UNIVERSITÀ DEGLI STUDI DI PARMA

Dottorato di Ricerca in Tecnologie dell'Informazione

XXIV Ciclo

Thermal and electro-thermal modeling of electronic devices and systems for high-power and high-frequency applications

Coordinatore:

Chiar.mo Prof. Marco Locatelli

Tutor:

Chiar.mo Prof. Roberto Menozzi

Dottorando: *Mirko Bernardoni*

Gennaio 2012

to my Mother

and to all the real Friends

of mine

Table of Contents

List of Figures

List of Tables

Introduction

The work covered in this thesis is focused mainly on four topics.

Chapter 1 focuses on the use of the Finite Element Method (FEM) to study the thermal dynamics in power devices, power electronics boards, and assemblies. The work is focused on thermal and fluid-thermal aspects of the components of a Switching Mode Power Supply used to supply detectors for High-Energy Physics Experiment performed at CERN, where the power supply has to be designed in order to meet stringent thermal constraints.

Chapter 2, which is the central chapter of this thesis, is focused on the self-consistent, electro-thermal modeling of power devices. A primary concern in device modeling is the correct evaluation of the device channel temperature. In particular, while plenty of accurate electro-thermal, large-signal device models are available nowadays, the *thermal* aspect is often overlooked or poorly accounted for in the simulation process. Although it is possible to develop accurate thermal models by means of FEM, for instance, these do not lend themselves to be inserted into circuit simulators. A solution has been found in the Lumped-Element (LE) thermal modeling, in which thermal resistances and thermal capacitances are used to model the thermal dynamics of a given device. The approach shown here is based on the physical dimensions and material characteristics of the device, and it accounts for three-dimensional effects, non-linearities in thermal conductivities, complex boundary conditions, modeling both thermal steady-state and transients, *without using fitting parameters*. This approach has been applied to several device structures, and compared with FEM simulations, with excellent results. When this approach becomes impractical due to the complexity of the structure, empirical compact Foster thermal networks have been used, and a simple automated way to optimize the fitting process will be shown.

Chapter 3, which focuses on *thermo-reflectance* thermal measurement, describes the work performed at Centre for Device Thermography and Reliability (CDTR), H. H. Wills Department of Physics, University of Bristol (UK). The development

of a bench to perform high-spatial resolution thermal measurement by the *thermo-reflectance* contact-less method is described here. This technique is based on the surface reflectance change with temperature.

Chapter 4 is focused on a secondary activity aimed at the development of an electronic bench to generate high-voltage pulses (kV amplitude), with nano-seconds time duration, for biomedical applications; in particular the final goal is to perform studies on the possibility of reactivating *apoptosis* in cancer cells. This is still work-in-progress and this chapter describes the activity carried out so far, and the future developments.

Three appendixes complete the thesis: the first appendix is a complement to Chapter 1, in which the FE method is used to create a preliminary model describing the accumulated damage of soldering joints in electronic components (the so called *fatigue*); the second appendix is the Matlab code to implement the FFT routine used to post-process the data obtained from *thermo-reflectance* measurements, described in Chapter 3; the third appendix is the list of publications.

Chapter 1

Thermal and thermo-fluid dynamics FEM modeling of power modules

This Chapter deals with a numerical thermo-fluid-dynamic study of power assemblies; in particular the study will be focused on the thermal behavior of a *Switching Mode Power Supply* (SMPS) which has been the main topic of a project carried out by the Universities of Parma, Cassino, Milano and Padova in the context of the PRIN 2007 project. In particular, the task was the complete re-design of the SMPS used to supply detectors in the ATLAS experiment at CERN (Genéve). This chapter will show the simulation procedure which has been followed to obtain a quantitative description of the thermal behavior of the entire assembly.

1.1 Introduction

PRIN 2007 was a project in which the task was, basically, the re-design of a SMPS with stringent thermal constraints. In this section, some technical details about the SMPS will be given, for the sake of completeness (even if this part has been mainly developed by the University of Milano) and to frame the problem. The particular features of this SPMS can be summarized as follows:

- ♢ the ability of working in the presence of high magnetic fields (typical of high-energy experiments);
- ♢ the ability of working with high power density, giving rise to very low temperature increase, with liquid cooling capability (not to heat the detectors, which are very close to the SMPS itself);

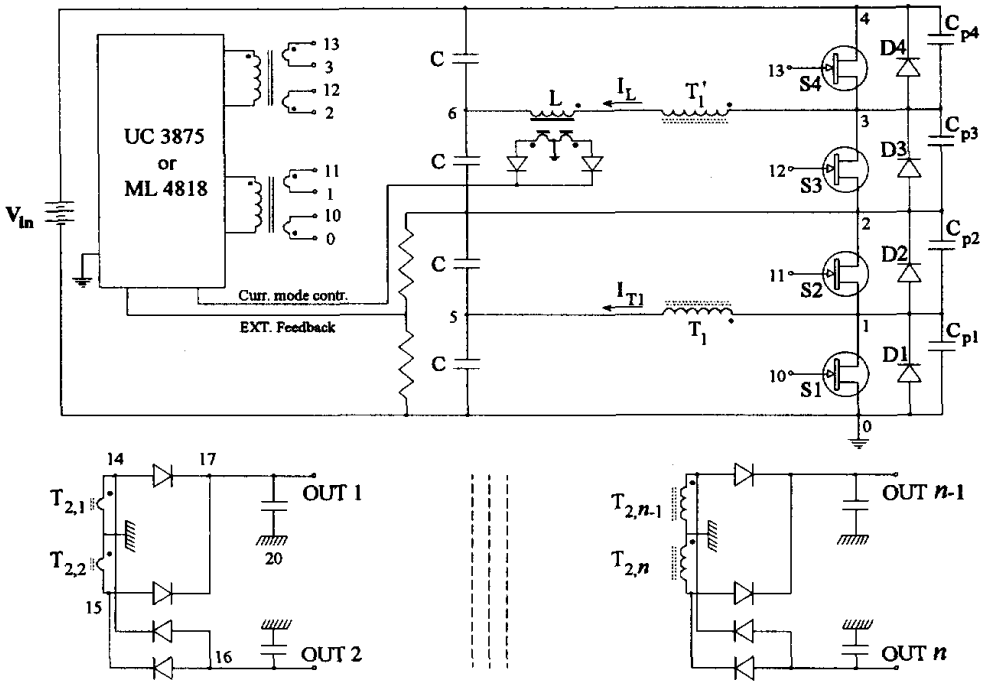


Figure 1.1 – Schematic of the SMPS (reprinted from [1]).

♢ the extremely high reliability, achieved by the introduction of redundancy (due to the location and to the impossibility of stopping the experiments for servicing defective parts).

The electrical schematic of the SMPS used in this project is shown in Figure 1.1: this SMPS is basically a phase-shift converter, in which one couple of transistors (S_1, S_3) is turned on in opposite phase with the other couple (S_2, S_4), and viceversa. Further details about the operation of this SMPS can be found in [1]. The conversion ratio of this converter is 280:14. The details that will be given below are just to introduce the project, since the chapter will discuss mainly the tasks of the University of Parma. The fluid-mechanical specifications of the assembly are the following, as set by the partners of the project:

♢ the power supply case is made of 1510 steel, 2 mm thick, with external dimensions of 150 mm × 285 mm × 402 mm;

♢ the number of modules in a case is three;

- ♢ the overall power dissipated by the modules is 800 W with a 2+1 redundancy; it means that, during normal operation conditions, all 3 modules are working and each one dissipates a power of 267 W; in case of failure of one modules, the other two have to supply the overall power, thus dissipating 800/2 = 400 W each. This is the worst case situation.

A mandatory step in the design is the liquid-cooling system, in this case a water-cooled aluminum heat-sink for each module in the case. The cooling specifications are the following:

- ♢ the water flow is 1.9 l/min, the maximum pressure drop is 0.35 bar, the inlet water temperature is 18 °C and the maximum temperature for the outgoing water is 25 °C;

- ♢ the heat-sink is an aluminum one, with a thickness of 15 mm and a channel diameter of 5 mm.

The theoretical power that can be extracted from such a heat-sink is given by:

$$Q = C_p \times \Delta T \times F = 4186 \cdot (25 - 18) \cdot \frac{1.9}{60} \simeq 928\,\mathrm{W} \tag{1.1}$$

where C_p is the specific heat of water in $[\mathrm{J}/(\mathrm{kg} \cdot \mathrm{K})]$ at constant pressure, ΔT is the temperature difference between outgoing and incoming water in [K], and F is the water mass flow (mass per unit time) in [kg/s]. Since the maximum dissipated power is 800 W, i.e., below the theoretical limit, the problem is now *how* to design the board to achieve as uniform as possible a temperature distribution over the board itself (avoiding areas in which the dissipated power gives rise to high temperature increases, or hot spots, that can be dangerous for the sensitive instrumentation close to the SMPS). In Figure 1.2, the structure of a single case containing three SMPSs with their heat-sinks is shown. Once the problem is fully defined, the next step is to develop a numerical model of the object under study. The tool used to develop the numerical model is the Finite Element Method (FEM). The object under study changes at every design step, because at the beginning the attention has been focused just on the case with the three boards (as in Figure 1.2, in which no details are given at the board contents), while in further design steps many additional details have been made available, and the attention has been focused on different components. The next sections will illustrate this approach, which follows the temporal evolution of the design (some parts will be intentionally neglected, since they were found not to be relevant to the design process).

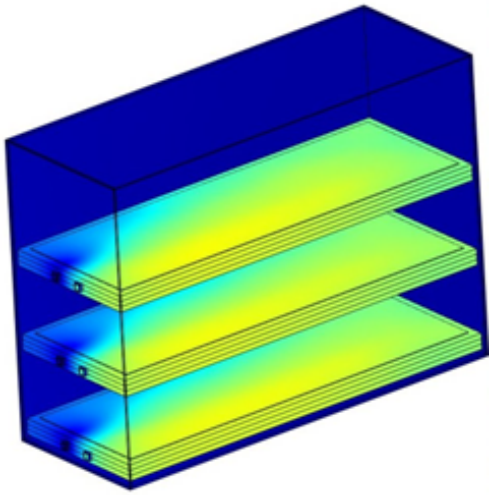


Figure 1.2 – The metal case with the three power supply modules and heat-sinks.

1.2 The thermo-fluid dynamics simulation process

In this part, the approach followed to describe the thermo-fluid dynamics inside the case will be shown. The device studied is a structure like the one shown in Figure 1.2, in which the power density is *not* representative of the effective power distribution, since the layout of the board (and thus the power sources that heat the board itself) is not known at this stage of the design. The aim of this stage is, basically:

- ♢ to learn and improve the capability of *how* to build a numerical model in which the thermal dynamics and the fluid motion dynamics are simulated at the same time;
- ♢ to test the performance of the simulator, especially in solving fluid-dynamic problems;
- ♢ to reduce the computational complexity as much as possible, for reasons that will be explained shortly.

First of all, a brief review of the physical equations related to the analyzed problems is following.

1.2.1 Heat equations

The thermal dynamics is described by means of the Fourier's law of heat propagation in solids, which can be written as follows for the most general case:

$$\varrho \cdot C_p \cdot \frac{\partial T}{\partial t} = \nabla(\lambda \cdot \nabla T) + q_g \tag{1.2}$$

where the meaning of different symbols is: ϱ is the density of the material in [kg/m^3], C_p is the specific heat of the material in [J/(kg $\cdot$ K)], T is the temperature in [K], λ is the thermal conductivity of the material in [W/(m $\cdot$ K)], and q_g is the heat generated per unit volume, expressed in [W/m^3]. The operator ∇ is a compact way for defining divergences and rotors, and can be written as a vector[1] as follows:

$$\nabla = \mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z} \tag{1.3}$$

where it is supposed to use a cartesian reference system, in which $\mathbf{i}, \mathbf{j}, \mathbf{k}$ are the versors of the x, y, z axes, respectively.

In some cases, Eqn. (1.2) can be simplified. In particular, in the case of study, the term q_g is null, transients are not of interest (thus, time derivatives goes to zero), and the thermal conductivities of materials are considered isotropic (it means that, for each single domain, λ does not depend on the direction of heat propagation). Eqn. (1.2) can thus be reduced to the following:

$$0 = \nabla^2 T. \tag{1.4}$$

Eqn. (1.4) is the equation that will be used to define the thermal problems in the rest of the study.

Once the equation has been defined for the domain, one must fix some boundary conditions, basically to define the "inputs" of the system. There are several types of boundary conditions, which describe how the body exchanges heat with the surrounding environment. There are three types of boundary conditions (T is the temperature over the surface):

fixed temperature in which we are going to fix the temperature over a surface of a body: $T = T_0$;

fixed heat flux in which the thermal flux is fixed through a surface. Matematically: $q = q_0 + h(T - T_{\text{ref}})$, where q is heat flux per surface unit in [W/m^2], q_0 is the imposed flux, T_{ref} is a reference temperature, and h is a heat exchange coefficient, or *convective* coefficient, that accounts for the fluid motions near the surface of interest; the physical concepts behind this simple condition are important, and then they will be explained shortly;

[1] It can be applied both to vectors and scalars, with both the scalar or vectorial product.

radiative convection in which the surface and the environment exchange heat by means of radiation, basically emitting/receiving electromagnetic waves. It is expressed by:

$$Q = \varepsilon \cdot \sigma \cdot A \cdot ({T_S}^4 - {T_C}^4) \tag{1.5}$$

where Q is the power supplied by the body in [W], ε is a coefficient – the emissivity of the body (1 for black bodies, 0 for white bodies), σ is the Stefan-Boltzmann constant equal to $5.67 \times 10^{-8}\,[\mathrm{W/(m^2K^4)}]$, and A is the area in $[\mathrm{m^2}]$ of the surface which is irradiating towards the ambient. T_S and T_C are the surface and the surrounding environment temperatures, respectively.

The sign for the heat flux is conventional, since the heat exchange can be bi-directional.

The convective boundary condition deserves to be explained in full detail. Consider, for example, an horizontal hot surface in contact with air. The air close to the surface tends to be hot, and to flow away from the surface (thus removing heat from the surface), and when the air cools down it flows again on the surface, and so on. The meaning of the parameters T_{ref} and h can be easily understood after some simple observations. It is obvious that the convective motions are related to a restricted volume of fluid, in particular that close to the surface: it can be assumed that, far enough from the surface, the fluid is in quiet, and also the temperature of the fluid itself is constant. This is the meaning of T_{ref}. Then, if the fluid is changed (for example water instead of air), or the temperature difference is increased, or a fan is used to force the fluid motion, the efficiency of heat exchange is modified. This is the meaning of h, which can itself be temperature dependent, which accounts for the *efficiency* of the thermal exchange. In general, then, $h = h(T)$; some values[2] can be found in [2, 3]. The coefficient h can vary over a wide range, from $2-25\,[\mathrm{W/(m^2K)}]$ for natural convection up to $10^5\,[\mathrm{W/(m^2K)}]$, for the case of liquid cooling with condensation (far improving the efficiency).

Another brief observation about the sign of the heat flux in the boundary conditions. Let us consider for example the convective boundary condition, here rewritten (in a slightly different form, assuming the flux q_0 null and considering the area $A\,[\mathrm{m^2}]$ of the surface: thus, the power is in [W] instead of being per unit area):

$$Q = h \cdot A \cdot (T - T_{\mathrm{ref}}) \tag{1.6}$$

and let us consider Figure 1.3; if we consider a positive flux with the same direction of the vector normal to the surface, then a heating process is described by a negative flux (since $T_{\mathrm{ref}} > T$ and then the direction of the flux is *opposite* to the direction of the surface versor), while a cooling process is described by a positive flux (since $T_{\mathrm{ref}} < T$). This can be against common sense, since in practice a positive flux is considered to

[2]The determination of the coefficient h is a very important step, as will be shown later on when the discussion will be focused on how to study the behavior of the heat-sink.

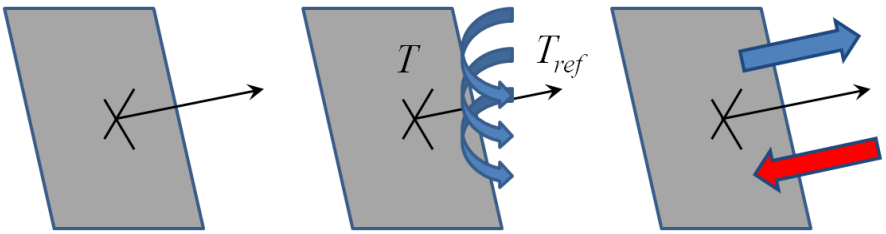


Figure 1.3 – A surface with its normal vector (left). Convective motions arise when the surface temperature T is different from the fluid reference temperature T_{ref} (center). A cooling process is is described by a positive flux (blue arrow, right), while a heating process is described by a negative flux (red arrow, right).

be impinging on a surface (and thus heating the surface), while a negative one is considered to cool down a surface; but, it is only a matter of convention, and then when setting the model for describing our system these conventions have to be accounted for.

1.2.2 The Navier-Stokes equations

The motion of a fluid is described by the Navier-Stokes equations:

$$\rho\frac{\partial \mathbf{u}}{\partial t} - \nabla \cdot \eta(\nabla \mathbf{u} + (\nabla \mathbf{u})^T) + \rho \mathbf{u} \cdot \nabla \mathbf{u} + \nabla p = \mathbf{F} \tag{1.7}$$

$$\nabla \mathbf{u} = 0 \tag{1.8}$$

where ρ denotes the specific mass of the fluid in [kg/m^3], $\mathbf{u}$ is the velocity field in [m/s], η is the viscosity of the fluid in [N $\cdot$ s/m^2], and p is the pressure in [Pa]. The second equation is the continuity equation. This set of equations are also known as *incompressible Navier Stokes equations*, because the fluid is considered uncompressible. There exists another set of equations, called *Weakly-Compressible Navier Stokes equations*, which describes the fluid motion considering the fluid slightly compressible. In this latter set of equations, the continuity equation and the momentum equation are considered:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \cdot \mathbf{u}) = 0 \tag{1.9}$$

$$\rho\frac{\partial \mathbf{u}}{\partial t} + \rho \mathbf{u} \cdot \nabla \mathbf{u} = -\nabla p + \nabla \cdot \left(\eta(\nabla \cdot \mathbf{u} + (\nabla \cdot \mathbf{u})^T) - (\frac{2}{3}\eta - \kappa_{\text{dv}})(\nabla \cdot \mathbf{u})\mathbf{I} \right) + \mathbf{F} \tag{1.10}$$

which is a formulation to describe the motion of a Newtonian fluid with an added term k_{dv} (this term represents a condition of non-thermodynamic equilibrium of each fluid

particle with the surrounding ones), which is set to zero by default. A fluid is called Newtonian when the relationship between the shear stress τ (the stress perpendicular to the cross section, in this case in the same direction of propagation of the fluid) [Pa] and the strain rate [s^{-1}] is the following:

$$\tau = \eta \frac{du}{dt} \tag{1.11}$$

which in practice states that the viscosity of the fluid is constant (an example of Newtonian fluid is water, while toothpaste is a non-Newtonian fluid). This model is useful when the fluid has different density, and this description is coupled with the thermal dynamics since it allows to describe, for instance, the variation of density of the fluid with temperature.

1.2.3 Modeling the liquid-cooling heat-sink

It is well known that the fluid-dynamics phenomena are the most difficult to model, due to the complexity of the equations. Numerical simulators are especially developed for solving fluid-dynamics problems. If the dependence of the fluid properties on the fluid temperature is also accounted for, the problem formulation becomes more complex and numerical convergence is more difficult to reach. In addition, these problems require a lot of computational power (8 GB RAM or more, multicore processors, and so on) and, often, some numerical "tricks" have to be included for facilitating the convergence of the problem.

While numerical simulation can actually be useful in particular situations (like microchannel heatsink, or other innovative structures), the case of study of a typical liquid-cooled heat-sink is not convenient to be studied with a full detailed fluid-dynamics simulation, mainly for two reasons:

1. liquid-cooled heat-sinks are commonly designed without the use of FEM simulators, because there are some well known relationships to design the heat-sink depending on the maximum temperature, pressure drop, and so on;

2. the coefficient of thermal exchange h can be easily estimated from the characteristics of the fluid motion, that are known once the design of the heat-sink is fixed (for example, the diameter of the channel, the flux, etc. . .).

It is useful to recall some basics concepts of fluid-dynamics, useful to get a simple model of the thermal exchange (basically, to estimate the value of the coefficient of thermal exchange h). A deeper description can be found in [2]. In particular, a set of *adimensional numbers* is used to describe the properties of a fluid, and this set is shown here.

The coefficient h contributes to the *Nusselt* number, defined as follows:

$$\mathrm{Nu} = \frac{hL}{k} \tag{1.12}$$

where k is the thermal conductivity of the fluid and L is the equivalent diameter of the channel, which is defined as $4S/P$, with S the cross section and P the perimeter of the channel. If the channel section is circular with diameter d, L is actually the diameter:

$$L = \frac{4 \times \frac{1}{4}\pi d^2}{\pi d} = d. \tag{1.13}$$

The calculation of the Nusselt number is fundamental to obtain the thermal exchange coefficient h. Usually, the Nusselt number is a function of the *Reynolds* number and the *Prandtl* number in the case of forced convection[3]. The case of a liquid-cooled heat-sink falls clearly in the forced convection case. The *Reynolds* number is used to describe the regime of fluid motion (laminar, transition, turbulent), and is defined as follows:

$$\mathrm{Re} = \frac{\rho U L}{\mu} = \frac{UL}{\nu} \tag{1.14}$$

where U is the average speed of the fluid in the channel, ρ is the density of the fluid, μ is the dynamic viscosity $[\mathrm{Pa} \cdot \mathrm{s}] = [\mathrm{kg}/(\mathrm{m} \cdot \mathrm{s})]$, $\nu = \mu/\rho$ is the kinematic viscosity $[\mathrm{m}^2/\mathrm{s}]$, and L is the equivalent diameter of the channel. The *Prandtl* number is defined as follows:

$$\mathrm{Pr} = \frac{\mu c_p}{k} \tag{1.15}$$

where μ is the dynamic viscosity, c_p is the specific heat at constant pressure $[\mathrm{J}/(\mathrm{kg}\cdot\mathrm{K})]$, and k is the thermal conductivity of the fluid.

Once the Prandtl and the Reynolds number are known, the relationship between them and the Nusselt number is usually given as (a, b, C are constants depending on the particular case of study):

$$\mathrm{Nu} = f(\mathrm{Re}, \mathrm{Pr}) = C \cdot \mathrm{Re}^a \cdot \mathrm{Pr}^b; \tag{1.16}$$

this method is actually simpler than the solution of the Navier-Stokes equations: it allows to focus on the thermal dynamics of the problem, since forced convection is, under certain aspects, simpler than natural convection: in the latter, the motion of the fluid arises *exclusively* from thermal gradients (and from the force field which is acting in the opposite direction), while in forced convection the motion of the fluid is mostly fixed by the external mechanical device (like a pump, for instance), and is

[3]The Reynolds number has to be replaced by the *Grashof* number in the case of natural convection.

less influenced by the thermal gradients. For the case of study, we used the following relationship, which is taken from [2] and holds in the case of a turbulent pipe flow:

$$\mathrm{Nu} = \frac{(f/8)(\mathrm{Re} - 1000) \cdot \mathrm{Pr}}{1 + 12.7\sqrt{f/8}(\mathrm{Pr}^{2/3} - 1)} \tag{1.17}$$

where f is the friction factor of the pipe given by

$$f = \frac{1}{(1.82 \log_{10} \mathrm{Re} - 1.64)^2}. \tag{1.18}$$

The above equations are a more recent formulation for the particular case of study with respect to Eqn. (1.16). The value of h obtained from the former relationships has been used for the thermal modeling of the water-cooled heat-sink. The important aspect on which to focus is the application of these basic concepts of fluid-dynamic to bypass the simulation problems arising from the complicated formulation of the Navier-Stokes equations. This allows to reduce the computational weight of the model remarkably or, on the other hand, to increase the detail of the simulation but focusing only on the thermal dynamics, since all the fluid-dynamics is described by means of the equivalent coefficient h. The underlying approach is to reduce the thermal description of the system to a thermal *conduction* problem, which can be easily solved by FEM simulators. This concept will be fully exploited in the next chapter, when the Lumped Element approach for the description of the thermal behavior of electron devices will be illustrated.

1.2.4 FEM thermal modeling of power assemblies

Once the type of active device used in the power supply has been chosen, the development of a FEM model that describes the thermal (static) behavior of the device is straightforward. The transient behavior is beyond the scope of this study. In particular, rather than the device itself, the thermal behavior is mainly dictated by its package (like TO-220, D^2PAK, and so on). In this phase of the work the attention is focused on the study of which board technology is the best thermal management, and which kind of mounting technology is the best for extracting heat from the device. This allows to have a quantitative overview of the thermal performance of power assemblies in sealed enclosures, like the ones used for high-energy physics experiments [4].

The aim of this section is to build a detailed three-dimensional thermal model of the the active devices used in the power supply. Two packages have been studied: the TO-220 and the D^2PAK, as shown in Figure 1.4. Note from Figure 1.4 that the TO-220, originally designed for a THT (*Through Hole Technology*) mounting, can be easily adapted to an SMD (*Surface Mount Technology*) mount simply by bending the pins. In fact, the D^2PAK is the SMD commercial evolution of the TO-220. A detailed

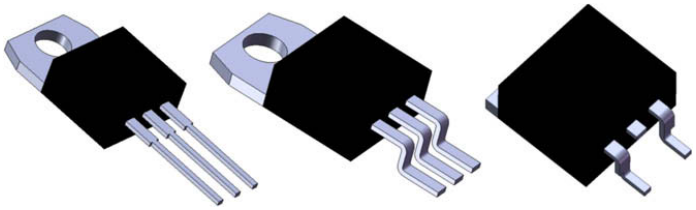


Figure 1.4 – Types of packages studied: from left, a TO-220, a modified TO-220 to be SMD mounted, and a D^2PAK.

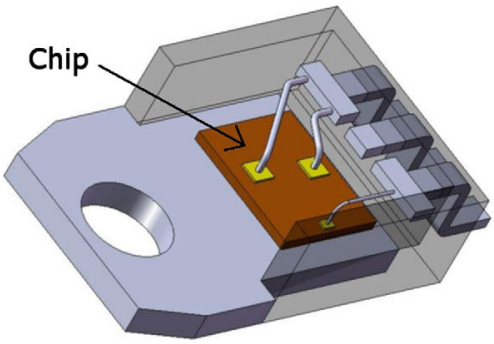


Figure 1.5 – Inner structure of the modeled device.

model of the inside of the device is given in Figure 1.5. The modeled wires are the two wires that connect the source and the gate with the die. The drain is electrically connected with the copper flange underneath the die; from a thermal point of view the heat is mainly exchanged by the copper flange (details about the electrical connection between the aluminum flange and the drain contact were not given). The heat is generated in the silicon die (a dissipated power of 20 W has been assumed). To fix the boundary conditions, it is necessary to define the configuration in which the device will work, in particular which technology will be used for assembling the power supply. In particular, four topologies have been considered (each one mounted on a heat-sink):

1. standard FR4 board, with a D^2PAK device cooled from the bottom;
2. insulated-metal-substrate (IMS) board, especially developed for high-power assemblies, with a D^2PAK device;
3. "hybrid" solution using the modified version of the TO-220 device, in which the device is back-mounted in order to have the aluminum flange up with top-side cooling; electrical insulation between the heat-sink and the aluminum flange is provided by a thin foil of silicone;

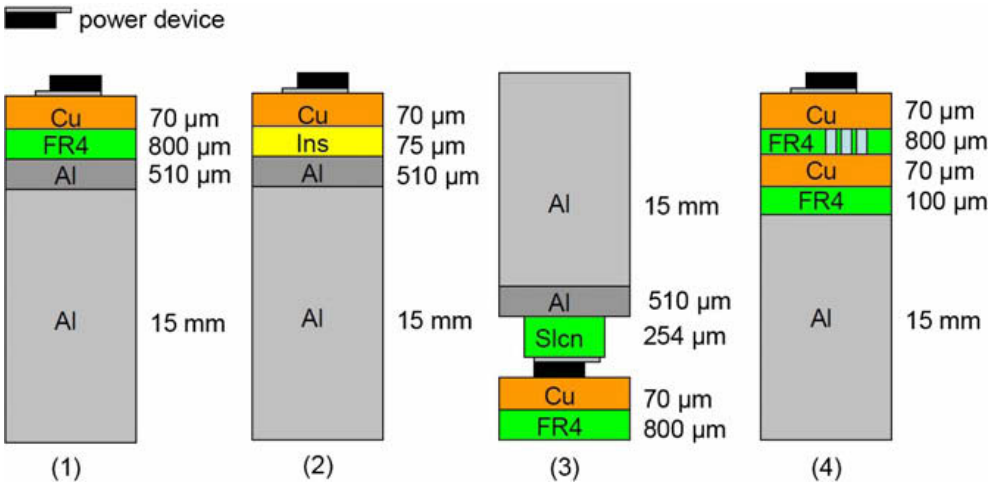


Figure 1.6 – Vertical cross-section of the four board topologies here studied: from the left: (1) standard FR4, (2) IMS board, (3) hybrid configuration with silicone layer insulation, (4) standard FR4 with thermal vias.

4. double-copper layer FR4 board, with thermal vias underneath the device.

All the four configurations are shown as schematic cross sections in Figure 1.6. All feature the presence of a single device and the same dissipating area of $26 \times 40\,\mathrm{mm}^2$. The silicone layer (in the hybrid solution), instead, covers only the drain flange area. The boundary conditions for heat dissipation are the following:

- ♢ the device exchanges heat in two ways: heat exchange from the surfaces in contact with free air is described by a convective boundary condition, in which the reference temperature is fixed at $27\,^{\circ}\mathrm{C}$ and a coefficient of thermal exchange of $3\,\mathrm{W}/(\mathrm{m}^2 \cdot \mathrm{K})$ has been used for both horizontal and vertical surfaces, since air flow is strongly hampered by the reduced free space;

- ♢ the surfaces in contact with the board exchange heat by conduction, i.e, the boundary condition is a continuity condition;

- ♢ the bottom surface of the heat-sink is fixed at $27\,^{\circ}\mathrm{C}$ (a fixed temperature boundary condition is imposed – note that this surface is the top surface for the hybrid structure);

- ♢ the vertical surfaces of the board are adiabatic, since it is assumed that the simulated section is surrounded by identical elements.

Material	Thermal conductivity $\lambda[\mathrm{Wm^{-1}K^{-1}}]$
Silicon die	149
FR4	0.3
Insulating film (IMS)	2.2
Aluminum contacts	237
Aluminum heat-sink and baseplate	150
Package enclosure resin	0.8
Silicone layer	2

Table 1.1 – Values of thermal conductivities λ used in the simulation.

In these structures, the thinnest layer is 70 µm thick (copper tracks), while the thickest layer is the aluminum heat-sink (15 mm). This calls for careful meshing, since the number of meshing elements (and thus of degrees of freedom) can soar when a thin layer is in contact with a thick layer, due to the continuity of the mesh between subdomains and the necessity of having quite a high number of elements in thin layers (which are the ones with highest thermal gradients). To overcome this problem, each domain is meshed separately to keep the level of complexity of the simulation at a reasonable value.

For static thermal behavior, only the thermal conductivity of the materials of the various domains is needed. The values of thermal conductivities λ are shown in Table 1.1.

If one wants to be confident in the simulation results, a model has to be tuned on measurements. To do so, a test board has been built and a model of this test board has been developed. In particular, parameters like the thermal resistance between the silicon die and the drain flange, and the coefficient of thermal exchange from the top of the device must be obtained by fitting measured results for building a reliable model. The test board is shown in Figure 1.7. For determining the coefficient of thermal exchange it is necessary to measure the external temperature at different points over the device, while to calculate the thermal resistance between the silicon die and the drain flange we must use an indirect temperature measurement technique, in which the gate-source voltage V_{GS} is monitored at different ambient temperatures in the absence of self-heating, in order to obtain a calibration curve between V_{GS} and the temperature T; the curve $T = f(V_{\mathrm{GS}})$ is then used to get the internal temperature of the device (the die temperature) once the V_{GS} is measured under self-heating conditions. While the external temperature measurement is straightforward (it is sufficient to position the thermocouples at the desired locations to measure the temperature, helping the thermal contact by using thermal paste), the procedure followed for the measurement of the die temperature needs to be explained.

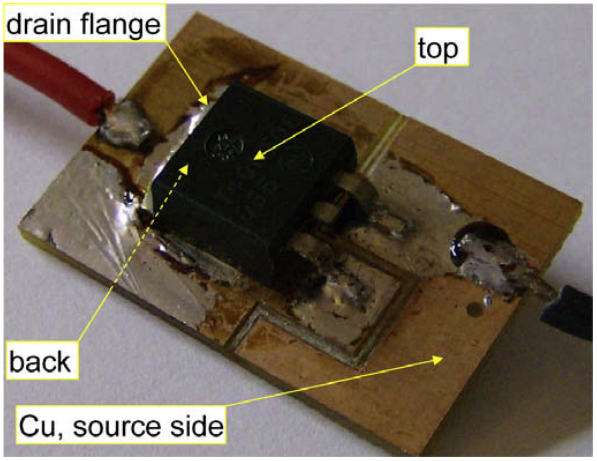


Figure 1.7 – Test board developed for FE thermal model tuning, with indication of the temperature measurement points.

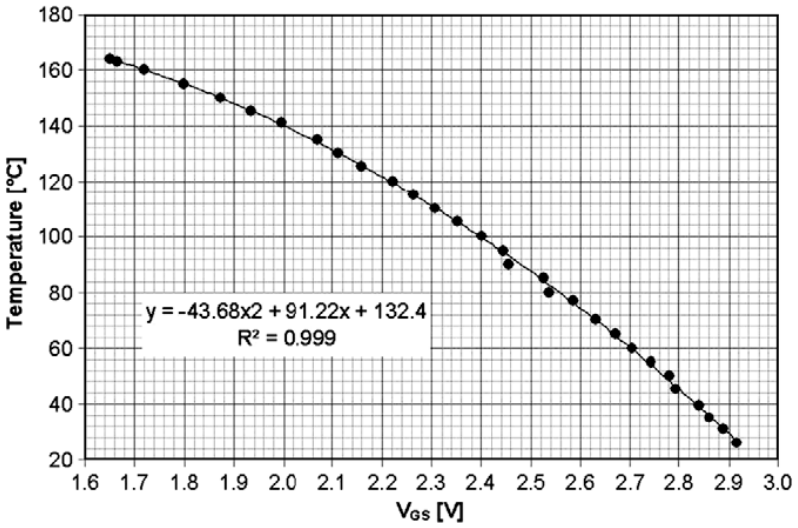


Figure 1.8 – Calibration curve $T = f(V_{\mathrm{GS}})$ obtained for the device under test.

1.2.5 Determination of the die temperature and thermal model tuning

First of all, a thermo-sensitive electrical parameter (TSP, *thermo-sensitive parameter*) has to be chosen: in this case, the MOSFET is connected in a diode configuration (with gate and drain short-circuited) and $V_{\mathrm{GS}} = V_{\mathrm{DS}}$ is monitored. Then, the device is

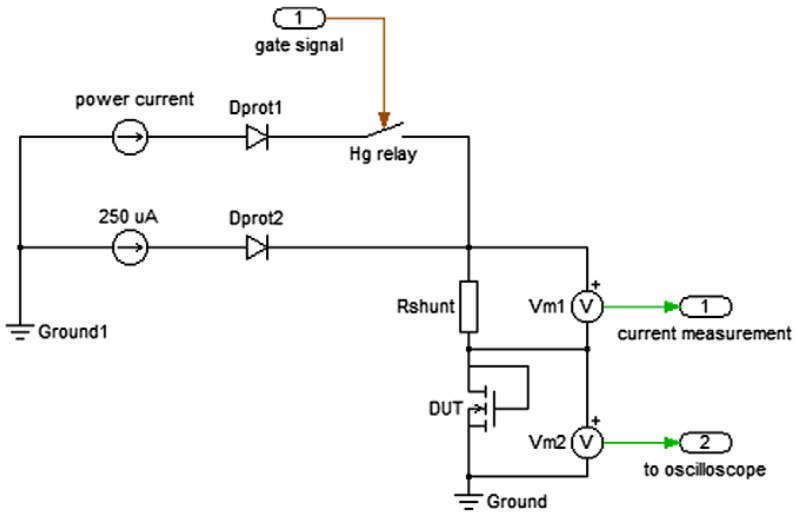


Figure 1.9 – Electrical setup for measuring the junction temperature of the device under test.

biased with a current small enough to avoid self heating (for our MOSFET, 250 µA). This is necessary because the calibration curve is obtained by putting the device in a climatic chamber, stepping up the temperature at values $T_1, T_2, \ldots$ while maintaining the device on (with a constant $I_D = 250\,\mu\text{A}$). At each temperature step, the whole device has to reach the fixed temperature, so that the measured voltage $V_{\text{GS}}(T_1), V_{\text{GS}}(T_2), \ldots$ are actually depending on the die temperature. A calibration curve is shown in Figure 1.8. Then, measurement of the die temperature is achieved by biasing the device at a *power current* level (where the device self-heats). Then, the *power current* is abruptly switched off, so that only the measurement current of 250 µA is flowing. Right after switching, V_{GS} is measured. The idea underneath this technique is that right after switching off the power current, V_{GS} is still given by the die temperature before switching, thanks to the finite thermal capacitance of the die. The faster the acquisition, the more precise the temperature measurement (the cooling of the device is negligible). Note that this technique can yield the average temperature of the die, while hot spots (if present) cannot be detected. In our set-up, the acquisition is done roughly 20 µs after switching off the power current; the thermal constant of the die is roughly 300 µs (estimated by simulation), so the error due to the acquisition delay is negligible. The measurement setup is shown in Figure 1.9.

The measurement of the different temperatures over the test board has been repeated for different dissipated power levels; at each power, different measurements have been taken to improve the precision, so that the uncertainty on the average

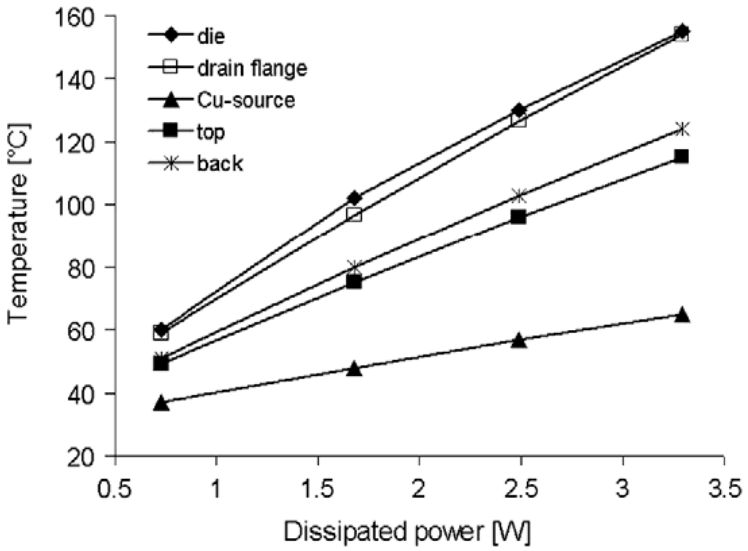


Figure 1.10 – Measured temperatures versus MOSFET dissipated power. The ambient temperature is 26 °C.

Parameter	Value
h_{up}	$35\,\mathrm{Wm^{-2}K^{-1}}$
h_{side}	$15\,\mathrm{Wm^{-2}K^{-1}}$
h_{down}	$27\,\mathrm{Wm^{-2}K^{-1}}$
h_{case}	$74\,\mathrm{Wm^{-2}K^{-1}}$
k_{resin}	$0.56\,\mathrm{Wm^{-1}K^{-1}}$
$k_{\mathrm{drain-Cu}}$	$35\,\mathrm{Wm^{-1}K^{-1}}$

Table 1.2 – Fitting parameters used for tuning the test-board FEM model.

value is within 1 °C. The graph showing the temperatures as a function of dissipated power is shown in Figure 1.10.

Starting from the temperature measurements obtained by thermocouples (external measurements) and the temperature of the die, the FEM 3D model of the test board has been tuned. The model, which is shown in Figure 1.11, features six fitting parameters, shown in Table 1.2. The fitting parameters are the thermal conductivities of the case plastic lid and of the thin layer between the drain flange and the copper area on which the device is mounted, and the coefficients of thermal exchange between different areas of the test board and the surrounding air. The comparison between the model and the measurements gives a quantitative evaluation of the fitting process.

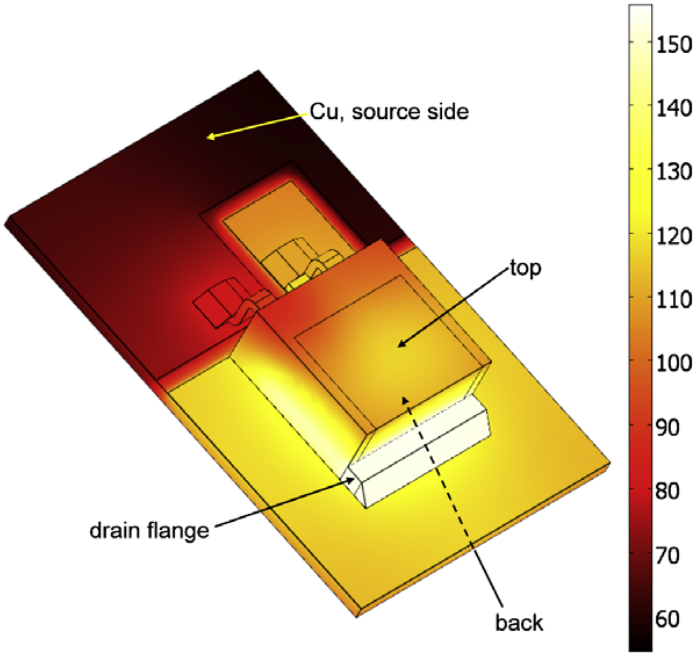


Figure 1.11 – Simulated thermal distribution of the test board (temperatures in °C), obtained by 3D FEM model. The dissipated power is 3.3 W, and the measurement points are pointed out by the arrows.

The results are shown in Table 1.3. Having this shown how the model was tuned, the aim of the next section is then to show a comparison between different assembly solutions, to obtain quantitative information about the best board mounting technology for high-power assemblies with stringent thermal constraints.

1.2.6 Comparison between different solutions

The next step is to build different FEM models to compare different board technologies. In this case, the approach is different, because some parameters cannot be known *a priori* – like the coefficients of thermal exchange. This is not a real issue when comparing different solutions, since all the parameters that are not related to the board technology (convective coefficients, for instance) are kept constant in the different simulations, so that the only difference between the models is actually given by the board technology and not from other sources of cooling. The temperature map so obtained is not tuned on a physical structure, but choosing sensible parameters and dissipated powers it is possible to have some useful indications about the best

Location	T_{meas} [°C]	T_{sim} [°C]
Silicon die	155	156
Cu, drain side	154	153
Cu, source side	65	63
D.U.T. top surface	115	117
FR4 back surface	124	125

Table 1.3 – Comparison between measured and simulated temperatures at various points of the test board. The ambient temperature is 26 °C and the DUT dissipates 3.3 W.

board technology that has to be used for a given circuit. The simulation results for the structures of Figure 1.6 are shown here.

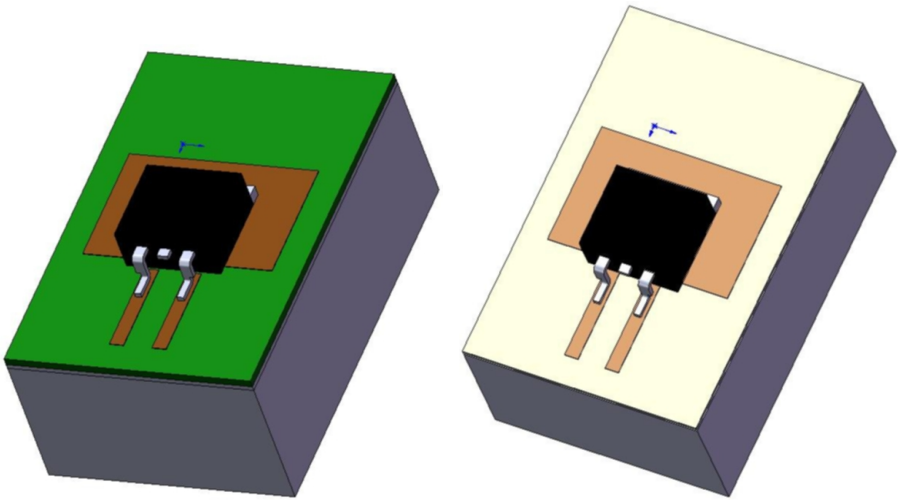


Figure 1.12 – The standard FR4 technology solution (left) and the IMS solution (right).

Structure A: standard FR4 The first structure is shown in Figure 1.12 (left). It is a single copper layer FR4 board, the most common technology for electronic boards, with a D^2PAK device mounted on it. The FR4 board thickness is 0.8 mm (chosen instead of the more common 1.6 mm to improve the thermal exchange); the board is mounted on 510 µm-thick aluminum baseplate. The assembly is placed on a 15 mm-thick aluminum heat-sink. When dissipating 20 W, the die reaches the temperature of 237 °C, as can be seen in Figure 1.13. The specific interfacial thermal conductance h_{if} between package and board is a figure of merit defined as the ratio between the thermal conductivity of the board material and the thickness of the board itself: the higher h_{if}, the better the thermal exchange between device and heat-sink[4]:

$$h_{\text{if}} = \frac{k}{t} = \frac{0.3}{800 \times 10^{-6}} = 375 \, \frac{\text{W}}{\text{m}^2\text{K}}. \tag{1.19}$$

The die reaches very high temperatures, suggesting that this solution is well suited only for low power.

Structure B: Insulated Metal Substrate (IMS) The structure shown in Figure 1.12 (right) is the so called IMS, in which the device sits on an electrically insulating, but thermally conductive thin-film (the thickness is 75 µm). Simulation results for the case

[4]Note that a substrate with low thermal conductivity, but small thickness, can achieve a high interfacial thermal conductivity.

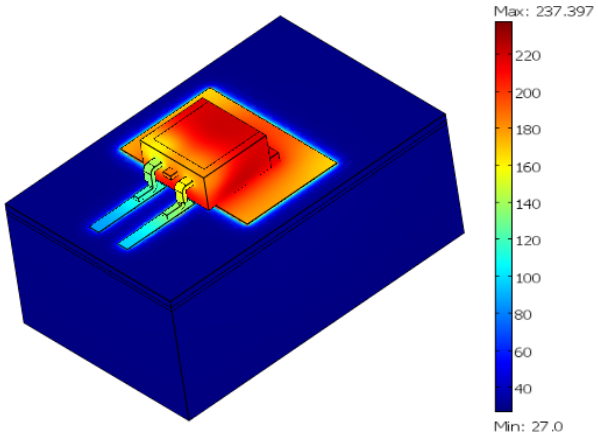


Figure 1.13 – Simulation results of the FR4 structure. The die dissipates 20 W. Temperatures in °C. The simulation features 206000 degrees of freedom. The die temperature is 240 °C.

of 20 W dissipation are shown in Figure 1.14. Note that, in this case, the interfacial thermal conductance is much higher than for structure A:

$$h_{\text{if}} = \frac{k}{t} = \frac{2.2}{75 \times 10^{-6}} = 29300 \, \frac{\text{W}}{\text{m}^2\text{K}}, \tag{1.20}$$

which means a far better thermal exchange between device and heat-sink, thanks to the very thin film between the two parts. Obviously, heating is strongly reduced with respect to case A. Note that, even if the thermal conductivity of the layer is not that high, its very small thickness makes this solution very efficient from the thermal point of view. A drawback of this solution, from an electrical point of view, is that it allows only one-layer PCBs, since copper can be deposited only on one side.

Structure C: Hybrid solution Figure 1.15 shows the hybrid solution. Here, a TO-220 device is SMD mounted, by bending the pins. Note that the heat-sink is on the top, so that the back of the device is mounted upwards. The thermal contact between the drain flange and the heat-sink is made by a thin silicone film of thermal conductivity $k = 0.3\,\text{Wm}^{-1}\text{K}^{-1}$ and 254 µm thickness. The interfacial conductance is:

$$h_{\text{if}} = \frac{k}{t} = \frac{3}{254 \times 10^{-6}} = 11800 \, \frac{\text{W}}{\text{m}^2\text{K}}, \tag{1.21}$$

an intermediate value between solutions A and B. Simulation results confirm this observation, since the temperature reached by the die is 68 °C, when dissipating 20 W, see Figure 1.16.

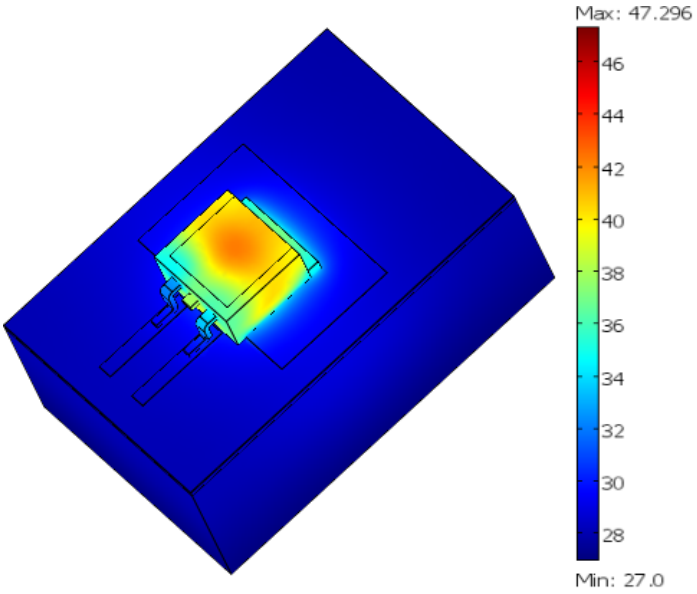


Figure 1.14 – Simulation of the IMS structure. The die dissipates 20 W. Temperatures in °C. The simulation features 267000 degrees of freedom. The die temperature is 47 °C.

Structure D: FR4 board with thermal vias In this structure, shown in Figure 1.17, thermal vias are placed under the drain flange, exploiting the higher conductivity of copper with respect to FR4. This solution can be well suited in a Direct-Bond-Copper assembly (DBC), in which one layer of copper is used for electrical routing and the other is used to improve the thermal exchange between the board and the heat-sink. Note that, in our case, a thin layer of FR4 (100 µm) is used between the bottom copper layer and the heat-sink, to avoid electrical short-circuits. Usually vias are not filled by copper, while in our model they are (for meshing simplicity reasons): this could lead to overestimating the cooling efficiency. To avoid this, the thermal conductivity of the copper filling the vias has been reduced in order to account for the effective area through which heat is flowing:

$$A_{\text{cond}} = \frac{\pi(\phi_{\text{ext}}^2 - \phi_{\text{int}}^2)}{4} \tag{1.22}$$

in which A_{cond} is the effective area conducting the heat flow, ϕ_{ext} is the external diameter of the via, ϕ_{int} is the internal diameter. Using $\phi_{\text{ext}} = 0.25\,\text{mm}$, with a copper thickness of 70 µm, the internal diameter is $\phi_{\text{int}} = 0.25 - 2 \times 0.07 = 0.11\,\text{mm}$,

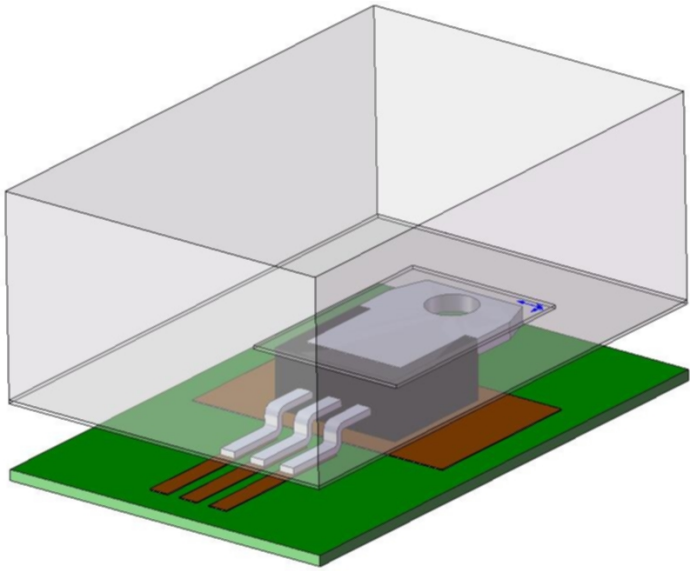


Figure 1.15 – The hybrid solution.

obtaining

$$A_{\text{cond}} = \frac{\pi(\phi_{\text{ext}}^2 - \phi_{\text{int}}^2)}{4} = 0.0395\,\text{mm}^2 = 0.806 \times A_{\text{ext}}. \tag{1.23}$$

It is therefore enough to reduce the conductivity of the vias copper by a factor $\alpha = 0.806$ to account for the fact that the vias are not filled. Assuming that the heat flux is uni-directional (which is a reasonable assumption, given the geometry of the via), the thermal resistance of the via is given by:

$$R_{\text{via}} = \frac{t}{k_{\text{Cu}} \cdot A_{\text{cond}}} = \frac{t}{\alpha k_{\text{Cu}} \cdot \dfrac{A_{\text{cond}}}{\alpha}} = \frac{t}{\alpha k A_{\text{ext}}} = \frac{t}{k_{\text{red}} A_{\text{ext}}} \tag{1.24}$$

in which k_{red} is the reduced thermal conductivity of the copper, whose original value is k_{Cu}. An example of simulation results, with the device dissipating 20 W, is shown in Figure 1.18. The effectiveness of the solution is proven by the temperature reached by the silicon die, which is around 100 °C, far less then the first solution with standard FR4 board. In the end, it is possible to see that (not surprisingly) the IMS solution is the best suited for improving the cooling of electronic boards, since it is the solution that features both high thermal conductivities and low thicknesses. On the other hand, the standard FR4 solution is the worst, since it features high thicknesses and low thermal conductivity materials. In between fall the hybrid solution and the thermal

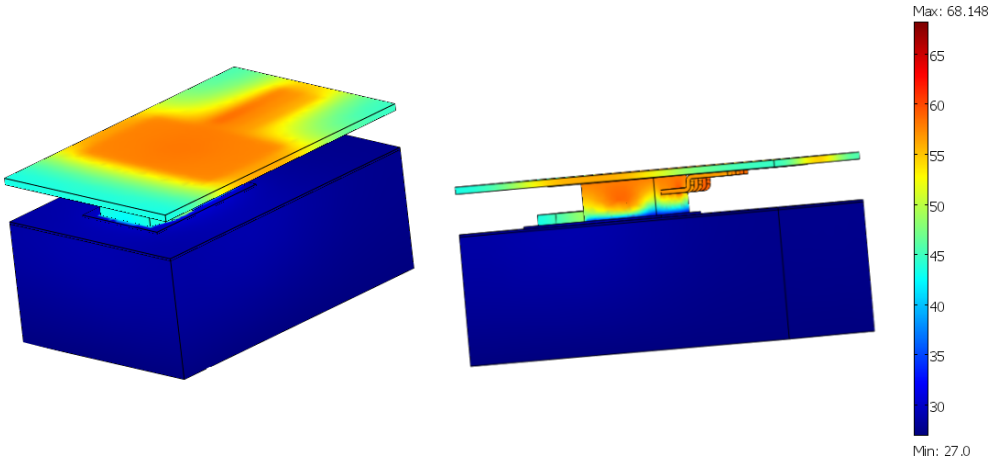


Figure 1.16 – Simulation of the hybrid solution. The die dissipates 20 W. Temperatures in °C. The simulation features 218000 degrees of freedom. The die temperature is 68 °C.

vias solution. The hybrid solution has the advantage that the heating part is not in contact with the electronic board (the plastic case is in contact with the copper, so the electrical insulation is guaranteed), which gives some degrees of flexibility in copper track routing (all other solutions are mainly suitable for electrical single-layer routing).

This comparison is an example of how the FEM can be successfully used to evaluate different technologies, even if some simulation parameters cannot be easily known *a priori*.

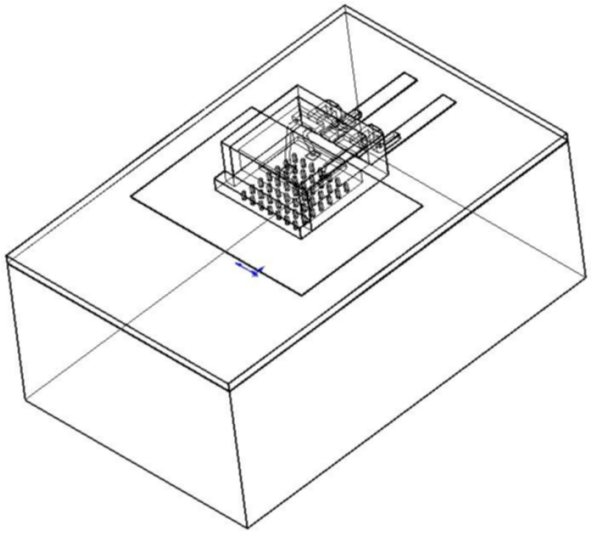


Figure 1.17 – Thermal vias structure.

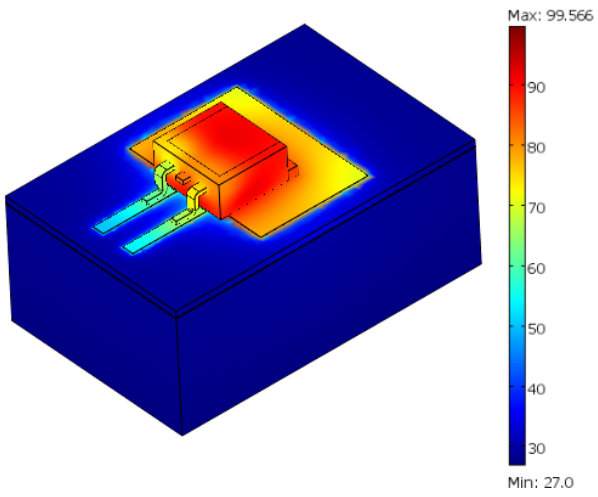


Figure 1.18 – Simulation of the structure with thermal vias. The die is dissipating 20 W. Temperatures in °C. The structure has 430000 degrees of freedom. The die temperature is 100 °C.

1.2.7 Planar transformer and surface modeling

An important component in the SMPS is the transformer [5], necessary for transferring power from the primary to the secondary side (at low voltage and high current) of the power supply. Since the free space in the assembly is limited, a very high-power density solution has to be designed to meet the project specifications. The study of this component, which will be fully described below, exploits a particular feature of the FEM simulator, that is the possibility of modeling very thin three-dimensional bodies as they were bi-dimensional, because the temperature gradient in the thickness direction is negligible. This is a very useful feature, since it allows to describe these domains, which are often responsible of the increasing of the degrees of freedom of the overall model, with a low number of elements. Indeed, this happens when the dimensions of adjacent domains are very different, a situation easily found in electronics: the simple FR4 board is a case in point, since the board can be as thick as 1.6 mm and the copper tracks are usually 0.035 mm thick. If the copper tracks were modeled as three-dimensional bodies, in order to achieve continuity of the mesh between the two adjacent domains, the increase of the model degrees of freedom would be unmanageable.

On the other hand, the possibility of considering the copper tracks as bi-dimensional elements (in fact, the gradient through the thickness of the tracks is negligible) actually helps to keep the model manageable and, if needed, to increase the number of degrees of freedom in other regions. This is particularly useful in a structure like the one shown below, since a very high number of domains are thin layers that will be modeled as surfaces, instead of three-dimensional bodies. The feature offered by the simulator is called *highly conductive thin layer* [6], and it is sufficient to specify the thermal conductivity of the material (k, [$\mathrm{Wm^{-1}K^{-1}}$]) and the thickness. Obviously, the thinner and more thermal conductive is the domain, the more accurate is the approximation.

The transformer here studied is interesting from a thermal point of view, because the windings are made with thin PCBs, instead of using copper windings as usual for the primary and the secondary side. This allows to build a very compact structure (to meet the free space constraints), and at the same time to achieve high power density. This transformer has to deliver 1.5 kW at a frequency of 100 kHz. We have designed and built a prototype for lower power, in which the number of windings is far smaller than the original one (mainly due to technology limits), but with the same topology. A photograph of the transformer we built is shown in Figure 1.19.

The transformer is made of a special *kool-mu* material, especially suited for high magnetic field environments; there are two E-shaped cores facing an I-shaped core each, obtained by milling the E-shaped cores. Since the magnetic material is very brittle, the milling process results in a rough surface, so that the air-gap between the E- and the I-shaped cores is distributed, and not easy to estimate. We used eight

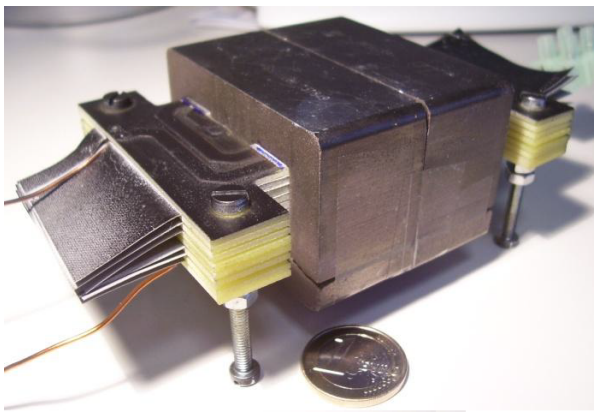


Figure 1.19 – Photograph of the planar transformer built at the University of Parma.

double-layer PCBs to make the windings: six boards are used for the primary side, each PCB featuring two windings per face (four windings per PCB), all of them connected in series. The secondary side of the transformer is made of two PCBs, with a single winding per face (for a total of four windings over two PCBs), and they are connected in parallel. The voltage ratio is 24:1. This transformer is operated in a square-wave mode by means of an IGBT H-bridge with 120 V DC bus, 20 kHz switching frequency, and the secondary connected on a 0.72 Ω resistive load. The bridge is controlled by a PIC16LF84A that generates the switching signals (comprehensive of blanking times) that drive a couple of half-bridge IR2113 drivers. The equation below is the power balance for the transformer:

$$P_{\text{in}} - P_{\text{out}} = P_{\text{Cu1}} + P_{\text{Cu2}} + P_{\text{core}} \tag{1.25}$$

in which P_{in} is the power entering the primary side, P_{out} is the power exiting the secondary side (and dissipated by the resistive load), P_{Cu1} is the power dissipated by the primary windings, P_{Cu2} is the power dissipated by the secondary windings, and P_{Core} is the power dissipated by the magnetic core. The effective section of copper in which the current flows coincides with the entire section of the copper tracks, since the skin effect is negligible:

$$\delta = \frac{1}{\sqrt{\pi \mu_0 \sigma f}} \tag{1.26}$$

where δ is the skin effect depth (determining the effective depth in which the current flows over the copper surface), μ_0 [Hm^{-1}] is the magnetic permeability of the copper (which is equal to that of air), σ [Sm^{-1}] is the copper electrical conductivity and f [Hz] is the switching frequency. In our operating conditions ($\mu_0 = 1.256 \times 10^{-6}\,\text{Hm}^{-1}$,

$\sigma = 6 \times 10^7\,\mathrm{Sm}^{-1}$, $f = 20\,\mathrm{kHz}$), the skin depth is $\delta = 470\,\mu\mathrm{m}$. This means that the current distribution on the conductor cross-section decreases proportionally to this law:

$$\mathrm{e}^{-t/\delta} \tag{1.27}$$

where t is the thickness of the conductor. For a $35\,\mu\mathrm{m}$ track, the skin effect is then negligible, so that the DC measured resistance can be used to evaluate the losses in the AC regime, by means of the following relationships:

$$P_{\mathrm{Cu1}} = R_{\mathrm{DC1}} \cdot I_{1\mathrm{rms}}^2 \tag{1.28}$$

$$P_{\mathrm{Cu2}} = R_{\mathrm{DC2}} \cdot I_{2\mathrm{rms}}^2 \tag{1.29}$$

and, knowing P_{in} and P_{out} by measurements, the core losses can be easily extracted:

$$P_{\mathrm{core}} = P_{\mathrm{in}} - P_{\mathrm{out}} - P_{\mathrm{Cu1}} - P_{\mathrm{Cu2}}. \tag{1.30}$$

Measurement setup

The next step is the setup of the measurement bench. The thermal measurements were performed by means of a FLIR A325 infrared (IR) thermocamera. To achieve a uniform emission on the transformer surface, we painted it black, so that the emission coefficient can be considered (with good approximation) close to 1. Calibration by a thermocouple thermometer was first performed. Then, the transformer is operated under standard conditions and the surface temperature map is captured. The ambient temperature was $28\,^{\circ}\mathrm{C}$ with natural convection cooling. A thermal map measured on the transformer top surface is shown in Figure 1.20. The electrical conditions are: $V_{1\mathrm{rms}} = 120\,\mathrm{V}$, $I_{1\mathrm{rms}} = 2.1\,\mathrm{A}$, $P_{\mathrm{in}} = 43\,\mathrm{W}$, $P_{\mathrm{out}} = 25.7\,\mathrm{W}$, $P_{\mathrm{Cu1}} = 7.3\,\mathrm{W}$, $P_{\mathrm{Cu2}} = 3.9\,\mathrm{W}$. The core losses were estimated to be $P_{\mathrm{core}} = 6.1\,\mathrm{W}$.

Once the FEM model of the transformer has been tuned on the basis of these measurement results, it can be used to check the effectiveness of different cooling solutions. As far as our transformer structure is concerned, the only way for the heat to be dissipated is by natural convection from the top surface of the core. Since the free space between the columns of the core is filled by the boards where the windings sit, no convection can take place there. Thus, the only fitting parameter we used to tune the simulation was the coefficient of natural convection h on the top surfaces of the transformer. We found a value of $h = 14.5\,\mathrm{Wm}^{-2}\mathrm{K}^{-1}$. Other values used in our model to define the properties of the domains are the thermal conductivities of FR4 ($k = 0.3\,\mathrm{Wm}^{-1}\mathrm{K}^{-1}$), of copper ($k = 400\,\mathrm{Wm}^{-1}\mathrm{K}^{-1}$), of the core's ferrite ($k = 80\,\mathrm{Wm}^{-1}\mathrm{K}^{-1}$) and of silicone foils ($k = 0.78\,\mathrm{Wm}^{-1}\mathrm{K}^{-1}$). These were taken from the FEM simulator material library, and were not used as fitting parameters.

In particular, we focused on careful meshing for the transformer windings, trying to keep the number of degrees of freedom down to a manageable level. This was

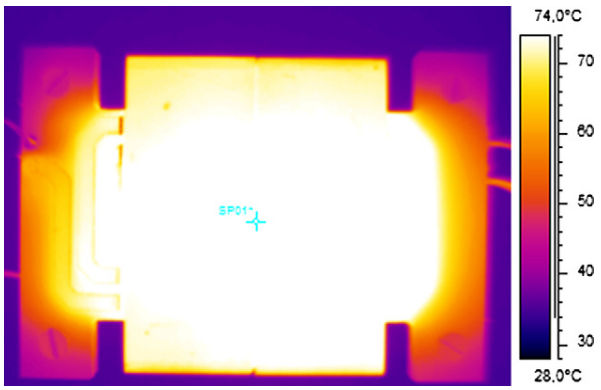


Figure 1.20 – Thermal map of the planar transformer operated at $V_{1\mathrm{rms}} = 120\,\mathrm{V}$, $I_{1\mathrm{rms}} = 2.1\,\mathrm{A}$. The ambient temperature is $28\,^{\circ}\mathrm{C}$ and the maximum temperature is $74\,^{\circ}\mathrm{C}$.

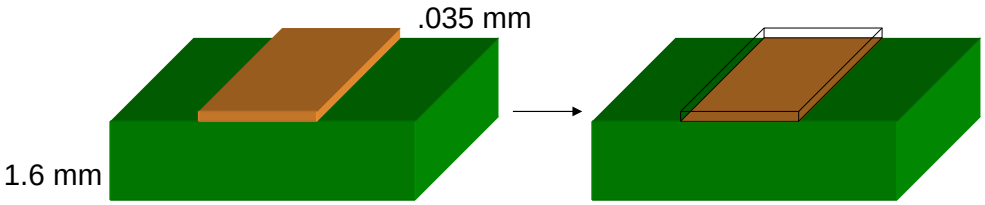


Figure 1.21 – Replacing a three-dimensional domain – a copper track in this particular case – with an equivalent two-dimensional layer, which is basically the footprint of the track on the underlying domain.

achieved by exploiting the FEM simulator feature called *high thermally conductive thin layer*, which allows to describe a thin and thermally conductive layer like a surface, since the gradient in the thickness direction can be neglected. In our structure, this feature has been used extensively, because we had to model both the primary and the secondary windings. This feature "shrinks" the three-dimensional domain into a surface, thus eliminating the thickness dimension (which has to be specified to the simulator, together with the thermal conductivity of the material). A graphical explanation is given on Figure 1.21. The thinner the layer, the better the approximation. Mathematically, the equation solved on the boundary (note that the domain is no longer present) is the following [6]:

$$d_S \rho_S C_{p,S} \frac{\partial T}{\partial t} + \nabla_t(-d_S \underline{k}_S \nabla_t T) = q_{\partial\Omega} - q_\Omega = -q_S \tag{1.31}$$

where:

- ♢ ρ is is the layer density in $\mathrm{kg/m^3}$;
- ♢ $C_{p,S}$ is the layer heat capacity in $\mathrm{J/(kg\,K)}$;
- ♢ k_S is the layer thermal conductivity tensor in $\mathrm{W/(m\,K)}$;
- ♢ d_S is the layer thickness in m;
- ♢ $q_{\partial\Omega}$ is the heat flux from the surroundings into the layer in $\mathrm{W/m^2}$;
- ♢ q_Ω is the heat flux from the layer into the subdomain in $\mathrm{W/m^2}$;
- ♢ q_S is the net outflux of heat through the top and bottom faces of the layer in $\mathrm{W/m^2}$.

The operator ∇_t is the operator ∇ projected on the surface of the thin layer.

1.2.8 Simulation results

Figure 1.22 shows the simulation results of the structure operated under the same conditions as used to obtain the thermal map shown in Figure 1.20. Due to the presence of two symmetry planes, only one fourth of the structure needs to be simulated. The match between measurements and simulation is very good, the difference between the maximum temperatures on the core surface being 1 °C. The maximum temperature is reached on the inner windings and is 85 °C.

The power losses in the core are frequency-dependent, and core manufacturers specify the dependence of core losses on the switching frequency as a polynomial relationship. We increased the switching frequency up to 30 kHz, a value at which the inner temperature reaches 100 °C. This is assumed to be a safe operating value, because the insulation resin between adjacent windings starts to degrade significantly only above this value, whereby the reliability of the transformer is compromised, because of possible short-circuits between windings. We estimated a value $P_{\mathrm{Core}} = 11.3\,\mathrm{W}$ under these operating conditions (being the other dissipated power levels the same as in the previous simulation), and the corresponding simulation results are shown in Figure 1.23. At this frequency, the skin effect is still negligible, so that the copper losses are unchanged.

The influence of output power has been evaluated next, keeping the frequency at its original value of 20 kHz: this allows to separate the heating contributions due to the frequency and to the output power, respectively. Also in this case, the output power was chosen in such a way as to keep the maximum inner temperature of the windings below the safe value of 100 °C. This corresponds to an output current of 8 A: under this load condition, the power dissipated by the secondary windings is 7 W

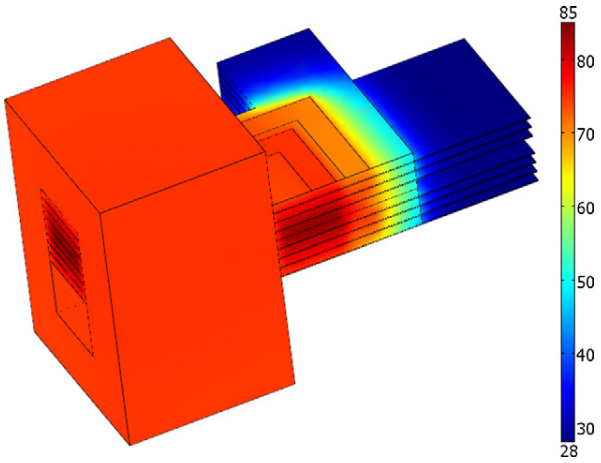


Figure 1.22 – Simulated transformer thermal map (temperature in °C); $P_{\mathrm{Cu1}} = 7.3\,\mathrm{W}$, $P_{\mathrm{Cu2}} = 3.9\,\mathrm{W}$, $P_{\mathrm{Core}} = 6.1\,\mathrm{W}$. The ambient temperature is 28 °C. The peak surface temperature is 75 °C.

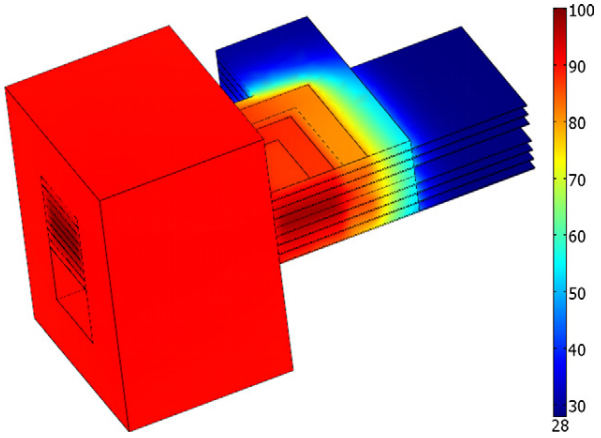


Figure 1.23 – Simulated transformer thermal map (temperature in °C); $P_{\mathrm{Cu1}} = 7.3\,\mathrm{W}$, $P_{\mathrm{Cu2}} = 3.9\,\mathrm{W}$, $P_{\mathrm{Core}} = 11.3\,\mathrm{W}$. The ambient temperature is 28 °C. The peak temperature is 100 °C and it is reached in the secondary windings.

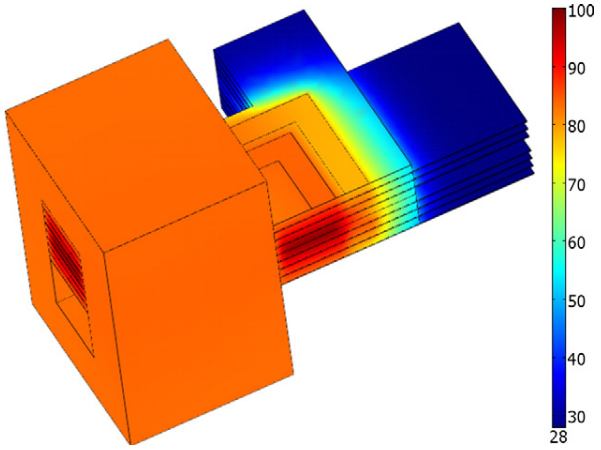


Figure 1.24 – Simulated transformer thermal map with output current $I_{\text{out}} = 8\,\text{A}$(temperature in °C); $P_{\text{Cu1}} = 7.3\,\text{W}$, $P_{\text{Cu2}} = 7\,\text{W}$, $P_{\text{Core}} = 11.3\,\text{W}$. The ambient temperature is 28 °C. The peak temperature is 100 °C and it is reached in the secondary windings.

(the increase of primary current and, consequently, of losses in the primary copper windings, was neglected since the magnetizing current is still the major contribution to the primary current). Simulation results are shown in Figure 1.24. Note that the hottest hot spot is in the secondary windings.

The simulation results shown so far are representative of different operating conditions, with the physical structure of the transformer being fixed. The FEM model can also be fruitfully exploited to investigate the effectiveness of different cooling solutions, obtained by changing the layout of the structure: three examples of different mounting solutions will be shown. In the first one, the magnetic core is connected to the heatsink by means of a thermally conductive, electrically insulating layer made of silicone (this is necessary to reduce the coupling of the magnetic field between the core of the transformer and the aluminum heat-sink), see Figure 1.25. The silicone layer is 10 mm thick and its thermal conductivity is $3\,\text{W}/(\text{m}\cdot\text{K})$ [7]. The second example shows the influence of electrically insulated thermal layers, i.e., copper layers inserted between the electrical layers with the only aim of removing heat from the inner part of the structure (which has been shown to be the hottest part of the transformer), see Figure 1.26. These layers are interleaved between primary and secondary windings, and they are thermally connected with the heat-sink by means of screws. The third solution shows the combined effect of the former two solutions, see Figure 1.27.

In the first solution, the copper losses were kept at the original level ($P_{\text{Cu1}} = 7.3\,\text{W}$, $P_{\text{Cu2}} = 3.9\,\text{W}$), while the switching frequency was increased up to 100 kHz

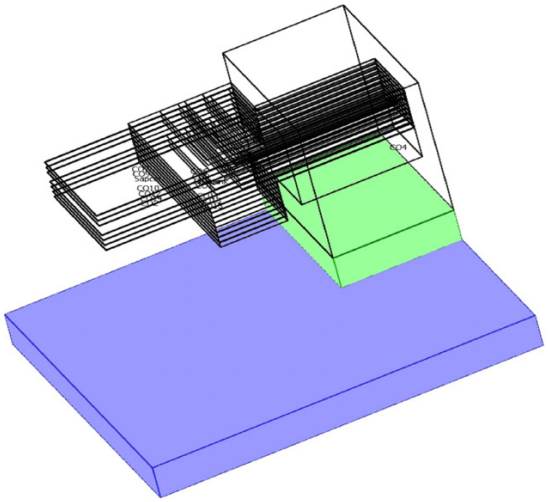


Figure 1.25 – 3D geometry of 1/4 of the planar transformer with aluminum baseplate (blue) and silicone gap-filler (green).

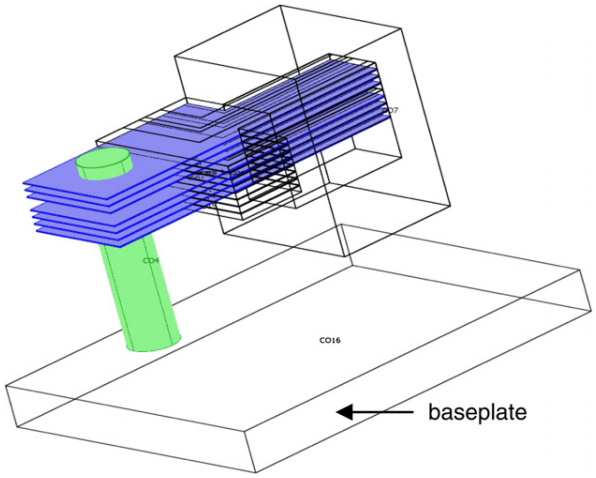


Figure 1.26 – 3D geometry of 1/4 of the planar transformer with aluminum baseplate, thermal bridge (green) and copper thermal layers (blue).

in order to increase the core losses, while at the same time remaining below the 100 °C limit in the copper windings. The skin effect at 100 kHz is still negligible (skin depth $\delta = 210\,\mu m$), meaning that copper losses remain constant at different frequencies. The only effect of the frequency change is the increase of core losses,

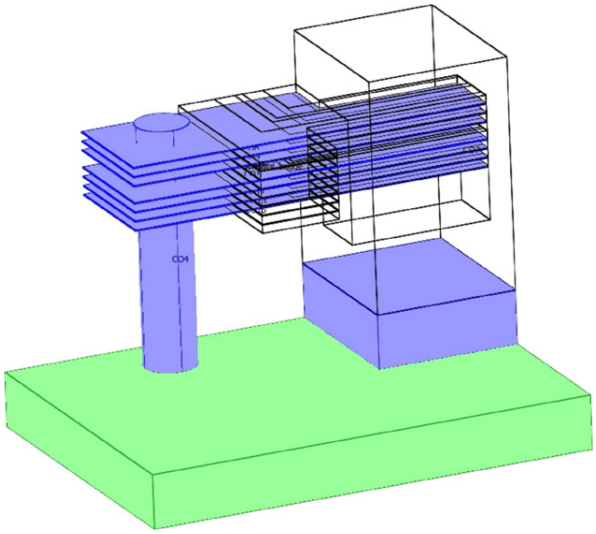


Figure 1.27 – 3D geometry of 1/4 of the planar transformer with aluminum base-plate, copper thermal layers, and silicone gap-filler.

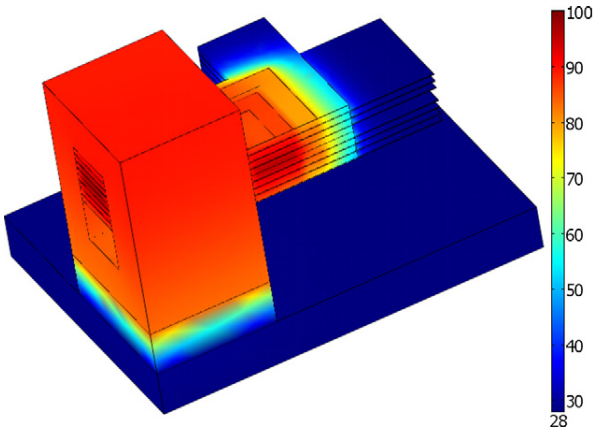


Figure 1.28 – Simulation results of the structure of Figure 1.25. $P_{\mathrm{Cu1}} = 7.3\,\mathrm{W}$, $P_{\mathrm{Cu2}} = 3.9\,\mathrm{W}$, $P_{\mathrm{Core}} = 65.3\,\mathrm{W}$. Temperature in °C. Ambient temperature is 28 °C.

estimated to be 65.3 W at 100 kHz. Simulation results are shown in Figure 1.28. The cooling effect of thermal bridges between thermal layers and aluminum heat-sink is shown in Figure 1.29. The peak temperature (98 °C) is still located in the secondary

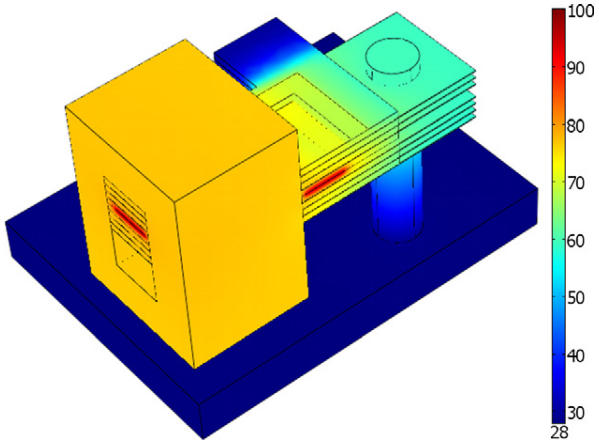


Figure 1.29 – Simulation results of the structure of Figure 1.26. $P_{\mathrm{Cu1}} = 7.3\,\mathrm{W}$, $P_{\mathrm{Cu2}} = 32.9\,\mathrm{W}$, $P_{\mathrm{Core}} = 6.1\,\mathrm{W}$. Temperature in °C. Ambient temperature is 28 °C.

windings. For the configuration shown in Figure 1.26, the power dissipated by the secondary windings has been increased to reach the maximum safe operating value of 100 °C in the windings. The secondary copper windings have been simulated to dissipate 32.9 W, which corresponds to an output current of 17.3 A. This solution is particularly effective in removing the heat generated by the windings, while it does not have a significant effect on the core temperature, since the thermal path between core and thermal layers is very resistive.

The cooling effect due to the combination of the two former solutions is shown in Figure 1.30. This solution can achieve reliable operation under both high output power and high frequency. Simulation results of the structure operated with $P_{\mathrm{Cu1}} = 7.3\,\mathrm{W}$, $P_{\mathrm{Cu2}} = 26.4\,\mathrm{W}$, $P_{\mathrm{Core}} = 65.3\,\mathrm{W}$ are shown in Figure 1.30. Under these operating conditions, the maximum temperature does not exceed 100 °C, which is therefore compatible with reliable operation.

1.3 Conclusions

In this section, the FEM method has been extensively applied to solve problems related to the thermo-fluid and thermal dynamics, in the power electronics field of applications. In particular, the method has been tested on different structures, including water-cooled heat-sinks, electron devices and planar transformers. Advantages and disadvantages of this method have been highlighted: if the FEM method allows to describe in a very accurate way the structures to be simulated, on the other hand it has

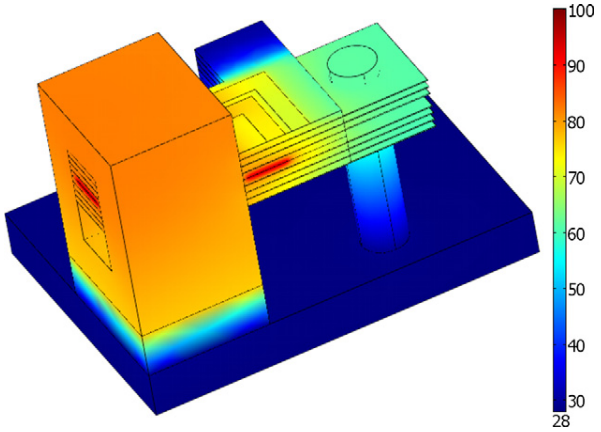


Figure 1.30 – Simulation results of the structure of Figure 1.27. $P_{\mathrm{Cu1}} = 7.3\,\mathrm{W}$, $P_{\mathrm{Cu2}} = 26.4\,\mathrm{W}$, $P_{\mathrm{Core}} = 65.3\,\mathrm{W}$. Temperature in °C. Ambient temperature is 28 °C.

shown serious difficulties in solving fluid-dynamics problems (i.e., the water-cooled heat-sink model): this problem has been bypassed by applying some basic relationships of theoretical fluid-dynamics, in order to reduce the fluid-dynamic problem to a static thermal conduction problem, which is far easier to solve. This approximation holds especially in the case of forced convection.

The FEM tools has been shown to be an effective tool in improving the thermal design in power electronics: in the early stage of the design, it can give design guidelines about which board technology is the best suited for removing the heat generated by the active components, and in this work four technologies have been studied: standard FR4, Insulated Metal Substrate, a hybrid solution, and a standard FR4 with thermal vias underneath the device. Quantitative considerations have been done about the effectiveness of heat removal of the different solutions, and the IMS technology has found to be the most effective one under this aspect.

The impact of different heat removal solutions on the thermal behavior of a planar transformer has been investigated as well. Besides the study of different solutions about *how* to remove the generated heat, an investigation of the influence of the *operating conditions* like switching frequency and output current was performed. Some design solutions to improve the heat removal from the transformer have been simulated: first, the effect of a silicone gap filler to improve the cooling of the core was investigated, as well as then the presence of thermal layers to improve the cooling of copper windings; in the end, the effectiveness of the arrangement featuring both the cooling solution has been evaluated.

Chapter 2

Electro-thermal modeling and characterization of electron devices

The modeling process of modern electron devices often lacks under one aspect: the accuracy of the thermal behavior of the component itself. The electrical behavior of electron devices is strongly dependent on the (channel) temperature, which depends both on the electrical characteristics of the device and on the surrounding environment (in other words, on *how* the generated heat can flow out of the component). If plenty of accurate *electrical* models are present in literature, on the counterpart these models are often coupled with very simplistic thermal models, which basically account for the thermal path between the source and the surrounding environment by means of a limited number of thermal RC couples (arranged in a Cauer network, for example) connected together. In this Chapter, an approach to the development of accurate, although compact, thermal models is presented, and their coupling with electrical device models is shown. The concept of *self-consistency* of a model resides in the mutual dependence of the channel temperature and the electrical behavior of the device, and its application will be shown in this context.

2.1 A review of some basic concepts

In this section, some basic aspects of semiconductor devices are recalled; these concepts are commonly covered in device physics textbooks, like [8, 9]. Is well known that, for semiconductor materials, a forbidden energy band is present between the valence and the conduction bands. The height of this forbidden band is called *energy gap*. Electrons cannot exist in the forbidden band. The band gap is a very important feature of the semiconductor used to build the devices: it determines the robustness

of the device to breakdown, to temperature, and so on. In between of the top of the valence band and the bottom of the conduction band falls the Fermi level E_F: mathematically speaking, it represents the energy level at which the probability of finding an electron is one half (provided there is an allowed state at that energy level). It can be evaluated considering the Fermi-Dirac distribution, which represents the probability of occupation of a state of energy E:

$$f(E) = \frac{1}{\mathrm{e}^{\frac{E-E_F}{kT}} + 1} \tag{2.1}$$

where $k = 1.38 \times 10^{-23}$[J/K] is Boltzmann's constant and T is the absolute temperature. A schematic drawing of the band structure for a semiconductor material is shown in Figure 2.1. The position of the Fermi level indicates how the semiconductor

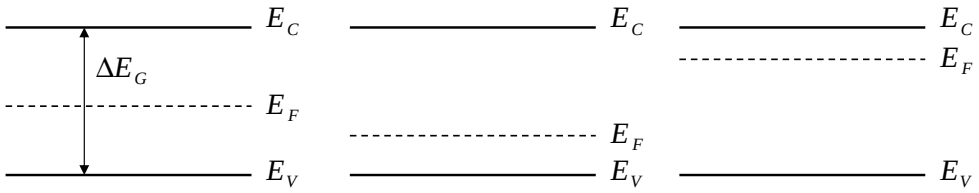


Figure 2.1 – Schematic drawing of a band structure for a semiconductor. From left: the case of an intrinsic semiconductor, a p-doped semiconductor, and an n-doped semiconductor. E_V is the top of the valence band, E_F is the Fermi level, and E_C is the bottom of the conduction band. ΔE_G denotes the bandgap.

is doped. The doping concentration is usually expressed in [cm^{-3}], and common values of doping concentrations range from 10^{14} to $10^{19}\,\mathrm{cm}^{-3}$. The doping is a process increasing the conductivity of the semiconductor, by inserting shallow energy states slightly above the valence band (p-doping) or slightly below the conduction band (n-doping). These states, due to their shallowness, are prone to release carriers (electrons or holes depending on their type) that are going to increase the conductivity of the semiconductor.

For the states to be effective in improving the conductivity of the semiconductor, they must be very close to the edge of the corresponding energy band. When an energy state is close to the middle of the energy gap, it is called a *deep state* or *deep level*, and acts in the opposite way, attracting and holding carriers rather than releasing free carriers. Due to this behavior, these states are called *traps*. Traps can be located at interfaces between materials, at the surface, or in the bulk as well. In a first approximation, if the concentration of filled electron traps is N_T and the number

of donor impurities is N_d, the concentration of free carriers can be expressed as

$$n \simeq N_d - N_T. \tag{2.2}$$

Once an electron is trapped, a significative amount of time my be needed for the electron to be released from the trap: this time can range from milliseconds up to seconds. This is an important aspect for microwave operation, since trapped carriers cannot participate to current conduction. The effect of deep levels can be noticed in the DC or low-frequency characteristics of the device, rather than the microwave operation, since in the last case the shift of the electric field direction is too fast for the trapped electrons to respond.

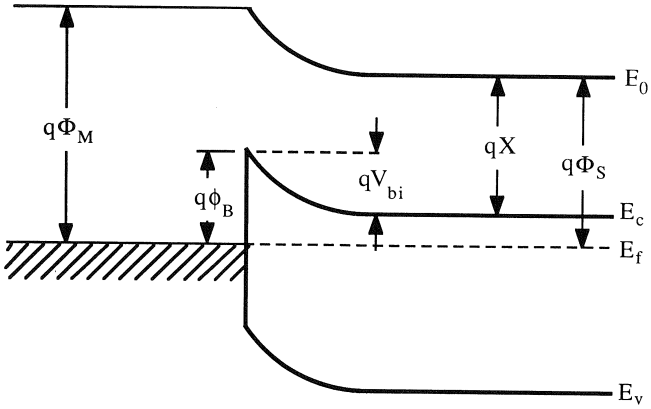


Figure 2.2 – Band diagram of a Schottky junction under no-bias condition – reprinted from [8].

Since the device analyzed in this Chapter are heterojunction FETs (it means that two materials with different crystalline structure are in contact), the description of a Schottky barrier can be an useful review. This is actually an heterojunction, since a metal and a semiconductor are in contact. Usually, the semiconductor work function ϕ_S is lower than the metal work function ϕ_M, so that the resulting band diagram is that in Figure 2.2 (in the case of contact between a metal and an n-doped semiconductor). The work function of a material is the energy between the vacuum level and the Fermi level. If $\phi_S < \phi_M$, before contacting the metal and the semiconductor the electrons of the latter are more energetic than the electrons of the former, and after the two materials are put in contact there will take place a transfer of electrons from the semiconductor to the metal (as the shape of the energy bands indicates). The metal is the *anode* of the device, and the semiconductor is the *cathode*, so that the application of a positive bias between metal and semiconductor decreases the height of the barrier to the electron flow from semiconductor to metal.

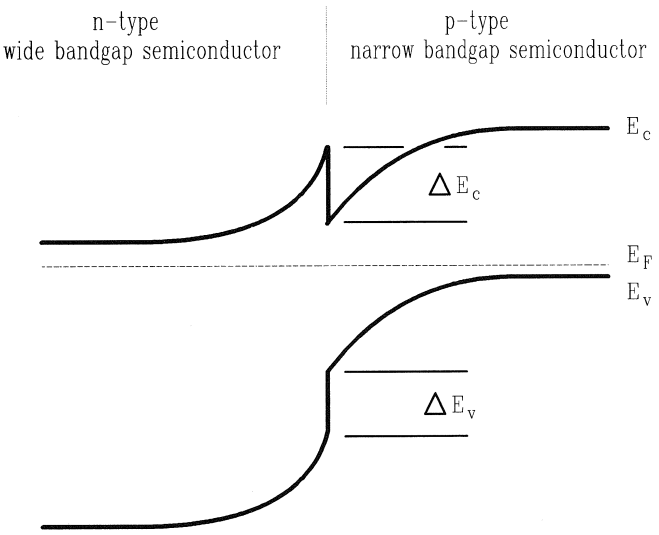


Figure 2.3 – Band diagram of an heterojunction – reprinted from [8].

Usually, the term *heterostrucuture* is used when a contact between two different semiconductors is achieved. The great advantage of building heterostructure devices is that the performance achievable with this kind of devices is superior with respect to silicon devices in terms of high-doping analog behavior. The drawbacks are mainly related with high cost, less mature technologies, and worse reliability. A band diagram of a *p*-*n* heterojunction is shown in Figure 2.3, in which the *p*-doped material has a smaller band gap than the *n*-doped material. It turns out immediately that, since the band gaps are different, a discontinuity in conduction and valence bands will exist. They are indicated with ΔE_C and ΔE_V, respectively. The devices studied here belong to the family of JFETs, in which the *gate* terminal modulates the electron density (and thus the conductivity) of the semiconductor region in between the other two contacts, the *drain* and the *source*. The description of a standard JFET can be found in many electronics textbooks (see for example [10]). Actually, the model of a silicon JFET is quite simple, being described (in the saturation region) by the following:

$$I_D = I_{DSS}\left(1 - \frac{V_{\mathrm{GS}}}{V_P}\right)^2 \tag{2.3}$$

where V_P is the pinch-off voltage and I_{DSS} the device current at $V_{\mathrm{GS}} = 0$; on the other hand, the analytical description of a HEMT device is more complex.

A specific modeling challenge is to describe in a *physics-based* and *self-consistent* way both the electrical and the thermal dynamics of the device. This Chapter is

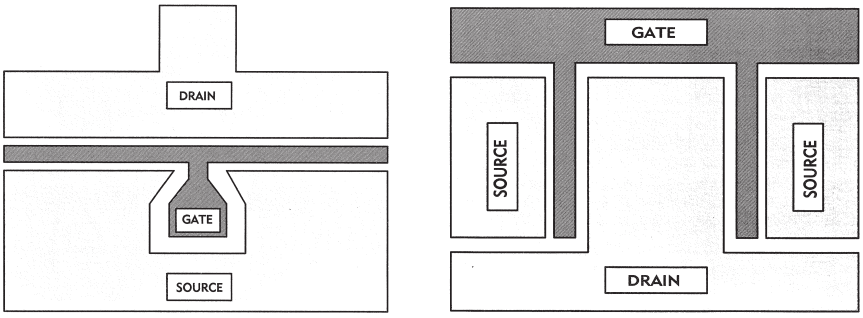


Figure 2.4 – Two typical geometries used in MESFET fabrication: T-shaped (left), standard cell (right) – reprinted from [8].

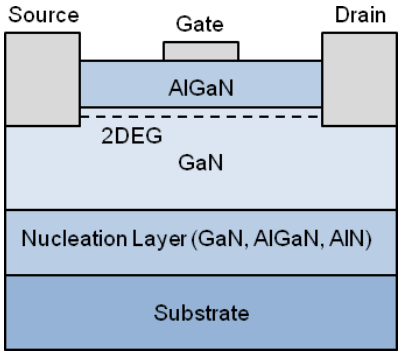


Figure 2.5 – Cross section of a basic AlGaN/GaN HEMT structure.

expressly focused on this aspect. For the sake of completeness, Figure 2.4 shows two typical layouts of MESFETs[1].

2.2 The HEMT device

This section aims at giving some useful concepts about the working principle and main features of *High Electron Mobility Transistors*. These devices are very interesting from the point of view of high-frequency performance, but their reliability is still an issue. The reliability of the device is related to the operating temperature, and accurate thermal modeling and characterization is mandatory. To better understand how the device works, a review of its physics is given here with specific focus on AlGaN/GaN devices. The structure of a basic AlGaN/GaN HEMT is shown in Figure 2.5. The

[1] MEtal Semiconductor Field Effect Transistor.

channel here is the dashed line called 2DEG, which stands for *2-dimensional electron gas*, since the layer in which electrons are confined is really thin. It is interesting how the 2DEG channel is created, since in AlGaN/GaN HEMTs it exists also in the absence of intentional doping of the AlGaN. It is useful to recall the meaning of *surface states* and the phenomenon of *Fermi level pinning* [9]. The band diagram relative of a metal-semiconductor junction (Schottky junction) is shown in Figure 2.6, in which E_0 is the vacuum energy level (the energy of an electron which is not influenced by the material).

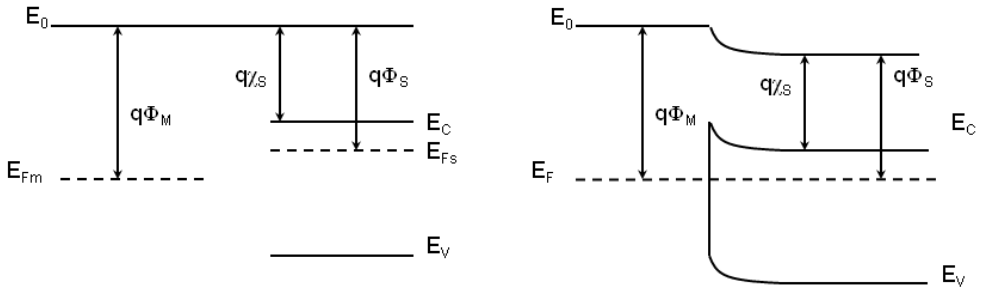


Figure 2.6 – Separated metal/semiconductor band diagrams (left) and band diagram of a Schottky barrier (right) at the thermal equilibrium. E_0 is the vacuum energy level.

Surface states are additional states allowed for electrons, available on the surface of the semiconductor but not in the bulk (basically, only the electrons towards the bulk of the crystal are bonding electrons, and not the ones towards the surface of the crystal). These states, which are called *Shockley-Tamm states*, can also be due to the presence of oxygen on the surface of the semiconductor. These states present a density peak in correspondence of an energy level roughly $E_G/3$ above the top of the valence band E_V. These states can be of *donor* type, if positive when empty and neutral when occupied, or *acceptor* type, if negative when occupied and neutral when empty. Since the density of these states is very high, even a large amount of charge which is moved across the surface implies that the Fermi level moves only slightly, as if it was *pinned*: this is in fact called the *Fermi level pinning* effect.

The theoretical analysis [11] considers the following sources of charge in an AlGaN/GaN HEMT: (i) the negative charge due to the concentration n_s of electrons in the 2DEG channel; (ii) the charge induced by spontaneous polarization in the AlGaN layer, respectively on the surface ($-\sigma_{\mathrm{PZ}}$) and on the interface with GaN ($+\sigma_{\mathrm{PZ}}$); (iii) the integrated sheet charge due to ionized donors in the AlGaN layer, called σ_{AlGaN}; (iv) the charge due to ionized surface states, denoted as $\sigma_{\mathrm{Surface}}$, and (v) the charge due to the buffer layer, σ_{Buffer}. Some considerations are then made on electrostatic

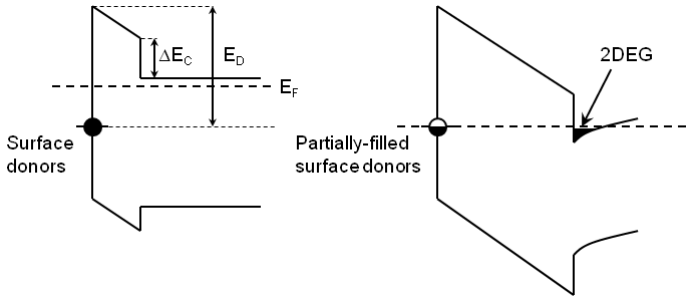


Figure 2.7 – Band diagram of the AlGaN/GaN system showing the surface donor model in the case of undoped AlGaN barrier. (left) The AlGaN thickness is lower than the critical thickness. (right) The AlGaN thickness is greater than the critical thickness necessary for the creation of the 2DEG [11].

basis. The system has to be neutral, if no external electric field is applied. In addition, the dipole due to the polarization does not change the charge balance, as the notation used by the author is confirming ($\pm\sigma_{\mathrm{PZ}}$). The charge in the buffer layer has to be negative, otherwise the 2DEG channel will not be confined at the AlGaN/GaN interface (if positive, the channel would be attracted in the buffer layer, instead of being "pushed" away from the buffer). This rules out the possibility for the 2DEG to be made of thermally generated electrons from the buffer, since in this case the charge left would be positive (it has been just shown that the charge in the buffer has to be negative). From a technological point of view, a well-designed FET calls for a low charge in the buffer layer, so it is reasonable to entirely neglect the buffer charge. The charge balance equation that can be written considering the previous statements is thus the following:

$$\sigma_{\mathrm{Surface}} + \sigma_{\mathrm{AlGaN}} - qn_S = 0 \tag{2.4}$$

that, if rewritten as follows,

$$qn_S = \sigma_{\mathrm{Surface}} + \sigma_{\mathrm{AlGaN}} \tag{2.5}$$

states that the concentration of electrons in the 2DEG channel is the sum of the concentration of ionized donors in the AlGaN layer plus the concentration of ionized donors on the surface of the AlGaN layer. A graphical explanation is given from Figure 2.7. It has been shown that the formation of the 2DEG channel is due to the presence of AlGaN surface donor states, and that for a nominally undoped $Al_{0.34}Ga_{0.66}N$/GaN structure the 2DEG channel forms only when the barrier thickness exceeds 35 Å.

2.2.1 Large signal models of HEMTs

This section aims at giving some details about the large-signal, electro-thermal modeling of HEMTs. In this section, the device temperature will be assumed as a parameter (i.e., no details will be given on how the temperature is calculated, which will be described later).

First of all, there are two big families of models: empirical and physics-based. The former are only aimed at fitting the actual device characteristic, while the latter aim at describing the device behavior based on the physical characteristics of the device itself. Since physics-based models are very heavy under a computational point of view, and thus not really useful in circuit design, they are not extensively used in this field. Conversely, in the field of microwave circuit design, the use of empirical models is preferred. An overview of the different models can be found in [8], in which several examples of the models derived from the two different approaches are described.

One of the first models being available for circuit simulations has been the Curtice model for MESFETs [8, 13]. It describes the drain current as function of V_{DS} and V_{GS} as follows:

$$I_{\mathrm{DS}}(V_{\mathrm{GS}}, V_{\mathrm{DS}}) = \beta(V_{\mathrm{GS}} - V_{\mathrm{T0}})^2(1 + \lambda V_{\mathrm{DS}})\tanh(\alpha V_{\mathrm{DS}}) \tag{2.6}$$

where $V_{\mathrm{GS}}, V_{\mathrm{DS}}$ are the device voltages and $\beta, V_{\mathrm{T0}}, \alpha, \lambda$ are model parameters. Eqn. (2.6) can be split in three components:

1. the term $\beta(V_{\mathrm{GS}} - V_{\mathrm{T0}})^2$ accounts for the square-law behavior of the drain current with respect to V_{GS}; the parameter β has the units of $[\mathrm{A/V^2}]$.

2. the term $1 + \lambda V_{\mathrm{DS}}$ models the device output conductance which is varying with V_{DS} (due to the channel lenght modulation);

3. the term $\tanh(\alpha V_{\mathrm{DS}})$ is used because it approximates well the characteristic $I_{\mathrm{DS}}(V_{\mathrm{DS}})$ (i.e., the output characteristic of the device) between the linear and the saturation regions.

Eqn. (2.6) describes the DC behavior of the device. The modeling of capacitances is derived from the theory of Schottky junctions; the gate-source and gate-drain capacitances are described as

$$C_{GS,\mathrm{GD}} = C_{GS0,\mathrm{GD0}}\left(1 - \frac{V_{GS,DS}}{V_{\mathrm{bi}}}\right)^{-\frac{1}{2}} \tag{2.7}$$

where V_{bi} is the built-in potential of the Schottky gate and $C_{GS0,DS0}$ are the zero-bias capacitances (gate-source or gate-drain, respectively).

Starting from this model, several models have been further developed. The electrical model we used (which is fairly common in HEMT device modeling) is shown in Figure 2.8, and it has been developed on the basis of the model proposed by [12],

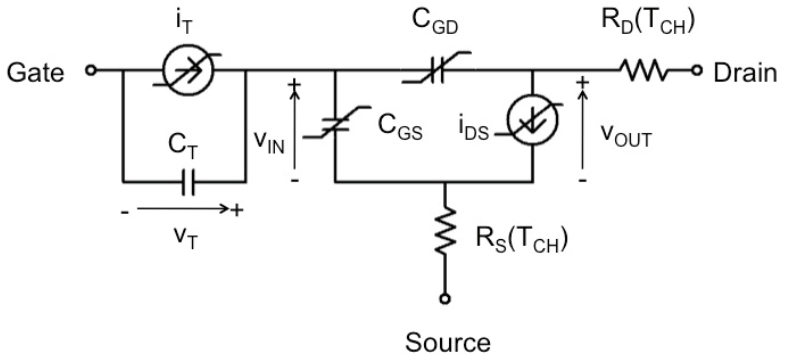


Figure 2.8 – HEMT large signal model schematic. I_T and C_T describe the trap behavior.

which is in turn an evolution of the model developed by Curtice and Ettenberg [13, 14]. Basically, the Curtice-Ettenberg model is defined by the following relationships:

$$I_{\mathrm{DS}} = (A_0 + A_1 v_1 + A_2 v_1^2 + A_3 v_1^3)\tanh(\gamma V_{\mathrm{DS}}) \tag{2.8}$$

$$g_m = \tanh(\gamma V_{\mathrm{DS}})(A_1 v_2 + 2A_2 v_1 v_2 + 3A_3 v_1^2 v_2) \tag{2.9}$$

$$g_{\mathrm{DS}} = (A_0 + A_1 v_1 + A_2 v_1^2 + A_3 v_1^3)\mathrm{sech}^2(\gamma V_{\mathrm{DS}})\gamma - \beta V_{\mathrm{GS}}(A_1 + 2A_2 v_1 + 3A_3 v_1^2)\tanh(\gamma V_{\mathrm{DS}}) + 1/V_{\mathrm{DS0}} \tag{2.10}$$

where

$$v_1 = V_{\mathrm{GS}}(1 + \beta(V_{\mathrm{DS0}} - V_{\mathrm{DS}})) \tag{2.11}$$

$$v_2 = 1 + \beta(V_{\mathrm{DS0}} - V_{\mathrm{DS}}) \tag{2.12}$$

The voltage v_1 is used to model the increase of pinch-off voltage with V_{DS}, β is used to control the change of pinch-off voltage, V_{DS0} is the value of V_{DS} (with the device working in the saturation region) in which A_i coefficients are evaluated; γ is an empirical parameter. Parasitic inductances are not accounted for in this model. They are important in the RF behavior of the component, but since the thermal dynamics is low-pass filtered by the thermal inertia of the system, the effect of inductances on the temperature oscillations is negligible. The current generator I_T and the capacitance C_T describe the effect of traps based on the approach of Rathmell and Parker [15, 16]. An extensive description of the model is given in [17], and a summary of the main features is given here:

trap modeling in GaN devices two kinds of traps are present, *bulk traps* and *surface traps*. When the device is pulsed, the former type acts like a high-pass filter,

while the latter behaves like a low-pass filter. This feature is accounted for by including the current generator I_T and the capacitor C_T, the former accounting for the rate of charging/discharging of traps and the latter accounting for the charge captured by the trap. The following relationship holds:

$$I_T = \omega_0 C_T \left((V_0 - v_T) - v_T \exp(q v_I / (k T_{\mathrm{CH}}) \right) \tag{2.13}$$

where

$$\omega_0 = A_T {T_{\mathrm{CH}}}^2 \exp(-E_{\mathrm{AT}} / (k T_{\mathrm{CH}})) \tag{2.14}$$

is the trap characteristic frequency, E_{AT} is the activation energy of the trap; the constants A_T and E_{AT} can be extracted from temperature-dependent measurements of dispersive phenomena. The voltage v_I is the difference between the trap level and the Fermi level, and is modeled as a function of drain and gate potentials:

$$v_I = h_0 + h_G v_{\mathrm{GS}} + h_D v_{\mathrm{DS}} \tag{2.15}$$

in which the signs of the coefficients h_0, h_G, h_V depend on the nature of the trap (donor or acceptor) and on the trap location (bulk or surface);

temperature dependence the model is strongly dependent on temperature: in the current generator i_{DS}, in the trap model, and in the parasitic resistances R_S and R_D. To exploit these features of the model, the channel temperature has to be evaluated with good accuracy and in a self-consistent way.

Summarizing, this section has shown that, nowadays, the circuit designer has both powerful simulation tools and accurate device models, which can be fruitfully employed in circuit simulation and design: non-linearities, hysteresis effects, physical aspects are accounted for by the family of models shown here. The next step is to find a way to correctly describe the thermal behavior of the system, in order to correctly evaluate the channel temperature T_{CH}.

2.3 Lumped elements modeling techniques

The previous section has covered the techniques for modeling HEMT devices under an electro-thermal point of view. A basic concept which needs to be highlighted is that *the device temperature influences very strongly the electrical characteristics*. In commonly used circuit simulators like SPICE, the temperature dependence of the electrical characteristics of a device is accounted for by the model, but usually the temperature is a *parameter*, i.e. a value which is fixed at the beginning in the simulation setup, and this value cannot change during the simulation. This is a strong approximation, which often leads to uncorrect evaluation of the device performance,

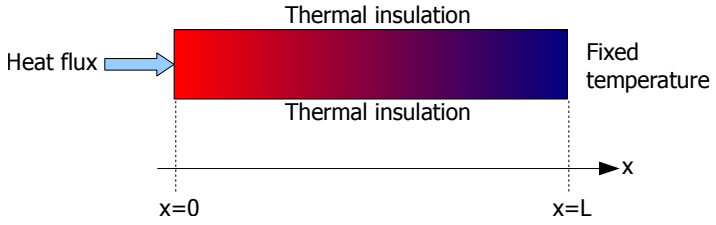


Figure 2.9 – A case of a bar with a heated end and the opposite end kept at a constant temperature. All other surfaces are insulated.

especially when studying power amplifiers (PAs). This calls for accurate thermal modeling of the device. Here we are thus looking for a way to describe, *given a fixed amount of power*, how the temperature develops in the device structure. At a first glance, we have already faced this kind of problems, and a powerful solution has been found in the Finite Element Method (FEM). In this case, this approach is not suitable, mainly for two reasons:

1. FEM models are heavy from a computational point of view, although this problem is made less and less critical by the continuous progresses in PC capabilities;

2. the FEM *thermal* model is very hard or impossible to interface with compact CAD models.

The second reason is the most powerful push towards the development of techniques for simple, computationally light and physics-based thermal modeling of HEMTs structures, and the *Lumped Elements Technique* is the most efficient replacement of the FEM model, since it can describe with good accuracy the static and dynamic thermal behavior of the devices under analysis. The advantage of Lumped Elements Models (LEMs) is that they can be embedded into circuit CAD simulators, since they are made of the same elements (resistances and capacitances) which are used in the electrical part of the circuit, with the only difference that they are *thermal resistances* and *thermal capacitances*.

This section will show examples of LE thermal modeling of different device structures. It will be shown that these models can satisfactorily replace FE models to give physics-based descriptions of the device thermal behavior; in addition, and unlike FE models, they can be easily integrated in standard circuit CAD tools.

2.3.1 Where the LEM idea comes from

Let us consider the Fourier's equation of heat transfer, here rewritten in the static case (i.e., assuming time derivatives equal to zero) and considering an isotropic material:

$$\nabla^2 T = 0 \tag{2.16}$$

which can be further simplified in the case of uni-directional (along x direction, for instance) heat transfer (this is the case of a bar with length L with one end tied at a constant temperature and the heat flux flowing into the other end in $x = 0$; all other surfaces are thermally insulated, as shown in Figure 2.9):

$$\frac{\partial^2 T}{\partial x^2} = 0 \tag{2.17}$$

yielding:

$$T(x) = ax + b \tag{2.18}$$

where the constants a, b can be determined by imposing the boundary conditions, which are the fixed temperature at the end of the bar $T(L) = T_0$ and the heat flux in $x = 0$, that is $-\partial T/\partial x|_{x=0} = -q/k$, assuming a fixed heat flux equal to $q\,[\mathrm{W/m^2}]$. Then, it is easy to find that the solution is

$$T(x) = T_0 + \frac{q}{k}(L - x). \tag{2.19}$$

The heat flux q is defined per unit area; let us include the cross section A by defining $Q = A \cdot q$; it is possible to write the previous equation as

$$T(x) = T_0 + \frac{Q}{A \cdot k}(L - x). \tag{2.20}$$

It is immediate to recognize the formal analogy between the voltage drop across a resistor in which an amount of current is flowing and the heat equation solution. It is straightforward to define a *thermal resistance* as follows:

$$R_\theta = \frac{T(x = 0) - T(x = L)}{Q} = \frac{L}{A \cdot k}. \tag{2.21}$$

This simple example shows how the heat transfer can be described by means of electrical *lumped elements*, on the basis of the following formal analogies: the temperature $T\,[\mathrm{K}]$ is the voltage $V\,[\mathrm{V}]$, the power $Q\,[\mathrm{W}]$ is the current $I\,[\mathrm{A}]$, and the thermal resistance expressed in [K/W] is represented by the electrical resistance in $[\mathrm{V/A} = \Omega]$.

In order for these formulas to be meaningful, the heat flux has to be *unidirectional*. Note that, in this case, it does not matter how long is the bar, meaning that a single element (one resistance) is enough to calculate exactly the temperature at the

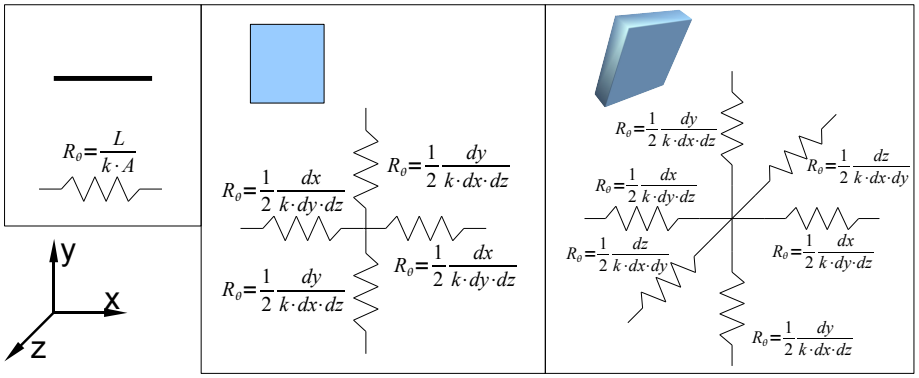


Figure 2.10 – Types of elements suitable for LEM thermal modeling. One-dimensional element (left), two-dimensional element (center), three-dimensional element (right).

heating point. This is obviously *not* the case in real structures. Let us consider a two-dimensional structure on the xy plane: actually, heat can flow in all directions over the plane, but each direction can be decomposed in its components, one parallel to the x axis and one parallel to the y axis. This can therefore be modeled, with some approximation, by a bi-directional resistance network. The concept can be extended to space, by means of a three-dimensional resistance network, as shown in Figure 2.10.

The advantage of this kind of modeling is that it can be embedded into circuit simulators and standard CAD tools, since it describes the thermal behavior of the system by means of electrical components. The extension of the modeling technique to the dynamic behavior, in order to describe the thermal transients of the structure under test, is straightforward. In the general case of a three-dimensional block, a thermal capacitance is connected to the central point of the resistance network, with a value given by

$$C_\theta = \rho \cdot C_p \cdot dx \cdot dy \cdot dz \tag{2.22}$$

where $\rho\,[\mathrm{kg/m^3}]$ is the density of the material, $C_p\,[\mathrm{J/(kg \cdot K)}]$ is the specific heat at constant pressure, and dx, dy, dz are the dimensions of the block, respectively. A block of the type just described is shown in Figure 2.11-(a). It has to be noted that the block shown in Figure 2.11-(a) can present some drawbacks when it is used to describe a part of a system when the incoming power is a boundary condition (i.e., a current generator is connected to a terminal of the network). Imagine to apply a power step; the temperature of the node where the generator is connected will have a step as well, since the temperature of that node will not be clamped by the capacitance:

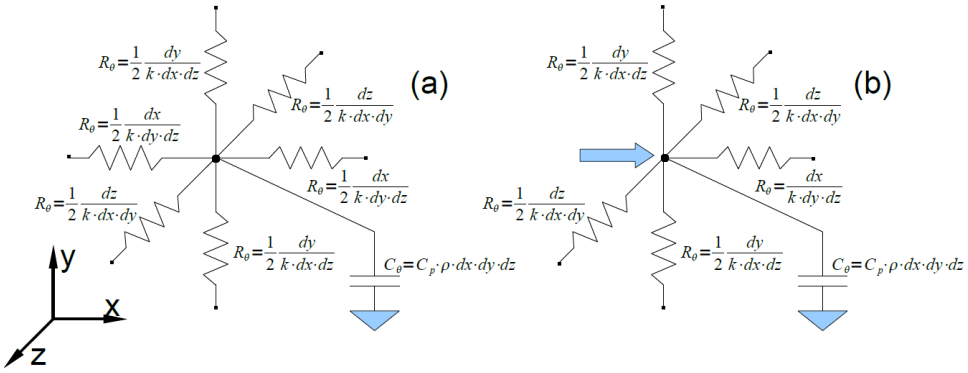


Figure 2.11 – Three-dimensional blocks used to model temperature transients; inner block (left) and border block (right) with indication of heat direction (arrow).

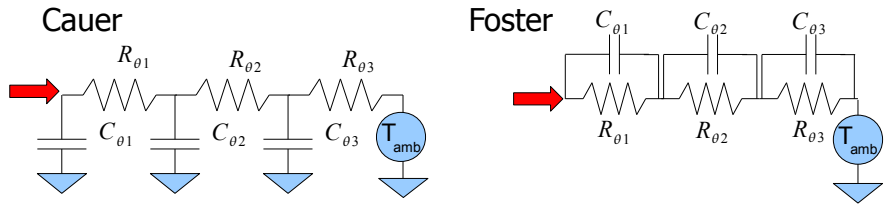


Figure 2.12 – Cauer and Foster thermal networks.

the capacitive effect starts from the central node, which is not the one to which the power is input. It is thus useful to replace the block of Figure 2.11-(a) with the one shown in Figure 2.11-(b), in which the central node and the input power node are merged together, thus using one resistance instead of two in the x direction (the one between the central node and the input power node is set to zero while the other one is doubled). This can also be understood remembering that the Cauer network is a $C_\theta - R_\theta$ ladder network, like the one shown in Figure 2.12. The heat flux is entering a node, the temperature of which cannot have discontinuities, since it is connected to a capacitance.

A critical step in the LEM approach is to find the minimum number of blocks that can describe the overall device with the required accuracy. Even if the meshing operation can be time-consuming, mainly due to the fact that it has to be done manually by the designer, the computational speed of a dynamic LEM model is far higher

than the speed of an equivalent FEM model (while the time needed for computing the static behavior can be of the same order of magnitude). Some practical aspects of building thermal networks will be given in the next sections.

A very important aspect of this approach is that its accuracy depends on mesh quality *only*, since no fitting parameters are needed. This observation will be fully supported by the application examples that will follow in the next sections.

The Cauer and Foster networks, shown in Figure 2.12, are still the simplest and most popular way to describe the dynamic thermal behavior of the temperature even for complex subsystems. While the Cauer network is more physically meaningful (each node is somehow related to a temperature which can be associated to a region in the physical system), the values of the components of a Foster network are easier to compute, but the only physically meaningful temperature is that of the first node, in which the power is flowing; considering a 3-stage Foster network, the time response to a power step is:

$$T(t) = R_{\theta 1}(1 - \mathrm{e}^{-t/\tau_1}) + R_{\theta 2}(1 - \mathrm{e}^{-t/\tau_2}) + R_{\theta 3}(1 - \mathrm{e}^{-t/\tau_3}) \tag{2.23}$$

which can be generalized to the case of an N-stage network:

$$T(t) = \sum_{i=1}^{N} R_{\theta i}(1 - \mathrm{e}^{-t/\tau_i}) \tag{2.24}$$

where $\tau_i = R_{\theta i}C_i$. Unlike in the LEM approach described above, the determination of coefficients $R_{\theta i}$ and τ_i (the determination of capacitance is then straightforward) is a matter of pure fitting. No mesh is involved in this case: the ladder network is one-dimensional, independently of the actual geometry of the system under test, and the number of stages depends only on the accuracy requested. Usually, the time response of an electron device cannot be fit in a reasonable way by using just one stage, and it is common to use at least 3 stages (or even more) to describe the thermal behavior with accuracy.

An interesting question is the following: *how* can one determine the "right" number of stages to be used in the Foster network in such a way as to obtain a good fit? Even if [20] has demonstrated how to determine the thermal constants in a system by using a deconvolution approach, we have developed a simple method to obtain compact thermal networks with the smallest number of stages necessary to achieve a good fit of the thermal response. Basically, it all starts from the following consideration: each thermal response can effectively be described by a relationship like (2.24). If we consider a single term, and we differentiate the term with respect to the *natural*

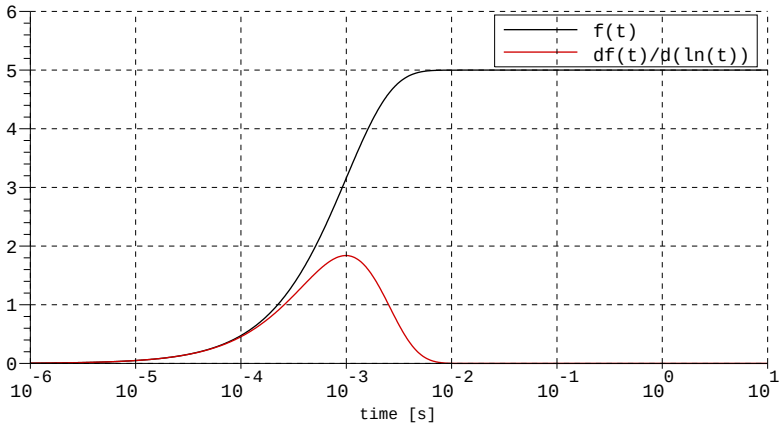


Figure 2.13 – The term $5(1-\mathrm{e}^{-t/10^{-3}})$ and its derivative.

logarithm of time, we have:

$$\frac{df(t)}{d\ln t} = \frac{df(t)/dt}{d\ln(t)/dt} = t\frac{df(t)}{dt} = t\frac{d(R_\theta(1-\mathrm{e}^{-t/\tau_1}))}{dt}$$
$$= \frac{R_\theta}{\tau}t\mathrm{e}^{-t/\tau} = h(t); \tag{2.25}$$

an example is shown in Figure 2.13. It can be easily demonstrated that derivative peaks at $t=\tau$, and the value of the maximum is R_θ/e. In fact, if we calculate the local maxima of the derivative, thus deriving $h(t)$ and finding its zeroes, we obtain:

$$\frac{dh(t)}{dt} = \frac{R_\theta}{\tau}\mathrm{e}^{-t/\tau}(1-t/\tau) = 0 \tag{2.26}$$

which is satisfied for $t=\tau$, and the value of $h(t=\tau)$ is R_θ/e. It means that, for a single time constant system, it is enough to differentiate the temperature response with respect to the natural logarithm of time, calculate the local maxima of the derivative, and the parameter values R_θ and τ are easily obtained. The method can be easily extended for systems with more than one time constant. First, the number of peaks in the temperature response is determined. This is also the number of stages needed to fit the transient waveform. This simple method cannot detect time constants too close to one another; they will simply be merged in a single term, with an equivalent R_θ and an equivalent τ. This is likely to be the real scenario, so a way to determine the values for $R_{\theta i}$ and τ_i is needed, in order to avoid manual tuning, that can be time-consuming when the stages are three or more. It is possible to make the fitting process

automatic by using methods of least-squares optimization, once the fitting function, the list of fitting parameters, and the initial guess values for these parameters are given. This process is extremely simple and fast (for the algorithm to converge to a suitable solution) if the initial guess values for the fitting parameters are close to the best-fit values. Let us suppose that we detect 3 peaks. Thus, we will have 6 fitting parameters, $R_{\theta 1}, R_{\theta 2}, R_{\theta 3}$ and τ_1, τ_2, τ_3. A sensible choice for the initial guess values are the values obtained from the thermal transient derivative $h(t)$ as explained before, in particular:

$$\tau_1, \tau_2, \tau_3 \rightarrow R_{\theta_1} = h(\tau_1) \cdot \mathrm{e}, R_{\theta_2} = h(\tau_2) \cdot \mathrm{e}, R_{\theta_3} = h(\tau_3) \cdot \mathrm{e}.$$

An application has been developed in LabVIEW using the Levenberg-Marquardt algorithm (which is a standard algorithm to find the minimum of a function expressed as a sum of squares, expressly developed for functions with several parameters) to calculate the best-fit parameters. Results of this method will be shown in the next sections.

2.4 Self-consistent coupling of electrical and thermal models

In the previous sections we have described both temperature-dependent electrical large-signal models, and lumped element thermal models for HEMTs. The next step is then to merge them into a self-consistent electro-thermal model.

The relationship between dissipated power P and temperature T_{CH} can be written as:

$$P = V \cdot I = V \cdot I(T_{\mathrm{CH}}) = V \times I(T_{\mathrm{CH}}(V \cdot I)) \tag{2.27}$$

which suggests the idea of a "loop" in the determination of V, I, T_{CH}. Let us suppose we have a thermal network, made of thermal resistances and capacitances: this can be represented like a system in which P is the input and T_{CH} is the output. On the other hand, the large-signal, electro-thermal model of an electron device takes the temperature as an input together with the bias voltages, and outputs the current, and thus the dissipated power. The situation is depicted in Figure 2.14.

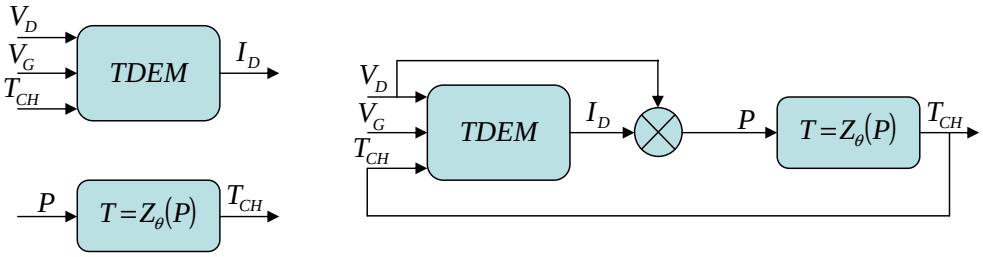


Figure 2.14 – Block diagram of a self-consistent electro-thermal device model. Starting from two independent systems, the first being the Temperature-Dependent Electrical Model (TDEM) and the second the lumped element thermal network Z_θ, their feedback connection describes the device behavior in the presence of self-heating.

2.5 Application cases

In the previous sections the theoretical concepts related to the large-signal, electro-thermal modeling of electron devices have been explained. In this section, a few applications will be considered. The first case study will be the application of these concepts to a power MOSFET; then, the attention will move to AlGaN/GaN HEMT devices, which require a complex electro-thermal model, and a thermal description that is all but straightforward, since they are made of multi-layer structures, and feature complex boundary conditions.

2.5.1 Power MOSFET on a heat sink

The first case study is the analysis of a power MOSFET that sits on a natural-air cooled heat sink. In this work, which is fully described in [21], we used the following model of the MOSFET[2]:

$$I_D = \begin{cases} \beta(V_{\mathrm{DS}}(V_{\mathrm{GS}} - V_T) - V_{\mathrm{DS}}^2/2)(1 + \lambda V_{\mathrm{DS}}) & \text{for} \quad V_{\mathrm{DS}} < V_{\mathrm{GS}} - V_T \\ \dfrac{\beta}{2}(V_{\mathrm{GS}} - V_T)^2(1 + \lambda V_{\mathrm{DS}}) & \text{for} \quad V_{\mathrm{DS}} \geq V_{\mathrm{GS}} - V_T \end{cases} \tag{2.28}$$

This fairly standard model has been coupled with a thermal network which models the physical structure of the assembly made of device, heat-sink and package. A photograph of the real system and an exploded view of the model are shown in Figure 2.15. Thanks to symmetry, only one half of the structure needs to be modeled.

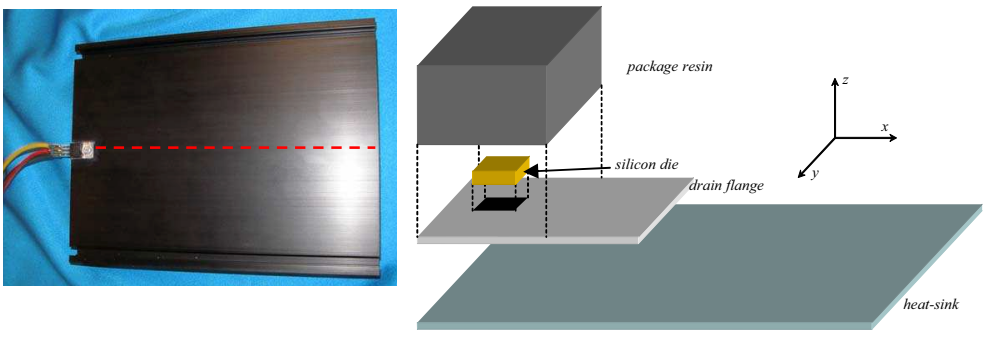


Figure 2.15 – Photograph (left) and exploded view (right) of the MOSFET assembly. The red line indicates the symmetry axis.

[2]The term $(1 + \lambda V_{\mathrm{DS}})$ is meaningful in the saturation region of the device; however, in circuit simulators, the term is present in both operating regions, in order to avoid discontinuities during the transition between the two regions.

At a first glance, the dependence of the electrical model (2.28) on temperature is not visible. Actually, two parameters in the model are strongly temperature-dependent: the threshold voltage V_T, and the electron mobility μ_n, which is defined inside the parameter β:

$$\beta = \mu_n C_{\text{ox}} \frac{W}{L}. \tag{2.29}$$

The temperature dependence of these two has opposite effects on the drain current: if the channel temperature increases, both V_T and μ_n decrease, but while the latter reduces I_D, the former increases I_D. This means that two opposite effects take place in the presence of self-heating, resulting in a significant deviation from the isothermal output characteristics, as will be shown later on.

If we want to build a physical, self-consistent electrothermal model of the MOSFET assembly, the first step is to build a lumped element thermal network that models the assembly. In order to keep the network within manageable dimensions, careful meshing was carried out. In particular, different building blocks were used in different parts of the assembly, based on the following considerations:

- ♢ the temperature over the heat source (i.e., the silicon die) can be considered uniform, and therefore it is modeled as a single node, in which the heat power is entering;
- ♢ heat is fundamentally forced to flow towards the heat sink, due to how the structure is arranged (the die is sitting on the drain flange, which is highly conductive, and in turn mounted on an aluminum heat sink); the heat dissipated by the top surface of the device is limited, so that this part can be modeled as a single resistance, since the heat flow can be assumed to be one-dimensional;
- ♢ conversely, the drain flange has to be modeled in a three-dimensional way, being the part in the assembly with the highest temperature gradient; basically, no assumptions or simplifications can be made *a priori* about which direction the heat flux will follow;
- ♢ the heat-sink is basically a large thin plate, made of aluminum, the thermal conductivity of which is high; thus, the thermal gradient in the vertical direction can be neglected, and two-dimensional elements can be used to model it;
- ♢ in order to reduce the number of elements, the size of the elements is increasing as they go further away from the heating source: this idea has been fruitfully exploited in the discretization of the aluminum heat-sink.

On the basis of the previous considerations, a mesh for the assembly has been generated, as shown in Figure 2.16. The different building blocks used in the lumped

element model are shown in Figure 2.17. Note that *the value of the thermal resistances is calculated based on the physical dimensions of the corresponding block and the thermal conductivity of the material of which the block is made of* (apart from some exceptions that will be highlighted later on). It means that *no fitting parameters* are involved in the thermal network.

The same aspects holds for the boundary conditions as well (even if the convective heat exchange coefficient can be considered as a fitting parameter, since it is usually unknown, and dependent on several aspects like surface orientation, emissivity, temperature, and so on). The building blocks used in this work are:

1. one-dimensional elements, where heat flux can be considered as one-dimensional;
2. two-dimensional elements, in which the temperature gradient can be neglected in one dimension; the presence of boundary conditions may need to be mod-

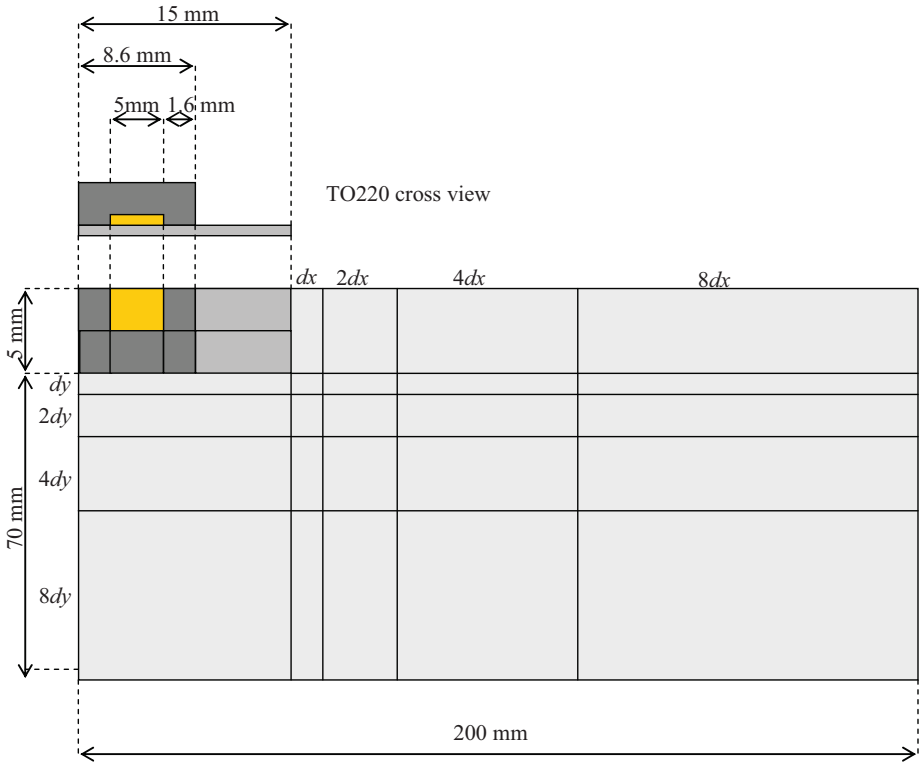


Figure 2.16 – Schematic 2-D view of how the assembly has been discretized. Thanks to symmetry, only one half of the structure needs to be modeled.

eled (e.g., the convective heat exchange between the heat-sink surfaces and the surrounding environment);

3. three-dimensional elements, used to describe the drain flange.

The convective boundary conditions have been modeled with resistances connected between the central node of the two-dimensional element and a node whose voltage is numerically equal to the ambient temperature. Given the dimensions of the element (only length ℓ and width w are important in this case), and given the coefficient of convective heat exchange h_{conv}, the value of the resistance modeling the boundary condition is given by

$$R_{\text{conv}} = \frac{1}{h_{\text{conv}} \cdot \ell \cdot w}. \tag{2.30}$$

The only exceptions to Eqn. (2.30) are the resistance that models the convective heat transfer between the top surface of the package and the ambient, which has been fixed at 388 K/W, and the contact resistance between the silicon die and the drain flange, which has been fixed at 1.21 K/W. A three-dimensional view of part of the lumped element model is shown in Figure 2.18.

It was said before that the electrical parameters dependent on temperature are the threshold voltage, V_T, and the electron mobility, μ_n. The electron mobility has been defined according to the Arora model [22]:

$$\mu_n = 88 \cdot T_n^{-0.57} + \frac{7.4 \cdot 10^8 \cdot T^{-2.33}}{1 + \dfrac{N}{1.26 \cdot 10^{17} \cdot T_n^{2.4}} \cdot 0.88 \cdot T_n^{-0.146}} \tag{2.31}$$

where $T_n = T/300$, with T [K] and N [cm^{-3}] is the doping concentration (here used as a fitting parameter with value $N = 3 \times 10^{15}\,\text{cm}^{-3}$); the term $C_{\text{ox}}W/L$ can be embedded in a single fitting parameter with value $\beta_0 = 6.05 \times 10^{-3}\,\text{A/V}^2$, which is not dependent on temperature. The term λ, which can be easily extracted from pulsed measurements (carried out by applying pulses 10 µs long), has been found to be $0.012\,\text{V}^{-1}$. The threshold voltage V_T has been modeled with a first-order temperature dependence:

$$V_T(T) = V_T(300)(1 - \alpha(T - 300)) \tag{2.32}$$

which takes the following form after the determination of the fitting parameters $V_T(300)$ and α:

$$V_T(T) = 4.22(1 - 0.006(T - 300)). \tag{2.33}$$

The last data necessary to set up the model are the thermal conductivities of the various materials and the coefficients of convective heat exchange on the top and bottom surfaces of the heat-sink. The thermal conductivities considered here were $200\,\text{W/(m} \cdot \text{K)}$ for aluminum, $148\,\text{W/(m} \cdot \text{K)}$ for silicon, and $0.78\,\text{W/(m} \cdot \text{K)}$ for the

epoxy resin. The determination of the contact resistance between silicon die and drain flange has been estimated by means of TSP measurements, as described in the previous chapter. For further details, refer to [21]. Here, the obtained results are described and analyzed.

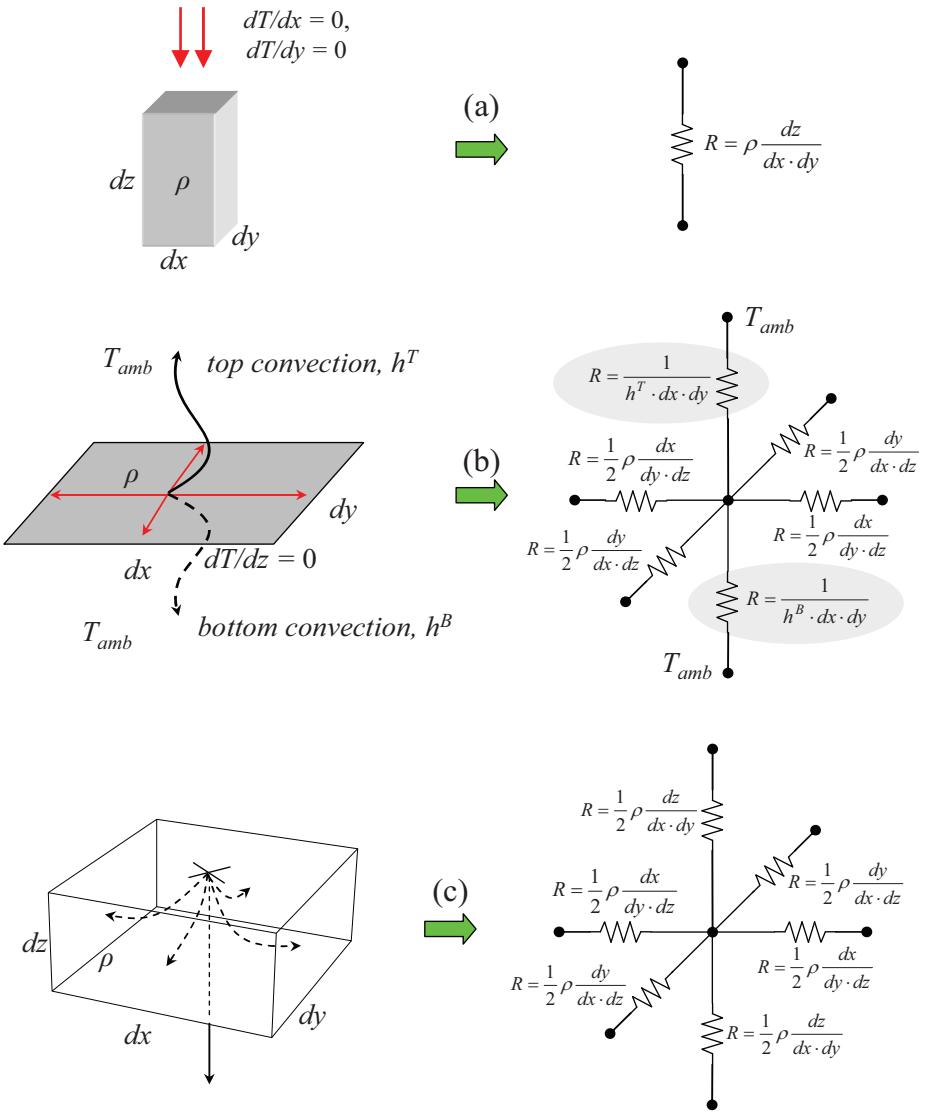


Figure 2.17 – An overview of the elements used to model the assembly, with the corresponding thermal networks. From top: one-dimensional element (a), two-dimensional element with convective boundary conditions (b), three-dimensional element (c).

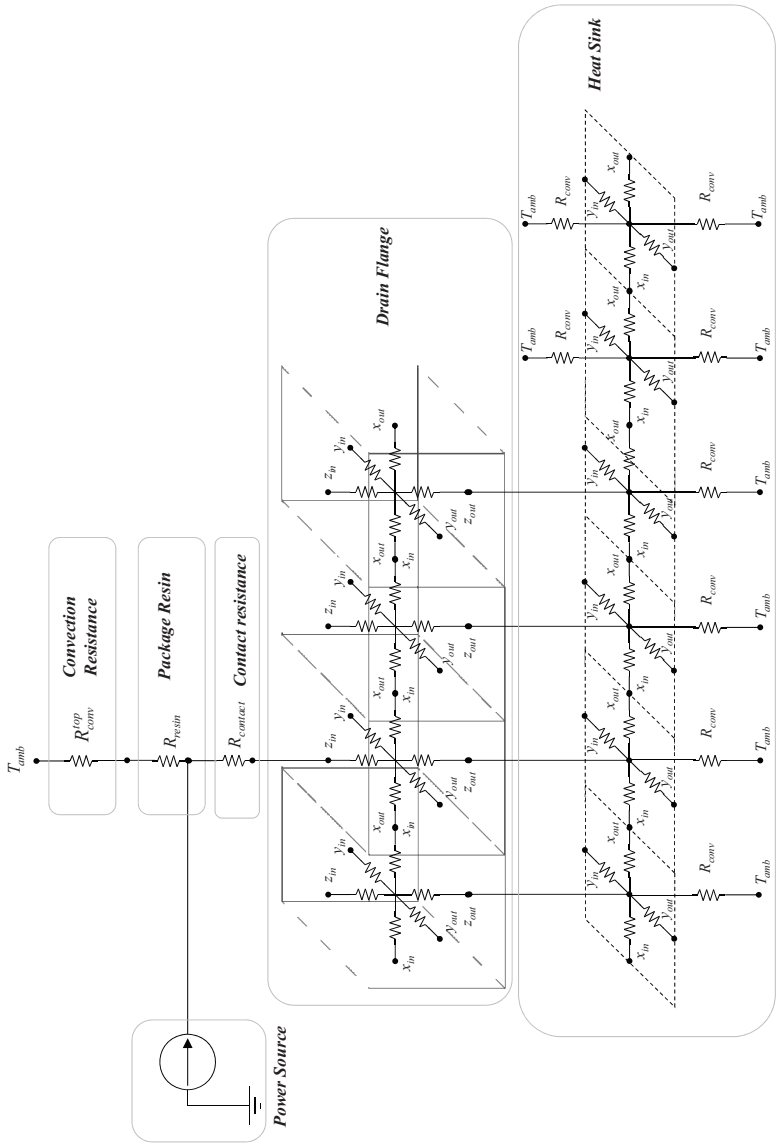


Figure 2.18 – A three-dimensional view of part of the lumped element thermal network. For the sake of clarity, only the part of the heat sink underlying the device is shown.

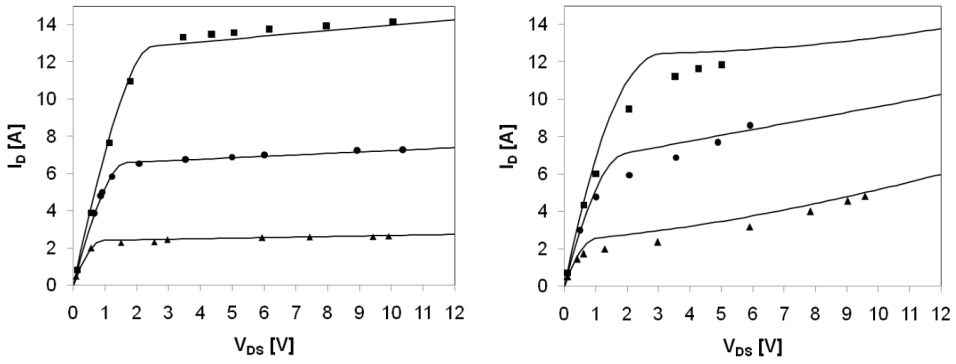


Figure 2.19 – Comparison between simulation results (lines) and measurements (points) in the case of pulsed (isothermal) measurements (left) and in the DC case (i.e., with self-heating) (right).

Test point	Measurement	FEM	LEM
Silicon die	86	86	86
Package top	75	77	76
Flange top	62	62	65
Flange bottom	n.a.	60	62
Heat-sink	52	52	54

Table 2.1 – Comparison between measured and modeled temperatures obtained from FEM and LEM models (in °C). The MOSFET dissipates 21.2 W, and the ambient temperature is 26 °C.

Figure 2.19 shows a comparison between modeled results and measurements both without and with self-heating. While the isothermal case features excellent agreement between simulations and measurements, in the presence of self-heating the match is less accurate, even if the results can be considered good, considering that the MOSFET model we utilized is one of simplest and that a limited number of fitting parameters has been used. The use of more complex models, like for example [23], will probably lead to better accuracy, especially in the self-heating case.

In order to test the accuracy of this electro-thermal model, a FE model has been developed. The comparison between the LE model and the corresponding FE model will be extensively used, since it allows to check if the discretization introduced in the LE model is fine enough. Modeling one half of the structure, thanks to symmetry, and using the same values for thermal conductivities and convective heat exchange coefficients as in the LE model, we obtained excellent agreement between the two models, as reported in Table 2.1. The geometry entities the temperature of which is evaluated

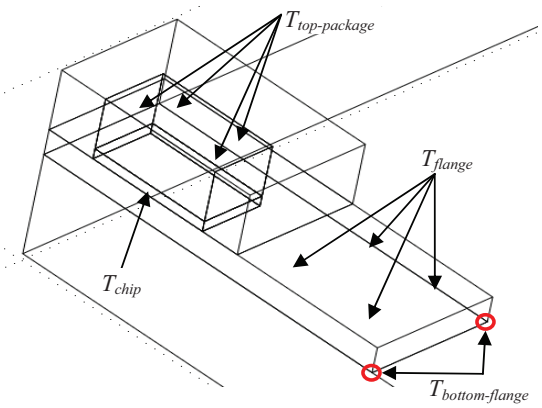


Figure 2.20 – A blow-up on the device with the points/surfaces/volumes used to evaluate temperatures from the FEM model.

are shown in Figure 2.20. The silicon die temperature, the flange temperature and the package top surface temperature were used for the fitting process. The silicon die temperature is averaged over the die volume while the flange and the top-package temperature are averaged over the respective surfaces. It must be pointed out that, except for the silicon die, where the difference between the average temperature and the peak temperature is not negligible (about 6°C, being the silicon die the heat source), the differences between single-point and surface-averaged temperatures on the top surface of the package and on the flange is always within $1 \div 2\,^{\circ}\mathrm{C}$.

This means that the *technique* followed in the LE discretization is correct, since, from a *numerical* point of view, the results given from the two models are the same – with the difference that while the FE model remains a self-standing black-box, the LE model has been embedded within an electro-thermal model of the electron device considered. The electro-thermal model model was implemented in MATLAB/Simulink, but it can be developed in ADS as well as in SPICE. Figure 2.21 shows the self-consistent feedback loop implemented in MATLAB.

Transient thermal modeling

The next step is to include the transient dynamics of self-heating, accounting for the thermal capacitances of the assembly by means of capacitors (whose values depend exclusively on the size of the discretization elements and on the material thermal characteristics).

This extension, with a refinement of the discretization of the thermal network as well, can be found in [24]. This work focuses in particular on the timescale that

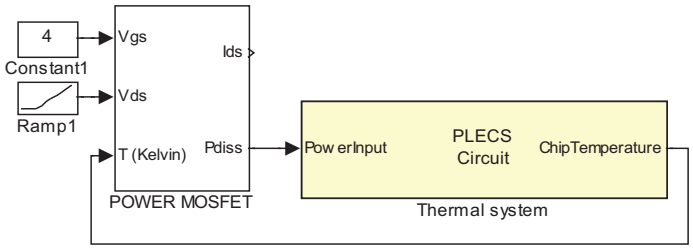


Figure 2.21 – Feedback connection of the electro-thermal MOSFET model and the LE thermal network, respectively.

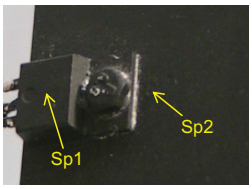


Figure 2.22 – A photograph of the device mounted on the heatsink, with indication of the two sensing points S_{p1}, S_{p2}.

the model is able to describe, ranging from microseconds up to tens of minutes, thus covering all the transient thermal dynamics which are present in the whole assembly. In this study, the temperature transient evolution has been captured by means of an IR thermocamera, using two sensing points S_{p1}, S_{p2} in order to tune the model, as shown in Figure 2.22. To uniform the emissivity coefficient on the overall surface, the whole assembly was painted black. The grid used to mesh the system is more accurate and geometrically regular then the one used in the static case, although using a limited number of elements, as it can be seen in Figure 2.23. Especially in the region close to the silicon die, the thermal network is sensibly different from the one shown in Figure 2.18. Here, the silicon die is modeled with a three-dimensional block in which the power is entering, rather than a single node as in the former thermal network. The building blocks are assembled as shown in Figure 2.24.

Bench for thermal transient measurements. As said before, thermal measurements were performed by means of a FLIR A320G IR thermocamera, capable of capturing videos with 320×240 pixels resolution with a frame rate up to 60 frames per second (fps). For our purposes, 9 fps were found to be enough. Thanks to the black painting, the emissivity of all the surfaces can be considered equal to 1. The ambient temperature was 26 °C, without forced ventilation. The silicon die temperature was measured

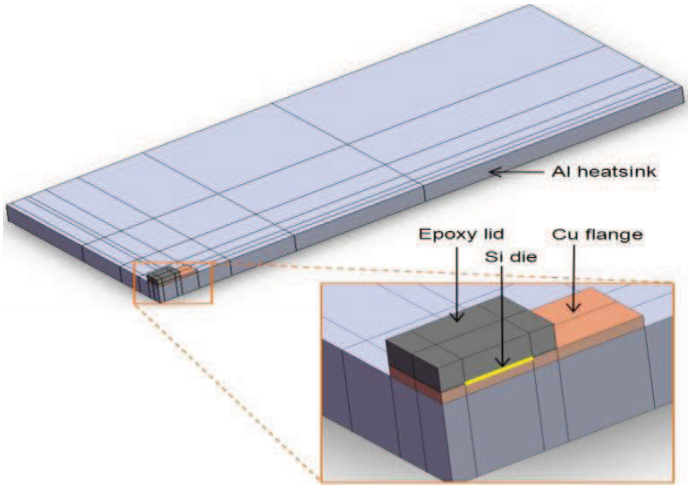


Figure 2.23 – The modeled test structure, with a blow-up of the device. The grid used to mesh the assembly is shown. Different colors correspond to different domains.

by an indirect method, using the $V_{DS} = V_{GS}$ as thermo-sensitive parameter (this method has been described in the previous chapter). The device is operated with a drain current high enough to self-heat the device (in our case, a drain current of 4.1 A was supplied to the device, with a dissipated power of 21 W). Starting from thermal equilibrium (the whole assembly is at the ambient temperature), the steady state is reached after roughly one hour, and then only a small current (1 mA) is supplied to the device by a HP4145 parameter analyzer. Then, V_{GS} is monitored at a sampling time of 1 s, allowing the HP4145 to store up to 1024 points. The IR thermocamera measures the temperature evolution of the two sensing points S_{p1} and S_{p2}: the sample frequency (9 fps) is high enough related to the time constants of the temperatures of the two sensing points.

The fastest transient in the assembly is that of the silicon die: it cannot be monitored by the IR camera, obviously, and neither from the HP4145, due to the low sample frequency. To this extent, two oscilloscopes were used in order to capture the transient behavior of V_{GS} (i.e., the die temperature) on different time scales: the first scope was set with 200 ns time division and 500 mV/div vertical resolution, in order to capture the first phase of the thermal response; the second scope was set with 1 ms time division and vertical resolution of 50 mV/div, in order to capture up to 2 s of $V_{GS}(t)$. The remaining slow part of the transient is captured by the HP4145. This allows to observe the time evolution of the die temperature on different time scales. The instruments arrangement is depicted in Figure 2.25.

Comparison between measurements and simulation. In the LE model we used 5 fitting parameters: (i) the convection coefficient at the bottom surface of the heat-sink, $39\,\mathrm{W/(m^2K)}$, (ii) the convection coefficient on all the other surfaces of the assembly, equal to $7\,\mathrm{W/(m^2K)}$, (iii) the thermal contact resistance (TCR) between the silicon die and the epoxy lid, $TCR_{\mathrm{Si-Lid}} = 10^{-6}\mathrm{m^2K/W}$, (iv) the TCR between silicon die and drain flange, $TCR_{\mathrm{Si-Flange}} = 2.6 \times 10^{-5}\mathrm{m^2K/W}$, and (v) the TCR between the drain flange and the heat-sink, $TCR_{\mathrm{Flange-Heatsink}} = 1.5 \times 10^{-5}\mathrm{m^2K/W}$ (these resistances were set as resistances per unit area). All the corresponding lumped elements are visible in the thermal network shown in Figure 2.24. The values of these parameters were found on the basis of the measured temperature of the two test points S_{p1} and S_{p2} and that of the silicon die. The comparison between measured temperatures and modeled temperatures is shown in Figure 2.26 and in Figure 2.27.

The model can be further employed to describe the behavior of the assembly when subjected to a given power cycle. As an example, Figure 2.28 shows the simulated behavior of the test structure when subjected to a train of 45 W pulses, with a period of 360 seconds and 50% duty-cycle; the transient response of the temperature of the three points is plotted versus time.

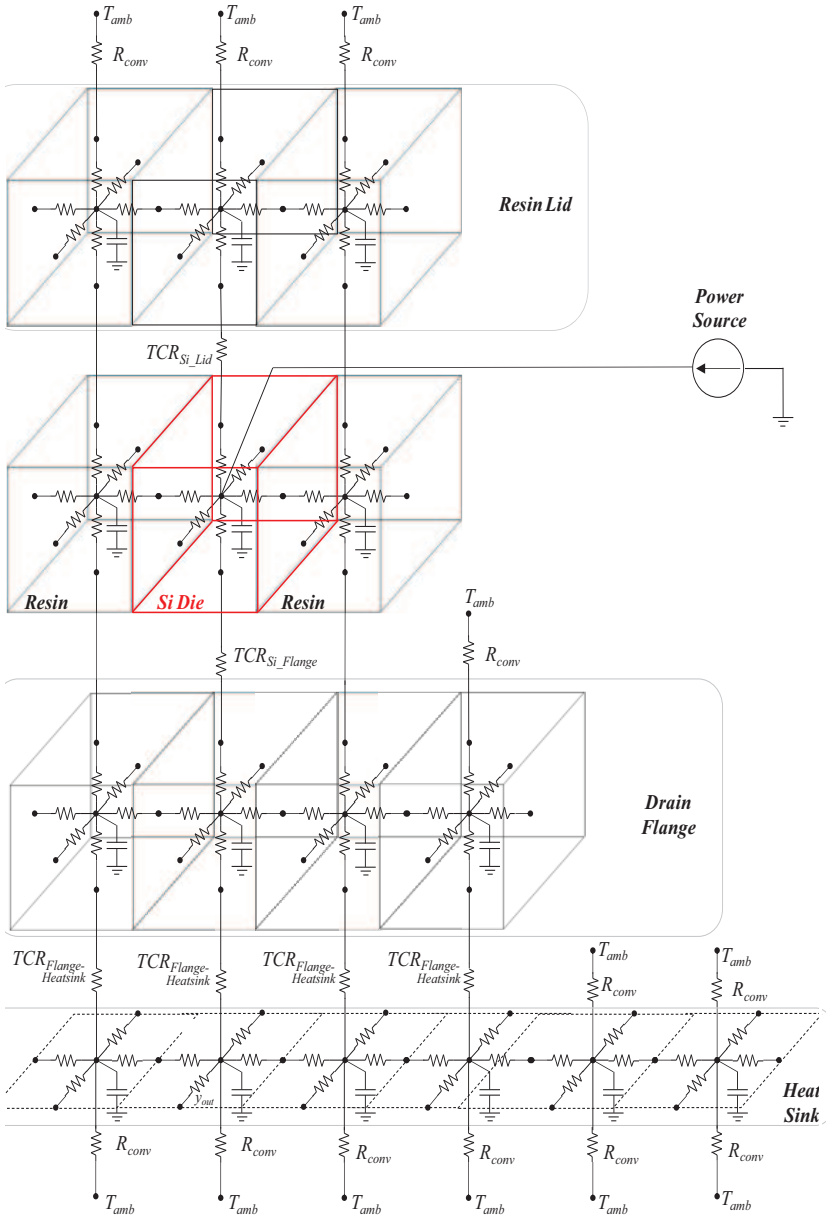


Figure 2.24 – A vertical slice of the *RC* lumped-element thermal network modeling the assembly shown in Figure 2.15. Each capacitance is connected between the block central node and thermal ground.

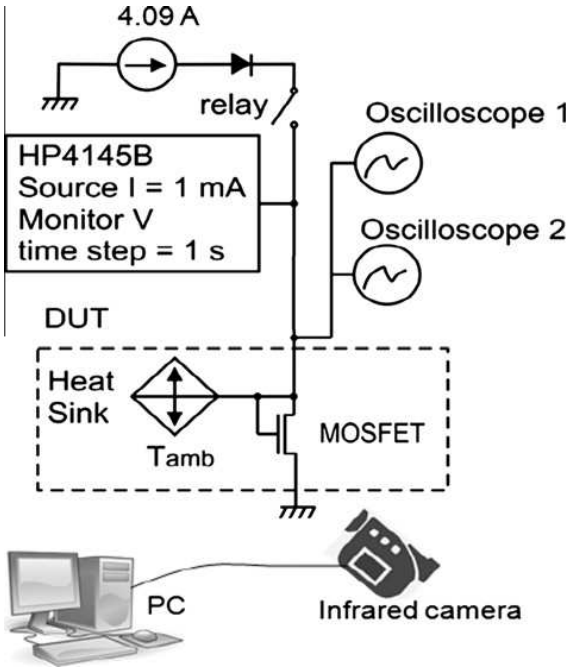


Figure 2.25 – Transient thermal measurement setup.

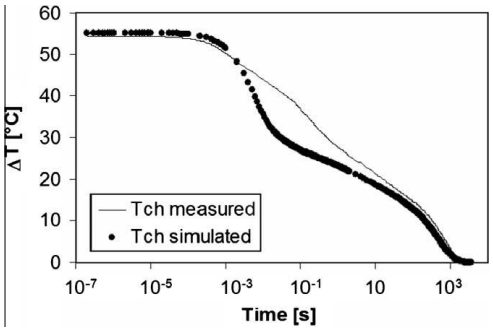


Figure 2.26 – Comparison between measured and simulated T_{CH} temperature increase transients.

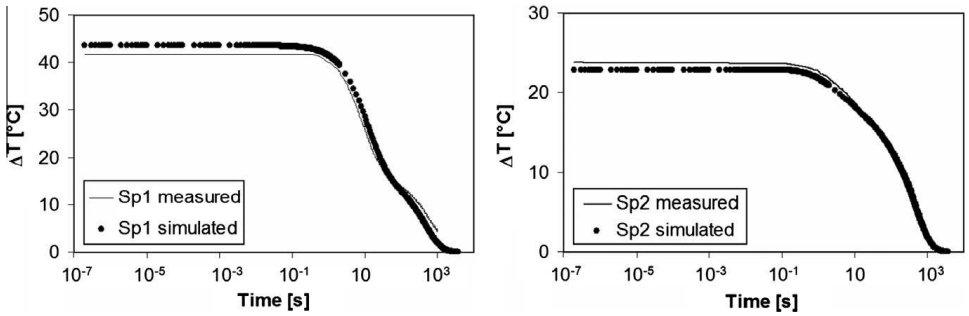


Figure 2.27 – Comparison between measured and simulated temperatures increase transients. S_{p1} temperature (left), S_{p2} temperature (right).

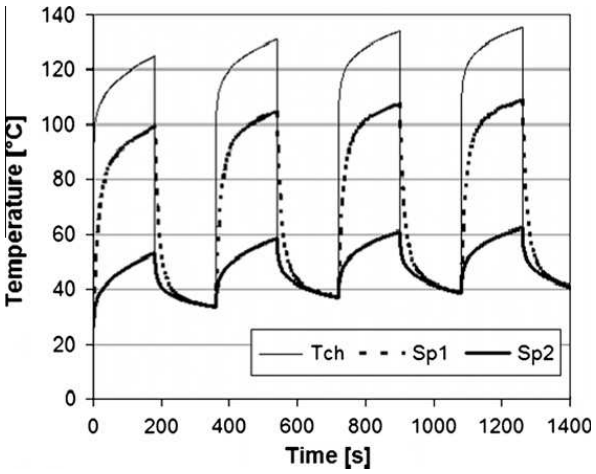


Figure 2.28 – Simulated temperatures of S_{p1}, S_{p2} and silicon die during a train of pulses with $P_D = 45\,\mathrm{W}$, $T = 360\,\mathrm{s}$, $d = 50\%$. Ambient temperature is $26\,^{\circ}\mathrm{C}$.

2.5.2 Thermal modeling of HEMTs

A similar approach has also been applied to thermal modeling of HEMT devices. In this case, the thermal dynamics is important because, although it is well below the range of microwave frequencies, it falls inside the band of the envelope frequency of complex telecommunication signals, and the thermal conductivies of the materials in the structure are strongly dependent on temperature as well: these devices work at high power density, and thus large temperature increases are expected. The dependence of thermal conductivity on temperature has to be accounted for, otherwise very inaccurate results will be obtained.

Several works have been published about this topic [17, 18, 19], in which the complexity of the thermal modeling progressively increased up to a full three-dimensional, non-linear LE model. This is the level on which this section will be focused on. In a chronological order, the first coupling of a large-signal, electro-thermal model of an AlGaN/GaN HEMT with a physical LE network which describes the thermal behavior of the device can be found in [17]. In this case, the LE network is two-dimensional, a satisfactory approximation if the heating fingers are almost as wide as the whole die. Figure 2.29 shows the two-dimensional thermal network used together with the large-signal, electro-thermal model of the HEMT. Each thermal contact injects a current which is numerically equal to the *power density* dissipated by each finger, and the resistances in the network are *specific* resistances.

One can write

$$\Delta T = P \times R_\theta = P \times \frac{1}{k} \cdot \frac{\ell}{S} = \frac{P}{S} \times \frac{\ell}{k} \tag{2.34}$$

where ℓ, S are the length and the cross-sectional area of a slice of material in which a power P is flowing through, giving rise to a temperature drop across the slice equal to ΔT. Thus, with the power density expressed in [W/m], the thermal resistances are expressed in [K · m/W]. The thermal capacitances are computed by multiplying the specific mass, the specific heat and the area of the cell they are describing. The most important features of the model [17] are:

1. *no fitting parameters* are involved in the thermal network: only the physical dimensions and physical properties of materials are used to determine the value of the elements;

2. a complete large-signal, electro-thermal model is coupled with a physical LE thermal network, while the usual Foster or Cauer networks with a much smaller number of elements are unrelated with the physical structure of the device;

3. the model allows to describe both the electrical and thermal *dynamic* behavior of the system, highlighting the temperature effects on traps, thermal non linearities, the presence of thermal hysteresis effects, and so on.

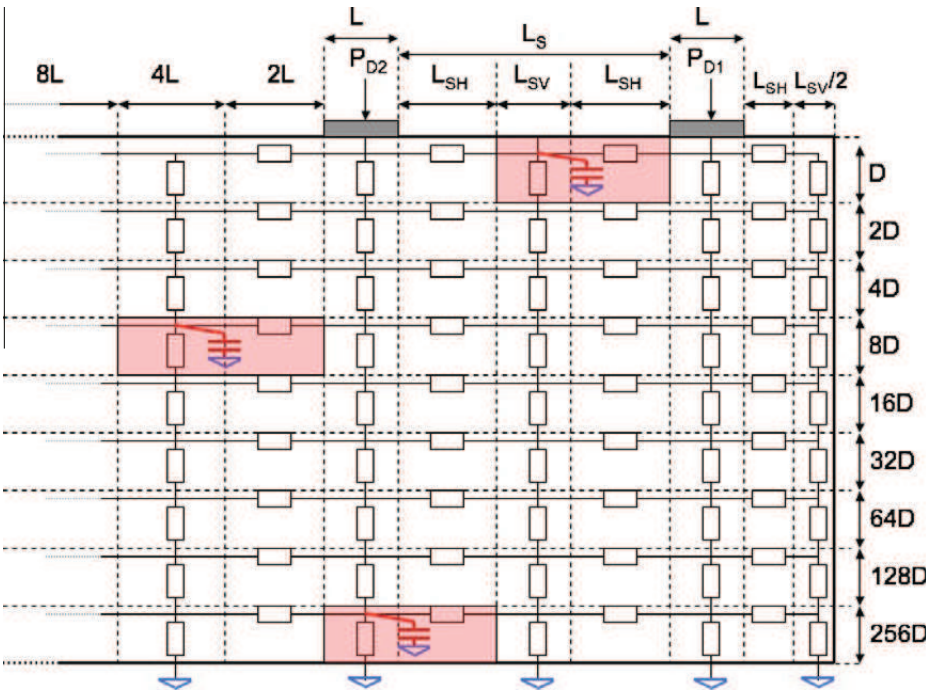


Figure 2.29 – Two-dimensional dynamic thermal network for an AlGaN/GaN HEMT [17]. Each node features a capacitance connected to ground, only three are shown for the sake of clarity.

Modeling results

The large-signal, electro-thermal HEMT model has been previously described, however, the model used here features some new characteristics. In particular, the drain current generator (see Figure 2.8) is the core of the DC part of the model: the voltage and temperature dependence have been improved by including a novel function for the dependence of i_{DS} on v_{in}, the coefficients of which are quadratic functions of channel temperature T_{CH}. The drain current is defined as:

$$
\begin{aligned}
&i_{\mathrm{DS}} = 0 \quad \text{for} \quad v_{\mathrm{in}} \leq V_{\mathrm{TH}} + V_{\mathrm{T0}} \\
&i_{\mathrm{DS}} = \left(A_0(v_{\mathrm{in}} - V_{\mathrm{TH}} - V_{\mathrm{T0}})^2 + A_1(v_{\mathrm{in}} - V_{\mathrm{TH}} - V_{\mathrm{T0}})^3\right) \cdot \\
&\quad \tanh\left((g_1(1 + g_{1t}T_{\mathrm{CH}} + g_{2t}T_{\mathrm{CH}}^2) + \right. \\
&\quad \left. g_{1d}(v_{\mathrm{in}} - V_{\mathrm{TH}} - V_{\mathrm{T0}})) \cdot v_{out}\right) \quad \text{for} \quad v_{\mathrm{in}} > V_{\mathrm{TH}} + V_{\mathrm{T0}} \\
&A_i = A_{i0} + A_{i1}T_{\mathrm{CH}} + A_{i2}T_{\mathrm{CH}}^2 \quad i = (0, 1)
\end{aligned}
$$

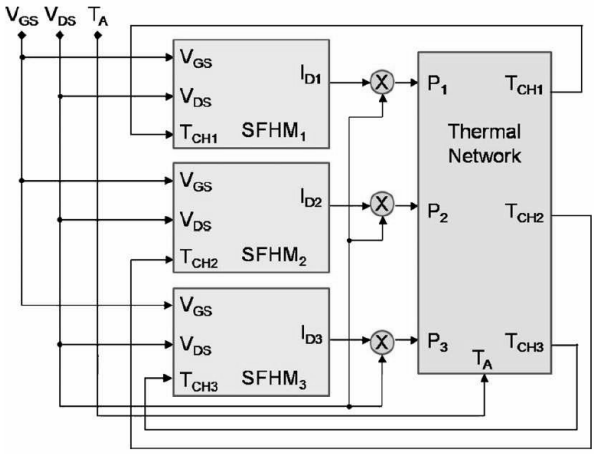


Figure 2.30 – The electro-thermal feedback loop. Each SFHM block is a single-finger HEMT model like the one in Figure 2.8.

where V_{TH} is the threshold voltage in the absence of traps, V_{T0} is the DC voltage across the capacitor C_T, corresponding to the DC occupancy of traps. Under dynamic conditions, V_T differs from V_{T0}: the latter instantaneously follows the applied bias, the former instead is bound to respond according to the trap characteristic frequency.

The channel temperature T_{CH} is the output of the lumped element thermal network of Figure 2.29, and an input to the i_{DS} generator, the blocks computing $R_S(T_{\mathrm{CH}})$ and $R_D(T_{\mathrm{CH}})$ based on measured data, and the trap model. The feedback loop connection is shown in Figure 2.30, in which each Single Finger HEMT Model (SFHM) like the one shown in Figure 2.8. These block are biased in parallel but work at different temperatures, computed in a self-consistent way through a thermal network like the one shown in Figure 2.29. Three fingers (i.e., half of a six-fingers device) have been considered here, but the model can be used with an arbitrary number of fingers. Single-finger 150 µm-wide AlGaN/GaN HEMT on sapphire have been manufactured (at University of California, Santa Barbara), measured and modeled at 200 K, 300 K and 400 K. The thermal conductivities used for sapphire are $70\,\mathrm{W/(m\cdot K)} \rightarrow (200\,\mathrm{K})$, $45\,\mathrm{W/(m\cdot K)} \rightarrow (300\,\mathrm{K})$, and $37\,\mathrm{W/(m\cdot K)} \rightarrow (400\,\mathrm{K})$; for GaN we use $282\,\mathrm{W/(m\cdot K)} \rightarrow (200\,\mathrm{K})$, $160\,\mathrm{W/(m\cdot K)} \rightarrow (300\,\mathrm{K})$, and $107\,\mathrm{W/(m\cdot K)} \rightarrow (400\,\mathrm{K})$. The temperature-independent specific heat capacitance is $2.56 \times 10^6\,\mathrm{J/(m^3\cdot K)}$ for sapphire and $3.00 \times 10^6\,\mathrm{J/(m^3\cdot K)}$ for GaN.

Referring to Figure 2.31 and Figure 2.32, the modeling results are described. The DC characteristics of the considered devices are shown in Figure 2.31-(a), (b), (c) at temperatures of 200 K, 300 K and 400 K, respectively. V_{GS} is swept from -4 V up to

0 V by steps of 0.5 V. The match is excellent in over the whole bias and temperature range.

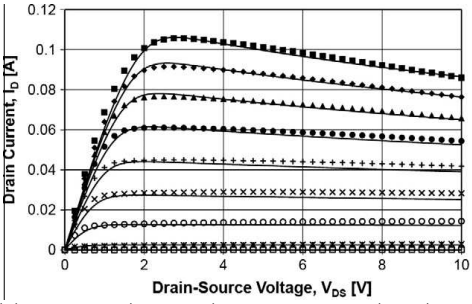
A modeled gate lag example showing the typical high-pass behavior of bulk traps is shown in Figure 2.31-(d), with and without the inclusion of the self-heating in the model. The drain current features an overshoot followed by a decaying transient due to the bulk donors being frozen in their positively charged state during the gate turn-on front, then slowly capturing electrons and tending to equilibrium according to the trap time constant. The modification of the trap dynamics due to the self-heating effect is clearly visible: the trap capture time is shortened, according to Eqn. (2.13) and Eqn. (2.14). The results of the same experiment in the case of a surface donor trap is shown in Figure 2.31-(e). The gate lag now displays the well-known low-pass behavior of surface traps, more clearly visible in the absence of self-heating. In the presence of self-heating, the trap emission time is remarkably shorter, and the current more rapidly tends to the steady-state value, until self-heating sets in about 100 µs after the pulse leading edge, and the current starts decreasing. It is worth pointing out the inaccurate trap dynamics in the case the self-heating is neglected.

In Figure 2.31-(f), the high pass-behavior of the bulk donor trap is shown on the transconductance frequency dispersion. Note the increase of the transition frequency (i.e., the decrease of the trap capture time) with increasing temperature. Figure 2.32-(a) shows results of the same modeled experiment in the case of a surface donor. The transconductance frequency behavior now shows the well-known low-pass behavior of surface traps. The difference between Continuous Wave (CW) and pulsed operation of GaN-based HEMTs is among the most relevant effects of traps and self-heating dynamics. Figure 2.32-(b) shows a comparison between DC and pulsed output curves, in the presence of the surface donor trap described above. The pulsed characteristics are obtained by applying 1 µs-long pulses from the quiescent bias point with $V_{GS0} = -4\,V$, $V_{DS0} = 8\,V$ (class B operation). The pulse duration is far less than the trap response time, and than the time required for appreciable self-heating, so the pulsed curves in Figure 2.32-(b) represent the HEMT output characteristic under isothermal conditions ($T_{CH} = 300\,K$) and with the donor occupancy frozen in its DC state ($V_{GS0} = -4\,V$, $V_{DS0} = 8\,V$), with the surface states occupied by electrons coming from the negatively charged gate.

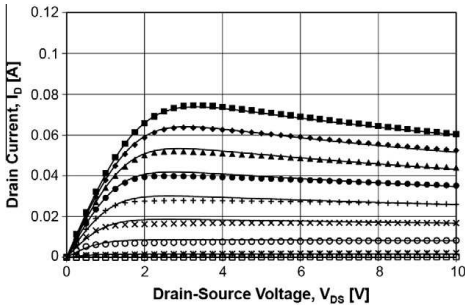
Another known consequence of dispersive effects is the observation of hysteresis during bias voltage and input power sweeps, and memory effects in general. This is illustrated for the modeled GaN HEMT by the example shown in Figure 2.32-(c), where the backward voltage sweep gives a different current from the preceding forward sweep. In this case, being the sweep duration 10 µs, the presence of a surface donor trap delays the current response, so that the backward sweep retains memory of the preceding device states.

Either or both thermal and trap-related memory effects can be significant, de-

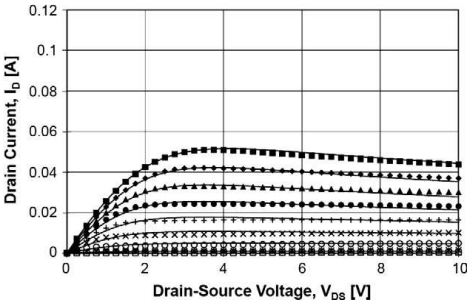
pending on the specific device features and the time scale. Figure 2.32-(d) shows the maximum normalized drain current excursion in the presence of hysteresis cycles such as that of Figure 2.32-(c). The maximum normalized drain current excursion is defined as the maximum difference between the forward sweep and the backward sweep, normalized to the final value of the forward sweep, and it is plotted versus the sweep duration. Sweeps below 1 µs are too fast for either traps or self-heating to respond, and thus the current excursion is zero; for sweeps in the tens to hundreds of µs range, trap dynamics is dominant, while self-heating is still marginal, and negative excursion are observed; for sweeps in the ms to to hundreds of ms range, the surface donor is practically in steady-state, and the slower self-heating dynamics is responsible for the hysteresis: the current excursion is now positive because during the backward sweep the device is hotter, thus lowering the current.



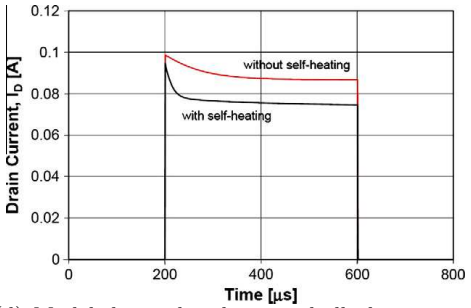
(a) Modeled (symbols) and measured (lines) DC characteristics at 200 K. $V_{GS} = -4$ to 0 V with 0.5 V steps.



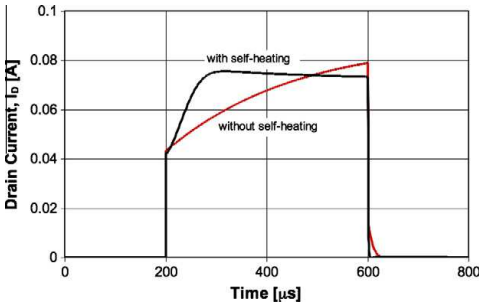
(b) Modeled (symbols) and measured (lines) DC characteristics at 300 K. $V_{GS} = -4$ to 0 V with 0.5 V steps.



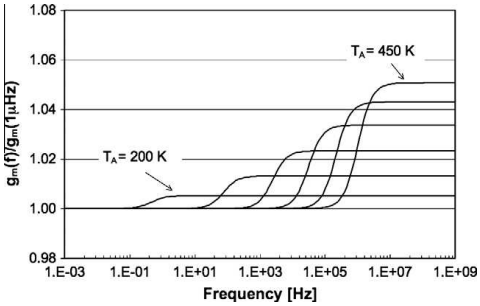
(c) Modeled (symbols) and measured (lines) DC characteristics at 400 K. $V_{GS} = -4$ to 0 V with 0.5 V steps.



(d) Modeled gate lag due to a bulk donor trap, with and without self-heating. V_{GS} is pulsed from -4 V to 0 V and then back to -4 V. $V_{DS} = 6$ V.

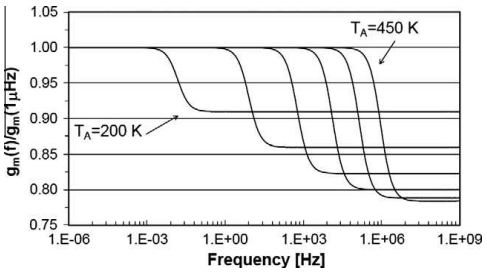


(e) Modeled gate lag due to a surface donor trap, with and without self-heating. V_{GS} is pulsed from -4 V to 0 V and then back to -4 V. $V_{DS} = 6$ V.

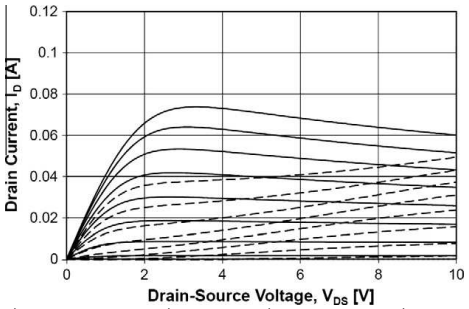


(f) Modeled transconductance frequency dispersion due to a bulk donor trap. Ambient temperature ranges from 200 K to 450 K in steps of 50 K. $V_{GS} = 0$ V, $V_{DS} = 4$ V.

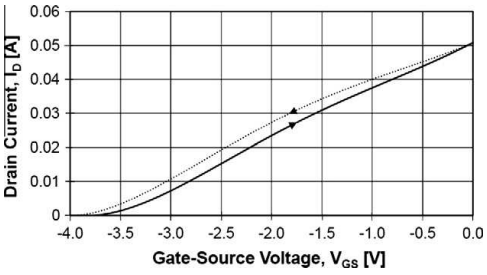
Figure 2.31 – Modeling results for a single-finger 150 µm-wide AlGaN/GaN HEMT on sapphire – part 1. Where not specified, ambient temperature is 300 K.



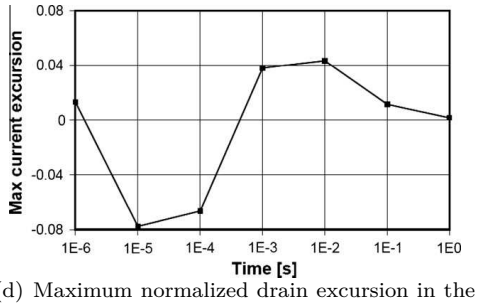
(a) Modeled transconductance frequency dispersion due to a surface donor trap. The ambient temperatures ranges from 200 K to 450 K in steps of 50 K. $V_{\mathrm{GS}} = 0\,\mathrm{V}$, $V_{\mathrm{DS}} = 4\,\mathrm{V}$.



(b) Modeled DC (solid lines) and pulsed (dashed lines) characteristics. 1 µs-long pulses have been applied from a quiescent bias point with $V_{\mathrm{GS0}} = -4\,\mathrm{V}$, $V_{\mathrm{DS0}} = 8\,\mathrm{V}$. $V_{\mathrm{GS}} = -4\,\mathrm{V}$ to 0 V, with 0.5 V steps. Ambient temperature is 300 K.



(c) Modeled drain current during a forward voltage sweep (solid line), immediately followed by a backward sweep (dashed line). The sweep duration is 10 µs. $V_{\mathrm{DS}} = 10\,\mathrm{V}$, $T_A = 300\,\mathrm{K}$.



(d) Maximum normalized drain excursion in the hysteresis cycle, as a function of the sweep duration, in the presence of a surface donor trap. $V_{\mathrm{DS}} = 10\,\mathrm{V}$, $T_A = 300\,\mathrm{K}$.

Figure 2.32 – Modeling results for a single-finger 150 µm-wide AlGaN/GaN HEMT on sapphire – part 2.

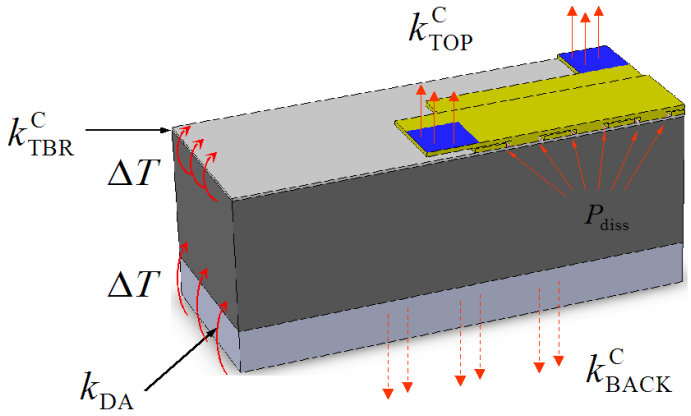


Figure 2.33 – View of the analyzed structure. Different materials are indicated by different colors. From the top: GaN, substrate (silicon or silicon carbide depending on different cases), and die-attach. Gold metallizations with blue pads for application of boundary conditions are also shown. The parameters used to model the boundary conditions are shown.

Further developments have been (i) a three dimensional description of self-heating and (ii) the inclusion of the temperature dependence of thermal conductivities.

First of all, it is useful to have a benchmark model with which to check the LE model. The development of a three-dimensional, non-linear (i.e., featuring temperature-dependent thermal conductivity) FEM model has been done [18]. Figure 2.33 depicts the structure that has been analyzed, with particular emphasis on the effect of the boundary conditions on the thermal behavior of HEMTs: different substrates, heat removal from the top contacts and from the backside of the device have been extensively analyzed. Also the influence of the die-attach material thermal conductivity has been taken into account. The Thermal Boundary Resistance (TBR), expressed in [m^2K/W], which is basically an interfacial resistance between the active layer (GaN) and the substrate (Si or SiC), has been taken into account; this parameter is often overlooked in thermal modeling. This resistance basically arises from imperfections – like lattice mismatch, for instance – between different materials, phonon scattering, and it can be modeled as a very thin layer (50 nm) with a finite thermal conductivity. For instance, a typical value for TBR between GaN and Si/SiC has been found to be $3.3 \times 10^{-8}\,m^2K/W$ [25]. Detailed results of this analysis are shown here.

We analyze a HEMT made of cells of 12 fingers each; each finger is 150 µm wide, and the separation between adjacent fingers is 30 µm. The cell's gate periphery is therefore 1.8 mm, but thanks to symmetry planes (which can be replaced by adiabatic

boundary conditions), only half of the fingers and half of the finger width need to be modeled. For the same reasons, a symmetric arrangement of cells allows to limit the study to a single cell. This work considers the two relevant cases of SiC and Si substrate. The simulated 12-finger cell is shown in Figure 2.33. It has a 2.5 µm-thick GaN layer on top of the substrate; we considered two values for the substrate thickness: 250 µm and 125 µm. We include in our simulations the often neglected TBR between GaN and substrate, as a 50 nm-thick interfacial layer with contact thermal conductivity k^C_{TBR}. The back of the substrate is stuck to a 40 µm-thick $Sn_{96}Ag_4$ die-attach layer with thermal conductivity k_{DA}. Top-side source and drain metals (another often overlooked feature that significantly impacts the thermal budget) are 4 µm-thick gold. Boundary conditions are adiabatic everywhere, except on the back of the die-attach and on the top metal pads, where contact conductivities, k^C_{BACK} and k^C_{TOP}, respectively, are varied to simulate different combinations between isothermal ($k^C = \infty$) and adiabatic ($k^C = 0$) conditions. GaN, SiC, and Si thermal conductivities are temperature-dependent.

DC simulation results. Figure 2.34 shows the temperature profile along a vertical line originating in the center of the hottest (innermost) finger, for the two SiC thicknesses and for two backside boundary conditions: the ideal isothermal case ($k^C_{\mathrm{BACK}} = 0$), and a realistic case where $k^C_{\mathrm{BACK}} = 3.6 \times 10^5\,\mathrm{WK^{-1}m^{-2}}$, corresponding to a case-to-ambient thermal resistance of 10 K/W for the 12-finger cell. The dissipated power is 5.4 W for the 1.8 mm cell (i.e., $P_D = 3\,\mathrm{W/mm}$); there is no heat exchange from the front-side pads ($k^C_{\mathrm{TOP}} = 0$). Note the sharp drop on the TBR ($k^C_{\mathrm{TBR}} = 3 \times 10^7\,\mathrm{W/(m^2K)}$) and the impact of the die-attach ($k_{\mathrm{DA}} = 45\,\mathrm{W/(m \cdot K)}$). The same analysis has been performed for a substrate made of Si. In the case of SiC, the substrate thickness plays a minor role in the self-heating process, being the thermal conductivity of SiC high; in the case of a Si substrate (which has a lower thermal conductivity with respect to SiC), instead, temperatures are higher and the substrate thickness makes a bigger difference. The effect of heat removal from the front-side metal pads is illustrated by Figure 2.35, for both SiC substrate thicknesses, in the case of $k^C_{\mathrm{BACK}} = 3.6 \times 10^5\,\mathrm{Wm^{-2}K^{-1}}$, $k^C_{\mathrm{TBR}} = 3.0 \times 10^7\,\mathrm{Wm^{-2}K^{-1}}$, and $k_{\mathrm{DA}} = 45\,\mathrm{Wm^{-1}K^{-1}}$. Moving from the case of adiabatic top $k^C_{\mathrm{TOP}} = 0$ to better heat removal from the top-side pads, the peak channel temperature $T_{\max}$ is drastically reduced.

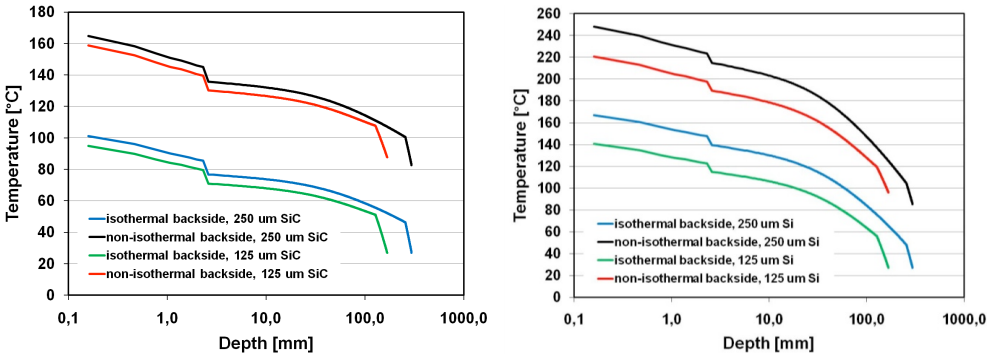


Figure 2.34 – Temperature profiles under the hottest (innermost) finger, for different SiC (left) and Si (right) substrates thicknesses and for ideal isothermal (300 K), and non isothermal ($k^C_{\mathrm{BACK}} = 3.6 \times 10^5\,\mathrm{W/(m^2K)}$) back-side boundary conditions. $P_D = 3\,\mathrm{W/mm}$.

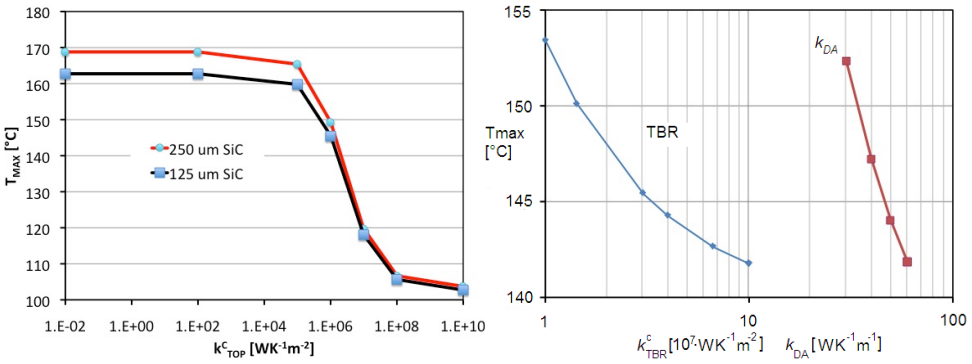


Figure 2.35 – (left) Effect of the front-side pad thermal conductivity on the peak channel temperature. The substrate is SiC, and the back-side is non-isothermal ($k^C_{\mathrm{BACK}} = 3.6 \times 10^5\,\mathrm{W/(m^2K)}$). (right) Effect of the GaN/SiC TBR (left curve, with $k^C_{\mathrm{TBR}} = 3 \times 10^7\,\mathrm{W/(m^2K)}$) and die attach thermal conductivity (right curve, with $k^C_{\mathrm{TBR}} = 3 \times 10^7\,\mathrm{W/(m^2K)}$) on the peak channel temperature, with $k^C_{\mathrm{TOP}} = 10^6\,\mathrm{W/(m^2K)}$, and $t_{\mathrm{SiC}} = 125\,\mu\mathrm{m}$. All data obtained for $P_D = 3\,\mathrm{W/mm}$ and $k^C_{\mathrm{BACK}} = 3.6 \times 10^5\,\mathrm{W/(m^2K)}$.

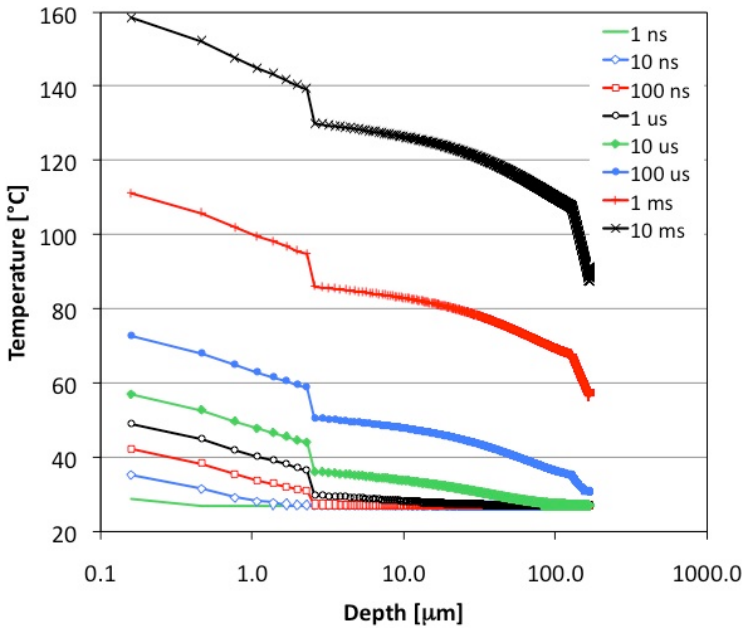


Figure 2.36 – Temperature profiles under the hottest (innermost) finger, for 125 µm-thick SiC substrate, at various times after the application of a power step ($P_D = 3\,\mathrm{W/mm}$). $k^C_{\mathrm{BACK}} = 3.6 \times 10^5\,\mathrm{W/(m^2K)}$, $k^C_{\mathrm{TOP}} = 0$, $k^C_{\mathrm{TBR}} = 3 \times 10^7\,\mathrm{W/(m^2K)}$, $k_{\mathrm{DA}} = 45\,\mathrm{W/(m \cdot K)}$.

Dynamic simulation results. The temperature profile along a vertical line originating in the center of the hottest (innermost) finger, for 125 µm-thick SiC substrate, at various times after the application of a power step, is shown in Figure 2.36.

A few facts are worth commenting upon. (i) The temperature of the active area starts rising 1 ns after the power step; after 10 ns it has increased by some 10 °C; these times can be expected to scale roughly with the gate periphery. (ii) 1 µs after the application of the power step, the heat front reaches the SiC substrate; up to this point in time (i.e., for very short-pulsed operation), the substrate material, whether SiC, or Si – or even sapphire for that matter – would not make a difference. (iii) The backside temperature starts increasing significantly some 100 µs after the power step; this indicates, for example, that in the case of pulses shorter than 100 µs with small duty cycle, heat removal from the backside is ineffective. (iv) Steady-state conditions are reached at about 10 ms.

The FEM model has shown, once more, to be a very powerful tool to study the

influence of several parameters on complex structures, both in steady state and during transients. However, the problem of this approach is that the FE model cannot be embedded into circuit simulators. A model that can be included into circuit-level CAD tools but that at the same time retains the physical information of the FEM model must therefore feature:

1. a three-dimensional lumped-element model of heat flow utilizing three-dimensional thermal networks, as previously illustrated;

2. inclusion of the non-linearity due to temperature dependence of thermal conductivities.

As far as point 2 is concerned, we know that temperature is, in the lumped-element network, a voltage, and thus a temperature-dependent resistor is actually a voltage-controlled resistor. More precisely, a resistor is a voltage-controlled current-source (VCCS), since Ohm's law states that:

$$I = V \cdot \frac{1}{R}, \tag{2.35}$$

which can be read as a current generator whose current is proportional to the applied voltage. Clearly, this is the simplest case of VCCS, but a resistance can always be defined in this way. The non-linearity in thermal resistance is due to the non-linearity in thermal conductivity. For instance, the relationships for thermal conductivities of GaN, Silicon, Silicon Carbide are of the form:

$$k(T) = k_0 \cdot \left(\frac{300}{T}\right)^{\alpha} \tag{2.36}$$

where T is the absolute temperature in [K], k_0 [W/(m·K)] is the thermal conductivity at $T = 300\,\mathrm{K}$, and $\alpha \geq 1$. Values (k_0, α) for GaN, Silicon and Silicon Carbide are $(160, 1.4)$, $(148, 1.3)$ and $(400, 1)$, respectively. It is useful to develop these expressions as polynomial series in order to improve the convergence of the simulator (with a least-square interpolation, for instance).

The inclusion of the non-linearity within the LE thermal networks can be done thanks to the configuration depicted in Figure 2.37. This network has been successfully implemented in the Advanced Design System (ADS) environment, from Agilent Technologies, which features a very powerful simulation engine (able to handle strong non-linearities) together with a useful user-defined device (basically, a VCCS that can be controlled by several voltages) that has been used in order to implement the resistor as in Figure 2.37. In ADS, these elements are called Symbolic Defined Devices (SDDs). The schematic of an SSD is shown in Figure 2.37. A two-port device is needed, since the first port implements the relationship $V = R_\theta \cdot I$ (i.e., $\Delta T = R_\theta \cdot P$),

while the second port is used only to sample the voltage in the thermal network (i.e., the temperature) that modulates the conductivity of the thermal resistance R_θ. For an n-port device, it is necessary to define n current-voltage port relationship, and each port current can be made dependent on all port voltages: with n currents $I_1, I_2, \ldots, I_n$ and n voltages $V_1, V_2, \ldots, V_n$, one can write

$$I_1 = f_1(V_1, \ldots, V_n), \quad I_2 = f_2(V_1, \ldots, V_n), \ldots \quad I_n = f_n(V_1, \ldots, V_n); \tag{2.37}$$

clearly, this is the most general case. In the case of a two-port SDD, one can write:

$$I_1 = f(V_1, V_2) = \frac{V_1}{R_\theta(V_2)}, \quad I_2 = 0; \tag{2.38}$$

it is worth commenting the above relationship. Port 1 is the port the terminals of which implement the electrical connections with other SDD blocks and with the electro-thermal model of the electron device, i.e., the *physical* connection in the thermal network. Port 2, instead, is just a *control* port, in fact the relationship $I_2 = 0$ means that no current is sunk by port 2, and only the voltage V_2 is sampled. Therefore, the use of port 2 is just that of sampling the voltage (i.e., temperature) in the thermal network on which the thermal resistance is depending. The above relationship can be thus rewritten as follows, where P is the power entering the thermal resistance R_θ, ΔT is the temperature drop across the thermal resistance, and T_{control} is the temperature that modulates the value of R_θ:

$$P = \frac{\Delta T}{R_\theta(T_{\text{control}})} = \frac{T_2^{(1)} - T_1^{(1)}}{R_\theta(T_{\text{control}})}. \tag{2.39}$$

where $T_1^{(1)}$ and $T_2^{(1)}$ are the absolute values of temperature at node 1 and node 2 for the considered thermal resistance, respectively. In order to make T_{control} modulate the value of R_θ, the voltage corresponding to the control temperature has to be wired to the control port, that is port 2. A two-dimensional building block can illustrate the application better. Four non-linear resistances, i.e., four SDDs are needed to build the block. Let the central node temperature be considered as the control temperature on which all four resistances depend. It means that, for each SDD, port 2 has to be wired to the central node voltage. Clearly, the SDD shown in Figure 2.37 models only *one* resistance: in order to describe two-dimensional and three-dimensional building blocks, four and six SDDs need to be used, respectively.

An example of *how* to implement a two-dimensional block is shown in Figure 2.37. In this case, we need four SDDs, and the control temperature T_{control} has been chosen as that of the central node: being the resistances dependent on the *absolute* temperature, each SDD has the 2nd terminal of the port 2 connected to the thermal ground (i.e., 0 K). This is highlighted in Figure 2.37-(c) with the dashed line. The continuous

line shows the connections between the ports 1 of each SDD: these are the connections needed to build the structure equivalent to the schematic shown in Figure 2.37-(d). It is worth noting that: (i) the so called *thermal modulation connection* is the connection that makes all the thermal resistances temperature-dependent; if that connection is fixed at 300 K, for instance, the temperature dependence is neglected, and the formulation of the thermal resistance reduces to $R_\theta(300)$; (ii) when a large-signal, electro-thermal device model is self-consistently coupled with a non-linear lumped-element thermal network, two feedback loops form in the model: the first links the large-signal, electro-thermal device model and the non-linear lumped-element thermal network, the second is the innermost loop inside each lumped-element thermal block, since the temperature of the central block depends on the resistance value and, in turn, the resistance value depends on the temperature of the central block.

In order to check the capability of the LE method, a benchmark structure has been developed and modeled by FEM and, once the corresponding LE model has been developed, the two models have been compared [19]. We modeled a $2 \times 150\,\mu\text{m}$ HEMT structure, which is made, from top to bottom, of: (i) 2.5 µm-thick top GaN active layer; (ii) 125 µm-thick SiC substrate; (iii) 50 µm-thick $Sn_{96}Ag_4$ die-attach layer. Thanks to symmetry, only one fourth of the structure needs to be modeled. The modeled structure is shown in Figure 2.38. This model accounts for the TBR as well, that has been modeled as 50 nm-thick layer with surface thermal conductivity $k^C_{\text{TBR}} = 3.0 \times 10^7\,\text{W}/(\text{m}^2\text{K})$. The number of gate fingers is at this stage limited to two, for simplicity. Larger number of fingers can be modeled. In this case, the back-side of the device is is assumed isothermal (300 K), but a convective boundary condition (i.e., non-isothermal) can be easily accounted for.

The studied structure is shown in Figure 2.38. It is a structure featuring two heating fingers, made of several stacked layers (GaN, SiC and die attach), with non-linear thermal conductivities. The shown structure actually models two 150 µm-wide fingers but, thanks to symmetry, only one fourth of the structure can be modeled, and so only one half of one finger is modeled. The three-dimensional lumped-element thermal network is made of building blocks like those in Figure 2.38. The temperature dependence of thermal conductivities of GaN and SiC, which introduces significant non-linearities in the Fourier equation, is accounted for in our model: the thermal conductivities used are $k(\text{GaN}) = 160 \cdot (300/T)^{1.4}\,\text{W}/(\text{m} \cdot \text{K})$, $k(\text{SiC}) = 400 \cdot (300/T)\,\text{W}/(\text{m} \cdot \text{K})$, $k(\text{DA}) = 45\,\text{W}/(\text{m} \cdot \text{K})$, T being the node temperature in [K]. The TBR is included in our model as a row of lumped resistances connecting the bottom GaN elementary blocks with the top SiC blocks. Thermal capacitances are connected between each node and thermal ground, and computed multiplying the unit-volume thermal capacitance, i.e., the product of specific heat capacity times the density of the material, by the element volume. The values of specific heat capacity we used are $470\,\text{J}/(\text{kg} \cdot \text{K})$ for GaN, $680\,\text{J}/(\text{kg} \cdot \text{K})$ for SiC, $240\,\text{J}/(\text{kg} \cdot \text{K})$ for die-attach, while densities are

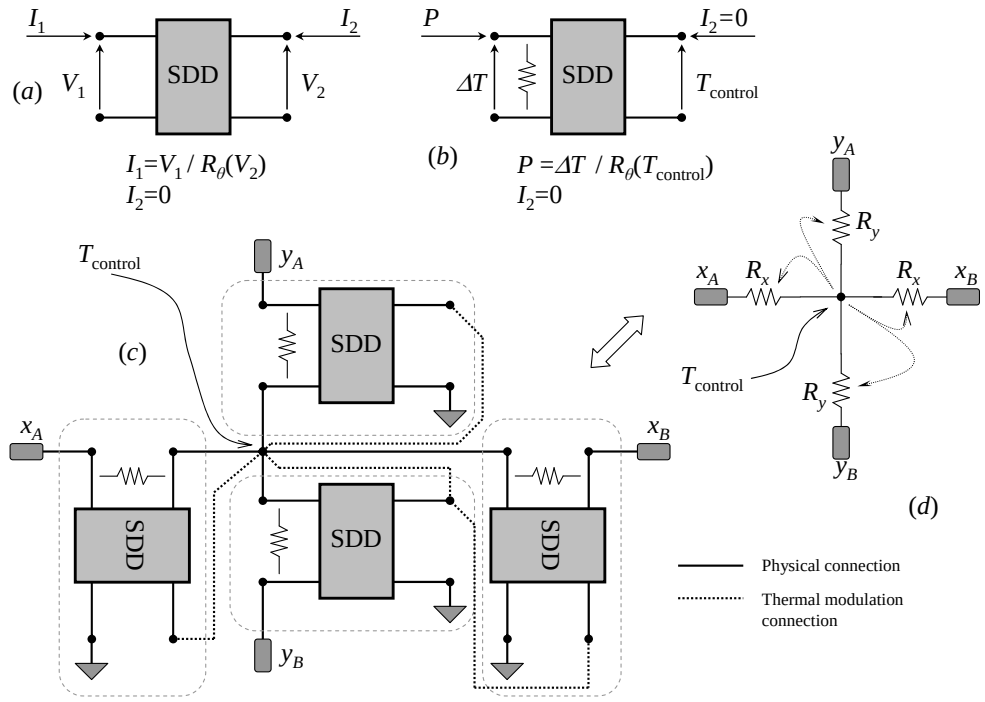


Figure 2.37 – How to use an SSD in order to implement a temperature-dependent resistance. (a) example of a two-port SSD. (b) port 1 implements Ohm's law (expressed in its thermal counterpart), while port 2 is used to sample the voltage in the thermal network corresponding to the control temperature T_{control}: this temperature modulates the thermal resistance R_θ. (c) an example of how to use four SDDs in order to describe a two-dimensional block, the schematic of which is shown in (d): the temperature of the central node modulates the four resistances in the building block.

6110 kg/m^3 for GaN, 3211 kg/m^3 for SiC, and 7330 kg/m^3 for die-attach.

The dimensions of the elements of volume in which the structure is discretized increase (exponentially) from the heating area towards the boundaries, which allows simulating large volumes with a relatively small number of elements. A total of 500 blocks was used to model the $165 \times 144 \times 177.5\,\mu\mathrm{m}^3$ structure; 50 additional thermal resistances connecting the GaN blocks with the SiC blocks simulate the TBR. The LE thermal model of the complete structure can thus be built in a totally modular fashion by instantiating the building block models, the resistance and capacitance values of which are automatically scaled based on the element dimensions, thermal conductance

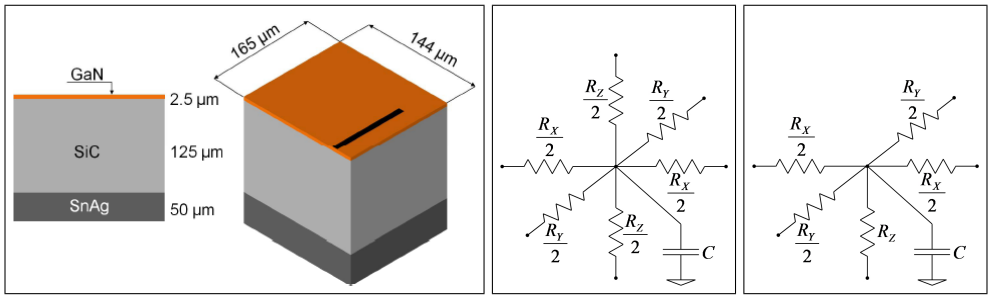


Figure 2.38 – The modeled 3D HEMT structure, not to scale (left). The structure represents a quarter of the real HEMT, because the two visible sidewalls are symmetry planes, modeled by adiabatic boundaries. The black line on the top surface represents the 75 µm-wide heating gate finger. (center, right) Building blocks of the 3D LE thermal network. The central block represents the generic element of volume, while the right block is used to model surface elements.

and specific thermal capacitance. Top contacts along the gate width inject currents equal to the dissipated powers. The value in volts of the voltage at each node gives the local temperature in [K]. This LE thermal network has been self-consistently coupled with the large-signal, electro-thermal HEMT model described in section 2.5.2. In the direction of the gate width, the 75 µm-wide heating finger (representing one half of the real 150 µm-wide gate finger) is partitioned into 3 elementary blocks, in order to take temperature and current non-uniformity along the finger into account; more than 3 finger sections can be considered for wider fingers or improved temperature resolution along the finger width. Multi-finger gates would also feature a larger number of blocks. The channel temperature T_{CH} for each of these 3 HEMT finger sections is the output of the LE thermal network, and an input to the i_{DS} generator. The thermal feedback loop is shown in Figure 2.30. In this case, each SHFM is an elementary HEMT model, representing one third of the 75 µm-wide finger (i.e., one sixth of the 150 µm-wide finger). These blocks are biased in parallel, but work at different channel temperatures, self-consistently computed by the lumped-element thermal network. In case of larger finger numbers, more SFHM blocks will have to be considered, but the model structure will be the same.

In order to show the excellent agreement between the two models (FE and LE), four temperature profiles are shown in Figure 2.39: (i) the vertical temperature profile below the center of the heating finger with a close-up (ii) to highlight the effect of the TBR, (iii) the thermal response to the application of a power step, and (iv) the temperature profile along the heating finger width. As it can be seen, the agreement between the FE model and the LE model is excellent under every point of view. The

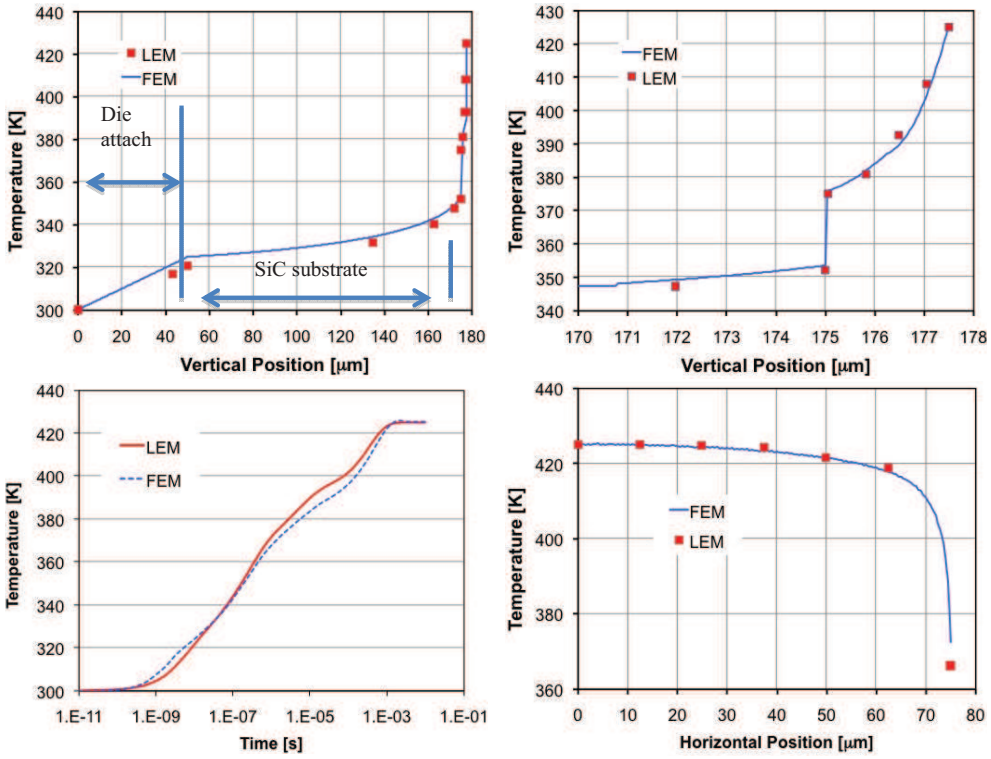


Figure 2.39 – (top left) Comparison between vertical temperature profiles obtained by FEM and LEM simulations. (top right) A close-up of the top part of the vertical profile showing the effect of the TBR. (bottom left) Comparison between LEM and FEM transient response after the application of a power step. (bottom right) Temperature profile along the finger width. In all cases, the steady-state dissipated power is 6.67 W/mm.

"temperature jump" due to the presence of TBR is visible in Figure 2.39.

The LE model has been found to be in excellent agreement with the FE model, and it has been coupled with the electro-thermal, large signal model of the HEMT device, described in section 2.5.2, in order to model the (remarkable) differences between isothermal working conditions and self-heated working conditions. Simulation results, obtained in the case of a class AB[3] amplifier working at 1 MHz, are shown in Figure 2.40. Notice the difference between isothermal and self-heated operating conditions: the current delivered to the output when self-heating is accounted for is

[3] A class AB amplifier is an amplifier working with a circulation angle ϕ between 180° and 360°.

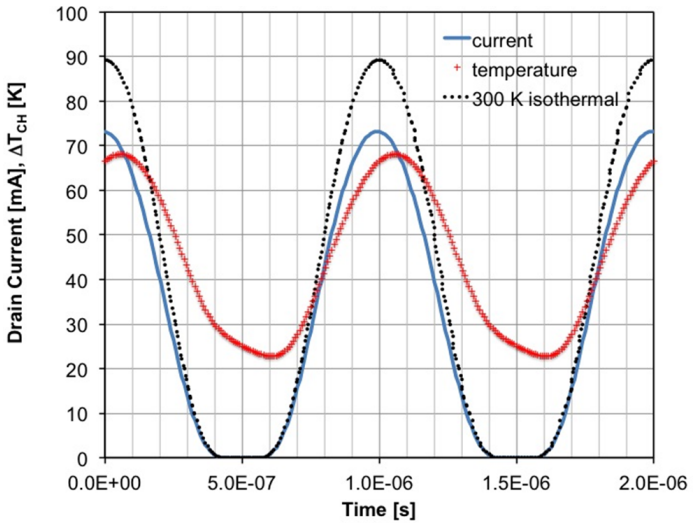


Figure 2.40 – Simulation results of the AB amplifier with and without self-heating. The channel temperature increase over ambient temperature is plotted as well, clearly showing the heating and cooling phases of the amplifier.

significantly lower than the current calculated in the isothermal case.

2.5.3 FinFET modeling

Another electron device whose physical structure is suitable to be modeled by the method previously shown is the FinFET. This kind of MOSFET is a solution to the demand of extremely scaled device, exploiting a three-dimensional gate geometry, which covers the gate on three sides. Thermal modeling of these devices is an interesting topic, since Silicon-On-Insulator (SOI) MOSFETs can overcome the limitations hindering the use of bulk FETs for gate lengths below 50 nm. Among the proposed geometries, FinFETs have a prominent role. In FinFETs, a thin Si fin lies on the SiO_2 buried oxide (BOX), with the gate oxide and poly-Si gate wrapped around the channel to improve gate control and limit short-channel effects. However, the BOX hinders heat removal, and makes self-heating a serious concern. The thermal behavior of FinFETs must therefore be carefully assessed for proper device design, modeling, and reliability estimation. Some published papers focus on purely thermal [26] or electro-thermal [27] numerical simulation using commercial software, as well as on analytical modeling [28]. Although these approaches have their own merits, none of them is amenable to integration in circuit CAD tools. An LE thermal network that

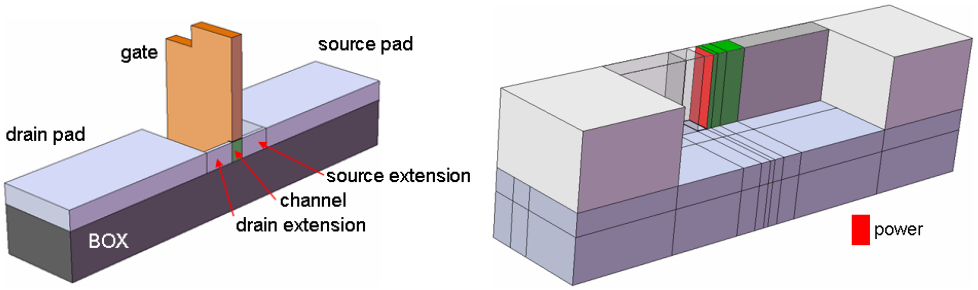


Figure 2.41 – (left) The FinFET structure. Only half of the device was modeled, since the vertical symmetry plane bisecting the FinFET along the channel length can be replaced by an adiabatic boundary. (right) The heating power is injected in the red area (channel end section); the channel is the green volume (gate omitted for clarity); the drawing shows the partitions for LE modeling.

models the self-heating in FinFETs has been developed, and it will be shown here. We validate our model by finite-element (FE) simulations. An example of FinFET, with a blow-up on the channel region, is shown in Figure 2.41. The BOX, Si film, gate oxide and poly-Si gate are 145, 66, 1.8 and 300 nm thick, respectively. The fin width is 22 nm (44 nm for the whole structure); the channel and source/drain extensions are 40 and 100 nm long, respectively. FE thermal simulations were performed using Comsol Multiphysics, with material properties taken from [28, 29, 30]. Self heating is simulated by injecting a power density of 8.26 MW/cm^2 on the end section of the channel (Figure 2.41, right). This is a reasonable approximation, since heat generation is strongly peaked and tightly confined [28]. In our model heat is dissipated only through the bottom of the BOX, but heat sinking surfaces can be considered on the top surface, too (e.g., source/drain pads). The lumped-element thermal network is made of building blocks like that in Figure 2.38 (center): each block models the dynamic thermal behavior of an element of volume. Thermal resistances are obtained by multiplying the thermal resistivity by the length of the associated heat path and dividing it by the path section; likewise, the thermal capacitance is given by the material specific heat, density, and element volume, as in the case of HEMT thermal modeling. We used a total of 70 blocks like that of Figure 2.38 (center) to model the FinFET structure of Figure 2.41; the heat flow through the 1.8 nm gate oxide is one-dimensional, and was modeled with 6 resistances (3 for the top and 3 for the sidewall of the fin). The LE thermal model has been validated by a FE thermal model. In Figure 2.42, two temperature profiles are shown: (i) the static temperature increase along the direction of the channel (right), where the LE temperatures are those of the central node of the elementary blocks, and (ii) the transient evolution

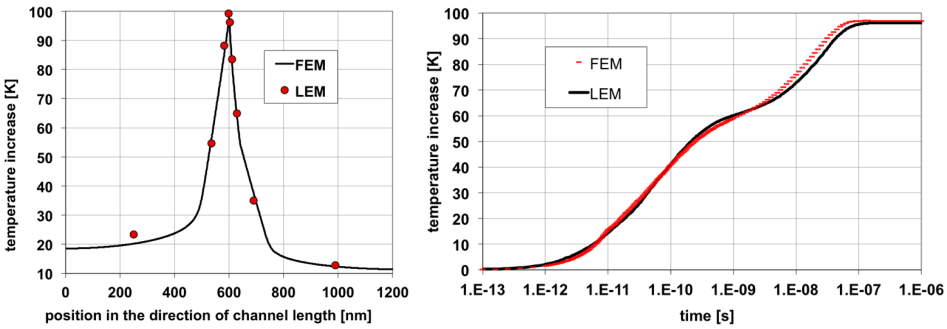


Figure 2.42 – Left: Comparison between FE model and LE model in the steady-state case. Right: temperature response following a power step, as given by the FE and LE models. In both cases, the match between the two models is excellent. No fitting parameters are used.

of the central node of the hottest building block in the LE network (i.e., the block which models the region of the channel in which the power flow is entering) versus the surface-averaged temperature of the channel section in the FE model. The agreement between LE and FE models is excellent, and it is once more worth pointing out that *no fitting parameters* are used in this approach.

2.5.4 Modeling of complex structures

The work shown in the previous sections has demonstrated how a LE thermal network can be built in order to describe the heat flow in a three-dimensional structure. In order to keep the discretization process (which, up to now, is still a manual step) simple, the structures under test were meshed in this way: after devising a two-dimensional mesh for the top surface of the structure, this mesh is extruded throughout the depth of the structure. Figure 2.43 explains this concept graphically. In general, in the vertical direction the mesh is finer close to the surface (where is the active volume of the device and therefore the heat sources) and coarser towards the bottom. This allows to simulate thick structures with a reasonable number of nodes. This approach, clearly, works well with simple structures, but the presence of contact pads, metallizations, air bridges, and so on, makes the generation of the mesh much more complicated.

The structure shown in Figure 2.44 represents a complete cell of a larger HEMT (the whole device is obtained by replicating this cell several times). Note the presence of a ground via, which connects the source metallization over the top surface of the device to the ground metallization on the back of the device. Actually, even a complex structure like that shown in Figure 2.44 could be discretized starting from a rectangular 2D mesh as described above: the real problem is that the number of

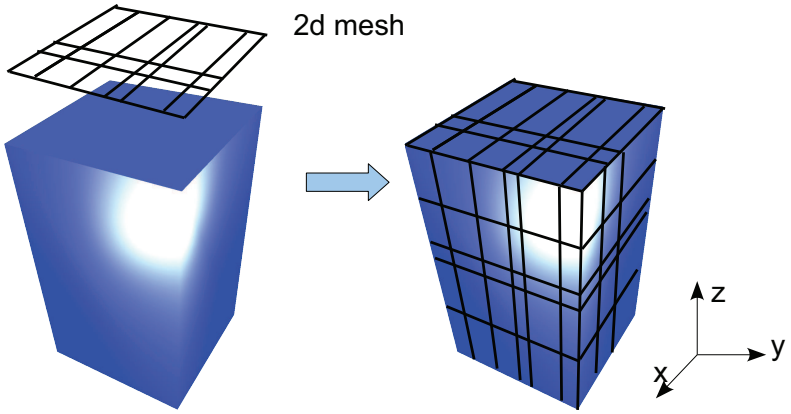


Figure 2.43 – Starting from a 3D domain and a surface 2D mesh, the discretization on the whole domain is simply achieved by extruding the 2D mesh throughout the z direction. The meshing can get coarser and coarser as depth increases in order to limit the number of nodes.

rectangles grows to unmanageable figures. It has to be remembered that these LE thermal networks are built manually, so the degree of complexity that can be handled efficiently is limited.

To overcome this problem, one possibility is to use $R_\theta C_\theta$ Foster networks in order to describe the thermal behavior of the individual fingers. This operation can be easily done starting from FE simulations of the temperature response to a power step of each finger; the derivative of temperature with respect to the natural logarithm of time is computed, and then the Foster's network is obtained, as previously explained (section 2.3). This is an empirical and not physically based approach to obtain compact thermal models for complex structures.

Unlike the physical LE networks described before, this method allows to describe precisely only one reference point, in our case the hottest point of the heating finger: the other nodes in the Foster's network can hardly be associated with specific physical locations in the modeled structure: the thermal network obtained in this case comes from a fitting process, and not from building a thermal network related with the physical structure of the device.

This empirical technique has been successfully applied to the structure shown in Figure 2.44. A blow-up of the HEMT surface is shown in Figure 2.45. The light blue lines are the gate metallizations, covering the active area. This structure has five fingers with 200 µm width and 1 µm length; each of them dissipates 0.5 W, i.e. the power density is 2.5 W/mm or 2.5 GW/m^2.

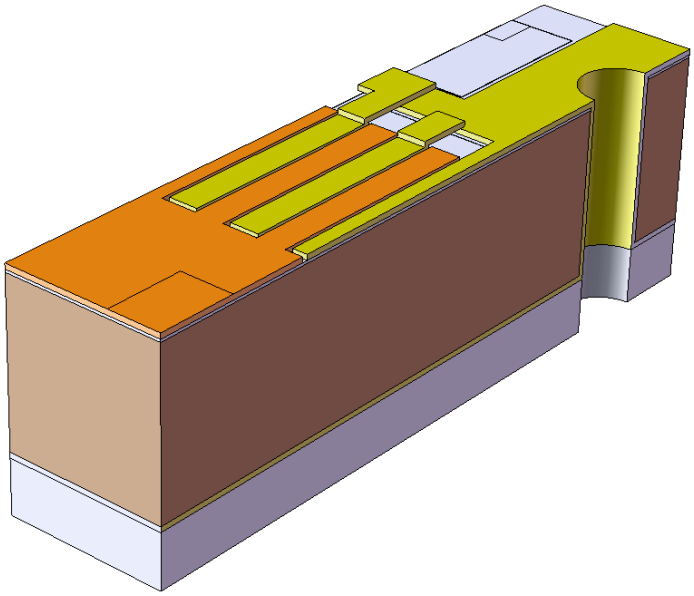


Figure 2.44 – The elementary cell of a large-periphery AlGaN/GaN power device, featuring drain, gate and source metallizations, and a source ground via.

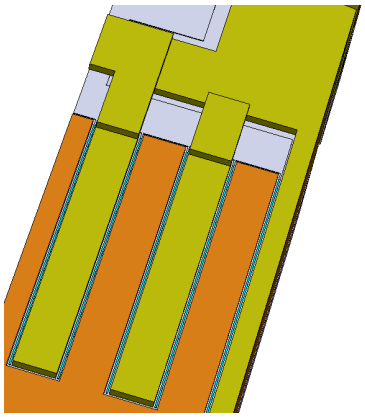


Figure 2.45 – A blow up of the HEMT surface cell. The light blue lines are the gate metallization, below which are the heating areas.

A FEM model was built and the temperature response to a power step of the central point of the finger (i.e., the hottest point) was simulated. FEM simulation

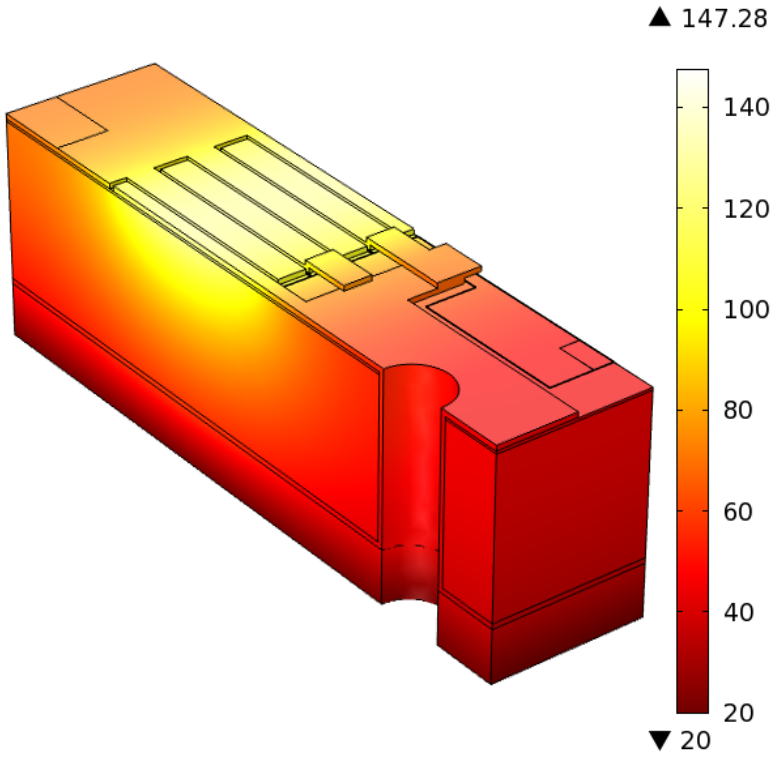


Figure 2.46 – FEM simulation results of the structure shown in Figure 2.44. The dissipated power for each finger is 0.5 W. Temperatures in °C.

results are shown in Figure 2.46. The $T(t)$ curve is then given as an input to the algorithm that extracts the time constants (developed in LabVIEW). From the values of the thermal resistances and time constants, it is straightforward to determine the values of the associated thermal capacitances. Even if the algorithm detected four time constants, hence four RC stages, two time constants nearly coincide, so only 3 RC stages are necessary. A comparison between the FE-simulated temperature response and the Foster network model is shown in Figure 2.47; the Foster network is made of three stages (R_i, C_i) whose values are $(52.2\,\mathrm{K/W}, 9.77\times10^{-8}\,\mathrm{J/K})$, $(48.3\,\mathrm{K/W}, 9.36\times10^{-7}\,\mathrm{J/K})$, and $(156\,\mathrm{K/W}, 2.81\times10^{-6}\,\mathrm{J/K})$. The three time constants are $\tau_1 = 5.1\,\mu\mathrm{s}$, $\tau_2 = 45.2\,\mu\mathrm{s}$ and $\tau_3 = 438\,\mu\mathrm{s}$.

Actually, the main advantages of this method are that it can be to a large extent automatic, and that the thermal network is very simple, made as it is of few resistors and capacitors. The drawback, as was said before, is that the only physically

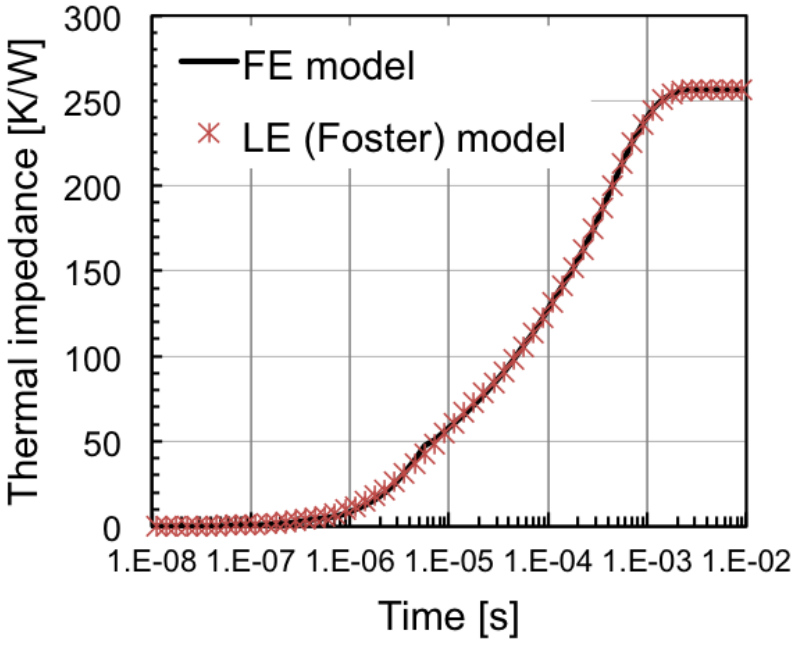


Figure 2.47 – Comparison between the thermal impedance calculated by the FEM model (continuous line) and that given by a 3-stage LE empirical Foster's network (stars). The profile corresponds to a power step of 0.5 W/finger and the temperature is taken at the hottest point of the finger.

meaningful temperature is that of the node where power is injected.

Although more complex methods are available (i.e., deconvolution methods [20]), the one shown here was found to be a good tradeoff between simplicity and accuracy.

2.6 Conclusions

This chapter has been focused on the self-consistent, electro-thermal modeling of electron devices: although the FE method is a very powerful tool in order to study the self-heating of an electron device, FE models are not amenable to insertion in circuital CAD tools to be coupled with a large-signal, electro-thermal model of an electron device. In order to bridge this gap, it has been shown that lumped-element thermal networks represent a suitable solution, since they can be used effectively with large-signal, electro-thermal models of electron devices for computing the channel temperature in a self-consistent way.

Several application examples of the LE technique have been shown, proofing the capability of this method to decribe satisfactorily three-dimensional structures, made of different materials with temperature-dependent thermal conductivities, and with complex boundary conditions applied. Power MOSFETs, HEMTs, and FinFETs LE thermal models have been developed, proofing that the approach is applicable to structures which feature completely different characteristic dimensions (several order of magnitude) and different materials. Both the steady-state and the transient thermal behaviors were studied, with excellent modeling results. In order to validate LE networks, their modeling results were compared with FE modeling results, and in all cases the agreement was found to be excellent.

In some cases the LE thermal network can be built by assemblying several building blocks whose element values depend only on block dimensions and the thermal properties of the material: that is, this is a *physically based* method: it is worth pointing out that (i) *no fitting parameters* were used in this modeling process, and (ii) the LE networks describes the thermal behavior of the *overall* structure.

When the structure is complicated for the mesh to be generated manually, the above method becomes not suitable. In this case, the use of *empirical* methods (i.e., a Foster thermal network) to describe the temperature of a single point in the structure, has been shown. Obviously, the LE thermal network so obtained is no longer related with the physical structure of the device, and the only meaningful temperature in the network is that of the node in which the power is entering. The advantages of this method are: (i) this method is automatic, and (ii) the LE thermal network so obtained is very compact.

Chapter 3

High spatial resolution thermal measurements

This section deals with the work carried out at the Centre for Device Thermography and Reliability (CDTR), H.H. Wills dept. of Physics, University of Bristol, UK. In particular, the development of the thermo-reflectance technique for temperature measurements on semiconductor devices, and its advantages, will be illustrated.

3.1 Contact and contact-less methods for temperature measurement

The temperature of a body can be measured by different methods, some requiring contact, some contact-less. In contact methods, a physical object (the sensor) contacts the device of which we want to measure the temperature, while in contact-less methods the temperature is measured without contacting the device, for example using laser light. Examples of non-contact methods are infrared thermoscopy (IR), Raman, and thermo-reflectance measurements, while temperature measurements by means of thermocouples fall among the contact methods.

Every method has advantages and drawbacks. Contact methods are simple to implement, usually do not require complicated instrumentation (e.g.,thermocouples), but they can strongly influence the measurement (basically because the mass of the thermocouple is not always negligible with respect to the mass of the measured device); they are very cheap, an aspect not to be forgotten. On the other hand, especially if high spatial resolution is needed (i.e. for measuring temperature maps over an electron device), it is mandatory to use non-contact thermal measurements techniques; an example is the Raman technique, which can achieve spatial resolutions of hundreds of nanometers. It was mentioned before that contact measurements can influence the

sample: actually, the laser-based techniques like Raman spectroscopy can influence the sample temperature as well, especially if the laser power is high enough to heat the device or to activate some opto-electronic process (i.e., carriers generation) that can strongly change the electrical behavior of the device.

After this brief overview about thermal measurements, the thermo-reflectance technique will be described, with particular emphasis on the technical details and issues faced during the development of the measurement bench.

3.1.1 Introduction and useful concepts

This section deals with the basics of thermoreflectance measurements process and the underlying physical concepts.

Reflectance at a glance. Let us suppose we focus a light beam on a surface: reflection by the surface will take place. The ratio between the reflected power and the incident power is defined as the *reflectance* of the body:

$$R = \frac{P_{\text{refl}}}{P_{\text{inc}}} \tag{3.1}$$

where P_{refl} and P_{inc} are the reflected power and the incident power, respectively. This number R obviously ranges between 0 (no reflection) and 1 (total reflection). This physical quantity mainly depends on surface material, wavelength of incident light, incident angle, and *surface temperature*. If all other factors are kept constant, R depends only on surface temperature. Note that this method allows to measure the surface temperature, while Raman measurements sample the average temperature of a volume extending some depth into the semiconductor. The reflectance is written as a function of temperature as follows:

$$R = R(T) \simeq R(T_0) + \frac{\partial R}{\partial T}(T - T_0) \tag{3.2}$$

in which the dependence of R on temperature has been approximated by a first-order Taylor expansion, as common in the literature. Further details can be found in [31, 32]. The main drawback of this technique is the *small* variation of reflectance with temperature, which is about $10^{-5} \div 10^{-3}$ of the reference reflectance $R(T_0)$ measured at the reference temperature T_0. From a practical point of view, some additional considerations need to be made. In general, for performing thermoreflectance measurements, the first datum to obtain is the *calibration coefficient* κ, which relates the variation of temperature with the variation of reflectance. Using photodetectors like photodiodes, a voltage proportional to the light power is obtained. Considering (3.2), it can be observed that:

$$R = \frac{P_{\text{refl}}}{P_{\text{inc}}} = \frac{V_{\text{refl}}}{V_{\text{inc}}} \tag{3.3}$$

in which V_{refl} and V_{inc} are voltages proportional to the reflected and incident power, respectively (since the photodetector is the same, the coefficient relating power and voltage is the same). Then, (3.2) can be rewritten as follows:

$$R(T) = \frac{V_{\text{refl}}(T)}{V_{\text{inc}}} = \frac{V_{\text{refl}}(T_0)}{V_{\text{inc}}} + \frac{\partial V_{\text{refl}}(T)/\partial T}{V_{\text{inc}}}(T - T_0); \tag{3.4}$$

it follows that:

$$V_{\text{refl}}(T) = V_{\text{refl}}(T_0) + \frac{\partial V_{\text{refl}}(T)}{\partial T}(T - T_0). \tag{3.5}$$

In general, the term $V_{\text{refl}}(T_0)$ can be affected by an offset, which can be measured in the absence of laser light (this can be due to the electronic amplifier, or to laser reflection background as well). In other words:

$$V_{\text{refl}}(T_0) = V'_{\text{refl}}(T_0) + V_{\text{os}}; \tag{3.6}$$

if (3.6) is substituted in (3.5) and the offset term V_{os} is subtracted from (3.5), we have:

$$V_{\text{refl}}(T) = V_{\text{refl}}(T_0)' + \frac{\partial V_{\text{refl}}(T)}{\partial T}(T - T_0). \tag{3.7}$$

If we divide both members by V'_{refl}, and make a linear fit [33] to the relation we obtain an estimate of the calibration coefficient κ,

$$\frac{V_{\text{refl}}(T)}{V_{\text{refl}}(T_0)'} = 1 + \frac{1}{V_{\text{refl}}(T_0)'} \cdot \frac{\partial V_{\text{refl}}(T)}{\partial T}(T - T_0). \tag{3.8}$$

where κ is the slope of linear fit (3.8):

$$\kappa = \frac{1}{R(T_0)}\frac{\partial R}{\partial T} = \frac{1}{V_{\text{refl}}(T_0)'} \cdot \frac{\partial V_{\text{refl}}(T)}{\partial T}. \tag{3.9}$$

3.2 The optical bench

In order to measure the reflectance of a surface, a well-characterized light source is necessary. To this purpose, we used a Renishaw system with a 532 nm laser source. A confocal microscope, in which laser light and white light can be channeled at the same time, is used to focus the laser on the sample under test. Before entering the microscope, the laser light is conditioned by a set of components (beam expanders, filters, mirrors, and so on) in order to change the size of the beam, the focus of the spot on the sample, so that the white light and the laser light are focused on the same horizontal plane. This is necessary in thermoreflectance measurements, since the spatial resolution of the technique is maximum when the laser beam is focused on the sample. To focus the sample, we use the white light: when the sample is in

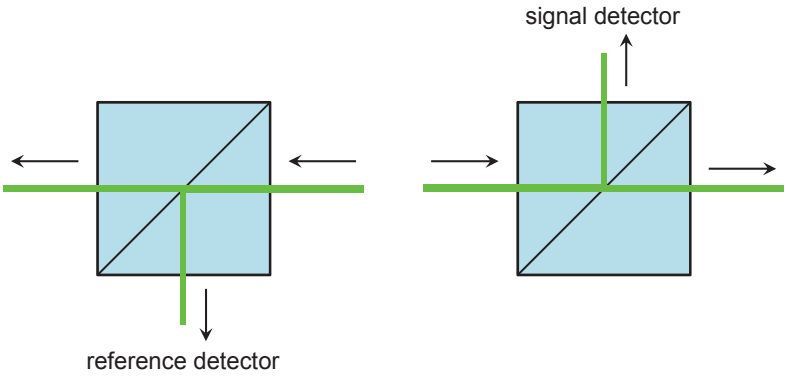


Figure 3.1 – Working principle of a beamsplitter. The primary incoming beam (left) is split into two components; the reflected beam (right) is split as well; the position of suitable detectors are shown.

focus under white light illumination, also the laser will be focused on the sample if the focal plane of both white and laser light is the same.

Since the information is carried by the reflected light (which is modulated by the temperature of the sample surface), it has to be collected and detected. To do so, a *beam splitter* has been used, since the *incident* light and the *reflected* light are traveling in the same direction, and the detector obviously cannot be placed on the light path (it would block the laser beam). By means of a beam splitter, the detector is placed along a direction perpendicular to the laser beam. The working principle of a beam splitter is graphically explained in Figure 3.1: it is a cube that divides the incoming light into two beams, one of which proceeds in the original direction, while the other is deviated by 90°. During normal operation, both the incident and reflected beams are present at the same time, and the measurement of the intensity of the incident light (*reference*) and of the reflected light (*signal*) can be made at the same time. The primary function of the optical bench is to separate the incident light from the reflected light. To detect the light, a photodiode with sensitive area of roughly $1\,\mathrm{mm}^2$ has been employed. Clearly, the reflected beam has to be focused on the sensitive area of the photodiode, in particular the beam has to be *smaller* than the active area of the sensor (to avoid problems arising from the gaussian shape of the beam over the active area, in the case the beam is larger then the sensor area), and – obviously – the beam has to hit the detector *inside* the active area; actually, the sensor has to be movable so that its active area can be aligned with the light beam. This means that other components are needed, in particular a lens (possibly with long focal length – 50 or 60 mm for example – because manual tuning of the laser beam size is easier) and an x-y micro-regulator which holds the detector (if z

Figure 3.2 – The beamsplitter cased in an aluminum box and mounted on a support to be fixed on the laser path (left). The detector is not shown here. The final arrangement used for thermoreflectance measurements (right): from left, the confocal microscope, the beamsplitter section, and the laser source.

is the direction of the beam, the photodiode has to move on the x-y plane). With a focusing (converging, plano-convex) lens and an x-y regulator, the beam size can be focused exactly inside the active area of the photodiode. The whole assembly and measurement bench are shown in Figure 3.2.

3.3 The electrical bench

Basically, the detection of light is achieved by means of a photodiode [34], a device that converts light into current according to a parameter, R_λ[A/W], called *responsitivity*. The photodiode can be used as a current source – this is actually the most common configuration; the electric model is shown in Figure 3.3. A current-voltage converter is then necessary to obtain a voltage proportional to the intensity of incident light. The I/V converter can be simply a resistor, or an opamp-based I/V converter.

The basic schematics of an I/V converter is shown in Figure 3.4. The design of a

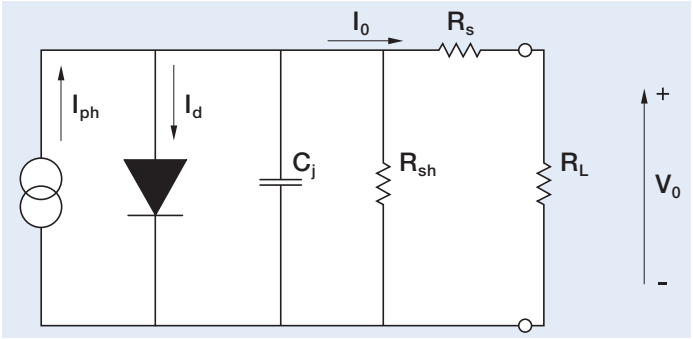


Figure 3.3 – Equivalent circuit of a photodiode.

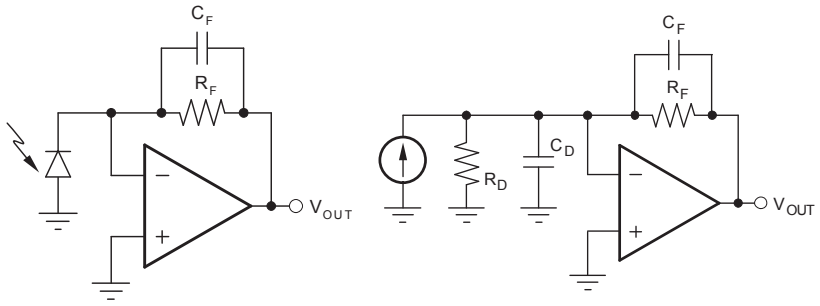


Figure 3.4 – Basic transimpedance amplifier (left). Transimpedance amplifier with a detailed model of the photodiode (from [35]) (right).

transconductance amplifier differs with respect to the design of a "standard" voltage amplifier. In particular, the first thing to consider is that, while in a voltage amplifier the bandwidth of the amplifier BW and the closed loop gain G_{cl} are related by the following (approximated) relationship:

$$\mathrm{GBP} = G_{\mathrm{cl}} \cdot \mathrm{BW} = \text{constant},$$

in a transconductance amplifier the above relationship does not hold. In particular, the bandwidth of the amplifier is dictated by the $R_F C_F$ low-pass filter in the feedback loop (for high-bandwidth amplifiers, the capacitance is the parasitic capacitance of the resistor, see Figure 3.4), while the limited bandwidth of the operational amplifier plays a minor role in determining the bandwidth of the I/V converter.

An analytical derivation of the transfer function of the transconductance amplifier is shown here. The circuit used to obtain the transfer function is shown in Figure 3.5.

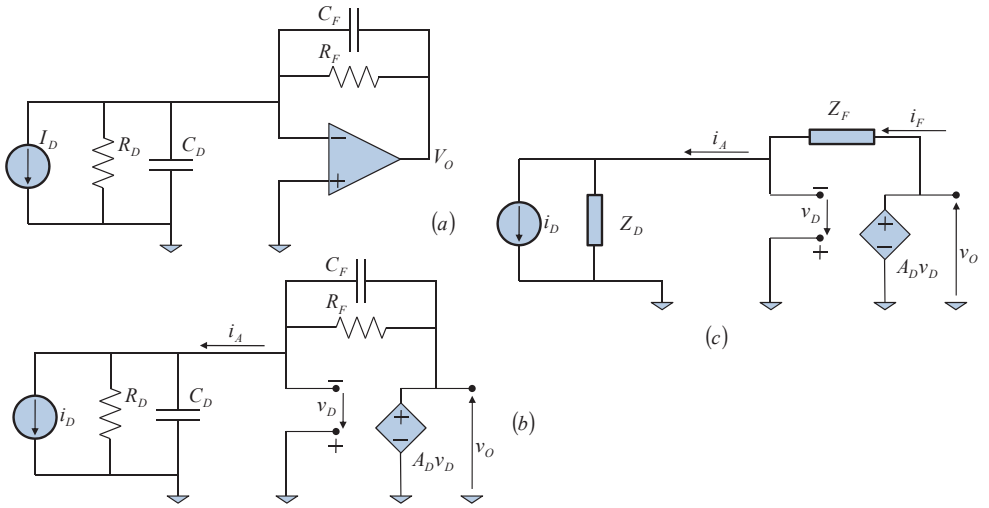


Figure 3.5 – Schematic for deriving the transfer function of the IV converter: (a) large signal circuit, (b, c) linearized circuit.

The finite open loop gain of the operational amplifier is described by the common single-pole transfer function:

$$A_D(s) = \frac{A_{D0}}{1 + s\tau_0}$$

where A_{D0} is the low frequency gain and τ_0 is the time constant corresponding to the corner frequency. Then, i_D is the input current, R_D and C_D are the shunt resistance and capacitance of the photodiode, R_F and C_F are the resistance and capacitance of the feedback network, and v_O is the output voltage. Let us define the following impedances:

$$Z_D = \frac{\dfrac{1}{sC_D}R_D}{\dfrac{1}{sC_D} + R_D} = \frac{R_D}{1 + sC_D R_D}, \quad Z_F = \frac{\dfrac{1}{sC_F}R_F}{\dfrac{1}{sC_F} + R_F} = \frac{R_F}{1 + sC_F R_F};$$

then, referring to Figure 3.5-(c), it is possible to write the following equations:

$$\begin{aligned} i_A &= i_D - \frac{v_D}{Z_D} = i_F = \frac{v_O + v_D}{Z_F}, \\ v_O &= A_D \cdot v_D. \end{aligned}$$

Combining the two equations above, it is straightforward to obtain the following:

$$v_O\left(\frac{1}{Z_F}+\frac{1}{Z_F\cdot A_D}+\frac{1}{Z_D\cdot A_D}\right)=i_D,$$

that is:

$$\frac{v_O}{i_D}=\frac{Z_F\cdot Z_D\cdot A_D}{Z_F+Z_D+Z_D\cdot A_D}.$$

Substituting for Z_F, Z_D, and A_D we get:

$$\frac{v_O}{i_D}=\frac{\dfrac{R_F}{1+sR_FC_F}\cdot\dfrac{R_D}{1+sR_DC_D}\cdot\dfrac{A_{D0}}{1+s\tau_0}}{\dfrac{R_F}{1+sR_FC_F}+\dfrac{R_D}{1+sR_DC_D}+\dfrac{A_{D0}}{1+s\tau_0}\cdot\dfrac{R_D}{1+sR_DC_D}},$$

which can be written in its final and complete form as follows:

$$\mathcal{H}_{cl}(s)=\frac{R_FR_DA_{D0}}{R_F+R_D+R_DA_{D0}}\cdot \frac{1}{1+\dfrac{R_F(\tau_D+\tau_0)+R_D(\tau_F+\tau_0)+R_DA_{D0}\tau_F}{R_F+R_D+R_DA_{D0}}s+\dfrac{R_F\tau_D\tau_0+R_D\tau_F\tau_0}{R_F+R_D+R_DA_{D0}}s^2}$$

in which, for simplicity of notation, the following time constants have been defined: $\tau_F=R_FC_F$, and $\tau_D=R_DC_D$. For a high bandwidth amplifier, usually the feedback resistance R_F is small compared with the diode's shunt resistance R_D, and since $A_{D0}\gg 1$, the previous relationship can be simplified as follows:

$$\mathcal{H}_{cl}(s)\simeq\frac{R_F}{1+\dfrac{R_F(\tau_D+\tau_0)+R_D(\tau_F+\tau_0)+R_DA_{D0}\tau_F}{R_D(1+A_{D0})}s+\dfrac{R_F\tau_D\tau_0+R_D\tau_F\tau_0}{R_D(1+A_{D0})}s^2}.$$

An amplifier used in an early stage of the project was the OPA703 from Texas Instruments [36], which features very low input bias current and very low power consumption. The circuit topology is quite simple, the output of the circuit is $V_o=R_FI_d$. The choice of components is crucial to achieve the desired performance. The first design parameter is the feedback resistance R_F. The higher the resistance (up to 10 GΩ, with careful layout), the higher the signal-to-noise ratio: since the gain of the amplifier is proportional to R_F, while the thermal noise of the resistor, which we assume to be the dominant noise source, is proportional to $\sqrt{R_F}$, the signal-to-noise ratio is (to first approximation) proportional to $\sqrt{R_F}$. A detailed analysis of noise in transimpedance amplifiers can be found in [37], while design guidelines for high-performance transimpendance amplifiers can be found in [38]. The main guidelines are the following[1]: use

[1]It will be shown later on that not all these guidelines can be strictly followed, in particular the use of batteries for powering instrumentation that has to work for a long period of time.

of low-noise power supplies (batteries); use of bypass capacitors close to the opamp supply pins; choice of an opamp with low input bias currents (the lower the bias currents, the higher the sensitivity of the amplifier) and low noise figure. The mimimum current which can be detected by the amplifier is limited by the input bias current of the opamp (it fixes the sensitivity of the amplifier). In the end, the whole circuit has to be sealed in a metallic shield, to improve the immunity from external noise sources. The technical details about the manufacturing of the first I/V converter are given below.

The opamp is powered between ground and −9 V by a single battery. Two versions of the same topology have been manufactured, the first with a high feedback resistance (selected choosing among the values 1.8 MΩ, 10 MΩ, and 100 MΩ), and the second with a fixed feedback resistance of 210 kΩ. The first version is used mainly for calibration,

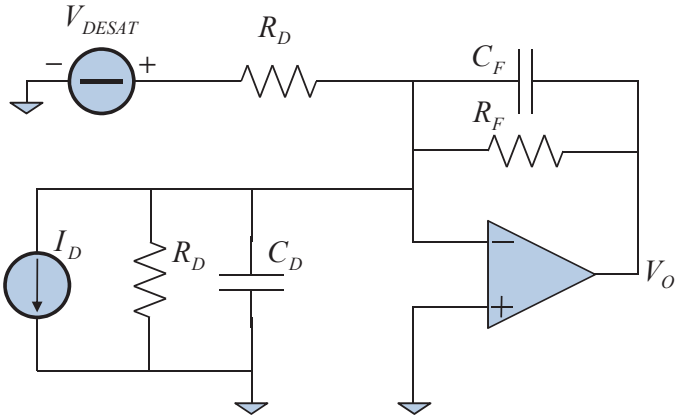


Figure 3.6 – Modified I/V converter with the addition of a second input V_{DESAT} to desaturate the opamp output.

i.e., to measure the $R(T)$ curve: the sample temperature is fixed by a Peltier cell and the reflectance is measured at different temperatures $T_1, T_2, \ldots$, thus obtaining a set of points $R(T_1)$, $R_(T_2), \ldots$; a very low laser power can be used, thanks to the high feedback resistance – which leads, on the other hand, to poor dynamic performance. The second version has been designed to be suitable for both transient measurements and for calibration: in practice, both the reflectance reference value and the variation of reflectance need to be available at the same time, and this can be achieved by using only one amplifier and splitting its output into DC (reference) and AC (temperature variation) components. It turns out that the chosen feedback resistor is a tradeoff between gain (for calibration, the higher the resistance, the better) and bandwidth (for transient measurements, the lower the resistance, the better). In reality, the simple

I/V converter topology presents some problems because, since the OPA656 is not a rail-to-rail opamp, the maximum[2] output voltage achievable is roughly $-0.8\,\mathrm{V}$, and for low laser power the amplifier output cannot go below $-0.8\,\mathrm{V}$ (for example, if the laser power produced an output of $-0.4\,\mathrm{V}$, in reality this value would be saturated at $-0.8\,\mathrm{V}$). This problem can be overcome by applying a second input, in order to de-saturate the output of the opamp. In this case, even when the laser power is zero or very low, the output of the opamp is always outside of the saturated region. The modified circuit is shown in Figure 3.6. Considering only the DC behaviour (neglecting the effect of capacitances), by applying the superposition principle to the inputs I_D and V_{DESAT} it is easy to show that the output voltage is given by

$$V_O = -I_D \cdot R_F - V_{\mathrm{DESAT}} \cdot \frac{R_F}{R_D} \tag{3.10}$$

and it is enough to choose V_{DESAT} and R_D to achieve, when $I_D = 0$, a value of output voltage greater (in absolute value) than the saturated one; for example it is possible to choose $V_O = -1 \div -1.2\,\mathrm{V}$. As an example, the values chosen in this case are $V_{\mathrm{DESAT}} = 9\,\mathrm{V}$, $R_D = 1.53\,\mathrm{M\Omega}$, which give an output of $-1.24\,\mathrm{V}$ when the input on the photodiode is naught.

Another important aspect for which the above modification was found to be necessary is that, with the present setup, the absolute value of reflectance of gold tends to decrease (in absolute value) with increasing temperature. This means that, if the initial value of the output voltage (with zero laser power) is not far enough from the saturated value, for high temperatures (and thus for lower values of reflectance) the signal could saturate again towards $-0.8\,\mathrm{V}$. Providing a high bias far from the saturated voltage value prevents this phenomenon.

An example of measured performance is shown in Figure 3.7, in which the output of the amplifier is monitored for quite a long time to detect the presence of initial drift effects and other non-idealities.

A simple but effective improvement is to use a linear regulator instead of a battery directly connected to the resistance R_D. This is convenient since the battery gets discharged while working, and the voltage is slightly changing during the measurement. If the battery is connected, for instance, to a 5 V linear regulator, the de-saturating voltage is much more stable in time.

As for every other instrumentation circuit, a warm-up time is required to ensure proper operation, in particular to let the first large drift settle and to stabilize the circuit itself, as can be seen from Figure 3.7. This initial drift can also be due to the settling time of the power supply used for generating the $9\,\mathrm{V}$ value of V_{DESAT}. Even after the warm-up, drift and offsets are still present. A good strategy can be to acquire as much data as possible, in order to have a good quantity of information that can

[2]The output voltage is negative, with a swing roughly between $-0.8\,\mathrm{V}$ and $-8.2\,\mathrm{V}$.

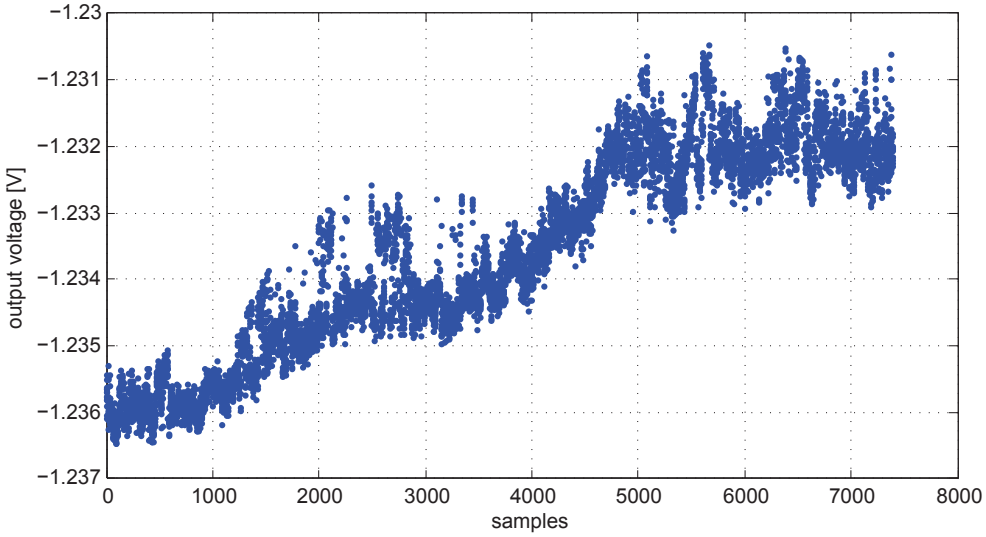


Figure 3.7 – Output voltage of the modified IV converter in function of time. Each point is sampled every 0.25 s. An effect of drift stabilizing can be noted after the first 5000 samples.

be successfully used in a post-processing phase for de-embedding the errors. A digital multimeter, properly driven by LabVIEW, for instance, can easily implement this task. In addition, if the measurement itself is fast, the drift in the time necessary for the measurement can be neglected: the only drift to account for is the one between different measurements, that can be significant if a long time passes between two subsequent measurements. This could be particularly true in the case of calibration, in which the stage has to reach the calibration temperature before the value of reflectance can be acquired.

Since the acquisition of different values of reflectance is a time-consuming task, the discharge of batteries is another problem to account for. This has been found to be an important source of error, since the measured reflectance at the initial temperature T_{init} of the temperature sweep (from T_{init} to T_{fin}) was largely different from the value of reflectance at the same temperature T_{init} at the end of the inverse temperature sweep (from T_{fin} to T_{init}). In addition, while during the first temperature sweep, a significant difference between reflectance values has been found for different temperature values, during the opposite sweep the values was clamped to a single value, and usually

$$|T_{\text{init}}(T_{\text{init}} \to T_{\text{fin}})| < |T_{\text{init}}(T_{\text{fin}} \to T_{\text{init}})| \tag{3.11}$$

suggesting a decrease in the amplifier power supply (since all optical components were

checked to be in the original state, with no defocusing effects that could change the measured value). To prove these considerations, a long-time acquisition task has been conducted, first of all to check the robustness of the battery powering the photodiode amplifier, and then to check the stability of the laser power, because also the laser power is fluctuating to some extent. To do so, a photodiode amplifier has been installed on the first reflective path of the beamsplitter (and this can be battery driven since the power consumption is really low), and the laser power has been recorded over some hours. Then, to verify the robustness of the battery, the amplifier was left to operate until complete depletion of the battery, acquiring reflectance values with the sample constantly kept in focus. The results are shown in Figure 3.8 for the laser power stability and in Figure 3.9 for the robustness of the battery-supplied amplifier. The precision of sampling time does not really matter; the important aspects are:

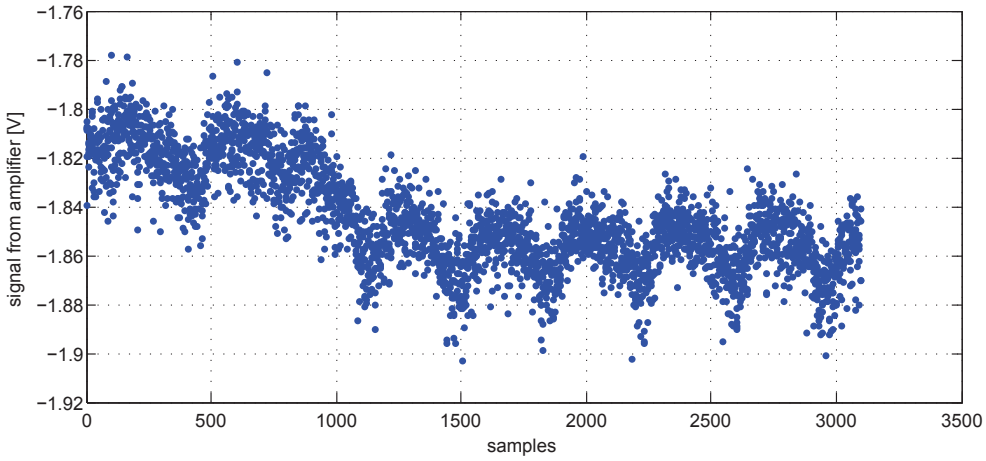


Figure 3.8 – Time behaviour of the laser power over some hours of operation. Samples are taken every 3 s.

♢ long time measurement cannot be performed by battery powered amplifiers, at least with the components used here;

♢ the laser power has to be monitored, since it features quite significant variations during measurements.

3.4 The complete setup

A significant part in the bench design is the firmware, developed in LabVIEW [39], that controls and synchronizes all the hardware components. Rather than giving a

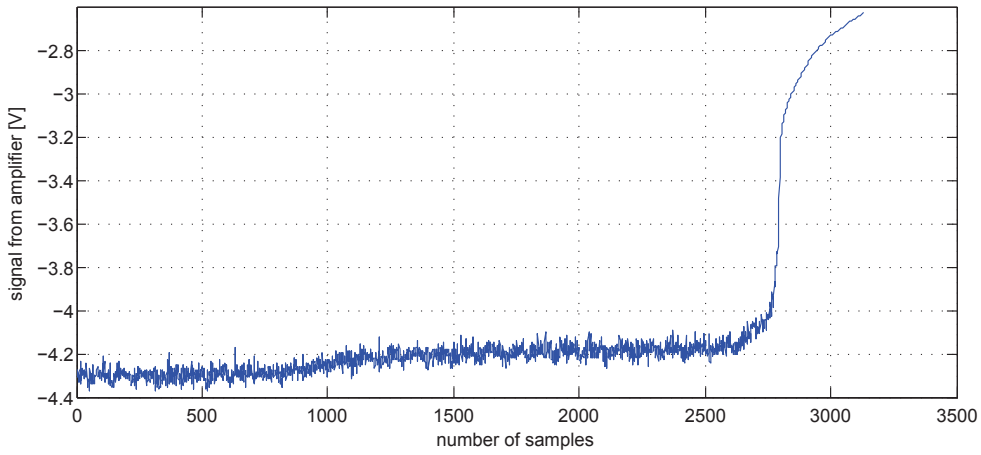


Figure 3.9 – Time behaviour of the battery-powered amplifier until complete depletion of battery. The overall observation window is roughly 10 hours.

detailed description of the various VIs (Virtual Instruments, the name of the applications developed in LabVIEW) used to control the measurement bench, the overall architecture of the bench will be described, focusing the attention on the role played by the firmware, and how it handles all the instruments while a thermoreflectance measurement is carried on.

Figure 3.4 shows the architecture of the thermoreflectance measurement bench. The hardware components are the following:

1. a confocal microscope with motorized stage, a webcam, and a Peltier cell (with a dedicated controller), used to set the temperature on the back of the sample;
2. measuring instruments, for instance an oscilloscope and a digital multimeter;
3. some optics in order to condition the laser light: a wheel-filter (which sets the attenuation of the laser light), beam expanders, filters, and mirrors;
4. an Acoustic-Optic Modulator (AOM), used to chop the light (driven by a TTL input signal);
5. an IR LED, used to illuminate the sample;
6. the previously described optical bench (only the beam splitter and the converging lens are indicated for simplicity);
7. one or two I/V converters for the photodetectors (if necessary, it is possible to monitor the laser beam right before the beam splitter);

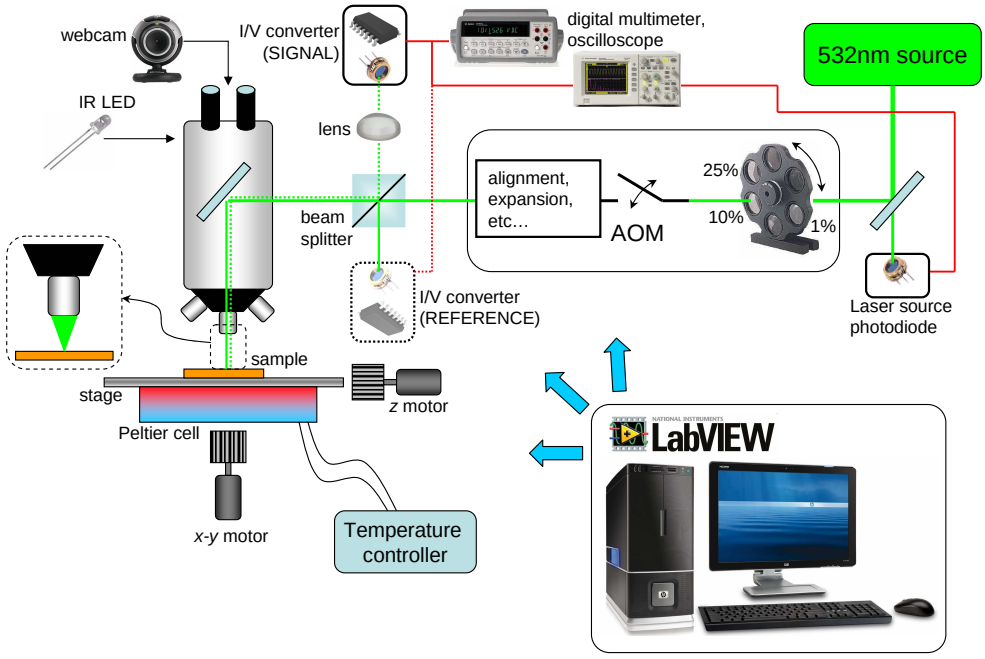


Figure 3.10 – The thermoreflectance measurement bench.

8. a single photodiode placed right after the 532 nm laser source, the output of which can be used to monitor the laser beam before the wheel-filter.

It has to be noted that the wheel-filter, AOM, and all the optics dedicated to beam alignment, beam expansion, and so on, can introduce some noise (i.e., reflections, change in beam shape, due to ambient temperature variations) thus making the laser light *after* this set of components a bit distorted with respect to the laser light *before* this set of components, and sometimes it can be useful to monitor the laser in both positions.

First of all, it is necessary to define the *position* on the sample surface where to perform the thermoreflectance measurement. This is done by moving the sample on the x-y plane: the webcam streams the images captured on the sample, while the x-y motor moves the stage (the sample sticks to the moving stage thanks to a vacuum pump). Once the right location has been found (the operator moves the stage by remote control), this area needs to be focused. The *focus routine* is implemented in LabVIEW: keeping the x-y position costant, the z coordinate is varied, and the *contrast* of the image is evaluated: once the maximum is found, the stage is set at the

z position corresponding to maximum image contrast. The number of steps and the total z-sweep can be selected by the user. The larger the number of points, the slower the routine; the smaller the step size, the finer the determination of the best focus point; the wider the step, the coarser the determination of the best focus point.

Then, the image is saved as the *reference image*: this helps the user to find (roughly) the location of the sample in case of measurements performed at different times (i.e., the sample is removed and then measured again); is necessary for the firmware to find *exactly* the reference position in case the sample has moved. This can be done intentionally by the user, for instance if another measurement has to be carried out, or it can be due to thermal drift during the measurement. A routine in LabVIEW has been developed in order to compare the image sampled by the webcam with the stored *reference image*; if the routine detects some differences, it moves the microscope stage until the sample comes back in the position corresponding to that of the *reference image*. This routine, thus, works on the x-y plane. The routine works as follows: let us suppose the sample drifted; then, if the routine is enabled, it takes a picture of the current position of the sample (preceded by a focusing routine, in order to avoid the detection of differences in the two images due to a different z location of the current image with respect to the reference image) and then the current image, *not the stage* – which would be far slower, is translated by one-pixel steps over rows and columns. The field of view of the routine can be changed (10 pixels, 20 pixels, and so on). Clearly, the time to execute the routine increases quadratically with the number of pixels, for instance a field of view of 10 pixels needs $10 \times 10 = 100$ comparisons. The wider the field of view, the slower the routine, the wider the drift that can be detected. Then, for each comparison, the difference between the reference image and the translated image is calculated, storing the values in a two-dimensional array. The minimum of this array is then found, and the coordinates of this minimum (in pixels) represent the *opposite* of the movement that has to be executed by the stage in order to place the sample in the position corresponding to the *reference image*. Pixels are proportional to microns by a scale factor dependent on image resolution and the magnification of the microscope objective lens.

Clearly, the drift can be larger than what the routine can detect in a single iteration. Instead of increasing the field of view of the routine, it is possible (and helpful in terms of execution speed) to set up the routine with a small field of view (e.g., 10×10 pixels over a 640×480 pixels image), but leave it running in a loop *until* the difference between the *reference image* and the current image is zero.

All these routines can be called separately, run in a loop separately, or grouped together in order to build more complex routines. For instance, a pseudo-code of a typical thermoreflectance measurement would be the following:

```
set(temperature) = 25°C;
run(focus);
```

```
store(reference_image);
while(measurement != stop) {
    check(position);
    if (position != reference) {
        run(focus);
        run(xy-adjustment);
    }
    else {
        activate(oscilloscope);
        save(oscilloscope_memory);
    }
}
```

3.5 Data analysis

The main concern when acquiring data from high-sensitivity instrumentation is the separation of signal from noise. In this case, the signal is very small compared with both the offset and the various sources of noise present in the system. For instance, the ambient temperature change influences all the mechanical parts (due to the different coefficients of thermal expansion), the air flow from air conditioners makes the microscope stage slightly vibrate, other sources of vibration are picked up by the cables that connect the instrumentation with the optical bench, and so on. Identifying all these causes is very hard, time-consuming and often useless, since they are very randomly distributed and, for instance, a disturbance present on a given day might not be present the following day. A better solution is then to increase the signal-to-noise ratio rather than trying to decrease the noise level (which cannot be reduced below a certain threshold); this can be done by increasing the signal power, or at least introducing some additional information known *a priori* in order to exploit them when post-processing the data. Working in the *frequency domain* instead of the *time domain* can be very useful, especially if the frequency of the signal is known. This is, basically, the same approach used in *lock-in detection*, in which a synchronous modulator is used to recover a small signal covered by noise (which cannot be practically distinguished from noise when looking at the sampled waveform in the time domain).

For instance, let us suppose we have to measure a signal x of amplitude 50 mV in a very noisy environment. Our model can include noise coming from the electrical network, $e(t)$ with amplitude 500 mV and fundamental frequency of 50 Hz, and random noise $n(t)$ due to different causes; let the sampling frequency be $f_s = 10\,\text{kHz}$. It is

easy to build such a model in MATLAB, where basically the signals are the following:

$$x = 0.05 \tag{3.12}$$

$$e(t) = 0.5\sin(2\pi 50t) \tag{3.13}$$

and the noise is generated by the `randn(r,c)` function, which generates a matrix of r rows and c columns of random numbers, taken from a normal distribution with mean value zero and standard deviation equal to one (*standard normal distribution*). The normal distribution with mean μ and standard deviation σ is the following:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{3.14}$$

It turns out immediately that such a signal can be hardly identified by just observing the time domain acquisition. If the amplitude is modulated at a certain known frequency f_m, such that we have a signal like this:

$$x(t) = x \cdot \sin(2\pi f_m t); \tag{3.15}$$

even if it is still hardly visible in the time domain, a *Fast Fourier Transform* on the acquired signal can be done, and a peak centered on f_m is visible. The time domain signals are shown in Figure 3.11. As we said before, the waveform of signal plus the noise sources does not allow to extract the information, even if it is modulated at a frequency f_m. Instead, the application of an FFT to the signal allows to distinguish between the signal $x(t)$, the line noise $e(t)$, and all the other noise contributions. The result of the FFT is shown in Figure 3.12.

All these concepts about data analysis have been recalled because they are indicating the guidelines to follow in order to improve the performance of the measurement bench. Basically, we are trying to measure a small signal in a very noisy environment. We have shown that, in general, working in the frequency-domain gives better results than working in the time-domain, especially if the signal to be measured has a frequency content which is known *a priori*. When, instead, the signal is DC or, even worse, its fundamental frequency is unknown, the signal cannot be measured.

For the thermoreflectance technique to be useful, a *calibration coefficient* is needed. This is related to the reflectance value in steady state conditions (i.e., the surface of the sample has to be kept at a constant temperature value). Here, we are trying to measure a small DC signal, and thus the measurement is not possible with the present setup. In addition, we performed *transient* thermoreflectance measurement, modulating the V_{DS} of the device in a square-wave fashion (1 kHz frequency, 25% duty-cycle) and capturing the reflectance waveform with a scope, just averaging over several acquisitions in order to improve the signal-to-noise ratio. In this case, the signal has a strong fundamental frequency, and its detection is far easier than that of

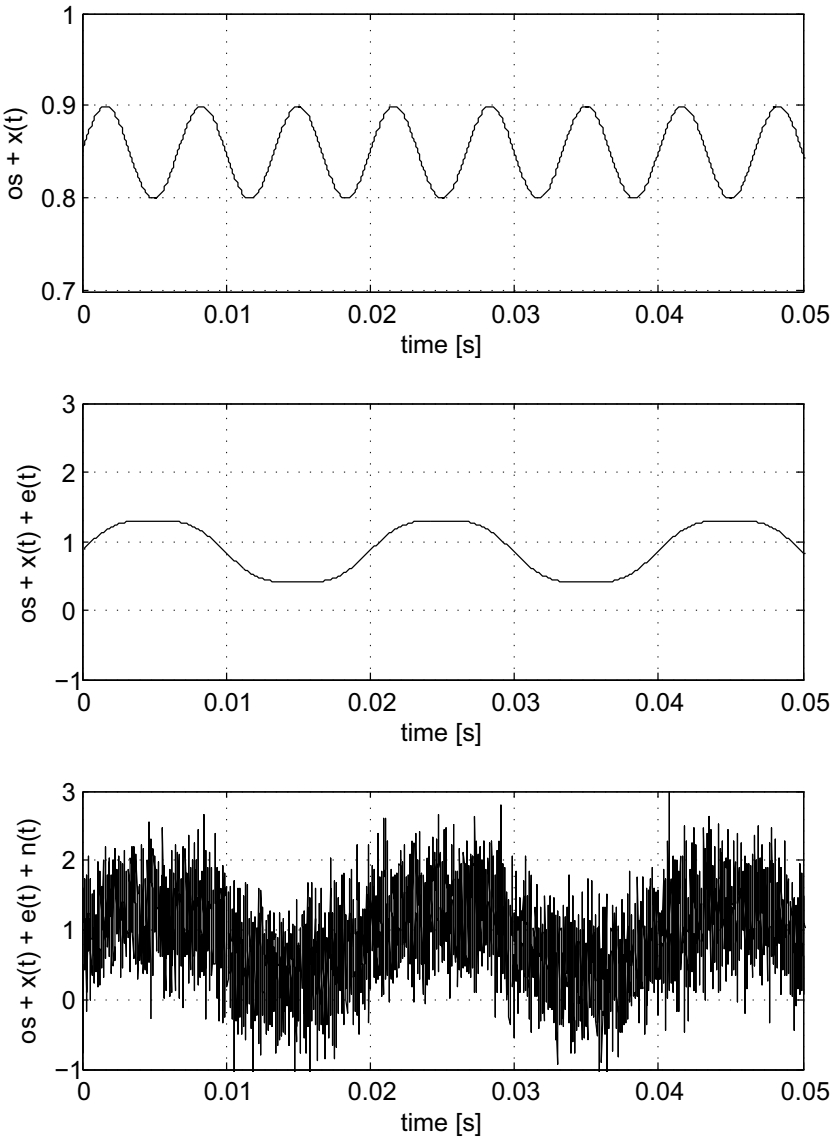


Figure 3.11 – (top) The signal $x(t)$ with offset os. (center) the sum of $x(t)$ and the line noise $e(t)$. (bottom) The sum $x(t) + os + e(t)$ plus random noise source.

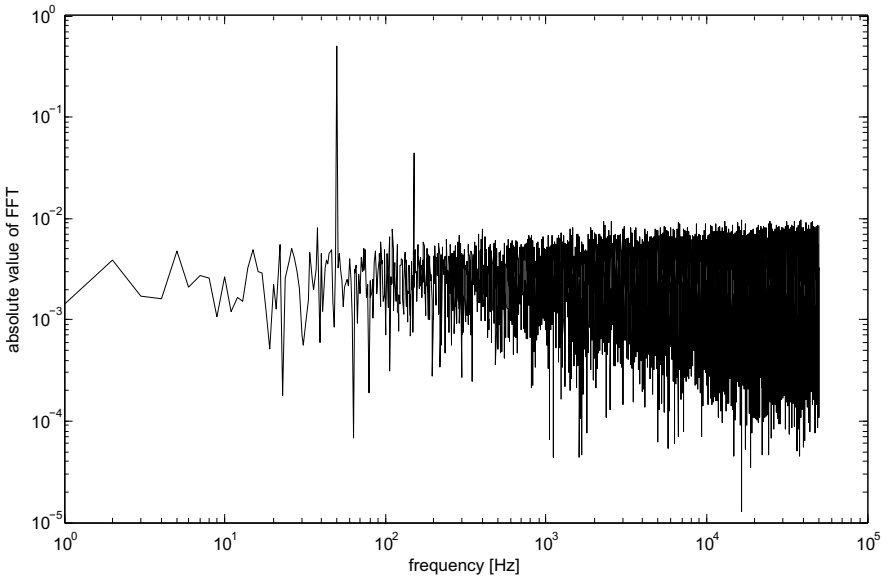


Figure 3.12 – FFT of the signal shown in Figure 3.11 (bottom). The two peaks corresponding to the line noise $e(t)$, centered at 50 Hz, and to the signal $x(t)$, centered at 150 Hz, are clearly visible.

the calibration coefficient. The measured waveform, compared with simulation results, is shown in Figure 3.13: the agreement between the two waveform shapes is excellent. In order to assign a temperature scale (i.e., measuring the calibration coefficient), it is necessary to modulate the thermoreflectance signal during calibration. Let us recall how the calibration coefficient κ is defined:

$$\kappa = \frac{1}{R(T_0)}\frac{\partial R}{\partial T} = \frac{1}{V_{\mathrm{refl}}(T_0)} \cdot \frac{\partial V_{\mathrm{refl}}(T)}{\partial T}; \tag{3.16}$$

the above relationship can be written in more useful form:

$$\kappa = \frac{1}{R(T_0)}\frac{\partial R}{\partial T} \simeq \frac{1}{R(T_0)}\frac{\Delta R}{\Delta T} = \frac{1}{V_{\mathrm{refl}}(T_0)} \cdot \frac{\Delta V_{\mathrm{refl}}(T)}{\Delta T}. \tag{3.17}$$

which states that, if the temperature sweep ΔT is applied at a certain frequency (it does not need to be very high, probably some Hz or tens of Hz would be enough), then the voltage sweep ΔV will be modulated at the same frequency and, consequently, easier to detect. This clearly requires a fast thermal stage, with a powerful Peltier cell, with careful mechanical design (in order to reduce as much as possible the thermal

deformations that would move the sample during the measurement), unfortunately not available at the moment.

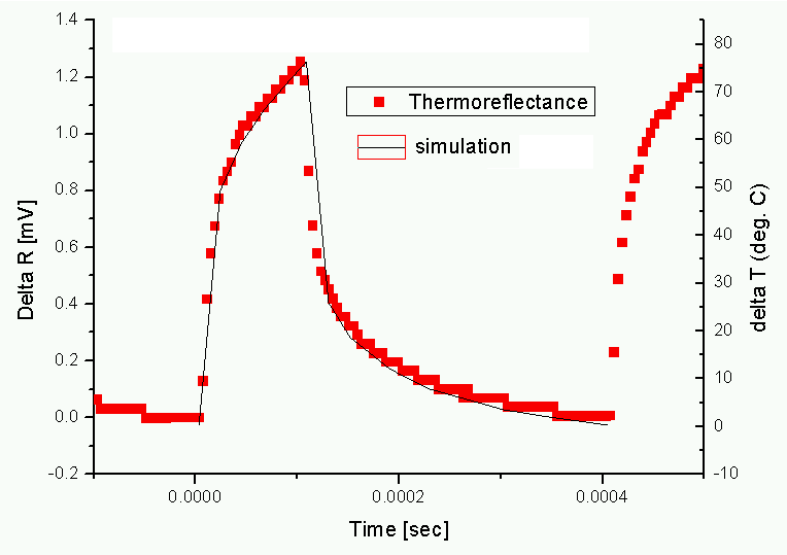


Figure 3.13 – Comparison between simulation results and termoreflectance measurement. The match between the two curves is excellent, but the reflectance signal still does not translate into a temperature scale.

3.6 Conclusions

This chapter has been focused on the development of a measurement bench to carry out surface temperature measurements on electron devices, exploiting the variation of the surface reflectance with temperature, the so called *thermoreflectance*.

This technique shows some advantages: (i) it is a contact-less method, (ii) it can measure temperature with a high-spatial resolution (in the order of microns), and (iii) it does not take any temperature average over the sample depth but it depends only on the sample *surface* temperature. The drawback of this technique is the low amplitude of the signal, i.e., the weak dependence of the reflectance on temperature.

A measurement bench has been designed and assembled, and LabVIEW firmware has been written in order to drive all the instrumentation remotely.

The thermoreflectance technique is well-suited for transient temperature measurement, in which the small signal is modulated and therefore easier to detect, thanks to the additional information that the signal is carrying (i.e., the fundamental frequency): the shape of the measured waveform was found to be in excellent agreement with thermal simulations, even if an absolute temperature scale is not available yet, due to the difficulties with the measurement of the calibration coefficient (i.e., the coefficient linking voltage and temperature); the noisy environment requires a modulation of the sample temperature even in the calibration phase, hence a fast thermal stage unavailable at the moment.

Chapter 4

Generation of high-voltage, nanosecond-duration electrical pulses for biomedical applications

This chapter deals with the development of a system for the generation of high-voltage (in the range of kiloVolts), short (in the range of tens of nanoseconds) pulses for non-thermal therapeutic treatment of biological cells and tissues.

4.1 Electroporation and nanoseconds pulse cell treatments

The treatment of biological tissues and cells with high-electric field pulses of short duration has been proven to have significant therapeutic effect [40, 41]. In particular, the electroporation process [42] consists in the application of pulses of intensity and duration resulting in the formation of small pores in the cell membrane; this is especially useful in case an external liquid has to be adsorbed by the cell.

Electroporation is not a recent treatment, since the first observations of electric-field-induced effects on mammalian cells can be found in the late 1950s [43].

In electroporation, the duration of the pulses ranges from some milliseconds down to few microseconds, while the amplitude ranges from hundreds of V/cm up to several kV/cm. It has to be noted that if *both* duration and amplitude increase, the energy delivered to the cells might get too high and, if the pulses were applied with a high repetition rate, the average power would increases and lead to cells overheating and destruction of the sample. The recent trend, thus, is to keep the delivered energy basically constant, by increasing the amplitude of the electrical field and, at the same time, decreasing the duration of the pulse for purely electrical, non-thermal treatment.

Following this trend, it has been found that when the duration of the pulses is on the order of 10^{-8} s or even less, different molecular processes are activated. At the

same time, the amplitude of the electrical field has to be greater than 10 kV/cm. The most attractive application is the induction of *apoptosis* in *cancer cells*.

Apoptosis is, in simple words, the programmed death of cells. This is a natural process and it is fundamental in order to keep the number of cells of a certain part of a mammalian organism under control. This natural process needs to be well controlled – which happens naturally in a healthy organism. Alterations of the apoptosis rate leads to health problems. An excess of apoptosis rate can be a source of degenerative illnesses, like Parkinson's disease; a defect in apoptosis rate, instead, is the process by which tumors grow. Clearly, the possibility of inducing apoptosis in cancer cells is of great scientific interest, and good results have been found in recent years [44, 45].

In this context, the aim of the following sections is to describe the design and the development of an electronic bench capable of generating such kind of pulses. Our bench has been designed to provide pulses with amplitude up to 5 kV and duration of 10 ns. As will be clearer later on, the amplitude of the electrical field applied to the load containing the cells suspension will be 50 kV/cm.

4.2 The design of the electrical bench

This section deals with the design of the electrical bench. After an overview of the architecture, each single part will be fully described.

4.2.1 The architecture of the bench

In this section, a detailed description of all the components of the bench will be given. We will start from the Pulse Shaping Network (PSN) and the load, moving backwards to the power supply section.

A schematic view of the bench is shown in Figure 4.1. The components shown in Figure 4.1 are currently being redesigned; the configuration depicted in Figure 4.1 is therefore not final.

In particular, the system has been designed to be powered by the electrical network; an AC/DC converter converts the 230 V_{ac} line input voltage to 300 V_{dc}, which is the input of the H-bridge converter, that generates a square-wave voltage with 300 V amplitude, 100 kHz frequency, and 50 % duty-cycle. These two converters make up the power section.

After the DC/AC converter a Diode-Capacitor (D-C) voltage multiplier generates a high DC voltage from the output voltage of the H-bridge. The D-C voltage multiplier topology [46] is an easy solution to generate high DC voltages starting from AC voltages, provided that the input frequency is high enough (this is the reason why the D-C multiplier is not powered directly by the network).

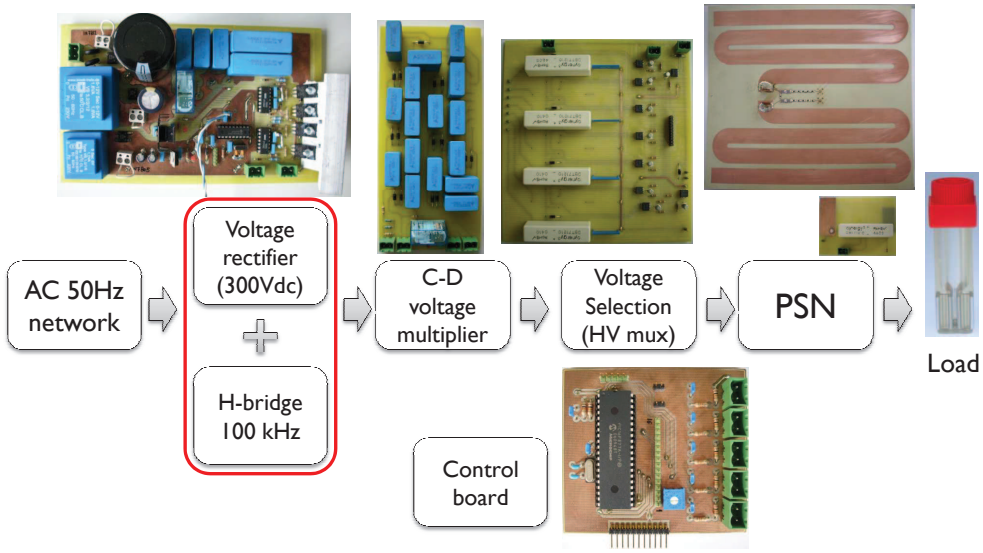


Figure 4.1 – Overview of the pulse generator bench.

The D-C multiplier can generate several output voltages, which can be selected by means of a high-voltage multiplexer, built with high-voltage relays. Then, this high DC voltage is used to charge the Pulse Shaping Network (PSN), which is made of a couple of transmission lines arranged in a *Blumlein* configuration [47]. In Figure 4.1, the Blumlein line is a microstrip. This solution is very compact, but (as will be explained clearly later) it requires a precise manufacturing process (i.e., wet etching should be used rather than milling), especially when employed with very high voltages. For this reason, the microstrip has been replaced by coaxial cables in the re-design phase: however, the line arrangement does not change.

Finally, the load to which the pulses are applied is a container for biological solutions, in which the cells are suspended (a *cuvette*). This container, visible in Figure 4.1, features a couple of electrodes (these containers are commonly used for electroporation treatments), which make the contact terminals of the load with the PSN.

After this brief introduction, each part of the bench will be described, in order to highlight its task, the design process, and its particular features. Since the design of the power section is strongly dependent on the design of the Blumlein line and the voltage multiplier, these two components will be fully described first.

The PSN. The Blumlein line is an easy way to generate a pulse with well-determined amplitude and duration.

The schematic diagram of a Blumlein line is shown in Figure 4.2. Basically, this arrangement is made of two identical lines, with characteristic impedance Z_0, and the same length. The load is connected between the two lines, and the load impedance Z_L has to be *twice* the line impedance ($Z_L = 2 \cdot Z_0$) in order to have the system matched.

The Blumlein line works as follows: the system is charged at a voltage HV, which determines the amplitude of the pulse; the switch S_d is then closed, thus shorting the line to ground. A pulse is then generated, and thanks to the reflections at the shorted end, at the load, and at the open end, a rectangular pulse with amplitude HV and duration equal to 2τ (where τ is the delay time of a single line) is delivered to the load.

A simple simulation has been performed with ADS and the results are shown in Figure 4.3. It is worth describing how ideal transmission lines are modeled in ADS. The ideal transmission line model is specified by giving the *electrical length* in degrees, the frequency at which the electrical length is evaluated, and the characteristic impedance. It is well known that the speed of light in vacuum is $c_0 \simeq 3 \cdot 10^8$ m/s, and in common transmission lines (e.g., coaxial cables) the phase velocity of electrical waves is $v = 0.66c_0 = 2 \cdot 10^8$ m/s: this is the velocity that ADS considers for ideal transmission lines. In our case, we use two transmission lines of 1 m, to set a 10 ns pulse. This can be shown remembering that

$$v = \nu \cdot \lambda, \tag{4.1}$$

where ν is the frequency and λ is the wavelength. Considering $v = 2 \cdot 10^8$ m/s, we can set the frequency $\nu = 200$ MHz in order to obtain a wavelength $\lambda = 1$ m. Since waves travel at $v = 2 \cdot 10^8$ m/s, the delay per unit length is equal to

$$\tau = \frac{\ell}{v} = \frac{1\,\text{m}}{2 \cdot 10^8\,\text{m/s}} = 5\,\text{ns/m} \tag{4.2}$$

or, in terms of electrical length, 5 ns/360°. In a Blumlein line, the pulse duration is twice the delay of either line, thus resulting in a 10 ns pulse on the load, as shown in Figure 4.3 (bottom).

The effect of load mismatch is easily accounted for by means of simulations. Figure 4.4 shows the shape of the pulse across the load in three different cases: line perfectly matched with load ($Z_L = 100\,\Omega = 2\text{Z}_0$), $Z_L = 50\,\Omega < 2Z_0$, and $Z_L = 150\,\Omega > 2Z_0$. When the load impedance is different from the matching value, some oscillations arise.

Another source of problems with Blumlein lines can arise from *the way* the pulse is measured. One could think of connecting two coaxial cables at the load terminals in order to measure the pulse, simply by connecting these two cables to an oscilloscope. This however modifies the shape of the pulse significantly, since the two coaxial cables used to connect the load with the scope act as *open stubs*, since the impedance of the scope is high (generally, 1 MΩ with a parallel capacitance of some $10 \div 15$ pF). The

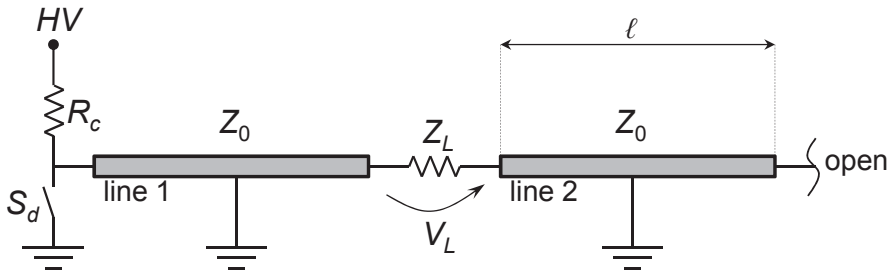


Figure 4.2 – A pulse generator made with a Blumlein line. The Blumlein arrangement is made of a couple of equal transmission lines, line 1 and line 2, with the same length ℓ and the same characteristic impedance Z_0. The lines are charged to the voltage HV by a charging resistance R_c, and discharged by closing the switch S_d.

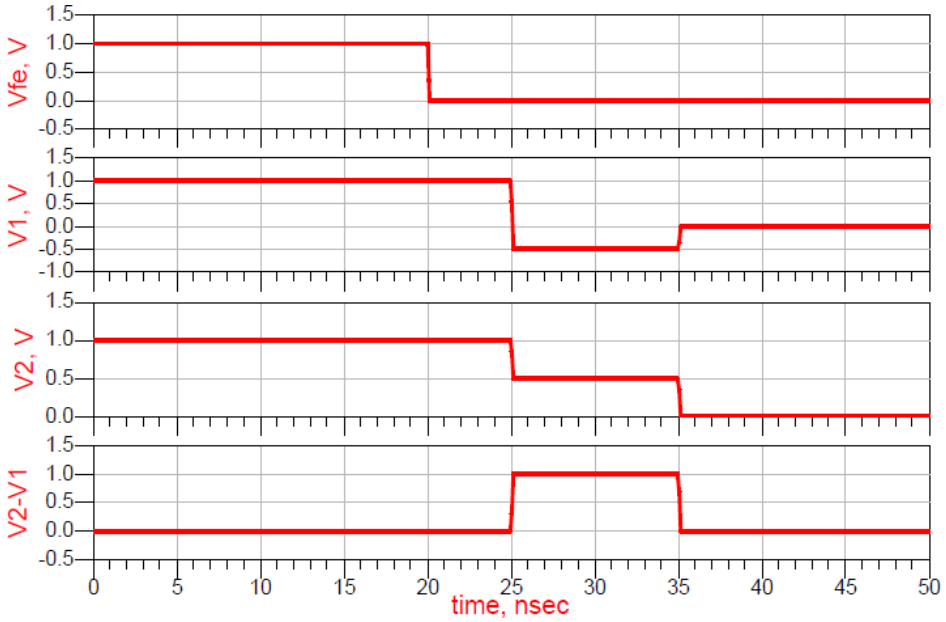


Figure 4.3 – Time-domain simulation results of a Blumlein line. From top: the pulse applied to the left end of line 1, the voltage at the left terminal of the load resistance, the voltage at the right terminal of the load resistance, and the voltage across the load (according to the notation in Figure 4.2). $\ell = 1\,\mathrm{m}$, $Z_0 = 50\,\Omega$, $Z_L = 100\,\Omega$.

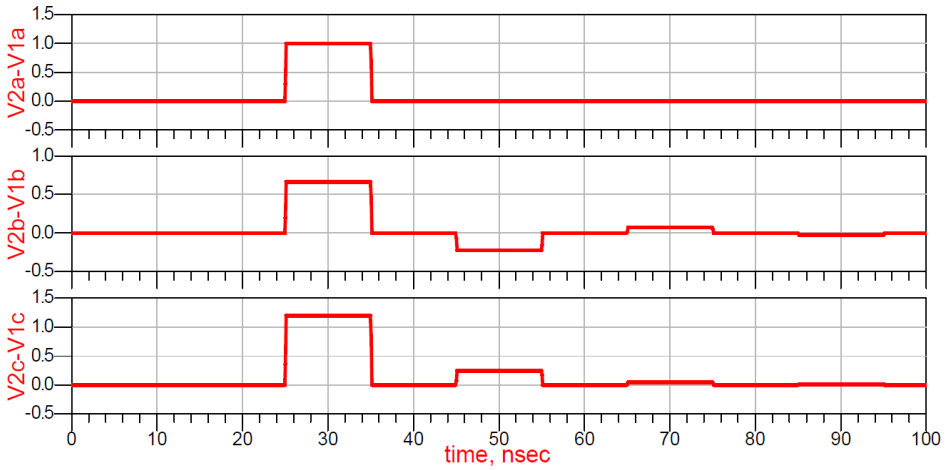


Figure 4.4 – The shape of the pulse across the load in a Blumlein line made of 50 Ω lines, for three different loads: (top) load perfectly matched (Z_L = 100 Ω), (center) $Z_L = 50\,\Omega$, (bottom) $Z_L = 150\,\Omega$.

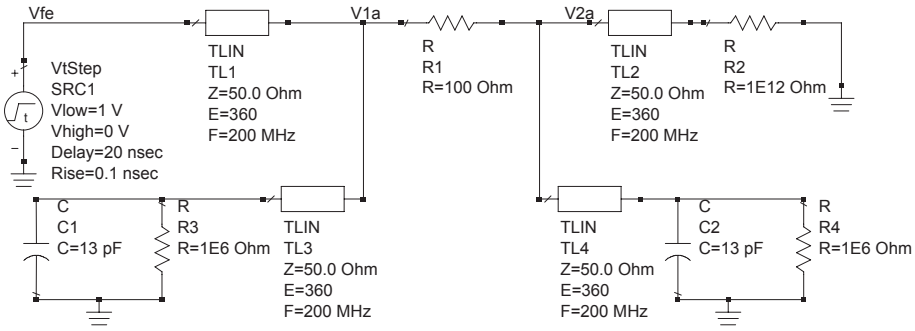


Figure 4.5 – Schematic used to simulate the effect of the connection of two coaxial cables (TL3 and TL4), to measure the pulse shape by an oscilloscope. The oscilloscope channel input impedance has been taken into account. $R_2 = 1\,\mathrm{T}\Omega$ is used to simulate the open circuit at the right end of transmission line 2 (Figure 4.2).

schematic used to simulate this condition is shown in Figure 4.5. The voltage generator simulates the falling pulse applied to the Blumlein line. In this case, we simulated a Tektronix oscilloscope in which the two channels have an input impedance given by the parallel of 1 MΩ resistance and 13 pF capacitance. The effect of this connection on the actual shape of the pulse is shown in Figure 4.6, in which the pulse is significantly

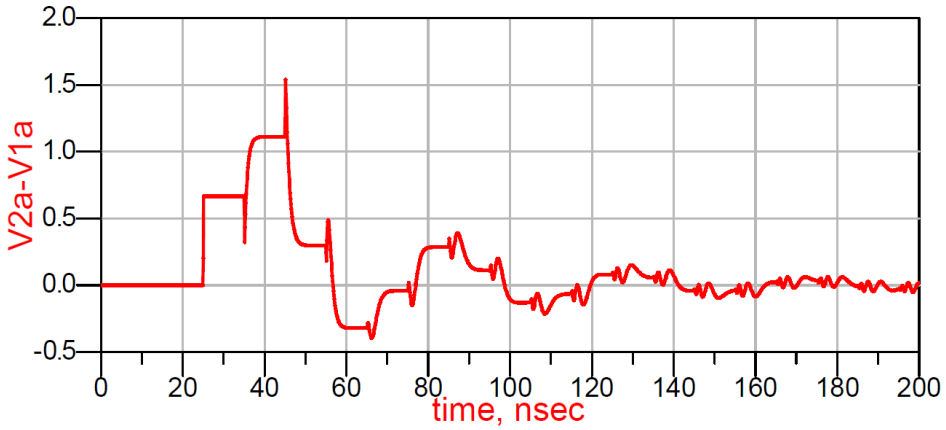


Figure 4.6 – Simulated effect of two additional coaxial cables used to measure the pulse in the circuit of Figure 4.5.

distorted. To avoid this distorting effect, one can use a resistive attenuator, with high input impedance, between each load terminal and the cables used to measure the pulse.

It is worth commenting on *how* the discharging switch has to be designed. This switch has to withstand the full charging voltage, and be able to short the line to ground through a very low-inductance path. A *reed* high voltage relay is a suitable solution, even if a mechanical switch limits the maximum frequency with which the pulses can be repeated ($1 \div 2\,\text{Hz}$ maximum frequency). In addition, in order to reduce the parasitic inductance of the discharging path, the use of ground planes surrounding the discharging path is mandatory.

For simplicity, the test of the PSN is currently being carried out using a $100\,\Omega$ resistor as the load. In the final system, the Z_L resistor will be replaced by the *cuvette*, and clearly the PSN has to be matched with the cuvette impedance Z_L^C (in general, $Z_L^C \neq Z_L$). For instance, if $Z_L^C = 50\,\Omega$, the characteristic impedance of the line must be $25\,\Omega$. The cuvette impedance Z_L^C depends on the distance d_e between its electrodes, the electrodes surface A_e, and the resistivity ρ_s of the solution in which the cells are suspended. Typical values are $d_e = 0.1\,\text{cm}$, $A_e = 1\,\text{cm}^2$, $\rho_s = 100 \div 500\,\Omega\text{cm}$, and the impedance is calculated as:

$$Z_L^C = \rho_s \cdot \frac{d_e}{A_e} = (100 \div 500) \cdot \frac{0.1}{1} = (10 \div 50)\,\Omega; \tag{4.3}$$

therefore, the PSN will have to be redesigned according to the characteristics of the cuvette used to perform the final experiments.

It has to be noted that Z_L^C is the *high-frequency* value of the cuvette impedance:

without going deeper into details, the actual impedance of the cuvette filled with biological solution can be modeled as an $R_S - C_{\mathrm{DL}}$ series load, in which the value of R_S is given by (4.3), and the capacitance C_{DL} (the value of which can be easily measured with an LCR; this value can be as high as some hundreds of nanoFarad or even more) is due to the creation of an *electrical double layer*, that is, a thin charge distribution between the electrode surface and the biological solution, which acts like a capacitor connected in series with the electrode[1]. This means that the actual impedance of the cuvette can be written (to first approximation) as a function of frequency f as:

$$Z_L^C(f) = \frac{1}{j2\pi f C_{\mathrm{DL}}} + R_S. \tag{4.4}$$

This *double-layer* capacitance C_{DL}, however, is not a problem when the load is pulsed: since the frequency spectrum of the applied pulse has significant high-frequency contents (tens of MegaHertz or even more), the impedance contribution due to C_{DL} can be neglected, and thus the high-frequency cuvette impedance Z_L^C can be considered as resistive. It is therefore possible to write that:

$$Z_L^C(f \uparrow\uparrow) \simeq R_S = Z_L^C; \tag{4.5}$$

this is the impedance which has to be considered for the design of the Blumlein line in the final configuration.

Once given d_e, the electric field E applied to the cell sample is given by

$$E = \frac{HV}{d_e}; \tag{4.6}$$

since the maximum value of HV that the Diode-Capacitor voltage multiplier can generate is roughly 5 kV (as it will be shown later on), it means that the maximum achievable E is 5 kV/mm, or 50 kV/cm.

The Diode-Capacitor voltage multiplier. This is the section that generates the high DC voltage charging the line. The D-C voltage multiplier is a ladder structure, where each stage is a peak detector made of two capacitors and two diodes. The schematic of a single-stage voltage multiplier, together with the waveforms of input and output voltages, is shown in Figure 4.7. The output voltage of a single cell is taken at the cathode of the second diode. To achieve higher voltages with the same power supply, several stages can be cascaded. The output voltage of each stage is available, i.e., for an n-stages voltage multiplier, made of $2n$ diodes and $2n$ capacitors, there are n DC voltages available.

[1]Therefore, the measurement of the conductivity of a solution cannot be performed in a DC fashion; to do so, AC measurements are performed usually, in order to lower the impedance of the double-layer capacitance until its module can be neglected with respect to the resistive part.

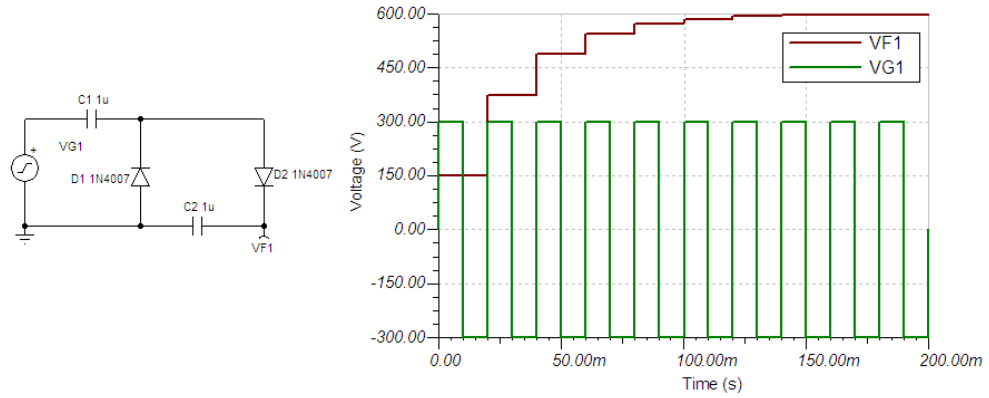


Figure 4.7 – (left) Single-stage voltage multiplier powered by a square wave voltage generator; (right) transient simulation results of input voltage (square wave with amplitude of 300 V and frequency 50 Hz) and output voltage.

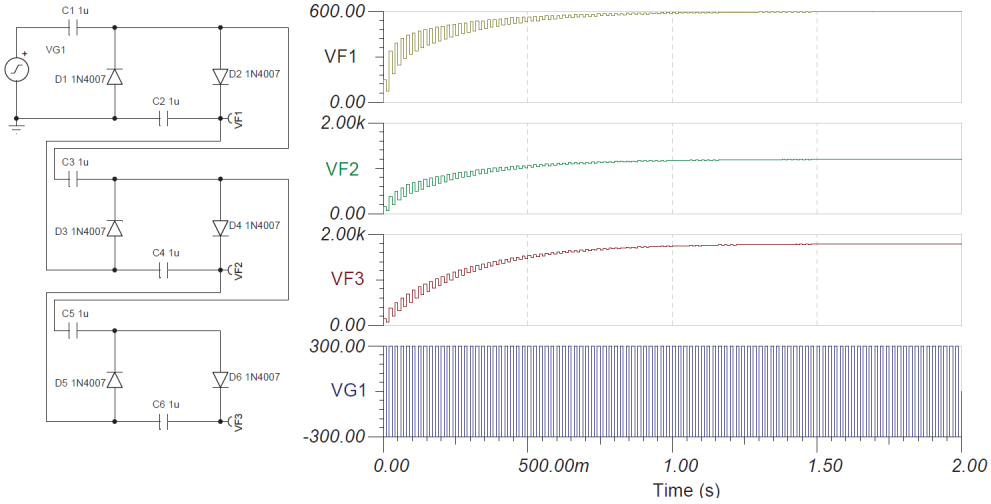


Figure 4.8 – (left) Schematic of a three-stage voltage multiplier; (right) simulated transient waveforms of the first, second and third output voltage, and voltage supply (from top to bottom).

An example of a 3-stage voltage multiplier is shown in Figure 4.8. It is worth noting that (i): as the number of stages increases, the time necessary to completely charge the voltage multiplier increases; (ii) the maximum voltage that the multiplier can supply (i.e., the voltage of the last stage) grows linearly with the number of stages:

starting from $V_{\mathrm{pk-pk}} = 600\,\mathrm{V}$, the output of the first stage is 600 V; the output of the second stage is 1.2 kV; the output of the third stage is 1.8 kV, and so on; in general the output of the n-th stage is $n \cdot V_{\mathrm{pk-pk}}$. In this configuration, the components have to be rated to withstand the $V_{\mathrm{pk-pk}}$ voltage (600 V in this case), independently of which stage they belong to. In particular, each diode and each capacitor has to withstand a voltage of 600 V.

In the previous cases, the frequency of the supply voltage was 50 Hz. Actually, with this frequency, a real voltage multiplier cannot work properly, unless it is made of one stage only. As the number of stages increases, the frequency of the power supply has to increase, otherwise the output impedance of each stage will be so high that even the connection of a digital multimeter (with 1 MΩ input impedance) will lower the output of the voltage multiplier: that is why the voltage multiplier cannot be connected to the network directly, but has to be powered by a high-frequency voltage source. This has been experienced in the laboratory: plugging in the multimeter is sufficient to discharge the multiplier. This effect can be simulated: for instance, let us consider the previously shown three-stage voltage multiplier: when it is powered with a 300 V, 100 kHz square-wave voltage source, it generates the right output of 1.8 kV; conversely, when the same multiplier is powered with a 300 V, 50 Hz square-wave, it cannot generate the correct voltage: the output voltage mean value is roughly 1.27 kV with a ripple of some 100 V.

The increase of input voltage frequency fixes the problem of the output impedance, but at the same time high current is needed to charge the voltage multiplier during each rise/fall of input voltage, and strong current peaks arise. Clearly, during the charging phase, the current is limited by parasitic inductances and resistances, which are hard to estimate and ineffective at suppressing the current surge. A solution is to implement a *soft-start* routine, in which the DC/AC converter generates short pulses at the beginning of the charging phase, and slowly increases the duty-cycle until the 50% final value is reached. To understand *why* this routine is effective, let us consider the variation of the current flowing through an inductance L (i.e., the parasitic inductance between the H-bridge and the voltage multiplier, VM):

$$\Delta I = \frac{\Delta V}{L} \cdot \Delta t = \frac{V_{\mathrm{Hbridge}} - V_{\mathrm{VM}}}{L} \cdot \Delta t, \tag{4.7}$$

where ΔV is the voltage applied to the inductance, and Δt is how long the charging voltage is applied. Let us consider the inductance initially discharged, as well as the voltage multiplier. Eqn. (4.7) states that the longer the time the charging voltage is applied, the higher is the current increase. If we, instead, apply short pulses to charge the voltage multiplier, the current peaks sunk by the voltage multiplier are lowered (the lower Δt, the lower ΔI). In addition to this, the voltage multiplier is getting more and more charged after each pulse (i.e., V_{VM} is increasing with time), meaning that the voltage across the inductance is getting lower and lower (clearly, V_{Hbridge} is

constant during the application of a pulse). Another way of looking at the charging process is in RMS terms: since the RMS value of the applied voltage slowly increases, the RMS current sunk by the voltage multiplier is lowered consequently, and thus the IGBTs in the H-bridge are less stressed.

Our voltage multiplier is made of 8 stages that allow to reach a maximum voltage of roughly 5.2 kV. The user can select different output voltages from the multiplier: 1.3 kV, 1.95 kV, 2.6 kV, 3.25 kV, 3.9 kV, 4.55 kV and 5.2 kV. All these voltages are selectable by a high-voltage multiplexer, featuring a high-voltage relay for each voltage multiplier output, which can connect or disconnect the selected output to the line. In our case, only the 1.3 kV, 2.6 kV, 3.9 kV and 5.2 kV output voltages are accessible (i.e., only these outputs are connected by the relays to the line, while the other intermediate voltages are not connected), mainly to avoid unnecessary expensive components in this prototyping phase. The voltage selection board is shown in Figure 4.1, where the multiplexer structure is clearly visible in the layout of the board. Each relay is driven by optocouplers, mainly for safety reasons.

The power supply section. The power supply section has been designed in order to meet some design constraints coming from the voltage multiplier. In particular, an AC/DC bridge rectifies the line voltage (230 V at 50 Hz sinusoidal voltage) in order to have a DC output of roughly $\sqrt{2} \cdot 230\,\mathrm{V} = 325\,\mathrm{V}$. An output capacitor bench ensures that the DC voltage does not decrease below a specified value when a load is connected. Since the relationship among the maximum output voltage ripple ΔV, the current sunk by the load I_{dc}, the network voltage frequency f, and the total bench capacitance C is

$$\Delta V = \frac{I_{\mathrm{dc}}}{2fC}; \tag{4.8}$$

it is easy to find that, at 50 Hz, to limit the ripple on the output voltage to 1% (roughly 3 V), with a maximum output current $I_{\mathrm{dc}} = 500\,\mathrm{mA}$, a capacitance of 2.5 mF is enough. This was obtained by paralleling $6 \times 470\,\mu\mathrm{F}$ electrolytic capacitors, rated for 450 V. Clearly, the bigger the capacitance, the bigger the *inrush* current sunk by the voltage rectifier. To limit this effect, a series resistance has been placed before the capacitor bench in order to slow down the charging process and thus limit the *inrush* current. The resistance is bypassed by a relay driven by a digital output of a microcontroller that implements a digital Schmitt trigger, in order to decide when bypassing the resistance. The output voltage is sampled by the A/D converter of the microcontroller, then this value is compared with a threshold set to 80% of the final output voltage. Once this threshold is exceeded, the relay is closed (i.e., the charging resistance is bypassed) in order to charge the capacitors quickly (since they are 80% charged, the inrush current is no longer a concern). The hysteresis of the Schmitt trigger (i.e., the difference between the threshold at which the relay is closed

and the threshold at which the relay is opened) is some percent of the final voltage. The threshold has been tuned empirically by testing the converter several times with different thresholds. The resistance used to charge the capacitors is 1.5 kΩ.

The output of the AC/DC converter is fed into the DC/AC converter, which is an H-bridge made with discrete IGBTs (600 V maximum collector-emitter voltage) driven by two IR2112 drivers (one for each leg). The driving signals are generated by a PIC microcontroller, which generates the signals for each leg, accounting also for the blanking time (500 ns). Due to frequency limitations of the microcontroller (maximum operating frequency 20 MHz, and 5 Mega Instructions-per-second), the maximum switching frequency that can be generated correctly is a bit higher than 100 kHz. A schematic view of the power section, made of the rectifier and the inverter, is shown in Figure 4.9.

The control board. Since the bench will have to work as a self-standing instrument, a local control board has been developed. At the moment the board is not used, since the bench is under a process of re-engineering, and some testing routines are being done by LabVIEW. However, the board has a user interface made of an LCD and several buttons allowing to (i) synchronize each single board with the others, (ii) select the voltage and the action type (e.g., to apply a limited number of pulses to the cuvette or run the bench continuously), (iii) stop the bench and put the high-voltage sections under safety conditions (i.e., discharge the line and the voltage multiplier: this is achieved by means of discharging relays which are used only when the bench is shut down).

The board features a PIC16F887A, and all the firmware has been written in C. The board is shown in Figure 4.1.

Table 4.1 summarizes all the features of the bench.

4.3 Conclusions

This work described in this section is still in progress, and in particular the bench has been re-designed under several points of view, in order to improve the reliability and robustness of each single part. This section will give some conclusive remarks about the results obtained at the time of writing, and the future developments.

The attention has been focused especially on the power section, that has to generate the 300 V square wave at 100 kHz to supply the voltage multiplier. The power section and the local power supplies (5 V and 12 V for the on-board logic and drivers) have been cased in different boxes, in order to facilitate the interconnection between the different sub-systems. A photograph of the power section is shown in Figure 4.10.

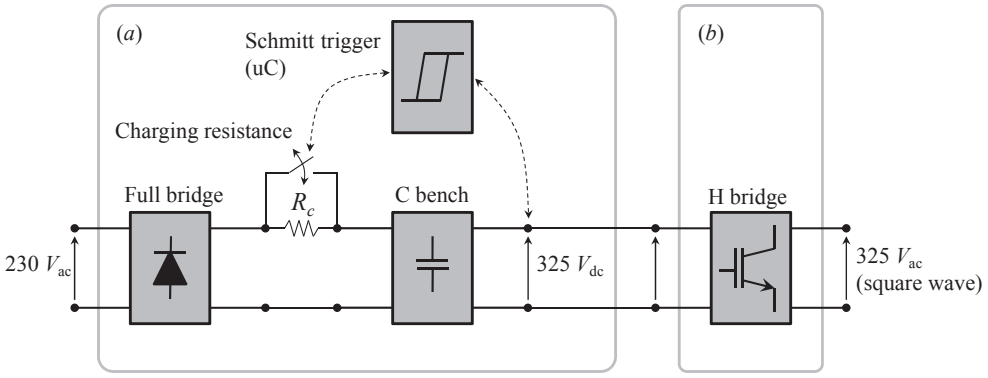


Figure 4.9 – A system view of the power section: (a) the rectifier; (b) the inverter.

Components	Task	Significant features/problems
Pulse Shaping Network	pulse generation	Blumlein arrangement
Discharging switch	discharge the PSN	HV switch, low inductance
HV mux	select the voltage from the voltage multiplier	made of several HV switches
D-C voltage multiplier (VM)	generate high voltages	need of high frequency voltage source
H-bridge (DC/AC converter)	high-frequency voltage source for VM	soft-start routine to charge the voltage multiplier
AC/DC converter	DC bus for H-bridge	inrush current-limiting circuit

Table 4.1 – Summary of the the various components in the bench.

The plastic case features all the connectors needed for proper operation of the whole power section, from the logic signals to the power input and output. The power section has been tested and it works correctly. The logic signals are still generated by an acquisition board driven by LabVIEW, for ease of debugging operations (this board will be replaced by the control board featuring the PIC16F877A, with the developed firmware running on it).

The power section has been connected to a 2 kΩ load, properly cooled to avoid excessive heating of the power resistors. The power load can be seen in Figure 4.11. It is made of two series-connected 1 kΩ power resistors, mounted on a metal foil, which is in turn connected to a CPU heat sink, cooled by a fan. The power source supplies a current of roughly 150 mA_{rms}, which corresponds to roughly 45 W dissipated by the load.

The next development will be the connection of the voltage multiplier in order to

check its proper operation and the robustness of the various components to the high-voltage. The voltage multiplier has been already tested, and it has been found to be able to generate up to roughly 4.9 kV, but some problems were found; in particular, the H-bridge microcontroller is reset when the H-bridge is working, suggesting a problem of EMI (Electro-Magnetic Interference) generated by the H-bridge itself. This effect was not experienced in the test phase, in which the load was resistive. EMI aspects have to be further investigated.

The last steps will be the connection of the Blumlein line, the measurement of the generated pulse, and the replacement of the acquisition board with the control board.

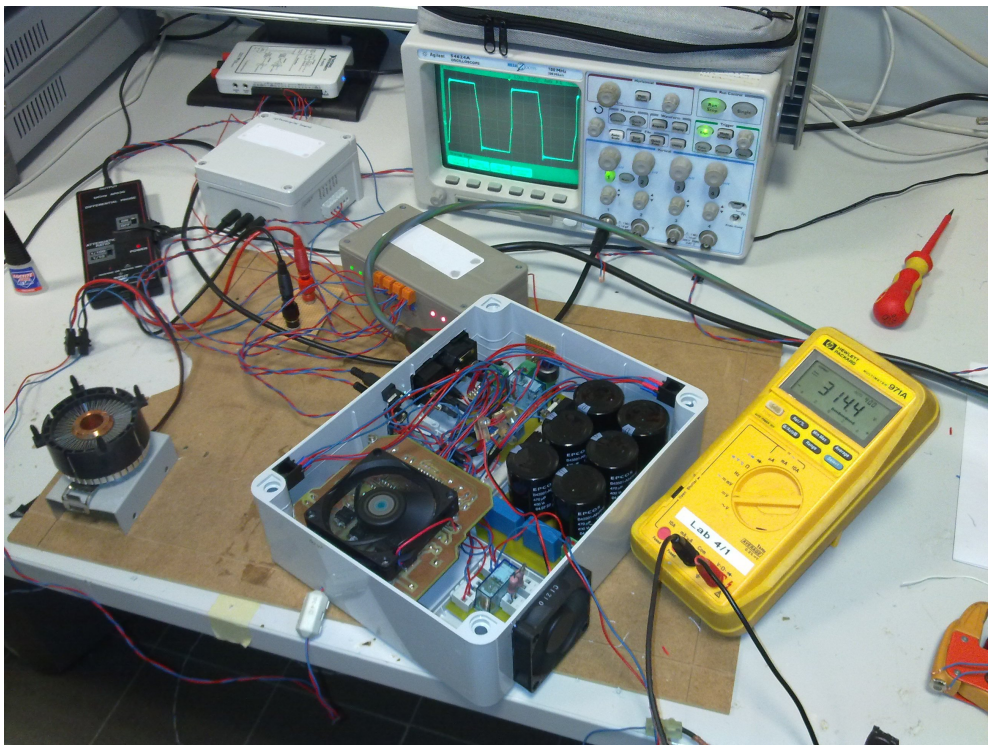


Figure 4.10 – An overview of the power section cased in a plastic box.

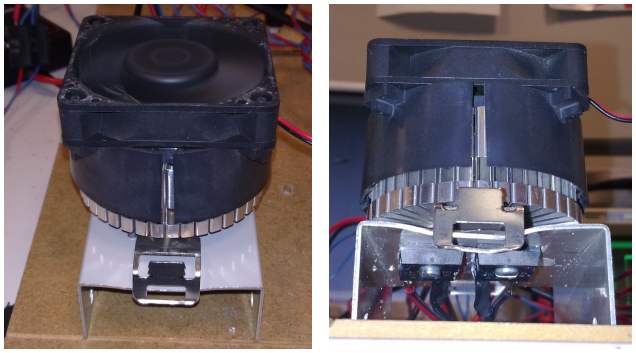


Figure 4.11 – The power load used to test the power supply: (left) top view of the fan used to cool down the resistors; (right) a bottom view in which the two TO247-packaged 1 kΩ power resistors are visible.

Appendix A

FEM modeling of thermomechanical stress in electronic assemblies for fatigue-oriented analysis

A.1 Introduction

This appendix is focused on the development of a FEM model that describes the damaging process of solder joints in power electronic assemblies; in general, each mechanical part is subjected to cyclic stress during its working life, and this continuous alternate stress can damage the component even if the amplitude of a single mechanical stress is far below the maximum strength of that given component. This progressive damaging process is well known in *mechanics of materials* and it is called *fatigue*.

Clearly, solder joints in power electronic assemblies play two roles: they ensure proper electrical connection between the device and the rest of the board, as well as *mechanical* connection between the considered device and the electronic board. This means that they are object of mechanical stress, like *thermomechanical* stress, vibrations (think about electronics for automotive), just to list some.

The FEM model shown here is at an early stage in modeling the *fatigue* phenomenon in solder joints. The description of the model will be preceded by a brief review about the mechanical concepts related to this topic.

A.2 Some concepts of *mechanics of materials*

This section will recall some basics of mechanics. It is well known that, if a force is applied to a body, a distribution of stress will arise and some deformations in the body will occur. If a rod of length ℓ, made of a given material, is *loaded axially* on

one end while the other end is mechanically *fixed* (that is, a *mechanical constraint* is applied), it will change its length of a quantity $\Delta\ell$. This means that the length of the rod will be $\ell + \Delta\ell$ during the application of the load. Clearly, $\Delta\ell$ will be positive if the rod is pulled, while $\Delta\ell$ will be negative if the rod is compressed.

In *mechanics of materials* is more common to describe the state of a body in terms of *stress* σ [Pa] and *strain* ε [−], where these two quantities are defined as *force per unit area* [N/m^2] and *relative deformation* $\Delta\ell/\ell$, respectively. Clearly, the area to define the stress is the effective area A on which the force F is applied, so that $\sigma = F/A$. For an axially-loaded rod, A is its cross-sectional area.

The relationship between σ and ε is known as the *mechanical characteristic* of a given material, and these characteristics are usually obtained by performing mechanical tests on specimen like that of Figure A.3, where an example of simulated test is shown as well. The *elasticity module*, or *Young's module* E [GPa], relates the stress and the strain while the material is in the *elastic* region:

$$\sigma = E \cdot \varepsilon; \tag{A.1}$$

this means that if a stress is applied, the body deforms, and when the stress is no longer applied, the body *returns to its original shape* (i.e., the behavior of the material is *linear*). Clearly, the *amplitude* of the applied stress cannot grow above certain limits, that is the reversibility of the deformation is possible only *below* a stress limit. This stress limit is called *yield stress*, and once this limit is reached[1], the behavior of the material changes, and the material starts to *plasticize*. This means that the deformation occurred in this region is *permanent*, that is, if the applied stress is removed, only the deformation occurred in the elastic region goes back to zero, while the *plastic* deformation remains. An example of mechanical characteristic accounting for the elastic and the plastic region of a given material is shown in Figure A.1. The characteristic shown in Figure A.1 is the simplest that describes the non-linearity in material behavior. A model like this is usually called *perfectly elasto-plastic* model. More detailed models exist (for instance the *exponential hardening* model and the *Ramberg-Osgood* model), which describe in a more detailed and realistic way the mechanical characteristic of a given material; these models, however, are not described here. All the considerations that will follow are made considering materials described by a mechanical characteristic like that of Figure A.1.

It is worth noting that, once the plastic region of a material is reached, it is no longer possible to know the actual deformation even if the stress is known, because σ is no longer dependent on ε in the plastic region, being σ now defined by $\sigma(\varepsilon) = \sigma_y$.

All the considerations done up to now are based on the following hypothesis: the specimen made of the considered material is *axially loaded*, like that shown in

[1] This is a simplistic view of the mechanical properties of a material; actually, the mechanical characteristics are defined by several parameters, that will not be listed here in order to keep the treatment simple.

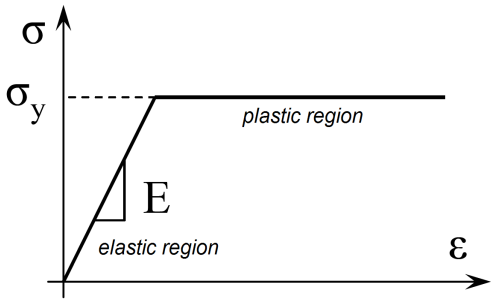


Figure A.1 – Perfect elasto-plastic stress-strain model for a given material.

Figure A.3. While the stress-strain state of axially-loaded bodies can be often determined manually, in the case of multi-axial loading it is mandatory to apply numerical methods, in particular the FE method.

In a general case, in order to define the *state of stress* of a point in the space, it is necessary to consider all the stress components in the space. They are 9, 3 *normal stresses* and 6 *shear stresses*. They are named σ_x, σ_y and σ_z, meaning normal stresses acting along x, y and z axis, respectively, and τ_{xy}, τ_{yx}, τ_{xz}, τ_{zx}, τ_{yz} and τ_{zy}, meaning that these shear stresses are acting on the plane normal to the axis denoted by the first subscript, and oriented along the axis identified by the second subscript. For instance, τ_{xy} is a shear stress over the plane normal to the x axis, and oriented along the positive y direction. In reality, by applying equilibrium equations of statics, it is possible to demonstrate that $\tau_{xy} = \tau_{yx}$, $\tau_{xz} = \tau_{zx}$, and that $\tau_{xz} = \tau_{zx}$. This means that only 6 components are necessary to fully describe the stress state of a point of a body in 3 dimensions, that is, 3 normal stresses and 3 shear stresses. The *stress tensor* σ_{ij} is a compact way to describe this concept:

$$\sigma_{ij} = \begin{pmatrix} \sigma_x & \tau_{xy} & \tau_{xz} \\ \tau_{yx} & \sigma_y & \tau_{yz} \\ \tau_{zx} & \tau_{zy} & \sigma_z \end{pmatrix} = \begin{pmatrix} \sigma_x & \tau_{xy} & \tau_{xz} \\ \tau_{xy} & \sigma_y & \tau_{yz} \\ \tau_{xz} & \tau_{yz} & \sigma_z \end{pmatrix}; \tag{A.2}$$

a graphic explanation of the various components of the stress tensor is shown in Figure A.2. Rather than considering each single stress components, it is useful to define an *equivalent* stress σ_e, named *Von Mises* stress, defined as follows:

$$\sigma_e = \frac{1}{\sqrt{2}}\sqrt{(\sigma_x - \sigma_y)^2 + (\sigma_y - \sigma_z)^2 + (\sigma_x - \sigma_z)^2 + 6 \cdot (\tau_{xy}^2 + \tau_{yz}^2 + \tau_{xz}^2)} \tag{A.3}$$

which will be very useful in evaluating simulation results for complex three-dimensional structures. At the same time, as for the stress tensor, a *strain* tensor is defined, and

it takes the following form:

$$\varepsilon_{ij} = \begin{pmatrix} \varepsilon_x & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{yx} & \varepsilon_y & \varepsilon_{yz} \\ \varepsilon_{zx} & \varepsilon_{zy} & \varepsilon_z \end{pmatrix}. \tag{A.4}$$

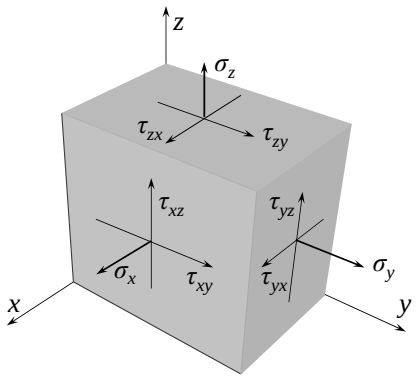


Figure A.2 – A small element on which a three-dimensional stress is acting.

A.2.1 Fatigue

Let us suppose we perform a yield test on a specimen, like that shown in Figure A.3 (together with all the steps usually followed to perform a mechanical FE analysis: once the specimen is drawn, it is discretized – mesh generation, mechanical loads and constraints are applied, and the model is solved). Once the yield stress is reached, if the deformation[2] applied to the specimen increases, the specimen will break shortly. This is actually one way to damage irreversibly a mechanical component, but this is not the only one. Damages can however grow up in mechanical components and bring them to rupture, even *each single applied stress* is far lower than the yield stress for that material. This phenomenon, in which a mechanical component fails after the application of a series of low intensity (far below the yield stress) repeated loads (not necessarily of the same amplitude), is called *fatigue*. A mechanical component can fail due to *fatigue damage* after some hundreds cycles, as well as after ten millions cycles. The mechanisms in the two cases are different, and therefore two types of fatigue are defined in mechanics: when the rupture of the component occurs below the (indicative) threshold of 10^5 cycles, we are talking about *Low Cycles Fatigue*,

[2]In the plastic region, the deformation increases significantly while the stress is almost constant.

LCF; when the rupture occurs above this threshold, we are in the case of *High Cycles Fatigue*, HCF.

Both processes activate, basically, micro-cracks (defects in material lattice, for instance) which are already present in the component before it is made working; near these micro-cracks, the stress is somehow "locally amplified" (*stress amplification factor*), and some localized plasticization occurs. Applying the load several times, this plasticized area grows up, and when the crack is grown up to a visible size (it is no longer on the micro-scale), the crack propagation is object of study of *fracture mechanics*. The difference between LCF and HCF is that, while HCF occurs with very low intensity loads, which are not forcing the component to work in the plastic region of the material (i.e., no visible plasticization occurs), the LCF is caused by loads which induce large area of plasticization on the component: this means that the yield stress is "globally" reached by the component (but the deformation is not sufficient to break the component), and not just locally near the initial defect (i.e., the micro-crack).

The goal of a fatigue analysis is to determine, *once known the stress distribution in a component*, the number of cycles N the component will last. This clearly can be done, once the *stress* distribution is known, in the HCF case, since stress and strain are proportional; in LCF, with the material that is plasticized, the *strain*, rather than the stress, has to be accounted: when the material is plasticized, the stress is almost equal to the yield stress, and thus the *strain* has to be measured. This is why the LCF is also known as *strain-based approach*, while the HCF is called as *stress-based approach*. While the HCF life is estimated by using the *Basquin* relationship (together with Wholer's diagrams), the LCF is evaluated by the *Coffin-Manson* law, that is the Basquin law counterpart.

The Basquin equation (HCF) takes the following form:

$$\sigma_a = \sigma'_f (2N_f)^b \tag{A.5}$$

where N_f is the number of load reversal (half of the number of cycles), σ'_f and b are material parameters, and σ_a is the *stress level* at which we want to evaluate how many cycles the component[3] will last.

The Coffin-Manson equation (LCF) takes the following form:

$$\frac{\Delta\varepsilon_p}{2} = \varepsilon'_f (2N_f)^c \tag{A.6}$$

where N_f is the number of load reversal (half of the number of cycles), ε'_f and c are

[3]This holds exactly if the component is *axially* loaded, as in the case of a specimen. Under multi-axial loading conditions, the number of cycles of the component is not necessarily equivalent to the number of cycles of the specimen made of that material and axially loaded, since the geometry plays a fundamental role in the stress-strain distribution.

material parameters, and $\Delta\varepsilon_p$ is the *strain* at which we want to evaluate how many cycles the component[4] will last.

As it can be seen, the Basquin and the Coffin-Manson equation are formally the same.

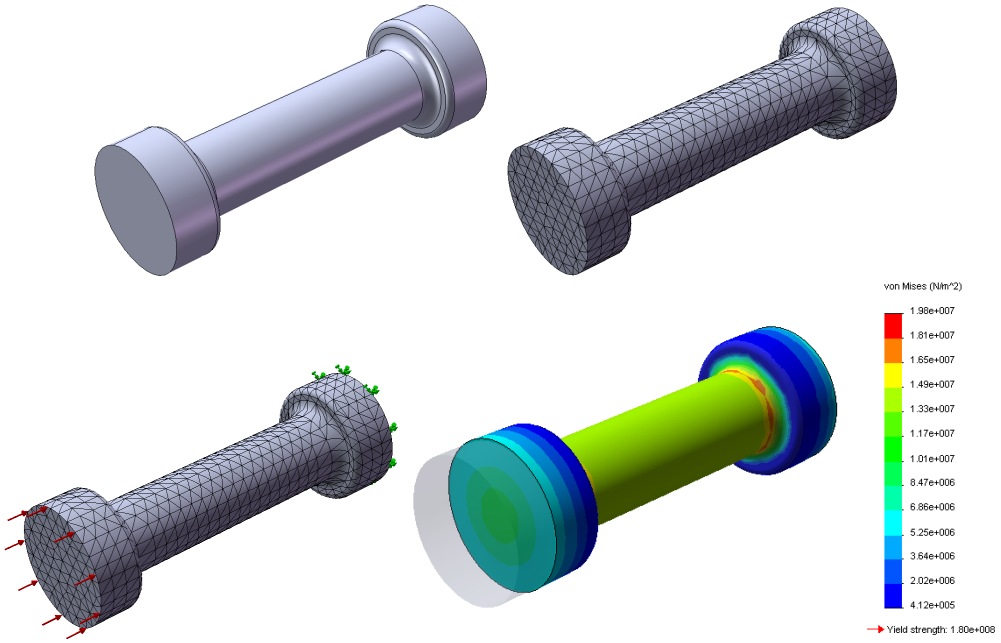


Figure A.3 – An example of mechanical analysis performed on a specimen: (top left) the specimen drawing; (top right) the discretization of the specimen for the FE analysis; (bottom left) the application of loads and constraints on the specimen; (bottom right) an example of Von Mises stress distribution over the structure after the solution of the FE model.

A.3 The LCF applied to solder joints reliability

The fatigue process in soldering joints damage is an LCF type [48, 49, 50]. This is due to the low value of *yield stress* of soldering alloys (i.e., the joints get plasticized quickly during thermal cycles). Electron devices used in power applications are subjected to power cycles, due to the power dissipated during their operation. These cycles are

[4]Clearly, what stated about the influence of geometry on HCF behavior holds for LCF as well.

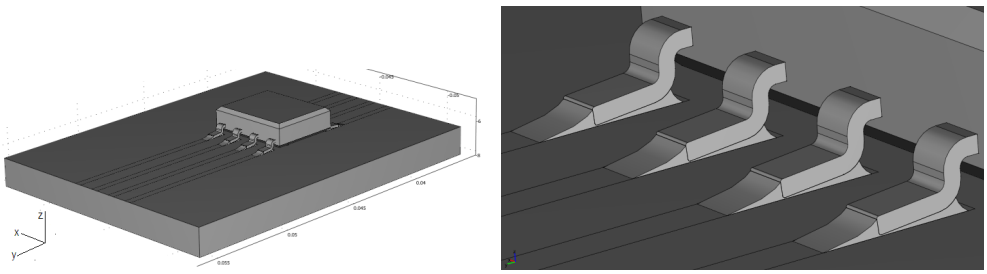


Figure A.4 – (left) View of the modeled structure. An SO-8 packaged device sits on a small FR4 board. (right) A blow-up of the structure highlighting the shape of the solder joints.

responsible of thermo-mechanical induced stress and strain [51], which can lead to solder joints failure due to LCF mechanisms.

Usually, the weakest part in electronic assemblies is the set of solder joints, and being these both the mechanical and electrical connection to the board, their reliability is of primary importance in evaluating the reliability of the whole circuit. In this work, a thermo-mechanical, non-linear[5] FEM model has been developed, and the Coffin-Manson law has been applied to solder joints, which have been assumed to be made of $Sn_{60}Pb_{40}$ solder alloy[6], discussing also about the limits of the model developed here.

A.3.1 The modeled system

The modeled system is shown in Figure A.4. An SO-8 packaged device (a MOSFET, for instance, even if the actual type of component is not important in this context) is mounted on a small FR4 board. The attention will be focused on the behavior of solder joints: usually they are the most stressed component, and at the same time the weakest components in the overall system. Particular attention has been put on the shape of the solder joints: their profile is similar to the profile obtained during a typical industrial soldering process [52]. A blow-up of the structure that highlights the shape of the solder joints is shown in Figure A.4 (right). When the device is working, the device self-heats, and every different part in the assembly will be at different temperature. Each material features an own Coefficient of Thermal Expansion (CTE), usually different from the CTE of the adjacent materials. Different deformations of adjacent bodies induce strain in the structure, since traction and/or compression occur

[5] A linear analysis is performed when all the materials are supposed to work in the elastic region. Accounting for plasticization requires a non-linear analysis.

[6] This alloy is no longer usable for soldering, due to RoHS directives in order to avoid leaded materials in electronic equipment.

Material	k [W/(m · K)]	CTE [K^{-1}]	E [GPa], ν	σ_y [MPa]
FR4	0.3	18×10^{-6}	22, 0.28	—
Epoxy	0.78	200×10^{-6}	1.8, 0.3	28
Copper	400	17×10^{-6}	110, 0.35	70
Silicon	163	4.15×10^{-6}	131, 0.27	—
$Sn_{60}Pb_{40}$	50	21×10^{-6}	10, 0.4	25

Table A.1 – Thermal and mechanical properties of the materials present in the assembly.

in each part of the overall assembly. Being the mechanical dynamics dependent on the thermal dynamics, the use of numerical methods to solve the problem is mandatory: in addition, the geometry is complex, and it has to be studied in its completeness. Since these components are subjected to thermal cycles, these cycles correspond to mechanical cycles as well. This leads to a cyclic strain, which tends to degrade the soldering joints. The study of this degradation falls in the LCF field, as said before, since usually a high degree of plasticization occurs in the soldering joints.

The model is set-up as follows:

1. the silicon die dissipates a fixed amount of power;
2. the temperature distribution in the assembly is calculated;
3. a non-linear analysis on stress and strain is computed;
4. the soldering joint with the highest degree of plasticization is found;
5. for this solder joint, the face with the highest degree of plasticization is found;
6. the maximum component of strain on this face is determined;
7. the Coffin-Manson law is applied to that joint, in order to evaluate the number of cycles before the starting of crack propagation.

The study of crack propagation falls in the field of *fracture mechanics*, and this topic is not covered here.

A.3.2 Material Properties and Boundary Conditions

The materials in the model are: FR4 (board), epoxy resin (device package), Copper (device pins and contacts), Silicon (die), and the $Sn_{60}Pb_{40}$ alloy (solder joints). Their thermal and mechanical properties are shown in Table A.1. They are taken from [53] and COMSOL material library. For FR4 and Silicon, the plasticization has been neglected.

It is common to describe the plasticity of the soldering joints only: in our case, however, when the yield stress (σ_y) is known for a certain material, it is used in the model. The model used to define the plastic behavior of materials is the perfectly elasto-plastic model, that describes the stress-strain relationship of a given material according to the graph in Figure A.1.

The silicon die is the heat source (the power is specified as power per unit volume, [W/m^3]). The assembly dissipates heat by convection through the horizontal surfaces of the board and of the device. The convection coefficient has been set to 30 W/(m^{2}K), which is representative of natural air convection cooling. The reference temperature (i.e., the ambient temperature) has been fixed to 25 °C. All the vertical surfaces were considered thermally insulated: these surfaces are small, and thus the heat exchange from these surfaces is negligible, that is they can be assumed to be thermally insulated for simplicity. From the mechanical point of view, the body is free except for the constraint applied to the highlighted border of the board, shown in Figure A.5. On that boundary, the board is mechanically *fixed*.

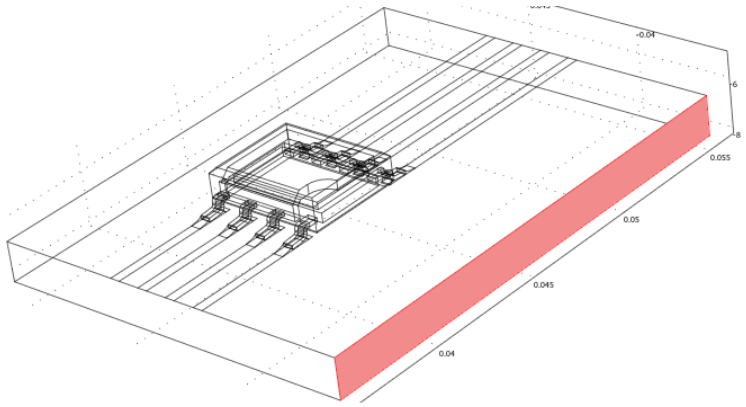


Figure A.5 – The mechanical constraint applied to the assembly: the board is fixed on the highlighted surface.

A.3.3 Cases of study

The model can be employed to study different operating conditions of the device. This will be useful to understand the flexibility of the model. In particular, we are interested in (i) modeling the actual operation of the device, when it is dissipating power in a cyclic way, and (ii) when the device is put in a *climatic chamber* in order to set the temperature sweep, for instance to reproduce a test according to the JEDEC standard (from −40 °C to 85 °C), simulating an accelerated life test. The second model is easily

obtained by imposing a fixed temperature on all the external surfaces of the assembly, and sweeping the temperature between the two limits. The latter is actually a model that can be easily compared with experimental results, since the temperature of the assembly is known (it is imposed by the climatic chamber).

When the dissipated power in the device changes, we are simulating a *power cycle*. When the temperature is changed from the external, i.e., with the system heated and cooled in a climatic chamber, we are simulating a *thermal cycle*.

A.3.4 How to analyze FEM simulation results

The theory about LCF is fully developed for single-axial stress: for instance, the mechanical tests on specimens are conducted by loading the sample along only one direction. In the case of multi-axial loading, there are different ways to apply the concepts of LCF to a three-dimensional, multi-axially loaded structure, without an universally accepted method. Here we follow a worst-case approach. The steps are the following:

1. determine the solder joint with the highest degree of plasticization: this can be done by evaluating the volume-averaged Von Mises stress in each single joint (the higher the average stress, the higher the degree of plasticization). Clearly, this is the joint in which a crack is most probable to develop;

2. determine the surface of joint (point 1) with the highest degree of plasticization; this can be done by evaluating the surface-averaged Von Mises stress over all the joint surfaces;

3. evaluate all the plastic components of total strain on the considered surface: these components are ε_{px}, ε_{py}, ε_{pz}, ε_{pxy}, ε_{pyz}, and ε_{pxz}. They must be evaluated in both the cases of maximum and minimum heat dissipation (note that these are two-dimensional strain distributions), i.e. $\varepsilon_{px}(P_{\max})$, $\varepsilon_{px}(0)$, $\varepsilon_{py}(P_{\max})$, $\varepsilon_{py}(0)$, and so on.

4. For each single strain component, compute the difference between the strain corresponding to the maximum power and the strain corresponding to the minimum power; they are $\Delta\varepsilon_{px} = \varepsilon_{px}(P_{\max}) - \varepsilon_{px}(0)$, $\Delta\varepsilon_{py} = \varepsilon_{py}(P_{\max}) - \varepsilon_{py}(0)$, and so on.

5. find the maximum (in absolute value) $\Delta\varepsilon_p = \max\{|\Delta\varepsilon_{px}|, |\Delta\varepsilon_{py}|, \dots\}$ of the strain differences, and apply the Coffin-Manson law to the solder joint. The absolute value is required since both positive and negative values can lead to crack initiation, due to positive deformation or negative deformation, respectively.

The Coffin-Manson law for $Sn_{60}Pb_{40}$ soldering alloy takes the following form:

$$\frac{\Delta\varepsilon_p}{2} = 0.325(2N_f)^{-0.5} \tag{A.7}$$

where $2N_f$ is the number of cycles, i.e., twice the number of full reversal, and $\Delta\varepsilon_p/2$ is half of the total applied strain during a cycle. The Coffin-Manson is applied in the following three cases of study: (i) the device dissipates 0.75 W, (ii) the device dissipates 0.15 W, and (iii) the device temperature is swept between $-40\,^\circ$C and $85\,^\circ$C. While the first two cases are representative of a typical operating condition of the device, the third is representative of the device put in a climatic chamber in order to perform an accelerated life test. Figure A.6 shows the numbered solder joints. As it will be illustrated below, usually the most stressed joints are the external ones, i.e., the joints 1, 4, 5 and 8: this can be intuitively explained observing that the external joints are the most free to move, and the more prone to deform, rather than the joints n. 2, 3, 6 and 7.

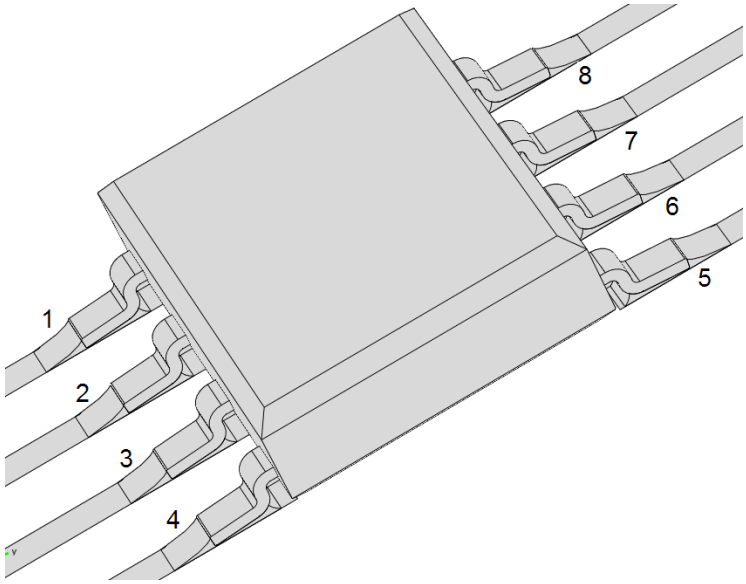


Figure A.6 – The solder joints numbered from 1 to 8, respectively.

Case 1: dissipated power $P_{\text{diss}} = 0.75\,\text{W}$

In order to generate several power cycles, the power in the Silicon die is set up as follows:

$$P_{\text{Si die}} = 0.75\sin^2(\pi \cdot t) \tag{A.8}$$

where t is a parameter which takes the following values: $t = 0, 0.25, 0.5, \ldots, 8$. In our case, the power has been defined as in (A.8) in order to have the power sweeping between 0 and 0.75 W (with intermediate values). A certain number of cycles is necessary for the structure to settle down: this is due to the hardening of the materials in the assembly, and the number of cycles depends on how the mechanical characteristic of materials have been modeled (perfectly elasto-plastic); in this model, two cycles are enough for the structure to stabilize.

The most stressed solder joints (i.e., that with the highest degree of plasticization) are the n. 1 and the n. 4, being the n. 4 slightly more stressed than the n. 1. This is reasonable, being two of the hottest solder joints and being also the more external. Thus, the analysis has been performed for that most stressed. The most stressed surface of that solder joint is that facing the device package, and on this face the maximum strain is ε_{pz}. The most stressed surface is clearly visible in Figure A.7. The maximum value of $\Delta\varepsilon_{pz} = 0.1162$, representative of strong strain along

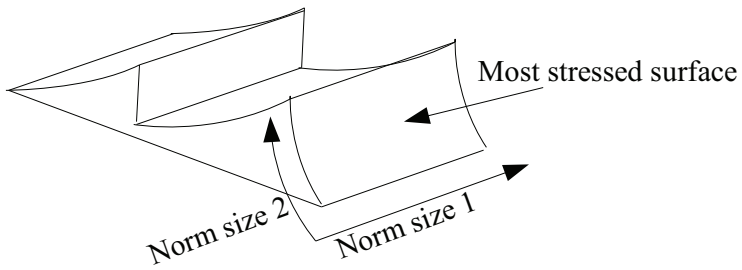


Figure A.7 – The most stressed surface of the most stressed solder joint in the case of an applied power cycle, pointed out by the arrow.

z direction: this is reasonable, since the device tends to lift from the surface of the board remarkably. A three-dimensional plot of $\Delta\varepsilon_{pz}$ versus normalized dimension 1 and dimension 2 of the solder joint surface is shown in Figure A.8. Using this value of $\Delta\varepsilon_{pz}$, the Coffin-Manson law returns $2N_f = 31$, which is far a low number of cycles for an electronic device. It has to be noted that, in this case, the temperature reached by the solder joints is roughly 130 °C, which is not a realistic operative condition for a circuit.

Case 2: dissipated power $P_{\text{diss}} = 0.15\,\text{W}$

Considered the results shown in the Case 1, the power has been reduced consequently, and applied to the Silicon die by the following law:

$$P_{\text{Si die}} = 0.15 \sin^2(\pi \cdot t) \tag{A.9}$$

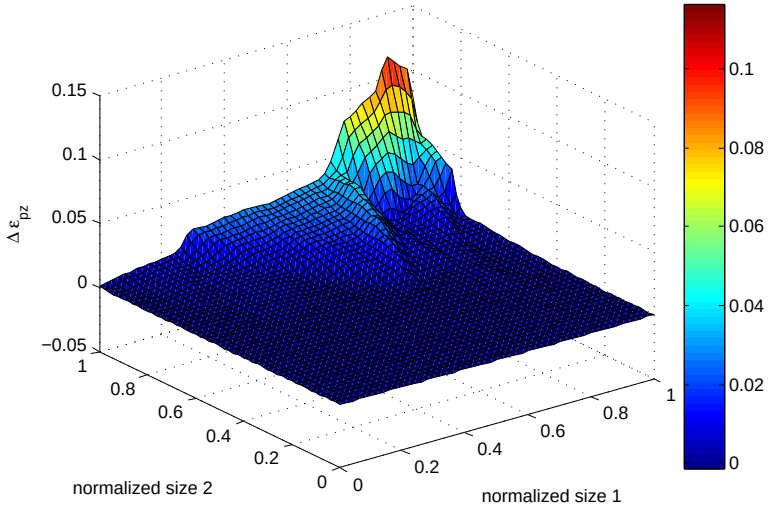


Figure A.8 – Three-dimensional plot of $\Delta\varepsilon_{pz}$ versus normalized dimension 1 and dimension 2 of the most stressed surface of the most stressed solder joint, when the device is dissipating 0.75 W.

where, in this case, the parameter t takes the values $t = 0, 0.5, 1, 1.5, 2$. Only maximum-dissipated-power and zero-dissipated-power conditions are considered, being the intermediate values not important in the analysis. When dissipating 0.15 W, the solder joints reach roughly 50 °C, a more realistic temperature value. Also in this case, the maximum strain occurs on the same face of the same solder joints of the Case 1, and the maximum strain $\Delta\varepsilon_{pz} = 4.2944 \times 10^{-4}$; with this value, the Coffin-Manson law returns roughly 2.6×10^6 cycles. This number is too high to be handled with the Coffin-Manson law, because this is more likely a high-cycle-fatigue (HCF) problem, to be studied with the Basquin relationship. A three-dimensional plot of $\Delta\varepsilon_{pz}$ versus normalized dimension 1 and dimension 2 of the solder joint surface is shown in Figure A.9.

Case 3: Device subjected to a thermal cycle

In this section, we are modeling an experiment in which the device is placed in a climatic chamber, and a temperature sweep from −40°C to +85°C is applied. In the model, the temperature is fixed on each surface of the system. In order to model a temperature sweep, an easy way to define it is:

$$T = 62.5 \cdot \sin(\pi \cdot t) + 22.5 \tag{A.10}$$

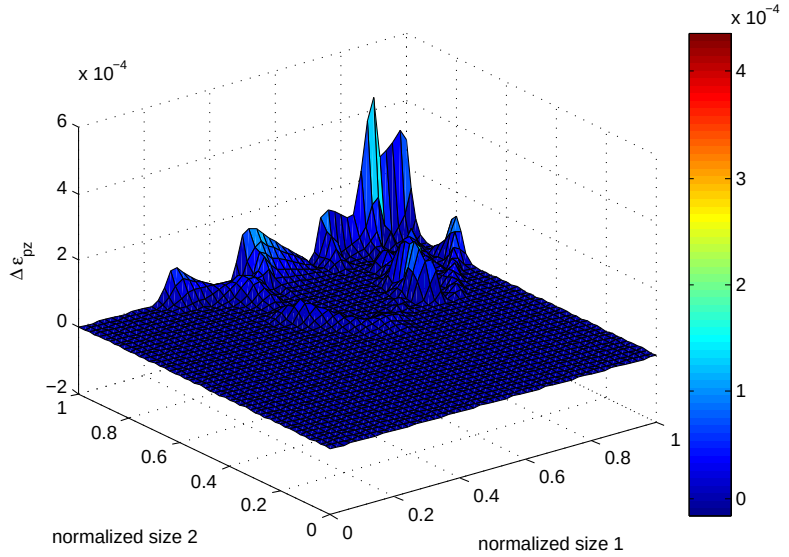


Figure A.9 – Three-dimensional plot of $\Delta\varepsilon_{pz}$ versus normalized dimension 1 and dimension 2 of the most stressed surface of the most stressed solder joint, when the device is dissipating 0.15 W.

where $t = 0.5, 1, 1.5, \ldots, 5.5$. It is easy to see that the temperature sweeps between -40°C and 85°C. For each temperature, the *steady-state* solution is evaluated, that is, the model does not describe the transient behavior of the assembly. Several temperature cycles are applied. Evaluating the volume-averaged Von Mises stress over the solder joints, the joints in position 4, 5 and 8 were found to be the most stressed, but with no significant differences: their volume-averaged Von Mises stress are really similar. So, we proceeded in evaluating the stress over the *surfaces* of these solder joints, in order to find the most stressed faces over the three joints. The lateral surface of solder joint 4, and the back surface of the solder joint 8 were found to be the most stressed. These surfaces are shown in Figure A.10. This model configuration is probably the most suitable to be supported by experimental results, since it is more easy to fix the temperature sweep on the whole system rather than turning on and off the device, in this case having unknown factors like convection coefficient, change in ambient temperature, and so on. The maximum strain values found in this case are the following: on the external face of solder joint 4, where the maximum strain was found to be $\Delta\varepsilon_{pxz} = 0.0155$, and on the back surface of solder joint 8, where the maximum strain was found to be $\Delta\varepsilon_{pz} = 0.0686$. Clearly, this means that in both cases the crack should start within few tens of cycles, being the maximum strain very similar to that obtained from the model in which the dissipated power was $P = 750$ mW. For the

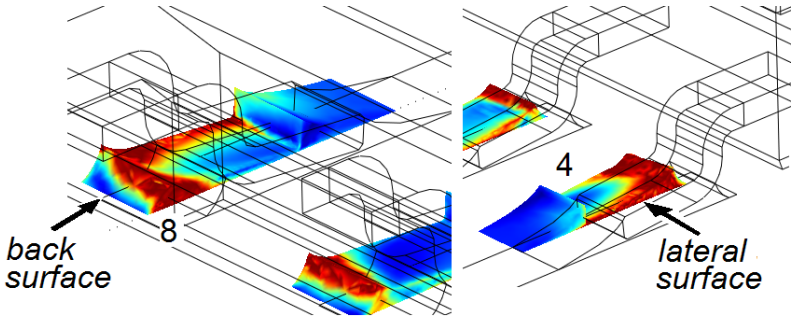


Figure A.10 – The most stressed surfaces in the structure, in the case of a thermal cycle. These are the lateral surface of the solder joint 4, and the back-surface of the solder joint 8. The red areas are that where the joints plasticize.

back surface of the solder joint 8, a two-dimensional plot of $\Delta\varepsilon_{pz}$ versus normalized dimension 1 and dimension 2 is shown in Figure A.11.

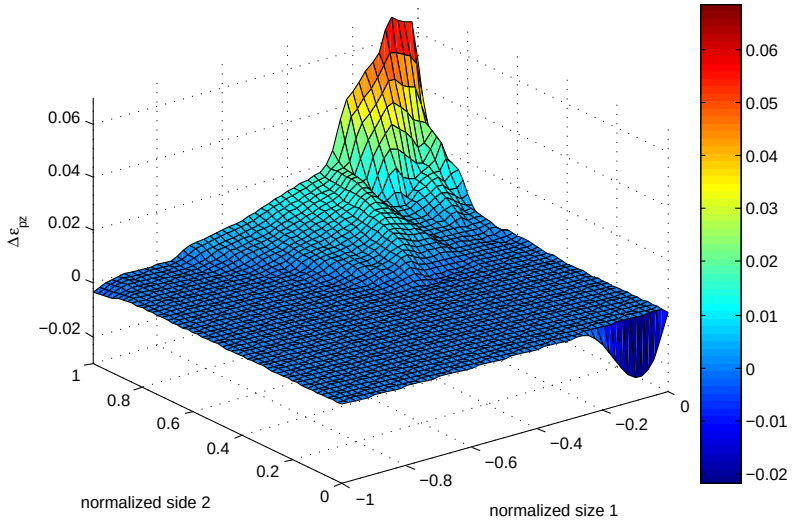


Figure A.11 – Three-dimensional plot of $\Delta\varepsilon_{pz}$ versus normalized dimension 1 and dimension 2 of the most stressed surface of the most stressed solder joint, after a thermal cycle.

A.3.5 Considerations about the model

Some final remarks about the model are given here. The dynamics studied by this model is quite complex, and being this an early model, it shows some limitations.

Crack process modeling

In solder joints the *creep* phenomenon [54, 55, 56] is the predominant one in crack initiation. *Creep* belongs to *visco-plastic* processes, it means that they depend on the amplitude of the plastic stress as well as the *strain rate*, i.e., the variation of applied strain with time $d\varepsilon/dt$. In this case, a transient simulation needs to be performed. In the model shown in this work, the time is present, but the simulation is not transient. This is because, even if the power is varying with time, the *steady-state* solution corresponding to a certain amount of power is evaluated. The transition between two different power levels is not accounted for. Actually, the Coffin-Manson law does not account for the strain-rate, but only for the strain amplitude. This is a first limitation of the model.

Mesh

Mainly due to computational limits, the mesh had to be kept at a coarse level. This can be seen in Figure A.12, in which only the mesh used to discretize the device is shown. Since the main attention has been focused on solder joints, the mesh of other domains has been kept at a coarser level. The model does not have symmetries (exact or approximate), since the temperature distribution over the assembly is not uniform (except from the case in which the same temperature is fixed on the whole assembly) and the induced strain are not uniform as well. The simulation of the overall system is necessary in order to find the most stressed solder joints. After that, a model describing only the solder joint and the pin should be developed, thus focusing the attention only on the most stressed part in the assembly. In this case, being the study reduced to a smaller system, a finer mesh can be generated.

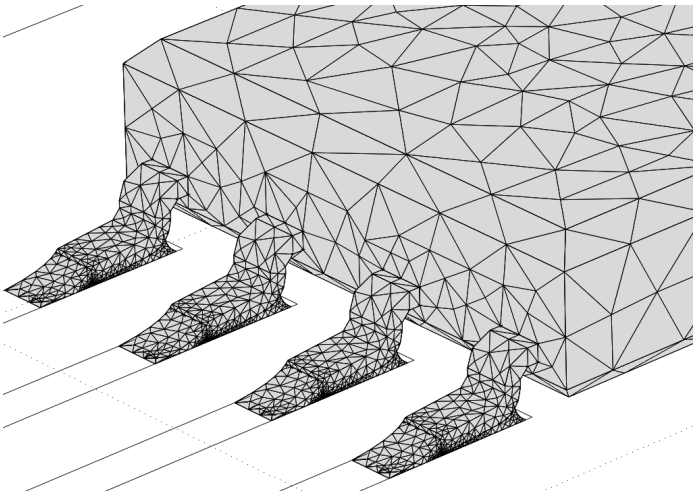


Figure A.12 – In inset of the structure showing the mesh related to the device only. The main attention has been focused on solder joints, thus the mesh of other domains has been kept at a coarser level.

Appendix B

FFT Matlab script

The MATLAB script that generates the signals and plots the results shown in Figure 3.11 is the following:

```
clear; clc;
N = 100000;
fs = 100E3;                % 100 kHz sampling
Ts = 1/fs;
Tobs = N*Ts;             % observation window
t = ([0:(N-1)]/N)*Tobs;
f0 = 150;
os = 0.85;                       % offset
x = 0.05*sin(2*pi*f0*t);         % our signal
e = 0.5*sin(2*pi*50*t);          % electrical noise
n = 0.5*randn(1,N);              % random noise
all = os + x + e + n;            % all sampled together
X = fft(all);
f = [0:(N/2-1)]/Tobs;
figure(1);
subplot(3,1,1); plot(t,os+x,'k'); grid on; axis([0 0.05 0.7 1]);
subplot(3,1,2); plot(t,os+x+e,'k'); grid on; axis([0 0.05 -1 3])
subplot(3,1,3); plot(t,all,'k'); grid on; axis([0 0.05 -1 3])
figure(2);
loglog(f,abs(X(1:N/2))/(N/2),'k'); grid on;
xlabel('frequency [Hz]');
ylabel('absolute value of FFT');
```

Appendix C

List of publications

Journals

[**J1**] M. Bernardoni, P. Cova, N. Delmonte, R. Menozzi, "Heat management for power converters in sealed enclosures: A numerical study," Microelectronics Reliability, vol. 49, pp. 1293-1298, 2009.

[**J2**] M. Bernardoni, N. Delmonte, P. Cova, R. Menozzi, "Thermal modeling of planar transformer for switching power converters," Microelectronics Reliability, vol. 50, pp. 1778-1782, 2010.

[**J3**] D. Mari, M. Bernardoni, G. Sozzi, R. Menozzi, G. A. Umana-Membreno, B. D. Nener, "A physical large-signal model for GaN HEMTs including self-heating and trap-related dispersion," Microelectronics Reliability, vol. 51, pp. 229-234, 2011.

[**J4**] P. Tenti, G. Spiazzi, S. Buso, M. Riva, P. Maranesi, F. Belloni, P. Cova, R. Menozzi, N. Delmonte, M. Bernardoni, F. Iannuzzo, G. Busatto, A. Porzio, F. Velardi, A. Lanza, M. Citterio, C. Meroni, "Power supply distribution system for calorimeters at the LHC beyond the nominal luminosity," Journal of Instrumentation (JINST), vol. 6, P06005, 2011, doi:10.1088/1748-0221/6/06/P06005.

[**J5**] P. Cova, M. Bernardoni, N. Delmonte, R. Menozzi, "Dynamic electro-thermal modeling for power device assemblies," Microelectronics Reliability, vol. 51, pp. 1948-1953, 2011.

Conferences

[**C1**] P. Cova, M. Bernardoni, F. Bertoluzza, "Feedback Control Simulation of Power Electronic Converters for Renewable Energies," Proc. International Conference on Clean Energy Power (ICCEP), June 9-11, 2009, p. 399-406, ISBN/ISSN: 978-1-4244-2544-0, doi: 10.1109/ICCEP.2009.5212024.

[**C2**] N. Delmonte, M. Bernardoni, P. Cova, R. Menozzi, "Thermal design of power electronic devices and modules," Proc. COMSOL Conf. 2009, Milano, Italy, Oct. 14-15, 2009, ISBN 978-0-09825697-0-2.

[**C3**] P. Cova, M. Bernardoni, "A MATLAB based approach for electro-thermal design of power converters," Proc. 6th Conference on Integrated Power Systems (CIPS) 2010, Nuremberg, Germany, March 16-18, ISBN 978-3-8007-3212-8.

[**C4**] D. Mari, M. Bernardoni, G. Sozzi, R. Menozzi, G.A. Umana-Membreno, B. D. Nener, "Compact modeling of GaN HEMTs including temperature- and trap-related dispersive effects," Proc. 2010 Workshop on Reliability Of Compound Semiconductors (ROCS 2010), Portland, OR, May 17, 2010, pp. 111-123.

[**C5**] M. Bernardoni, N. Delmonte, P. Cova, R. Menozzi, "Self-consistent compact electrical and thermal modeling of power devices including package and heat-sink," IEEE Proc. Int. Symp. Power Electronics, Electrical drives, Automation and Motion (SPEEDAM 2010), Pisa, Italia, June 14-16, 2010, pp. 556-561.

[**C6**] M. Bernardoni, N. Delmonte, R. Menozzi, "Modeling of the impact of boundary conditions on AlGaN/GaN HEMT self heating," Proc. 2011 Int. Conf. Compound Semiconductor Manufacturing Technology (CS-MANTECH), Indian Wells, CA (USA), May 16-19, 2011, pp. 229-232, ISBN 987-1-893580-17-6.

[**C7**] M. Bernardoni, N. Delmonte, G. Sozzi, R. Menozzi, "Large-signal GaN HEMT electro-thermal model with 3D dynamic description of self-heating," Proc. 41st European Solid-State Device Research Conf. (ESSDERC 2011), Helsinki, Finland, Sep. 12-16, 2011, pp. 171-174, ISBN 978-1-4577-0708-7.

[**C8**] M. Citterio, M. Riva, A. Lanza, M. Alderighi, S. Baccaro, P. Cova, F. Iannuzzo,G. Busatto, V. De Luca, A. Sanseverino, R. Menozzi, N. Delmonte, P. Tenti, G. Spiazzi, A. Paccagnella, S. Latorre, M. Bernardoni, "Power Converters for Future LHC," Proc. Topical Workshop on Electronics for Particle Physics TWEPP, Vienna (Austria), Sep. 26-30 2011.

Bibliography

[1] M. Legnani, P. Maranesi, G. Naummi, "SILC: A Novel Phase-Shifted PWM Converter," 1993, Fifth European Conference on Power Electronics and Applications.

[2] John H. Lienhard IV and John H. Lienhard V, "A heat transfer textbook," available for download at `http://web.mit.edu/lienhard/www/ahttv201.pdf`.

[3] N. Mohan, T. Undeland, W. P. Robbins, "Power Electronics," John Wiley and Sons, 3rd edition.

[4] M. Bernardoni, P. Cova, N. Delmonte, R. Menozzi, "Heat management for power converters in sealed enclosures: A numerical study," Microelectronics Reliability, vol. 49, pp. 1293-1298, 2009.

[5] M. Bernardoni, N. Delmonte, P. Cova, R. Menozzi, "Thermal modeling of planar transformer for switching power converters," Microelectronics Reliability, vol. 50, pp. 1778-1782, 2010.

[6] "Heat transfer module user guide," User manual of COMSOL ver. 3.5a.

[7] Bergquist Company. Gap pad comparison table. `www.bergquistcompany.com/pdfs /selectorGuides/GPComparison_Web_Table.pdf`.

[8] J. Michael Golio, "Microwave MESFETs and HEMTs," Artech House, 1991.

[9] R. S. Muller, T. I. Kamins, "Dispositivi elettronici nei circuiti integrati," Bollati Boringhieri.

[10] Paul R. Gray, Robert G. Meyer, "Analysis and design of analog integrated circuit, 2nd edition," Wiley, 1984.

[11] J. P. Ibbetson, P. T. Fini, K. D. Ness, S. P. DenBaars, J. S. Speck, and U. K. Mishra, "Polarization effects, surface states, and the source of electrons in AlGaN/GaN heterostructure field effect transistors," Applied Physics Letters, Vol. 77, Num. 2, 2000.

[12] Jong-Wook Lee and Kevin J. Webb, "A temperature-dependent nonlinear analytic model for AlGaN-GaN HEMT on SiC," IEEE Tran. Microwave theory and Techniques, Vol. 52, No. 1, January 2004.

[13] Walter R. Curtice, "A MESFET model for use in the design of GaAs integrated circuits," IEEE Trans. Microwave theory tech., Vol. MTT-28, 1980.

[14] Walter R. Curtice and M. Ettenberg, "A non-linear GaAs FET model for use in the design of output circuits for power amplifiers," IEEE Trans. Microwave theory and techniques, vol. MTT-33, no. 12, Dec. 1985.

[15] James G. Rathmell and Anthony E. Parker, "Circuit implementation of a theoretical model of trap centres in GaAs and GaN devices," Proc. SPIE, vol 6798, 2007.

[16] Anthony E. Parker and James G. Rathmell, "Comprehensive model of microwave FET electro-thermal and trapping analysis," Proc. Workshop on Applications in Radio Science (WARS 08), 10-12 Feb. 2008.

[17] D. Mari, M. Bernardoni, G. Sozzi, R. Menozzi, G. A. Umana-Membreno, B. D. Nener, "A physical large-signal model for GaN HEMTs including self-heating and trap-related dispersion," Microelectronics Reliability, vol. 51, pp. 229-234, 2011.

[18] M. Bernardoni, N. Delmonte, R. Menozzi, "Modeling of the impact of boundary conditions on AlGaN/GaN HEMT self heating," Proc. 2011 Int. Conf. Compound Semiconductor Manufacturing Technology (CS-MANTECH), Indian Wells, CA (USA), May 16-19, 2011, pp. 229-232, ISBN 987-1-893580-17-6.

[19] M. Bernardoni, N. Delmonte, G. Sozzi, R. Menozzi, "Large-signal GaN HEMT electro-thermal model with 3D dynamic description of self-heating," Proc. 41st European Solid-State Device Research Conf. (ESSDERC 2011), Helsinki, Finland, Sep. 12-16, 2011, pp. 171-174, ISBN 978-1-4577-0708-7.

[20] Vladimir Székely, "Identification by RC networks by deconvolution: chances and limits," IEEE Tran. on Circuits and Systems, vol. 45, no. 3, March 1998.

[21] M. Bernardoni, N. Delmonte, P. Cova, R. Menozzi, "Self-consistent compact electrical and thermal modeling of power devices including package and heat-sink," IEEE Proc. Int. Symp. Power Electronics, Electrical drives, Automation and Motion (SPEEDAM 2010), Pisa, Italia, June 14-16, 2010, pp. 556-561.

[22] N. D. Arora, J. R. Hauser, D. J. Roulston, "Electron and hole mobilities in silicon as a function of concentration and temperature," IEEE Trans. El. Dev., ED-29, pp. 292-295, 1982.

[23] J. Zarębski, K. Górecki, "The Electrothermal Large-Signal Model of Power MOS Transistors for SPICE," IEEE Trans. on Power Electronics, Vol. 25, No. 5, May 2010, p. 1265-1274.

[24] P. Cova, M. Bernardoni, N. Delmonte, R. Menozzi, "Dynamic electro-thermal modeling for power device assemblies," Microelectronics Reliability, vol. 51, pp. 1948-1953, 2011.

[25] A. Sarua et al., "Thermal Boundary Resistance Between GaN and Substrate in AlGaN/GaN Electronic Devices," IEEE Trans. El. Dev., vol. 54, Dec. 2007, pp. 3152-3158.

[26] W. Molzer et al., "Self Heating Simulation of Multi-Gate FETs," Proc. ESSDERC 2006, pp. 311-314.

[27] M. Braccioli et al., "Simulation of self-heating effects in different SOI MOS architectures," Solid-State Electron., vol. 53, pp. 445-451, 2009.

[28] S. Kolluri, K. Endo, E. Suzuki, K. Banarjee, "Modeling and Analysis of Self-Heating in FinFET Devices for Improved Circuit and EOS/ESD performance," IEDM07.

[29] S. Makovejev, S. Olsen, J. P. Raskin, "RF Extraction of Self-Heating Effects in FinFETs," IEEE Trans. El. Dev., vol. 58, no. 10, pp. 3335-3341, 2011.

[30] W. Liu, M. Asheghi, "Thermal conduction in ultrathin pure and doped single-crystal silicon layers at high temperatures,"J. Appl. Phys. 98, 123523 (2005); doi: 10.1063/1.2149497.

[31] C. Filloy, G. Tessier, S. Holé, G. Jerosolimski, D. Fournier, "The contribution of thermoreflectance to high resolution thermal mapping," Sensor Review, Volume 23, Number 1, 2003, pp. 35-39.

[32] M. Farzaneh, K. Maize, D. Luerben, J. A. Summers, P. M. Mayer, P. E. Raad, K. P. Pipe, A. Shakouri, R. J. Ram, Janice A. Hudgings, "CCD-based thermoreflectance microscopy: principles and applications," J. Phys. D: Appl. Phys. 42 (2009), pp. 1-20.

[33] S. Dilhaire, S. Grauby, W. Clayes, "Calibration procedure for temperature measurements by thermoreflectance under high magnification conditions," Applied Physics Letters, Vol. 84, No. 5, 2004.

[34] "Photodiode characteristics and application," UDT Sensor inc. Application note.

[35] "Compensate transimpedance amplifiers intuitively," Texas Instruments Application note.

[36] "OPA 703 datasheet," Texas Instruments (available on web).

[37] "Noise analysis of FET transimpedance amplifiers," Texas Instrument Application note.

[38] "Design photodiode amplifier circuit with OPA128," Texas Instruments Application note.

[39] National Instruments website and LabVIEW on-line support, `http://www.ni.com/it`.

[40] M. J. Heish, T. Salameh, I. Camarillo, R. Sundararajan, "Irreversible Electroporation Effects: A Drug-Free Treatment for Cancer," Proc. ESA Annual Meeting on Electrostatics, 2007.

[41] E. Maor, A. Ivorra, J. Leor, B. Rubinsky, "The Effect of Irreversible Electroporation on Blood Vessels," Tech. Cancer Research and Treatment, Vol. 6, No. 4, August 2007.

[42] J. C. Weaver, "Electroporation of Cells and Tissues," IEEE Tran. on Plasma Science, Vol. 28, No. 1, pp. 24-33, February 2000.

[43] R. Stämpfli, "Reversible breakdown of the excitable membrane of a Ranvier node," Anais da Academia Brasileira de Ciencias 30, 57-63 (1957).

[44] S. J. Beebe, P. M. Fox, L. J. Rec, K. Somers, R. H. Stark, and K. H. Schoenbach, "Nanosecond Pulsed Electric Field (nsPEF) Effects on Cells and Tissues: Apoptosis Induction and Tumor Growth Inhibition," IEEE Transactions on Plasma Science, Vol. 30, no. 1, Feb. 2002.

[45] Karl H. Schoenbach, Barbara Hargrave, Ravindra P. Joshi, Juergen F. Kolb, Richard Nuccitelli, Christopher Osgood, Andrei Pakhomov, Michael Stacey, R. James Swanson, Jody A. White, Shu Xiao, Jue Zhang, Stephen J. Beebe, Peter F. Blackmore and E. Stephen Buescher, "Bioelectric Effects of Intense Nanosecond Pulses," IEEE Transactions on Dielectrics and Electrical Insulation Vol. 14, No. 5; October 2007.

[46] C. K. Dwivedi, M. B. Daigavane, "Multi-purpose low cost DC high voltage generator (60 kV output), using Cockcroft-Walton voltage multiplier circuit," Intern. Journal of Science and Technology Education Research, Vol. 2(7), pp. 109-119, July 2011.

[47] J. H. Crouch, W. S. Risk, "A Compact High Speed Low Impedance Blumlein Line for High Voltage Pulse Shaping," Rev. Sci. Instrum. 43, 632 (1972); doi: 10.1063/1.1685710.

[48] C. Kanchanomai, Y. Miyashita, Y. Mutoh, "Low cycle fatigue behavior and mechanisms of a eutectic Sn-Pb solder 63Sn/37Pb," International Journal of Fatigue 24, (2002), 671-683.

[49] Chaosuan Kanchanomai, Yukio Miyashita, Yoshiharu Mutoh, "Low-Cycle Fatigue Behavior of Sn-Ag, Sn-Ag-Cu, and Sn-Ag-Cu-Bi Lead-Free Solders," Journal of Electronic Materials, Vol. 31, No. 5, 2002.

[50] John H.L. Pang, B.S. Xiong, T.H. Low, "Low cycle fatigue study of lead free 99.3Sn-0.7Cu solder alloy," International Journal of Fatigue 26, (2004), 865-872.

[51] P. Towashiraporn, K. Gall, G. Subbarayan, B. McIlvanie, B.C. Hunter, D. Love, B. Sullivan, "Power cycling thermal fatigue of Sn-Pb solder joints on a chip scale package," International Journal of Fatigue 26, (2004), 497-510.

[52] D. Romm, B. Lange, and D. Abbott, "Evaluation of Nickel/Palladium-Finished ICs With Lead-Free Solder Alloys," Texas Instruments Application Report SZZA024, January 2001.

[53] Gretchen A. Bivens, "Reliability Assessment Using Finite Element Technique," RADC-TR-89-281 In-House Report, 1989.

[54] X. P. Zhang, C. B. Yu, S. Shrestha, L. Dorn, "Creep and fatigue behaviors of the lead-free Sn-Ag-Cu-Bi and Sn60Pb40 solder interconnections at elevated temperatures," J Mater Sci: Mater Electron (2007), 18:665-670, DOI 10.1007/s10854-006-9078-3

[55] S. Ridout, C. Bailey, "Review of methods to predict solder joint reliability under thermo-mechanical cycling," Fatigue Fract Engng Mater Struct 30, 400-412, doi: 10.1111/j.1460-2695.2006.01065.x.

[56] Z.N. Cheng, G.Z. Wang, L. Chen, J. Wilde, K. Becker, "Viscoplastic Anand model for solder alloys and its application," Soldering and Surface Mount Technology, 12/2 (2000), 31-36.

Acknowledgments

The order in which people are cited here is absolutely not related to their importance during this path: everyone has played his role, in different situations but with the same importance. First of all my research group, and in particular Nicola Delmonte, Paolo Cova and professor Roberto Menozzi, for making up a very nice and skilled group to work within, under professional aspect as well as human aspect; Cristian, my university and PhD course mate, which has given me very precious technical assistance in several occasions, and the "digital people" of the building 4, for their nice company (Fabio T. and Andrea R.). A special acknowledgment to Andrea S. and Kévin M., for helping me in the development of the electrical bench for pulse generation during their thesis activity.

My acknowledgments go also to all the PostDocs and PhD students at the CDTR (Centre for Device Thermography and Reliability) research group, H.H. Wills Dept. of Physics, University of Bristol (UK), headed by prof. Martin Kuball, and in particular to James W. Pomeroy for his great support during my english experience.

Out of the academic sphere, I do have to thank my family, for supporting me during my academic path and the PhD course after, and in particular my mother, for her precious suggestions and support.

And for last (but not the least!), all the guys with which I have spent these years, in particular Marco M., "er dottò" Mattia M., Emiliano, Mattia C., Tommaso and Francesco (better known as "Penz", for us)... the musicians' and cyclists' crew.

To all these people a sincere *thanks*.