. 33, no. 9, pp. 711–718, 2002.

[32] E. Bosch, "Thermal compact models: an alternative approach," *IEEE Transactions on Components and Packaging Technologies*, vol. 26, no. 1, pp. 173–178, 2003.

[33] M.-N. Sabry, "Compact thermal models for electronic systems," *IEEE Transactions on Components and Packaging Technologies*, vol. 26, no. 1, pp. 179–185, 2003.

[34] J. V. Beck, K. D. Cole, A. Haji-Sheikh, and B. Litkouhi, *Heat Conduction using Green's Functions*. Hemisphere Publishing Corporation, 1992.

[35] H.-C. Wu, S. Mijalkovic, and J. Burghartz, "A referenced geometry based configuration scalable Mextram model for bipolar transistors," in *Proc. IEEE International Behavioral Modeling and Simulation Workshop (BMAS)*, pp. 50–55, 2006.

[36] *Comsol 3.5a Users' Guide*. Comsol AG, 2009.

[37] L. La Spina, H. Schellevis, N. Nenadović, and L. K. Nanver, "PVD aluminium nitride as heat spreader in silicon-on-glass technology," in *Proc. IEEE International Conference on Microelectronics (MIEL)*, pp. 338–341, 2006.

[38] F. Brotzen, P. Loos, and D. Brady, "Thermal conductivity of thin SiO_2 films," *Thin Solid Films*, vol. 207, no. 1-2, pp. 197–201, 1992.

[39] S. Shin, H. N. Cho, B. S. Kim, and H. H. Cho, "Influence of upper layer on measuring thermal conductivity of multilayer thin films using differential 3-ω method," *Thin Solid Films*, vol. 517, no. 2, pp. 933–936, 2008.

[40] N. Rinaldi, "Small-signal operation of semiconductor devices including self-heating, with application to thermal characterization and instability analysis," *IEEE Transactions on Electron Devices*, vol. 48, no. 2, pp. 323–331, 2001.

[41] N. Le Gallou, J. M. Nebus, E. Ngoya, and H. Buret, "Analysis of low frequency memory and influence on solid state HPA intermodulation characteristics," in *IEEE International Microwave Symposium Digest (MTT-S)*, vol. 2, pp. 979–982, 2001.

[42] S. Boumaiza and F. Ghannouchi, "Thermal memory effects modeling and compensation in RF power amplifiers and predistortion linearizers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, no. 12, pp. 2427–2433, 2003.

[43] T. Chen, W.-M. L. Kuo, E. Zhao, Q. Liang, Z. Jin, J. Cressler, and A. Joseph, "On the high-temperature (to 300 °C) characteristics of SiGe HBTs," *IEEE Transactions on Electron Devices*, vol. 51, no. 11, pp. 1825–1832, 2004.

[44] J.-S. Rieh, D. Greenberg, Q. Liu, A. Joseph, G. Freeman, and D. Ahlgren, "Structure optimization of trench-isolated SiGe HBTs for simultaneous improvements in thermal and electrical performances," *IEEE Transactions on Electron Devices*, vol. 52, no. 12, pp. 2744–2752, 2005.

[45] 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 Transactions on Electron Devices*, vol. 40, no. 11, pp. 1917–1927, 1993.

[46] W. Liu and A. Khatibzadeh, "The collapse of current gain in multi-finger heterojunction bipolar transistors: its substrate temperature dependence, instability criteria, and modeling," *IEEE Transactions on Electron Devices*, vol. 41, no. 10, pp. 1698–1707, 1994.

[47] W. Liu, "Thermal coupling in 2-finger heterojunction bipolar transistors," *IEEE Transactions on Electron Devices*, vol. 42, no. 6, pp. 1033–1038, 1995.

[48] W. Liu, "Failure mechanisms in AlGaAs/GaAs power heterojunction bipolar transistors," *IEEE Transactions on Electron Devices*, vol. 43, no. 2, pp. 220–227, 1996.

[49] W. Liu, A. Khatibzadeh, J. Sweder, and H.-F. Chau, "The use of base ballasting to prevent the collapse of current gain in AlGaAs/GaAs heterojunction bipolar transistors," *IEEE Transactions on Electron Devices*, vol. 43, no. 2, pp. 245–251, 1996.

[50] W.-C. Liu, K.-H. Yu, R.-C. Liu, K.-W. Lin, K.-P. Lin, C.-H. Yen, C.-C. Cheng, and K.-B. Thei, "Investigation of temperature-dependent characteristics of an n^+-InGaAs/n-GaAs composite doped channel HFET," *IEEE Transactions on Electron Devices*, vol. 48, no. 12, pp. 2677–2683, 2001.

[51] W. Liu and A. A. Balandin, "Temperature dependence of thermal conductivity of $Al_xGa_{1-x}N$ thin films measured by the differential 3ω technique," *Applied Physics Letters*, vol. 85, no. 22, pp. 5230–5232, 2004.

[52] E. K. Sichel and J. I. Pankove, "Thermal conductivity of GaN, 25-360 K," *Journal of Physics and Chemistry of Solids*, vol. 38, no. 3, pp. 330–330, 1977.

[53] G. A. Slack, L. J. Schowalter, D. Morelli, and J. A. Freitas, "Some effects of oxygen impurities on AlN and GaN," *Journal of Crystal Growth*, vol. 246, no. 3-4, pp. 287–298, 2002.

[54] B. E. A., von Muench W., and P. E., "Thermal conductivity and electrical properties of 6H silicon carbide," *Journal of Applied Physics*, vol. 50, no. 9, pp. 5790–5794, 1979.

[55] S. Pinel, A. Marty, J. Tasselli, J.-P. Bailbe, E. Beyne, R. Van Hoof, S. Marco, J. R. Morante, O. Vendier, and M. Huan, "Thermal modeling and management in ultrathin chip stack technology," *IEEE Transactions on Components and Packaging Technologies*, vol. 25, no. 2, pp. 244–253, 2002.

[56] J. Palacin, M. Salleras, J. Samitier, and S. Marco, "Dynamic compact thermal models with multiple power sources: application to an ultrathin chip stacking technology," *IEEE Transactions on Advanced Packaging*, vol. 28, no. 4, pp. 694–703, 2005.

[57] J. Burghartz, ed., *Ultra-thin Chip Technology and Applications*. Springer, 2010. I edn.

[58] D. Dawson, A. Gupta, and M. Salib, "CW measurement of HBT thermal resistance," *IEEE Transactions on Electron Devices*, vol. 39, no. 10, pp. 2235–2239, 1992.

[59] D. T. Zweidinger, R. M. Fox, J. S. Brodsky, T. Jung, and S.-G. Lee, "Thermal impedance extraction for bipolar transistors," *IEEE Transactions on Electron Devices*, vol. 43, no. 2, pp. 342–346, 1996.

[60] N. Bovolon, P. Baureis, J.-E. Muller, P. Zwicknagl, R. Schultheis, and E. Zanoni, "A simple method for the thermal resistance measurement of AlGaAs/GaAs heterojunction bipolar transistors," *IEEE Transactions on Electron Devices*, vol. 45, no. 8, pp. 1846–1848, 1998.

[61] M. Reisch, "Self-heating in BJT circuit parameter extraction," *Solid-State Electronics*, vol. 35, no. 5, pp. 677–679, 1992.

[62] W. Liu and A. Yuksel, "Measurement of junction temperature of an AlGaAs/GaAs heterojunction bipolar transistor operating at large power densities," *IEEE Transactions on Electron Devices*, vol. 42, no. 2, pp. 358–360, 1995.

[63] S. Marsh, "Direct extraction technique to derive the junction temperature of HBT's under high self-heating bias conditions," *IEEE Transactions on Electron Devices*, vol. 47, no. 2, pp. 288–291, 2000.

[64] R. Menozzi, J. Barrett, and P. Ersland, "A new method to extract HBT thermal resistance and its temperature and power dependence," *IEEE Transactions on Device and Materials Reliability*, vol. 5, no. 3, pp. 595–601, 2005.

[65] J. C. J. Paasschens, S. Harmsma, and R. van der Toorn, "Dependence of thermal resistance on ambient and actual temperature," in *Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting (BCTM)*, pp. 96–99, 2004.

[66] J. J. Sparkes, "Voltage feedback and thermal resistance in junction transistors," *Proceedings of the IRE*, vol. 48, no. 10, pp. 1305–1306, 1958.

[67] W. B. Joyce, "Thermal resistance of heat sinks with temperature-dependent conductivity," *Solid-State Electronics*, vol. 18, no. 4, pp. 321–322, 1975.

[68] S. Russo, V. d'Alessandro, L. La Spina, N. Rinaldi, and L. K. Nanver, "Evaluating the self-heating thermal resistance of bipolar transistors by DC measurements: A critical review and update," in *Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting (BCTM)*, pp. 95–98, 2009.

[69] S. Russo, L. La Spina, V. d'Alessandro, N. Rinaldi, and L. K. Nanver, "Influence of layout design and on-wafer heatspreaders on the thermal behavior of fully-isolated bipolar transistors: Part I - static analysis," *Solid-State Electronics*, vol. 54, no. 8, pp. 745–753, 2010.

[70] D. L. Blackburn and F. F. Oettinger, "Transient thermal response measurements of power transistors," *IEEE Transactions on Industrial Electronics and Control Instrumentation*, vol. IECI-22, no. 2, pp. 134–141, 1975.

[71] S. Russo, L. La Spina, V. d'Alessandro, N. Rinaldi, and L. K. Nanver, "Influence of layout design and on-wafer heatspreaders on the thermal behavior of fully-isolated bipolar transistors: Part II - dynamic analysis," *Solid-State Electronics*, vol. 54, no. 8, pp. 754–762, 2010.

[72] N. Nenadović, S. Mijalkovic, L. K. Nanver, L. K. J. Vandamme, V. d'Alessandro, H. Schellevis, and J. W. Slotboom, "Extraction and modeling of self-heating and mutual thermal coupling impedance of bipolar transistors," *IEEE Journal of Solid-State Circuits*, vol. 39, no. 10, pp. 1764–1772, 2004.

[73] A. Maxim and G. Maxim, "Electrothermal SPICE macromodeling of the power bipolar transistor including the avalanche and secondary breakdowns," in *Proc. IEEE Annual Conference of the Industrial Electronics Society (IECON)*, vol. 1, pp. 348–352, 1998.

[74] V. d'Alessandro, N. Nenadović, F. Tamigi, N. Rinaldi, L. K. Nanver, and J. W. Slotboom, "A novel SPICE macromodel of BJTs including the temperature dependence of high-injection effects," in *Proc. IEEE International Conference on Microelectronics (MIEL)*, vol. 1, pp. 253–256, 2004.

[75] F. M. De Paola, J. P. Nowakowski, V. d'Alessandro, and N. Rinaldi, "Fully automated electrothermal simulation using standard CAD tools," in *Proc. IEEE International Conference on Microelectronics (MIEL)*, pp. 451–454, 2006.

[76] J. P. Nowakowski, V. d'Alessandro, F. M. De Paola, M. Spirito, and N. Rinaldi, "Advances in electrothermal simulation of solid-state devices and circuits using commercial CAD tools," in *Proc. International Conference on Microtechnology and Thermal Problems in Electronics (MicroTherm)*, pp. 65–72, 2007.

[77] N. Rinaldi, "On the modeling of the transient thermal behavior of semiconductor devices," *IEEE Transactions on Electron Devices*, vol. 48, no. 12, pp. 2796–2802, 2001.

[78] K. Poulton, K. L. Knudesn, J. J. Corcoran, K.-C. Wang, R. L. Pierson, R. B. Nubling, and M.-C. F. Chang, "Thermal design and simulation of bipolar integrated circuits," *IEEE Journal of Solid-State Circuits*, vol. 27, no. 10, pp. 1379–1387, 1992.

[79] K. J. Negus, R. W. Franklin, and M. M. Yovanovich, "Thermal modeling and experimental techniques for microwave bipolar devices," *IEEE Transactions on Components, Hybrids, and Manufacturing Technology*, vol. 12, no. 4, pp. 680–689, 1989.

[80] S.-S. Lee and D. Allstot, "Electrothermal simulation of integrated circuits," *IEEE Journal of Solid-State Circuits*, vol. 28, no. 12, pp. 1283–1293, 1993.

[81] 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 Electronics*, vol. 38, no. 7, pp. 1409–1412, 1995.

[82] M. Pesare, A. Giorgio, and A. Perri, "Analytical approach for thermal and electrical design of multilayer structure integrated devices," *Electronics Letters*, vol. 36, no. 13, pp. 1120–1121, 2000.

[83] L. La Spina, V. d'Alessandro, S. Russo, N. Rinaldi, and L. K. Nanver, "Influence of concurrent electrothermal and avalanche effects on the safe operating area of multifinger bipolar transistors," *IEEE Transactions on Electron Devices*, vol. 56, no. 3, pp. 483–491, 2009.

[84] L. La Spina, V. d'Alessandro, S. Russo, and L. K. Nanver, "Thermal design of multifinger bipolar transistors," *IEEE Transactions on Electron Devices*, vol. 57, no. 8, pp. 1789–1800, 2010.

[85] J. Kim, A. Sadovnikov, T. Chen, and J. Babcock, "Safe operating area from self-heating, impact ionization, and hot carrier reliability for a SiGe HBT on SOI," in *Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting (BCTM)*, pp. 230–233, 2007.

[86] T. Vanhoucke and G. A. M. Hurkx, "Unified electro-thermal stability criterion for bipolar transistors," in *Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting (BCTM)*, pp. 37–40, 2005.

[87] L. Codecasa, "Thermal networks from phonon Boltzmann's transport equation: Part I - fundamentals," in *Proc. IEEE International Workshop on Thermal Investigation of ICs and Systems (THERMINIC*, pp. 135–135, 2007.

[88] V. d'Alessandro, I. Marano, S. Russo, D. Céli, A. Chantre, P. Chevalier, F. Pourchon, and N. Rinaldi, "Impact of layout and technology parameters on the thermal resistance of SiGe:C HBTs," in *Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting (BCTM)*, pp. 137–140, 2010.

List of publications

[1] S. Russo, L. La Spina, V. d'Alessandro, N. Rinaldi, and L. K. Nanver, "Thermal design of fully-isolated bipolar transistors," in *Proc. IEEE International Workshop on Thermal Inveatigation of ICs and Systems (THERMINIC)*, pp. 101–105, 2008.

[2] S. Russo, L. La Spina, V. d'Alessandro, N. Rinaldi, and L. Nanver, "Impact of aluminum nitride heatspreaders on the thermal impedance of silicon-on-glass BJTs," in *Proc. Semiconductor Advances for Future Electronics (SAFE/STW)*, pp. 501–505, 2008.

[3] L. La Spina, S. Russo, V. d'Alessandro, and L. Nanver, "Thermal design of multifinger bipolar transistors," in *Proc. Semiconductor Advances for Future Electronics (SAFE/STW)*, pp. 565–568, 2008.

[4] L. La Spina, V. d'Alessandro, S. Russo, N. Rinaldi, and L. K. Nanver, "Influence of concurrent electrothermal and avalanche effects on the safe operating area of multifinger bipolar transistors," *IEEE Transactions on Electron Devices*, vol. 56, no. 3, pp. 483–491, 2009.

[5] S. Russo, L. La Spina, V. d'Alessandro, N. Rinaldi, M. De Magistris, and L. K. Nanver, "Thermal transient behavior of silicon-on-glass BJTs," in *Proc. IEEE International Conference on Thermal, Mechanical and Multi-Physics simulation and Experiments in Microelectronics and Microsystems (EuroSimE)*, pp. 1–8, 2009.

[6] S. Russo, L. La Spina, V. d'Alessandro, N. Rinaldi, M. De Magistris, and L. K. Nanver, "Numerical analysis of the dynamic thermal behavior of RF bipolar transistors," in *Proc. of International Conference on Computational Methods for Thermal Problems (ThermaComp)*, pp. 345–348, 2009.

[7] S. Russo, V. d'Alessandro, L. La Spina, N. Rinaldi, and L. K. Nanver, "Evaluating the self-heating thermal resistance of bipolar transistors by DC measurements: A critical review and update," in *Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting (BCTM)*, pp. 95–98, 2009.

[8] L. La Spina, V. d'Alessandro, S. Russo, N. Rinaldi, and L. K. Nanver, "Electrothermal behavior of highly-symmetric three-finger bipolar transistors," in *Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting (BCTM)*, pp. 21–24, 2009.

[9] S. Russo, L. La Spina, V. d'Alessandro, N. Rinaldi, and L. K. Nanver, "Influence of layout design and on-wafer heatspreaders on the thermal behavior of fully-isolated bipolar transistors: Part I - static analysis," *Solid-State Electronics*, vol. 54, no. 8, pp. 745–753, 2010.

[10] S. Russo, L. La Spina, V. d'Alessandro, N. Rinaldi, and L. K. Nanver, "Influence of layout design and on-wafer heatspreaders on the thermal behavior of fully-isolated bipolar transistors: Part II - dynamic analysis," *Solid-State Electronics*, vol. 54, no. 8, pp. 754–762, 2010.

[11] L. La Spina, V. d'Alessandro, S. Russo, and L. K. Nanver, "Thermal design of multifinger bipolar transistors," *IEEE Transactions on Electron Devices*, vol. 57, no. 8, pp. 1789–1800, 2010.

[12] V. d'Alessandro, I. Marano, S. Russo, D. Céli, A. Chantre, P. Chevalier, F. Pourchon, and N. Rinaldi, "Impact of layout and technology parameters on the thermal resistance of SiGe:C HBTs," in *Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting (BCTM)*, pp. 137–140, 2010.

[13] N. Rinaldi, S. Russo, and V. d'Alessandro, *Thermal effects in thin silicon dies: simulation and modeling*. Chapter 23 in "Ultra-thin Chip Technology and Applications" edited by J. Burghartz, Springer, 2010. I edn.

List of abbreviations

1-D – one-dimensional

2-D – two-dimensional

3-D – three-dimensional

ADS – advanced design system; it is an electronic circuit simulator

b.c. – boundary condition

BCB – benzocyclobutene

BOX – buried oxide

BJT – bipolar junction transistor

CMOS – complementary metal-oxide-semiconductor

CTM – compact thermal model

DC – direct current

EE – Early effect

ET – electrothermal

FEM – finite element method

HBT – heterojunction bipolar transistor

HEMT – high electron mobility transistor

HF – high frequency

IC – integrated circuits

IFX – Infineon technologies

II – impact ionization

MEMS – micro-electro-mechanical systems

MOS – metal-oxide-semiconductor

MOSFET – metal-oxide-semiconductor field effect transistor

NDR – negative differential resistance

NID – network identification by deconvolution

PCB – printed circuit board

RC – resistance-capacitance network

RF – radio frequency

SDD – symbolically defined device

SEM – scansion electron microscopy

SOA – safe operating area

SiP – system-in-package

SoC – system-on-chip

SOG – silicon-on-glass

SOI – silicon-on-insulator

SPICE – simulation program with integrated circuit emphasis; it is an electronic circuit simulator

STM – ST Microelectronics

TEM – transmission electron microscopy

UTCS – ultra-thin chip stacking; it is technology

Acknowledgments

This thesis (and the research work beyond it) would have been not possible without the contribution and help of many people.
First, I want to express my sincere gratitude to my supervisor, Prof. Niccolò Rinaldi, who supported me during the past three years of research. His knowledge and suggestions have been of the utmost importance to fulfill my research activity.
Furthermore I would like to thank Dr. Vincenzo "Enzo ASAP" d'Alessandro for his constant help in the past three years, the fruitful discussions on solid-state devices, and the time spent working together that has provided a number of scientific results. I also want to especially thank Enzo for his patience and friendship.
I especially want to thank Grazia "Grace Rock" Sasso and Michele Riccio for the nice time spent together and the help they provided since we started this PhD "adventure" in 2007.
I would also like to thank my friends at DIBET, Ilaria, Jurek, Valeria, Maurizio, Pierluigi, Marino, Matteo, Luca, Lucio, Martina, Davide, Nicola, Mariangela, for the pleasant moments spent together.
I want to express my sincere gratitude to Prof. Lis K. Nanver, that supervised me during the time spent in Delft, for her valuable help, kind encouragement and the time she spent correcting my English.
Furthermore, I want to express my gratitude to Dr. Luigi La Spina for his friendship and help. Since the first day I arrived in The Netherlands I could always count on him.
I would also like to thank Dr. Luigi "Gigi" Mele for his friendship and all the profitable time spent together.
Moreover, I want to thank all my friends at DIMES for having made me feel at home: Mauro Marchetti, Agatha Šakić, Marco Spirito, Marko Mihailović, Alessandro Baiano, Fabio Santagata, Benjamin Mimoun, Bruno Morana, Gennaro Gentile, Vladimir Jovanović, Giampaolo Lorito, Francesco

Sarubbi, Olindo Isabella, Yann Civale, Michael Wank and all those I have forgotten.
Thanks are due to my whole family and especially to my mother for her patience and support.
Finally, I want to express my gratitude to my beloved Viola for her understanding, patience and unconditional help.