# D7.7

(original Deliverable No D7.11)

# Efficient Modelling of Electro-Thermo-Mechanical Stress During Fast Power Cycling Operation of an IC

| | |
|---|---|
| **Project title:** | iDev40: Integrated Development 4.0 |
| **Project start date:** | 1st of May 2018 |
| **Duration:** | 36 months |

| | |
|---|---|
| **Deliverable type:** | Report |
| **Activity and Work package contributing to the deliverable:** | Task 7.2.1 / WP 7, Task 2.4.1/ WP 2 |
| **Due date:** | 30.04.2020 |
| **Actual submission date:** | 29.04.2020 |

| | |
|---|---|
| **Responsible lead organization:** | Technical University Cluj-Napoca (UTC) |
| **Lead Author:** | Marius Purcar (UTC) |
| **Dissemination level:** | Public |
| **Reviewer:** | Cristian Boianceanu (IFRO) |
| **Revision Date:** | 27.04.2020 |
| **Document Reference** | D82 (original D86) |

This project has received funding from the ECSEL Joint Undertaking under grant agreement No 783163. The JU receives support from the European Union's Horizon 2020 research and innovation programme. It is co-funded by the consortium members, grants from Austria, Germany, Belgium, Italy, Spain and Romania.

## Authors (all contributors to the Deliverable)

| | Name | Chapters |
|---|---|---|
| Technical University Cluj-Napoca | Adrian Bojita | 2.4, 2.5, 2.6, 2.7 |
| | Luiza Dobre | 2.2 |
| | Ciprian Florea | 2.3, 3.4 |
| | Marius Purcar | 1, 2.2, 2.5 |
| Infineon Technologies Austria | Josef Schweda | 4.1, 4.2, 4.5 |
| Infineon Technologies Romania | Dan-Ionut Simon | 2.1, 3.1, 3.2, 3.3, 4.3, 3.5 |
| | Cazimir Bostan | 1, 2.3, 3.2 |

## Publishable Executive Summary

This deliverable focuses on activity performed in **tasks 7.2.1** (Design Methodology for Reliability Analysis and Yield Prediction) in the direction of reliability analysis and **task 2.4.1** (Methodology for Reliability Analysis and Yield Prediction). Both activities are related to **Use Case 11** (IC Design Methodology Based on Multi-physics Modelling and Simulation).

The deliverable is structured into 4 chapters:

- A general introduction presented the scope of the work and the partners involved
- Chapter 2 which presents the phase one (qualitative assessment) for the numerical efficiency of domain reduction methods for thermos-mechanical simulations
- Chapter 3 focused on the design of a test structure that can be used for the verification of the accuracy of electro-thermo-mechanical simulations of fast power cycling stress in integrated power devices, and on the design of a test system for fast power cycling conditions. The sanity check test results measured on the test structure are analyzed and is shown that the IC is suitable for use as test vehicle for further test in fast power cycling conditions.
- Chapter 4 presents a test structure and a simulation flow which will be used to assess stress in semiconductor devices in order to derive design rules for matching the characteristics of paired devices.

Activity in task 7.2.1 has started in December 2019 due to delays in the release of funds by the Romanian funding authority.

Activity in task 2.4.1 related to the simulation of stress in matching device pairs has started in December 2019.

### Key Words

Efficient electro-thermo-mechanical simulation, FEM, Fast Thermal Cycling, domain homogenization, domain simplification, non-conformal meshes, viscous layers, matching

# Contents

# List of Figures

# List of Tables

# 1 Introduction

This deliverable focuses on activity performed in **tasks 7.2.1** (Design Methodology for Reliability Analysis and Yield Prediction) in the direction of reliability analysis and **task 2.4.1** (Methodology for Reliability Analysis and Yield Prediction). Both activities are related to **Use Case 11** (IC Design Methodology Based on Multi-physics Modelling and Simulation).

To work focuses on the integration of a flow for electro-thermo-mechanical simulation of fast thermal cycling events in the IC design flow and the improvement of the simulation speed and usability of this flow. It allows the designer to analyze the reliability of a power stage, which undergoes fast thermal cycling, and understand how mechanical strain accumulates under this operating condition. The flow which is integrated is developed based on a proof of concept work done by IFRO in partnership with UTCN and is based on an electro-thermal simulation flow used by Infineon which is linked with a thermo-mechanical simulation tool.

Activity related to task **2.4.1** is focused on:

- improvement of the flow's integration with the overall IC design flow: numerical core compilation and installation on the server farm, installation of support software and communication modules with the existing design flow
- optimization of the temperature field mapping method for speed versus accuracy
- development of a dedicated plotting tool for simulation results
- extending the flexibility of the flow for use in other related analysis
- development of design rules for device matching

Activity in **task 7.2.1** is focused on testing various methods for improving the simulation speed and selection of those, which show a reasonable trade-off between speed and accuracy, for future integration into the productive flow. The tests are performed in two stages. First, a qualitative assessment of the method is performed to understand possible implementation issues and pitfalls. If phase one is successful, a second test on a more complex test structure is performed to assess the actual numerical efficiency and eventual accuracy trade-off and compared with measured data. Due to the delay in the start of activities funded by the Romanian state, the activity has started in December 2019 so phase one is still ongoing.

A test vehicle (IC) which will be used to check the accuracy of the different methods for improving the simulation speed which originally was planned for development in **task 7.2.1** was developed in **task 2.4.1**, due to the delay in the start of the activities funded by the Romanian funding authority in **work package 7**. This IC is crucial for the verification of the simulation results against test data in controlled conditions. A delay in in the fabrication of the IC would have put at high risk the possibility of verifying the validity of the results in **task 7.2.1** and of **UC 11**.

The team from Infineon Technologies Romania (IFRO) participating in this activity is part of the Smart Power EDA Tools & Methodologies group and is tasked with development of design methodologies and flows for power stage analysis and optimization. Its main focus is the development and integration of the flow for electro-thermo-mechanical simulation.

The team from Infineon Technologies Austria (IFAT) participating in this activity is part of the Physical Verification & Parametric DFM group and is tasked with the development of design rules in the context of Design for Manufacturability (DFM). The focus of the team is development of design rules for matching of devices which are impacted by mechanical stress due to the proximity of top-level thick metallization.

The teams from Technical University Cluj-Napoca (UTCN) is part of the Electrical Engineering Faculty and the Electronics, Telecommunications and Information Technology Faculty. The team from the Electrical Engineering Faculty focuses on numerical methods and modelling of electrical systems and within the project is tasked with the studying the speed accuracy trade-off of various methods for

reducing the numerical complexity of the electro-thermo-mechanical simulation. The team from the Electronics, Telecommunications and Information Technology Faculty is tasked in the project with the development of IC structure for detection of cracks in the metallization of DMOS devices operated in fast thermal cycling conditions.

Within **task 7.2.1** IFRO collaborated with UTCN in the study the speed accuracy trade-off of various methods for reducing the numerical complexity of the electro-thermo-mechanical simulation. In this direction, IFRO defines models' structures of varying complexity, which UTCN uses in a simulation environment compatible with the IFRO simulation flow to study the efficiency of different methods for reducing the numerical complexity analysis. Both teams then asses the speed accuracy trade-off. Part of these numerical analysis perfumed by UTCN are done on Infineon's server farm, for final evaluation of the simulation accuracy on the test structure design for UC 11. Also, in this task UTCN designs an circuit detection of cracks in the metallization of DMOS devices based on technology specification provided by IFRO.

Within **task 2.4.1** IFRO has developed and extension of the existing simulation flow to accommodate the requirements of IFAT for simulation of mechanical stress. This extension is then used by IFAT for performing simulation of various scenarios, which are used for deriving the DFM design rules.

After this introduction, the deliverable is structured in 3 parts. Chapter 2 starts with a presentation of the existing electro-thermo-mechanical simulation flow followed by the description of a simplified reference structure that will be used as a test vehicle for the assessment of the potential for numerical complexity reduction. The results of the analysis for five methods for numerical complexity reduction is than presented in the following subchapters.

Chapter 3 is dedicated to the test structure and test systems used for the experimental validation of the flow and of the some of the methods presented in Chapter 2. In subchapter 3.4 the initial test results measured on the test structure are presented.

Chapter 4 focuses on modelling of mechanicals stress influence on device matching. It presents the models and vehicle for this analysis and the extension of the existing simulation flow, which was implemented for this purpose.

At the end of the deliverable a set of unified conclusions is presented.

The activity within the two tasks which is not presented in this deliverable or which will take place in the final year of the project will be included in deliverable D2.16 IC Design Methodology based on multi-physics modelling and simulation and D7.8 Methodology for electro-thermo-mechanical simulation.

# 2 Models and methods for efficient electro-thermo-mechanical simulation

## 2.1 Electro-thermo-mechanical simulation flow

Power ICs for automotive industry are typically manufactured in BCD technologies (Bipolar-CMOS-DMOS)(Wagner et al., 2011). The functions accomplished by these ICs can range from low complexity, i.e. from simple switches without over-temperature or over-current protections, up to complex ASICs with configurable protected multichannel switches, voltage regulators, embedded communication (e.g. SPI) etc. However, no matter the complexity, all these ICs contain at least one power device, which is used as switching element. This is usually a DMOS (Double-Diffused MOS) transistor, which can drive loads with reactive behaviour. During repetitive switching events (e.g. switching on capacitive or off inductive loads), the power device withstands high power transients, thus it undergoes significant self-heating. In these conditions, the power device acts as a heat source, transferring heat into its surroundings, including its metallization structure. The constitutive materials undergo thermal expansion and metals can suffer plastic deformations. A heating-cooling thermal cycle is associated with each high-power transient event (fast thermal cycle). Thermal ratcheting of metals occurs as result of fast thermal cycling. In the end, the pressure applied due to cyclic incremental plastic deformation of metal lines leads to cracking of nearby oxide (Smorodin, et al., 2008). The metal from one line extrudes through the crack and touches another metal line (short-circuit), which leads to chip failure. This destruction mechanism will be referred to as IMD or ILD (inter-metal or inter-layer dielectric) cracking (Nguyen et al. 2002).

The safe-operating-area (SOA) of a power transistor is thermally limited (Dibra et al. 2011) by single-pulse operation at junction temperatures at the boundary of thermal instability (Spirito et al. 2002). However, an additional boundary on the SOA (see Figure 1) is set by multiple-pulse conditions (Bosc et al. 2000) at junction temperatures below the temperature corresponding to the thermal instability limit (depending on the requirement of number of cycles to failure). The multiple-pulse SOA boundary is currently calculated as function of the maximum and minimum temperature in the DMOS heat-source (Smorodin et al. 2008, Bosc et al. 2000, Nguyen et al. 2003).

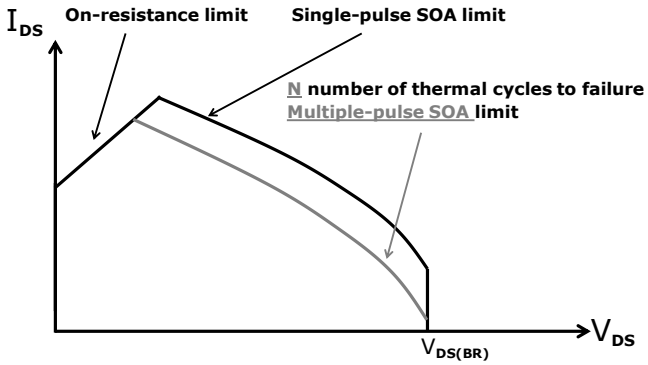



Figure 1: Schematic SOA bounds of DMOS transistors: single-pulse border corresponding to thermal instability limit (black) and multiple-pulse border corresponding to the fast thermal cycling limit (gray).

During semiconductor fabrication, the process steps require various environment conditions. For example, copper (Cu) can be deposited at high temperatures (e.g. 400°C), while other process steps require negative temperatures (e.g. -40°C). Therefore, the semiconductor can be subject to extreme temperatures from its first process step until the end of fabrication, which means that certain material layers will develop intrinsic mechanical stresses. The intrinsic stresses in the metallization can negatively influence the parameters of the electronic devices from the Silicon below. Therefore, it is necessary to determine the Silicon region that is most prone to such intrinsic stresses, in order to avoid placing sensitive devices in the proximity of this region. In order to estimate the intrinsic mechanical stresses for different metallization configurations, Finite Element Method (FEM) thermo-mechanical simulations are needed. The focus will be on the metallization system and the surface of the Silicon

chip. The simulation flow for determining the intrinsic stresses is a bit different than the one presented in this section. The differences will be described in Section 4.3.

As explained at the beginning of the section, power transistors can dissipate a significant amount of energy during load switching. This will cause internal heat dissipation and the junction temperature can rapidly swing from environment temperature to more than 350°C and back, at a rate of 1°C/µs. The fast temperatures cycles can lead to progressive mechanical degradation of the chip metallization. The degradation speed and the most likely failure region should be predicted in the early phases of chip design. There are several thermo-mechanically triggered failure mechanisms (i.e. IMD cracking (Smorodin et al. 2008), bond-wire lift-off (Lassnig et al. 2012), power metal delamination (Kravchenko, et al., 2013) etc.). The most likely failure mode is a combination of technology (i.e. materials, layer thicknesses) and operating conditions (e.g. environment temperature, junction temperature swing, temperature gradients). Therefore, it is equally important to accurately predict the temperature in the chip. Due to the non-uniform temperature distribution at the surface of the chip (Pfost et al. 2014) and in the metallization, a complete FEM electro-thermo-mechanical simulation methodology is needed (Simon et al. 2016). The focus will be on the dielectric failure due to progressive plastic deformation of the metal lines (i.e. IMD cracking failure mode).

This chapter will provide a description of the thermo-mechanical simulation flow used for investigating the IMD cracking failure mode. The methodology used for estimating the intrinsic stresses developed in the metallization is mostly similar and the differences will be detailed in Section 4.3.

Typically, a chip consists of many details, but these can be divided in three categories (see Figure 22):

- The Silicon die: it consists of Silicon bulk (from 60µm to 600µm) and the surface of the Silicon where the devices are built (a few micrometers thick), the plastic mold compound surrounding the chip, the leadframe and die attach
- The signal metallization (or thin metallization): it consists of thin metal lines used for routing and for carrying low power signals, the dielectric between them and via connections (a few micrometers thick)
- Power metal: it consists of a thick power metal (in the range of 3-20µm), the imide passivation and bond-wires

However, it should be mentioned that, for some analysis, certain chip details could be safely ignored. For example, the die attach is not relevant when analyzing the metallization system, from a mechanical point of view, while for the thermal analysis, it cannot be ignored because it adds a significant component to the junction to case thermal impedance.

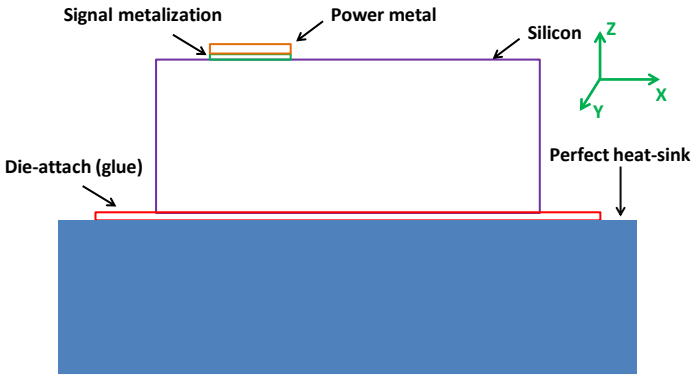



Figure 2: Schematic representation of the chip structure (mold compound is not drawn, as it surrounds all elements of the chip).

### 2.1.1 Electro-thermal simulation flow

From an electro-thermal point of view, there are not many simplifications that can be done. Most notably, there are bulky parts of the geometry that cannot be ignored: mold compound it acts as a thermal capacitance for transient power pulses, the die attach has also a significant contribution to the thermal impedance starting from the hundreds of microseconds' time domain. Only, bond-wires can be safely ignored if they are placed away from the heat-dissipating region. Hence, the focus of analysis is on the metallization and the power device itself.

The layout of the DMOS transistor metallization, manufactured in each BCD technology, is not standard and can vary from one product to another. A technology with two thin metal layers, corresponding to three layout variations, in three cross-sections, through the metallization system is schematically represented in Figure 33. The placement of electrical vias, from the metallization, influences the peak temperature (Pfost et al. 2014) and the temperature distribution in the metallization system. The layout of the power metal is also very important from the thermal point of view. Therefore, even if computationally intensive, all thin metal lines and the dielectric between them are considered in the thermal simulation. It is important to capture the exact temperature gradients in the thin metal lines in order to have an accurate thermo-mechanical simulation. It is very difficult to keep the vias in the simulation. Therefore, vias are grouped with the nearby dielectric and equivalent thermal properties are assigned to these new shapes, as explained in Section 2.3.

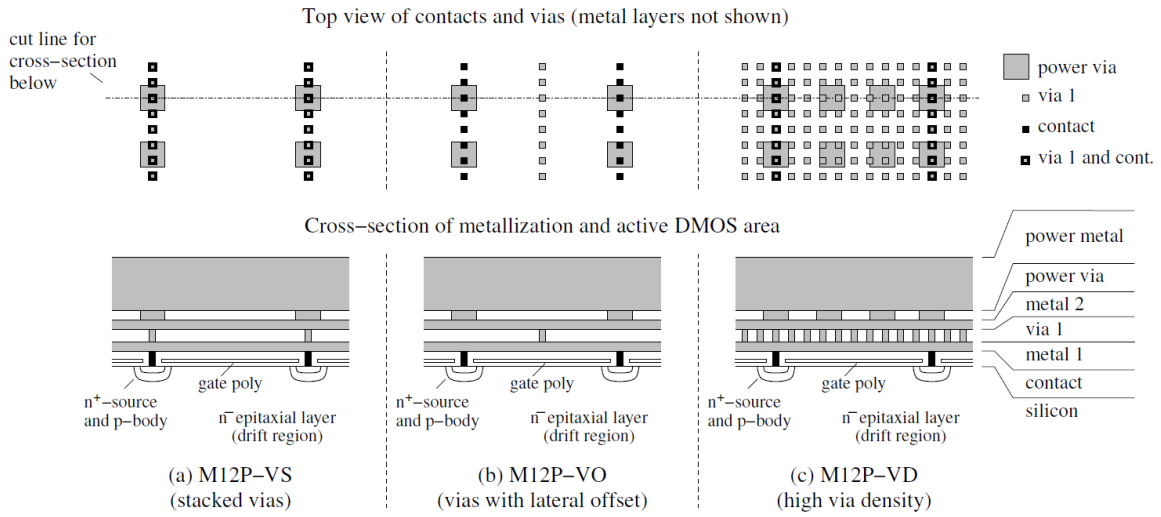



Figure 3: Various layout variations for the DMOS metallization system for a two-layer BCD technology. Via position and via densities are varied.

Such level of chip detail (in range of micrometers), coupled with bulky package details (in range of micrometers-centimeters) requires a robust approach. Infineon© tool, detailed in (Pfost & Joachim, 2008), based on FVM method with on a non-conformal mesh approach, is used to solve the electro-thermal problem.

The electrical and thermal problems are tightly coupled because the power density distribution in the heat generating region (i.e. DMOS device) is heavily dependent on temperature. Temperature distribution is, in turn, heavily dependent on power distribution. The power density distribution is dependent on the operating point of the transistor, with respect to the gate-source voltage, influences the peak temperature and temperature distribution, which, finally, affects the reliability of the power device. Therefore, the heat generating sources require accurate electrical models across a wide temperature range (-40°C - 100°C), as explained in (Pfost et al. 2013).

The electro-thermal analysis at chip level starts from the two-dimensional (2D) drawing (layout) of the product (or prototype). The tool is fully automated, and the chip geometry is automatically extracted

from 2D layout data. The following settings should be made before starting the electro-thermal simulation:

- Specify the source of the 2D geometry (which will be automatically extruded to create a three-dimensional (3D) geometry)
- Specify a power pulse to the desired heat generating source
- Specify other boundary conditions for the simulation (i.e. temperature of the heat source, which is typically located on below die attach, as in Figure 2)
- Assign an electrical model to the respective heat source
- Enable chip details which should be considered in the simulation: e.g. trenches, power metal, thin metal lines, die attach
- Specify material properties for the chip details which were selected in the previous step

### *2.1.2 Thermo-mechanical simulation flow*

From the mechanical point of view, not all the chip elements are relevant when analyzing the targeted failure mode: IMD cracking. The relevant parts of the chip, considered for simulation are the following: Silicon, signal metallization and power metal. The die attach is not taken into consideration because it is a soft material and it would imply insignificant mechanical stresses on the bottom of the Silicon substrate. The mold compound, which is a polymer, becomes very soft in the targeted temperature range (above 150°C) and it will not introduce significant stresses in the system. Finally, the vias in the signal metallization are also simplified, though they are not ignored altogether. This will be explained in detail in Section 2.3.

It is very difficult to obtain quantitative results from thermo-mechanical simulations at chip level due to the very high complexity of the metallization system and due to the lack of accurate mechanical material properties. Specifically, the level of detail in the signal metallization includes metal liners, vias and other material coatings. These details cannot be captured due to limitations like the maximum number of mesh elements and simulation efficiency. The material properties of passivated metal lines or coated vias are also difficult to measure. Moreover, the mechanical properties should be measured for a wide range of temperatures and for different loading rates. Unfortunately, the loading rates are very high, in our case, because the temperatures variations are very steep, and materials exhibit various behaviours as function of loading speed.

As in the case of the electro-thermal simulation tool, the chip level analysis starts from the product's layout, though the tool is not fully automated, for the moment. The simulation steps can be divided in three stages: pre-processing, solving and post-processing.

**Pre-processing stage:**

***Geometry data***: A typical thermo-mechanical simulation flow starts by specifying the 2D geometry data of the simulated structure (layout file). The 2D geometry data will be automatically extruded to a 3D geometry. This is a natural step, because it reassembles the manufacturing process, where thin films are frequently grown by electro-chemical deposition or physical vapor deposition, for example. However, the 2D layout data will need a few manual simplifications and preparations, due to the high number of features. For example, operations like tiling, chamfering and de-chamfering might be needed.

***Mesh controls and region definition file (Input file 1)***: Every setup contains a technology dependent file, which has the name of the layout layers involved in the mechanical simulation and the thickness of each layer. The file also contains mesh controls like the maximum and minimum element sizes in the metallization system and the maximum and minimum element sizes in the Silicon die. Also, the name of the regions involved in the simulation and the associated material properties are specified in input file 1.

***Mesh controls and region definition file (Input file 2):*** The second input file contains data about thermal stimuli, boundary conditions for the chip and post-processing options:

- Usually, only a smaller region of the chip, around the heat generating sources, needs to be simulated, while the rest of the Silicon chip can be safely ignored. This can be specified in the *input file 2*.
- Some structures are symmetric and the axis of symmetry (parallel to O*X* and/or OY in the global system of coordinates) can be specified in this input file to drastically reduce the dimension of the problem.
- The ambient temperature needs to be specified in order to compute an initial stress distribution. The stress-free temperature also needs to be set in the setup file.
- The stimuli are 3D temperature distributions from an electro-thermal simulation. Therefore, the name of the electro-thermal simulation setup needs to be specified in the *input file 2*. Not all the time steps from the electro-thermal simulation are needed for the thermo-mechanical setup. The time stepping in the electro-thermal simulation is much finer than needed for the thermo-mechanical simulation. Therefore, a list with time steps useful for the mechanical simulation should be provided as input. In order to capture the cyclic behaviour of the mechanical degradation, several heating-cooling cycles should be simulated, so the number of cycles is specified in the input file. It should be noted that the electro-thermal and thermo-mechanical problems are weakly coupled. In other words, the mechanical stresses do not influence the electro-thermal problem. Hence, only one thermal cycle needs to be computed by the electro-thermal simulated. The same temperature distribution will be cyclically applied to the mechanical simulation during 3-10 cycles.
- The user can provide post-processing options: the time steps for which the results should be retrieved from the solver. In addition to this, the user can also specify the coordinates for 2D contour plots, if desired.

***Model construction:*** The setting up of simulation input files, is followed by the first pre-processing step: construction of 3D computational domain from 2D layout data and mesh generation. A combination of extruded and non-structured mesh is generated started from 2D and 3D geometry data. A sample of the resulting mesh can be seen in Figure 4.

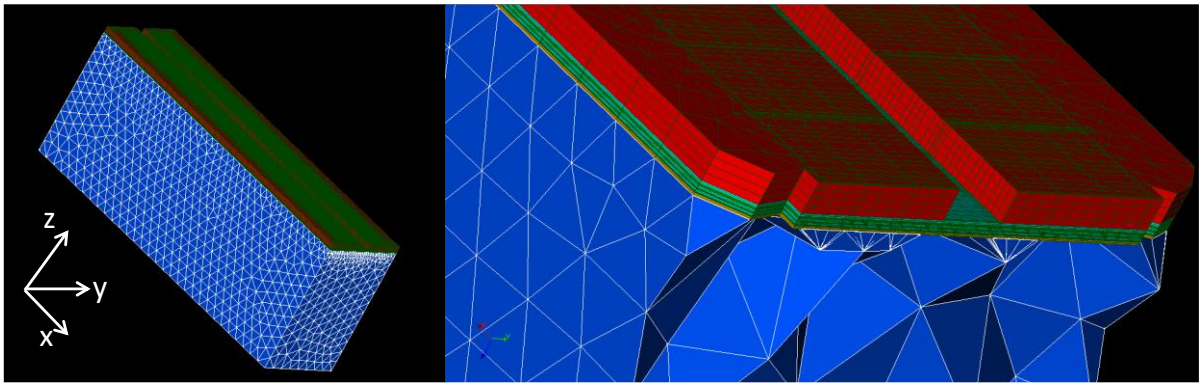



Figure 4: Mesh of a DMOS transistor active area: (left side) mesh of the whole simulated structure or active area; (right sider) zoom of a clipped view through power metal (colored in red), signal metallization or metal layers (colored in light blue), ILD (colored in green) and Silicon bulk (colored in blue).

***Temperature distribution transfer to mechanical mesh:*** The temperature distribution from the electro-thermal simulator needs to be transferred to the mechanical model. Temperature causes thermal dilatation and affects the strains in the structure, which in turn, influences the mechanical stress field. Therefore, each node of the mechanical mesh must have an associated temperature property, and the thermal strain will be computed with the following equation:

$$\varepsilon^{th}(T) = \alpha(T)\big(T - T_{def}\big) - \alpha\big(T_{ref}\big)\big(T_{ref} - T_{def}\big), \quad (1)$$

where $\alpha$ is the thermal expansion coefficient (CTE), $T_{def}$ is the temperature relative to which the CTE was measured and $T_{ref}$ is the reference temperature, i.e. the temperature at which the thermal strains are null. The temperature interpolation method can be specified in the *input file 2*. Currently, there are three temperature interpolation methods implemented: *method 1* is the fastest and least accurate, *method 2* is the slowest, but the most accurate and *method 3* is a trade-off between accuracy and speed. Depending on the granularity of the electro-thermal mesh and the size of the thermo-mechanical mesh, the user can select the best-suited method for interpolation. The implemented interpolation methods will be presented in deliverable D2.16.

**Solving stage:**

The input simulator file for *Code_Aster* solver is automatically created by scripting from settings found in *input file 1* and *input file 2*. The following boundary conditions which are applied to every simulation setup and the user has no control over them: zero displacements, are always applied to the nodes in the XY plane, at the bottom of the Silicon die.

The material models of various mesh regions are also automatically defined in the solver input file, based on the settings found in the first *input file 1*. For example, Silicon and oxides have an elastic behavior model (Silicon has anisotropic properties), while the behavior of metals is typically modelled with the elastoplastic non-linear kinematic hardening model of Chaboche. The parameters for the materials are stored in a database in XML format. The temperature distribution, which is applied as a load to the mechanical simulation, is read by the solver from a text file.

The solver in binary format provides the mechanical results, after the solving stage is completed.

**Post-processing stage:**

The output binary files contain 3D data with Von-Mises stresses and cumulative equivalent plastic strains. The results are provided at discrete in the 3-dimensional space. With the aid of scripting, the files are automatically post-processed by scripts that provide 2D continuous-contour plots for Von-Mises stresses and cumulative plastic strains. 1-dimensional stress and strain graphs are also possible to be obtained. Since the amount of data from the mechanical simulation is very large (a 3D mesh can have up to 5 million nodes), even in the case of 2D contour plots, a new result visualization tool was developed. The tool is much faster when dealing with large amounts of data, enhances the quality of colour representation and replaces the old tool, which is used for thermal results visualization. The tool will be presented in deliverable D2.16.

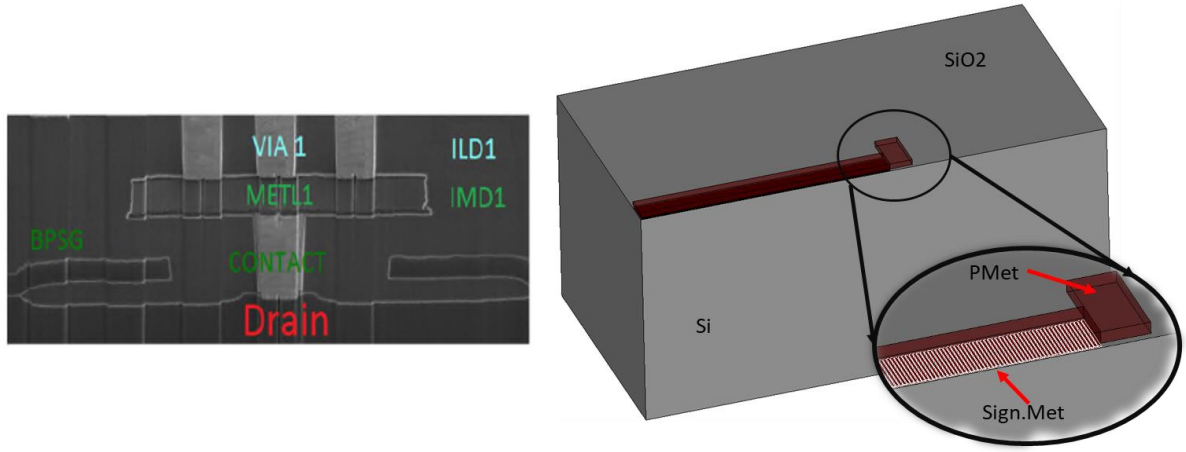
## 2.2 Overview of simulation models and methods

Upon the classification of main defect analysis and prediction methods in power integrated circuits, (Bojita et al. 2018, Bojita et al. 2019), for an efficient numerical simulation methodology (flow) of electro-thermo-mechanical processes, the following techniques have been investigated: (1) simplification of the computational models, (2) approximation of the topological elements such as contacts, via, metallization lines etc., (3) identification of equivalent thermal and mechanical material parameters in case of simplification or homogenization of the computational model, (4) definition of adequate boundary conditions for the extended computational models, (5) influence of the semiconductor substrate dimensions, (6) techniques for increasing the mechanical simulations efficiency.

### *2.2.1 Description of simplified reference structure*

The power IC devices, subjected to repetitive temperature cycles, leads to accumulation of plastic deformations over time, both at the upper level of the routing metallization system and power metallization level. The accumulation of plastic deformation eventually leads to the IC failure after a large number (e.g. 1e6) of temperature cycles. In case of DMOS power devices, crack formation between two adjacent signal metal lines and delamination of the power metal plates from the dielectric layer are the predominant failure mechanisms. DMOS technologies, predominantly use the

routing metallization system arranged on several layers interconnected through the sets of vias and contacts, Figure 5 a) (Rudolf et al. 2011, Pham et al. 2016). Therefore, efficient computational structures, which easily highlight the accumulation of plastic deformations inside the DMOS metallization structures are required in order to analyses the reliability of the power IC's.

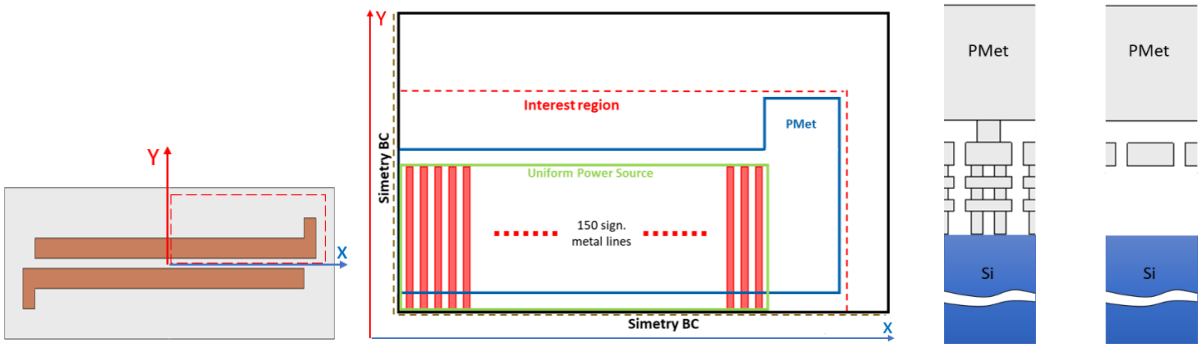



a) b)

Figure 5: a) Example of structure of the metallization and via system in DMOS circuits, b) 3D model, schematic representation.

The thermo-mechanical behavior of the IC can be studied using the FEM. The high geometric aspect ratio of the metallization lines and the number of the interconnection paths leads to many computational nodes (e.g. 1e8) and mesh elements (e.g. 5e8) of the finite discretization elements (mesh) inside the computational model. The accumulation of plastic deformations is more pronounced in the upper level of the metallization system due to the greater thickness than the lower levels (Smorodin et al. 2008).

In order to highlight the failure mechanism, a computational structure consisting of a silicon (Si) substrate of 400µm thickness is proposed. An insulating (dielectric) layer made of silicon dioxide ($SiO_2$) covers the Si substrate. The $SiO_2$ encapsulates the routing metallization (signal metallization) - composed of multiple metal line levels, interconnected by via groups in the vertical direction. Above the $SiO_2$ substrate, two power metal plates of 10µm thickness are placed and further connected with the signal metallization by via groups. The schematic model is presented in Figure 6 a) and b).




a) b) c) d)

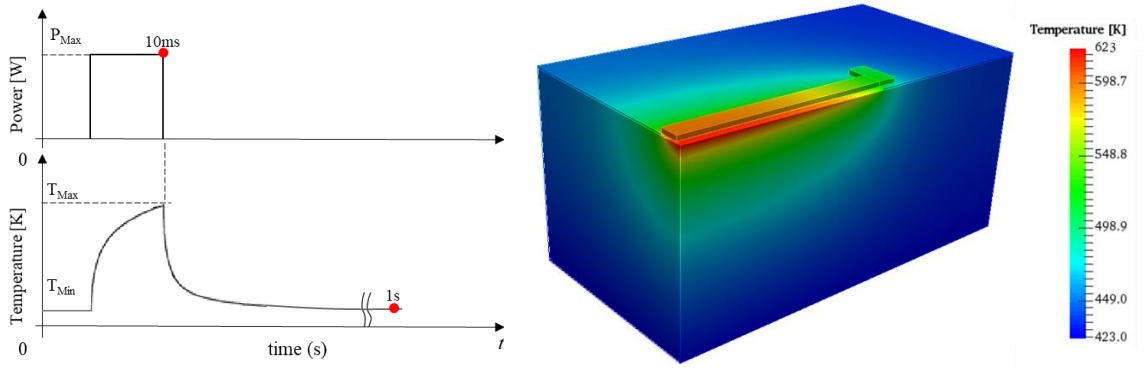
Figure 6: a) Computational model of the DMOS circuit – top view, b) reduced computational test structure used in mechanical analysis, c) to d) further simplified model (dimensions are not to scale).

Due to the symmetry of the structure, several simplifications can be performed on the model. Thus, by applying symmetry conditions, only a quarter of the computational model can be retain in the simulation, Figure 6 b).

Several simplifications are further made on the computational model, thus, only the topmost thin signal metal lines and the power metal plates are considered. The other levels of routing lines and vias are neglected, in order to reduce the computation time. This simplification is presented in Figure 6 c) to d).

### *2.2.2 Definition of thermo-mechanical stress conditions*

In order to generate the thermo-mechanical stress conditions, the temperature distribution in the model is varied by means of a power source placed at the Si/SiO2 interface. When a fixed temperature condition is imposed on the bottom face of the Si die, a temperature gradient is forced across the model. If a 10 $ms$ power pulse duration is applied, the temperature profile over time inside the model will evolve according to the graph presented in Figure 7 a). The applied heating-cooling cycle is repeated with the frequency of 1 Hz. The maximum temperature reached inside the model at the end of the power pulse is 623K and the temperature distribution extracted at the same time is presented in Figure 7 b).




a) b)

Figure 7: a) Power pulse applied to the source and the evolution of temperature profile over time in b) temperature distribution over the model at time t = 10ms.

## 2.3 Domain homogenization using equivalent properties

Sometimes the size of the three-dimensional (3D) domain that needs to be analyzed using finite element method (FEM) is significantly large and extremely heterogeneous (e.g. multiple physical subdomains with different material properties and complex geometrical structure comprising large numbers of shapes with high aspect ratios), leading to a mesh with a very large number of elements and rendering the simulation extremely time-consuming if not impossible in most cases. Such is the case for the metallization of modern power devices and integrated circuits. In order to overcome such obstacle, one needs to resort to homogenization, which consists in replacing certain regions of the highly detailed heterogeneous 3D structure with homogeneous region which are expected to behave identically as the original structure at macroscopic scale, provided that the distribution of the heterogeneities inside those regions is well known and provided that their contribution to the macroscopic behavior can be averaged. Consequently, equivalent material properties need to be extracted for the new region. First, a subdomain is selected for this analysis in such a manner that is representative for the entire structure (e. g. comprises all the essential geometrical/structural details), region referred to as Representative Volume Element (RVE), as shown in Figure 8. On this RVE certain stimuli are then applied through adequate choices of certain sets of periodic boundary conditions (PBCs) in order to extract the equivalent properties, be it thermal or mechanical. The extraction procedure is detailed separately for the thermal domain as well as for the mechanical domain in the following sections.

### 2.3.1 Extraction of equivalent thermal properties

The material properties with importance in thermal analysis are density - ρ [kg/m3], specific heat - $cp$ [J/kg·K] and thermal conductivity – $k$ [W/m·K]. A thermally equivalent domain must preserve the energy storage and transfer, e.g. the volume thermal capacitance and the thermal conductivity. The volume thermal capacitance is the quantity equal to the product between the volume, the density and the specific heat. Provided this, one must extract for the selected RVE, which has predetermined volume, the equivalent density and the equivalent specific heat.

The equivalent density can be computed as follows:

$$\rho_{eq} = \frac{m_{tot}}{V_{tot}} = \frac{\sum_i^n m_i}{\sum_i^n V_i} = \frac{\sum_i^n \rho_i V_i}{\sum_i^n V_i} = \sum_i^n \rho_i \frac{V_i}{V_{tot}}, \quad (2)$$

where $m_i$ and $V_i$ refer to the i-th mass/volume in the structure. The equivalent density represents a volume-weighted average of the various densities in the RVE. In a similar manner the equivalent specific heat is calculated as:

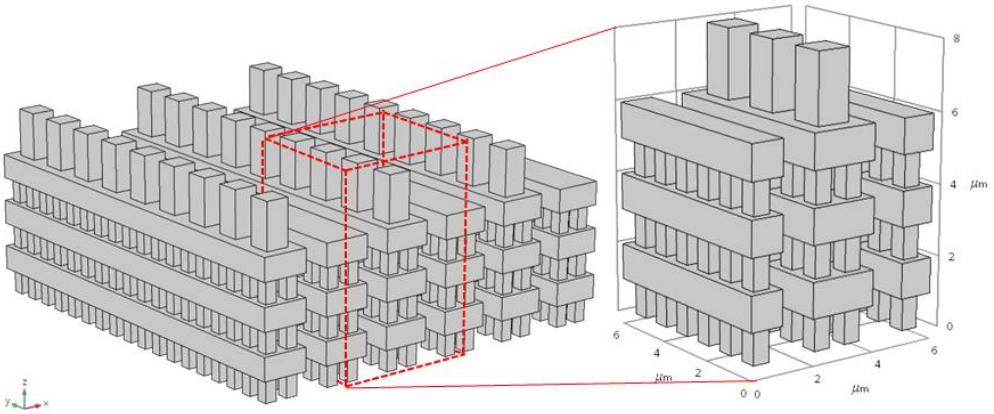


Figure 8: Typical detail example of a routing metallization found in power ICs which contains multiple metal lines interconnected by vias, all surrounded by an insulator (not shown in the picture) and a zoom of metal stack selected as RVE at the right.

$$c_{p,eq} = \frac{c_{tot}}{m_{tot}} = \frac{\sum_i^n c_i}{\sum_i^n m_i} = \frac{\sum_i^n C_{p,i} m_i}{\sum_i^n m_i} = \sum_i^n C_{p,i} \frac{m_i}{m_{tot}}, \quad (3)$$

where $C_i$ and $C_{p,I}$ stand for i-th thermal capacitance and specific heat respectively. The equivalent specific heat represents a mass-weighted average of the various specific heats in the RVE.

Unlike the equivalent density and the equivalent thermal specific heat, the equivalent thermal conductivity exhibits high anisotropy due to spatial distribution in the stack of the various materials with large differences in thermal conductivity. A higher thermal conductivity, closer to that of the metal, be it Al or Cu, is expected alongside the directions where there is a continuous heat flow path through the metal (see directions Y and Z in Figure 8) and a lower thermal conductivity, closer to that of the silicon dioxide, is expected alongside the directions where regions of high thermal conductivity and low thermal conductivity alternate (direction X Figure 8). Due to this alternation in thermal conductivity, for a unidirectional temperature gradient, there are significant heat flows in all three directions (Figure 9), therefore the series-parallel resistance formulations do not yield correct results. The correct approach is to determine the equivalent thermal conductivity with the aid of FEM simulation. The RVE is a cuboid with six faces. For a pair of two opposite faces a Dirichlet boundary condition is set (de Filippis 2012), forcing a fixed temperature difference between them as shown in Figure 9. All other four faces are set to be adiabatic. In this way, in a thermostatic simulation, the energy transferred between the two faces is directly proportional to the temperature difference between them and the thermal resistance of the cuboid in that direction.

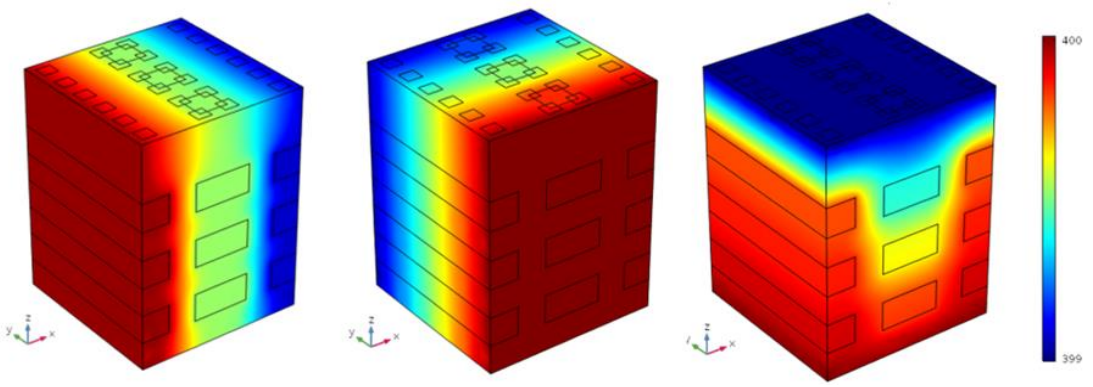



Figure 9: Temperature distributions resulting from setting a temperature difference on 1K on each pair of opposite faces one at a time. All other four faces are set adiabatic, so that the heat flows in one direction only.

The equivalent thermal conductivity can be extracted from the thermal resistance as follows:

$$R_{th,x} = \frac{\Delta T}{P} = \frac{1}{k_{x,eq}} \frac{\Delta x}{\Delta y \Delta z}, \tag{4}$$

where P is the total heat flux through one of the surfaces which has a Dirichlet boundary condition. The procedure is repeated on all three directions as shown in Figure 9. The simulations are carried out on the structure shown in Figure 8 presented in detail in Section 2.3.2. The geometrical details are listed in Table 3. The thermal properties of the original constituents are listed in Table 1. The constituents are assumed to be isotropic.

Table 1 Thermal properties of the constituents.

| | **Al** | **W** | **$SiO_2$** |
|---|---|---|---|
| **Thermal conductivity, k [W/m·K]** | 237 | 174 | 1.4 |
| **Specific heat capacitance, cp [J/kg·K]** | 904 | 132 | 730 |
| **Density, ρ [kg/m³]** | 2700 | 19350 | 2200 |

The extracted equivalent thermal properties are listed in Table 2.

Table 2 Equivalent thermal properties.

| | | |
|---|---|---|
| **Thermal conductivity, k [W/m·K]** | **$k_{ex}$** | 8.605 |
| | **$k_{ey}$** | 71.529 |
| | **$k_{ez}$** | 30.982 |
| **Specific heat capacitance, cp [J/kg·K]** | | 610.021 |
| **Density, ρ [kg/m³]** | | 3039.583 |

Since the standard metallization of power devices is periodic both in X and Y direction across the whole active area of the DMOS device, the metallization can be homogenized entirely.

### *2.3.2 Extraction of equivalent mechanical properties*

The material properties with importance in mechanical analysis are Young's Modulus – E [N/m²], Poisson's Ratio - ν [1] and Shear Modulus - G [N/m²]. The two domains that are equivalent from a mechanical standpoint are characterized by the same strain energy (V.-D. Nguyen, Apr. 2012):

$$U = \frac{V}{2} \bar{\sigma}_{ij} \bar{\varepsilon}_{ij} = \frac{V}{2} C_{ijkl} \bar{\varepsilon}_{kl} \bar{\varepsilon}_{ij}, \tag{5}$$

where $V$ is the volume of the domain (in this case, the RVE), $\bar{\sigma}_{ij}$ and $\bar{\varepsilon}_{ij}$ are the average stress and strain tensors and $C_{ijkl}$ is the stiffness tensor according to Hooke's Law:

$$\bar{\sigma}_{ij} = C_{ijkl} \bar{\varepsilon}_{kl}. \tag{6}$$

The validity of this method of is based on two assumptions which need to be satisfied. First, the microscopic stresses and strains inside the RVE must affect the macroscopic behavior only through their volume averages and secondly, any fluctuation in the stress and strain fields at macroscopic scale must be negligible inside the RVE (Kurukuri 2005, Li 2012). This implies that adopting the adequate size of the RVE has to be carried out with great care since the two conditions are competing. As a rule of thumb, for randomly distributed inhomogeneities, the RVE scale is chosen to be at least one order of magnitude larger than the size of the biggest inhomogeneity and at least one order of magnitude smaller than the size of the actual macroscopic structure. However, given that in this particular case the analyzed structure is the metallization of power devices and given certain scale factors (width of a the device/width of a metal line < 100) it is best to select an RVE scale as small as possible (for example, the source pitch) in order to full fill the second condition. The first condition is still fulfilled due to the periodic geometry of the metallization. The periodicity of the structure adds some flexibility to the proposed method, as the selection of the RVE is not unique, as shown in ( Xia 2006, Würkner 2014, Nguyen 2012, Nanakorn 2013).

The structure to be homogenized is expected to behave as a physically continuous body, therefore displacement continuity and traction continuity need to be satisfied at the boundaries of the RVE (Krukuri 2005, Xia 2006). The periodic boundary conditions are applied in the form of nodal equations. The displacement difference boundary conditions ensure the continuity of the displacement:

$$u_i^{k+} - u_i^{k-} = \varepsilon_{ij}^0 \left(x_j^{k+} - x_j^{k-}\right) = \varepsilon_{ij}^0 \Delta x_j^k, \tag{7}$$

where „k+" and „k-" identify the k-th pair of opposite parallel boundary surfaces of a repeated unit cell, $u_i^{k+/-}$ is the displacement field on the opposite surfaces, $\varepsilon_{ij}^0$ is the global average strain tensor of the periodic structure, $x_j^{k+/-}$ are points belonging to the opposite boundary surfaces. The traction continuity boundary conditions:

$$\sigma_n^{k+} - \sigma_n^{k-} = 0, \qquad \sigma_t^{k+} - \sigma_t^{k-} = 0, \tag{8}$$

where $\sigma_n$ and $\sigma_t$ are normal and shear stresses at the corresponding parallel boundary surfaces. Depending on the problem, displacement difference BCs might suffice and there is no need to explicitly define the traction continuity conditions. This can be verified by examining the stress and strain distributions after a shear deformation test. If the stress and strain fields observed at the opposite boundaries of the RVE are identical, then the RVE can be assembled as a physical continuous body and the assumption that the stress and strain fields inside the RVE affect the macroscale behavior only through their volume averages values holds valid. The stiffness tensor coefficients can then be calculated using the following equations:

$$\bar{\varepsilon}_{kl} = \frac{1}{V}\int_V \varepsilon_{kl} dV \tag{9}$$

$$\bar{\sigma}_{ij} = \frac{1}{V}\int_V \sigma_{ij} dV \tag{10}$$

$$\begin{bmatrix}\sigma_{11}\\ \sigma_{22}\\ \sigma_{33}\\ \sigma_{23}\\ \sigma_{31}\\ \sigma_{12}\end{bmatrix} = \begin{bmatrix} C_{1111} & C_{1122} & C_{1133} & C_{1123} & C_{1131} & C_{1112} \\ C_{2211} & C_{2222} & C_{2233} & C_{2223} & C_{2231} & C_{2212} \\ C_{3311} & C_{3322} & C_{3333} & C_{3323} & C_{3331} & C_{3312} \\ C_{2311} & C_{2322} & C_{2333} & C_{2323} & C_{2331} & C_{2312} \\ C_{3111} & C_{3122} & C_{3133} & C_{3123} & C_{3131} & C_{3112} \\ C_{1211} & C_{1222} & C_{1233} & C_{1223} & C_{1231} & C_{1212} \end{bmatrix} \begin{bmatrix}\varepsilon_{11}\\ \varepsilon_{22}\\ \varepsilon_{33}\\ 2\varepsilon_{23}\\ 2\varepsilon_{31}\\ 2\varepsilon_{12}\end{bmatrix} \tag{11}$$

The stiffness tensor for fully anisotropic materials has 21 independent coefficients symmetrical to the first diagonal ($C_{ijkl} = C_{klij}$). In order to determine each coefficient, the PBCs are applied in such a manner so that only one strain tensor component at a time is non-zero while all others are zero [Kur05], [Li12], [Ye]. For instance, for uniaxial strain applied in X direction, ε11 ≠ 0, ε22 … 2ε12 = 0:

$$\begin{bmatrix}\sigma_{11}\\ \sigma_{22}\\ \sigma_{33}\\ \sigma_{23}\\ \sigma_{31}\\ \sigma_{12}\end{bmatrix} = \begin{bmatrix} C_{1111} & C_{1122} & C_{1133} & C_{1123} & C_{1131} & C_{1112}\\ C_{2211} & C_{2222} & C_{2233} & C_{2223} & C_{2231} & C_{2212}\\ C_{3311} & C_{3322} & C_{3333} & C_{3323} & C_{3331} & C_{3312}\\ C_{2311} & C_{2322} & C_{2333} & C_{2323} & C_{2331} & C_{2312}\\ C_{3111} & C_{3122} & C_{3133} & C_{3123} & C_{3131} & C_{3112}\\ C_{1211} & C_{1222} & C_{1233} & C_{1223} & C_{1231} & C_{1212}\end{bmatrix}\begin{bmatrix}\varepsilon_{11}\\ 0\\ 0\\ 0\\ 0\\ 0\end{bmatrix} = \varepsilon_{11}\begin{bmatrix}C_{1111}\\ C_{2211}\\ C_{3311}\\ C_{2311}\\ C_{3111}\\ C_{1211}\end{bmatrix} \quad (12)$$

Practically, in each test one column of the stiffness tensor is determined by forcing a predetermined strain and then dividing the resulting average stress tensor components to the average strain. There is a total of six stationary mechanical simulations to be performed, three tests for normal traction and three for shear traction.

The simulations are carried out on a theoretical model which resembles very well the typical routing metallization structure found in power devices. The geometry is presented in Figure 8. The RVE is selected according to the aforementioned considerations. The geometrical parameters of the RVE are listed in Table 3. The materials used in this study are aluminum (Al) for metal the lines, tungsten (W) for contacts, vias and power vias, and silicon dioxide ($SiO_2$) for the insulator. The constituents are assumed to be linear-elastic and isotropic. The material properties are listed in Table 3.

Table 3 Geometrical parameters of the RVE

| **Parameter** | Length [μm] |
|---|---|
| Width of the RVE (x direction) | 6 |
| Depth of the RVE (y direction) | 6 |
| Height of the RVE (z direction) | 8 |
| Width of a contact/via | 0.5 |
| Height of the contact/via | 1 |
| Pitch of the contact/via | 1 |
| Width of the metal line | 2 |
| Thickness of the metal line | 1 |
| Pitch of the metal line | 3 |
| Width of the power via | 1 |
| Height of the power via | 2 |
| Pitch of the power via | 2 |

Given the geometry and the mismatch of the linear-elastic parameters of the constituents, some degree of anisotropy for the equivalent linear-elastic parameters is expected. A fully anisotropic stiffness tensor is assumed in the beginning.

Table 4 Material properties of the constituents.

| | **Al** | **W** | **$SiO_2$** |
|---|---|---|---|
| **Young's Modulus, E [GPa]** | 70 | 411 | 70 |
| **Poisson's Ratio, v [1]** | 0.35 | 0.28 | 0.17 |
| **Shear Modulus [GPa]** | 26 | 160 | 30 |

In order to apply periodic boundary conditions (PBCs), the mesh on each pair of opposite faces must be identical. The classical approach is to mesh one face and copy it onto the opposite one, one pair at a time, and then mesh the remaining domains. The meshed RVE is presented in Figure 10. A fine mesh with quadratic tetrahedrons was adopted for an adequate sampling of the stress and strain distributions, avoiding possible element locking.

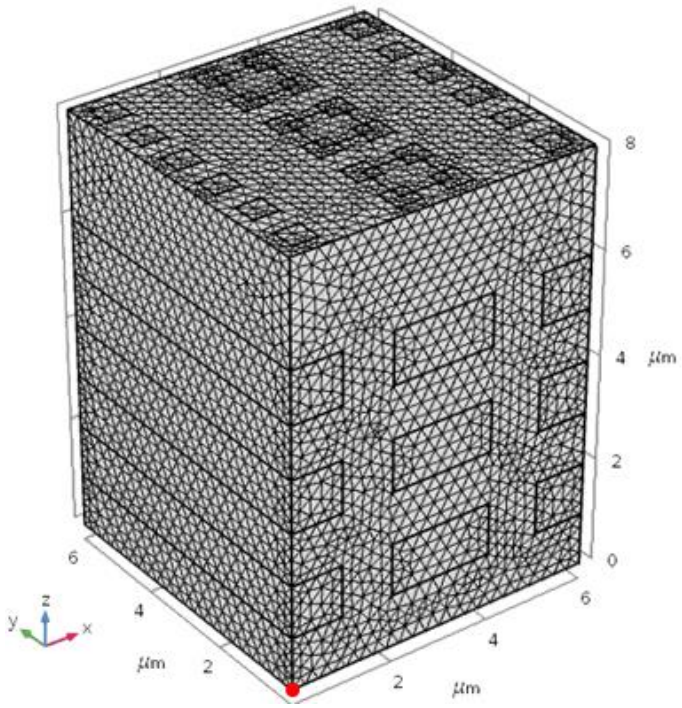


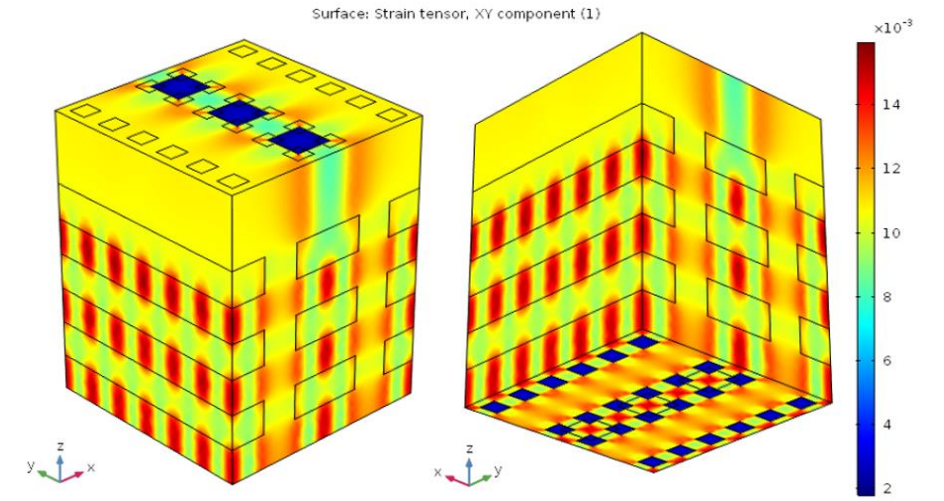

Figure 10: Meshed Model.

The first test that is performed is a shear deformation test on XY direction in order to verify if the periodic BCs assure the displacement and traction continuity conditions. The constraints are applied as follows: The node of coordinates (0,0,0), marked with a red dot in Figure 10 is constrained from movement in any direction to prevent the solid body from moving. Roller conditions are applied on boundaries normal to Z axis to prevent any deformation of those faces because the interfaces with the power metal or the silicon chip are not expected to suffer any significant deformation. The pair of opposite faces normal to X direction are forced to move alongside Y axis in opposite direction to each other with a prescribed displacement of 1% of the depth of RVE. The pair of faces normal to Y direction are forced to move alongside X axis in opposite direction to each other with a prescribed displacement of 1% of the width of RVE. The stress and strain distributions across the RVE are plotted in Figure 11. No discontinuities are observed at the boundaries, which indicate that the periodic BCs were applied correctly.

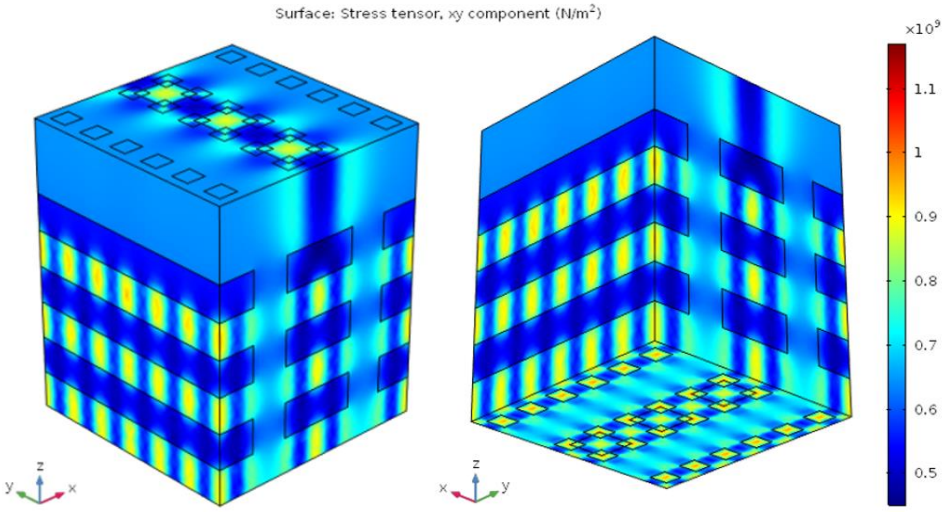
Now that the use of the PBCs has been validated, one can proceed with the rest of the six simulations as explained earlier. After the simulations are carried out only the elements on the first 3 lines and 3 columns and the elements on the first diagonal are non-zero, which indicates that the stiffness tensor is orthotropic rather than fully anisotropic:

$$[C] = \begin{bmatrix} 92.55 & 26.31 & 28.73 & 0 & 0 & 0 \\ 26.31 & 92.95 & 28.75 & 0 & 0 & 0 \\ 29.13 & 29.15 & 100.61 & 0 & 0 & 0 \\ 0 & 0 & 0 & 32.25 & 0 & 0 \\ 0 & 0 & 0 & 0 & 32.20 & 0 \\ 0 & 0 & 0 & 0 & 0 & 31.98 \end{bmatrix}. \quad (13)$$

The values symmetrical to the first diagonal have differences smaller than 2% which means that the calculated stiffness tensor is symmetrical, as expected from theory. Only the values on the first three rows and three columns and the remaining values on the first diagonal are non-zero, which indicates that the metallization is not fully anisotropic but orthotropic. This result is consistent with reality because the metallization has two symmetry planes: XZ and YZ as can be observed in Figure 8.




a)




b)

Figure 11: Shear load applied on XY direction. a): XY shear strain distribution, b): XY shear stress distribution.

From the inverse of the stiffness tensor - the compliance matrix S, the anisotropic Young's Modulus, Poison's Ratio and Shear Modulus can be derived.

$$[C]^{-1} = [S] = \begin{bmatrix} \frac{1}{E_x} & -\frac{\nu_{yx}}{E_y} & -\frac{\nu_{zx}}{E_z} & 0 & 0 & 0 \\ -\frac{\nu_{xy}}{E_x} & \frac{1}{E_y} & -\frac{\nu_{zy}}{E_z} & 0 & 0 & 0 \\ -\frac{\nu_{xz}}{E_x} & -\frac{\nu_{yz}}{E_y} & \frac{1}{E_z} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{G_{yz}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{G_{zx}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{G_{xy}} \end{bmatrix}. \quad (14)$$

The equivalent calculated material properties of the metallization are shown in Table 5.

Table 5 Equivalent material properties of the metallization

| | | |
|---|---|---|
| **Young's Modulus, E [GPa]** | $E_x$ | 79.41 |
| | $E_y$ | 80.73 |
| | $E_z$ | 86.48 |
| **Poisson's Ration, v [1]** | $v_{yx}$ | 0.21 |
| | $v_{zx}$ | 0.25 |
| | $v_{zy}$ | 0.23 |
| **Shear Modulus, G [GPa]** | $G_{yz}$ | 32.25 |
| | $G_{zx}$ | 32.20 |
| | $G_{xy}$ | 31.98 |

The extracted properties are found in between the original properties of the constituents, which is also consistent with reality. The extracted E of the composite is slightly higher (+14%) than that of the constituents, with the largest volume fraction (Al and $SiO_2$). The vias, which are made of tungsten, have an important influence because E is much larger than that of the aforementioned materials. The largest value of E is obtained along Z direction, parallel to the vias. The anisotropic v is in between the values of W and $SiO_2$. These two stiff materials seem to have the most pregnant effect on the transverse deformation of RVE: the W – through its geometrical distribution and the $SiO_2$ – through its high volume, thus preventing the Al from suffering any significant deformation, Figure 11. The shear modulus has similar values in all three directions. This is expected due to the relatively uniform distribution of the constituents inside the RVE. Similar simulations were repeated on a homogenous region with the new extracted linear-elastic properties and the strain energy was calculated in both cases for comparison (detailed structure vs homogeneous structure). For the homogeneous region a coarse mesh was used. The comparison is shown in Table 6. The strain energy is the same in both cases which indicates that the homogenization is valid. In addition, the homogenization allows for a coarser mesh, with a smaller number of elements, hence, a drastically reduced simulation time without affecting the simulation results.

Table 6 Comparison between the detailed region and homogeneous region.

| | | **Detailed region** | **Homogeneous region** |
|---|---|---|---|
| **Mesh** | | **Fine** | **Coarse** |
| **Nr. of elements** | | 302527 | 7632 |
| **Simulation time** | | 20 minutes | 4 seconds |
| **Strain energy components** | **Uxx** | 1.3183E-9 | 1.3183E-9 |
| | **Uyy** | 1.3384E-9 | 1.3385E-9 |
| | **Uzz** | 1.4488E-9 | 1.4488E-9 |
| | **Uyz** | 2.0156E-9 | 2.0156E-9 |
| | **Uxz** | 2.0126E-9 | 2.0125E-9 |
| | **Uxy** | 1.8419E-9 | 1.842E-9 |

Same simulations were carried out on a larger slice of the metallization, obtained by extending the original RVE by 3x3 on X and Y directions. The newly extracted stiffness tensor is:

$$C = \begin{bmatrix} 91.82 & 26.34 & 28.83 & 0 & 0 & 0 \\ 26.34 & 93.32 & 28.89 & 0 & 0 & 0 \\ 29.22 & 29.27 & 101.05 & 0 & 0 & 0 \\ 0 & 0 & 0 & 32.35 & 0 & 0 \\ 0 & 0 & 0 & 0 & 32.28 & 0 \\ 0 & 0 & 0 & 0 & 0 & 32.07 \end{bmatrix} \quad (15)$$

The differences with respect to the previous values are less than 0.5%, which means that the strain energy stored in both cases is the same. This indicates that the homogenization method can deliver good results at different length scales, which is an advantage of the proposed method. This method of homogenization can be used in conjunction with the one described in Section 2.4 to simplify regions

that exhibit predominant linear-elastic behaviour and require no specific investigation. The regions where the stress and strains distributions need to be analyzed in detail will be kept detailed.

## 2.4 Plastic or elastic model selection based on temperature distribution

The metallization system of the power IC's under repeated thermal stress conditions, is subjected both to plastic and elastic deformations.

Typical thermal and mechanical material properties used in the IC computational model are defined in Table 7.

Table 7 Material properties.

| | **Si** | **$SiO_2$** | **Al** |
|---|---|---|---|
| **λ** [W/ (m · K)] | eq. (2) | 1.5 | 237 |
| **ρ** [kg/ $m^3$] | 2330 | 2200 | 2700 |
| $c_v$ [J/ (kg · K)] | eq. (3) | 710 | 890 |
| **E** [MPa] (Young's Module) | 120 | 59 | 60 |
| **α** [1/K] | 4e-6 | 2.1e-7 | 2.2e-5 |
| **ν** (Poisson's ratio) | 0.18 | 0.2 | 0.334 |

FEM analysis of the thermo-mechanical problems involves the use of complex mathematical models that describe the plastic behaviour of metals. The nonlinear elastic-plastic kinematic hardening constitutive laws of the metals from the metallization system are defined according to the Chaboche model (Chaboche, 2008). Thus, the plastic behaviour is described by 8 internal variables. At the opposite end, the elastic behaviour is described by a single internal variable in the system matrix.

The assignment of the plastic behaviour, only to the mesh elements of the regions susceptible to undergo plastic deformations, is expected to finally conduct both in a reduction of the computational effort and a facile highlight of regions of high plastic deformations.

The methodology used in the assignment of plastic or elastic material properties to mesh elements is described in Figure 12. Initially all elements are initiated with elastic material properties. Temperature distribution obtained from thermal simulations is interpolated on the mechanical mesh. If the temperature exceeds the threshold imposed for plastic deformation, the corresponding ID of mesh element is added in the plastic behaviour element list.

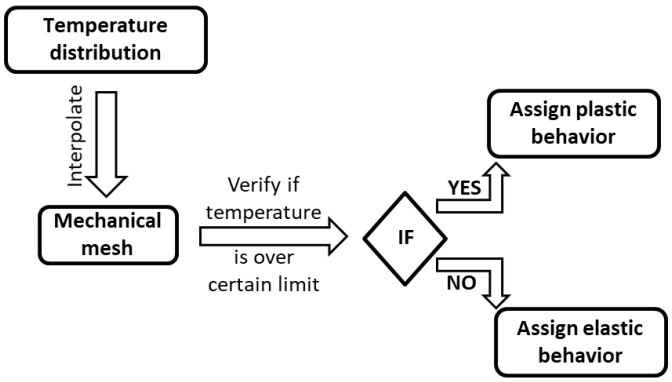



Figure 12: Methodology for assigning the plastic or elastic material properties to mesh elements.

For testing the procedure, a temperature distribution is applied according to Figure 13: a) to a simple test structure, defined by an Al bar. The mesh elements whose temperature exceeds 550K are added to the group of elements with plastic deformations, the rest are get elastic properties, see Figure 13: b).

For testing purposes, the same computational model was studied by applying a plastic behaviour over the entire computational domain, see Figure 13: c). The distribution of mechanical displacement for material properties applied as a function of temperature is presented in Figure 13: d). The mechanical displacement is identical for both studied cases in Figure 13: c) and Figure 13: d).

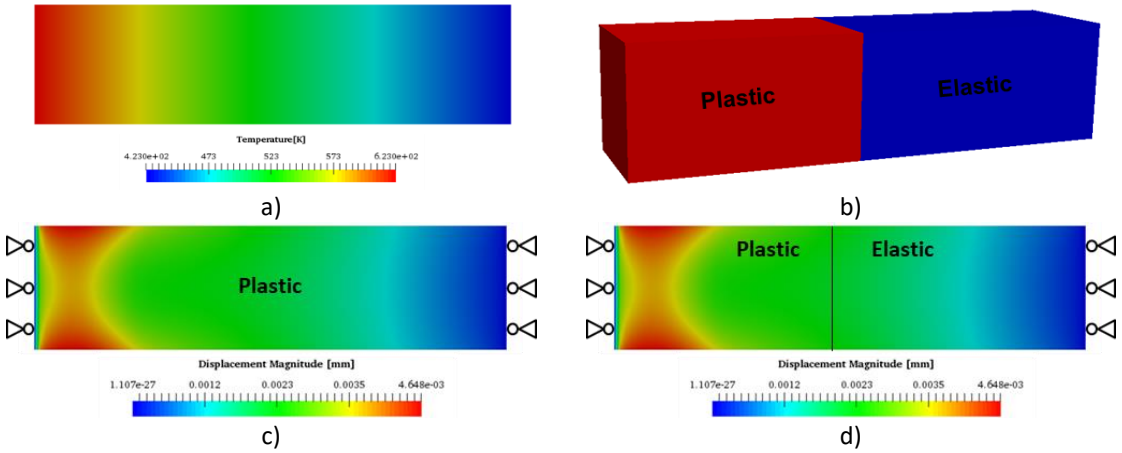
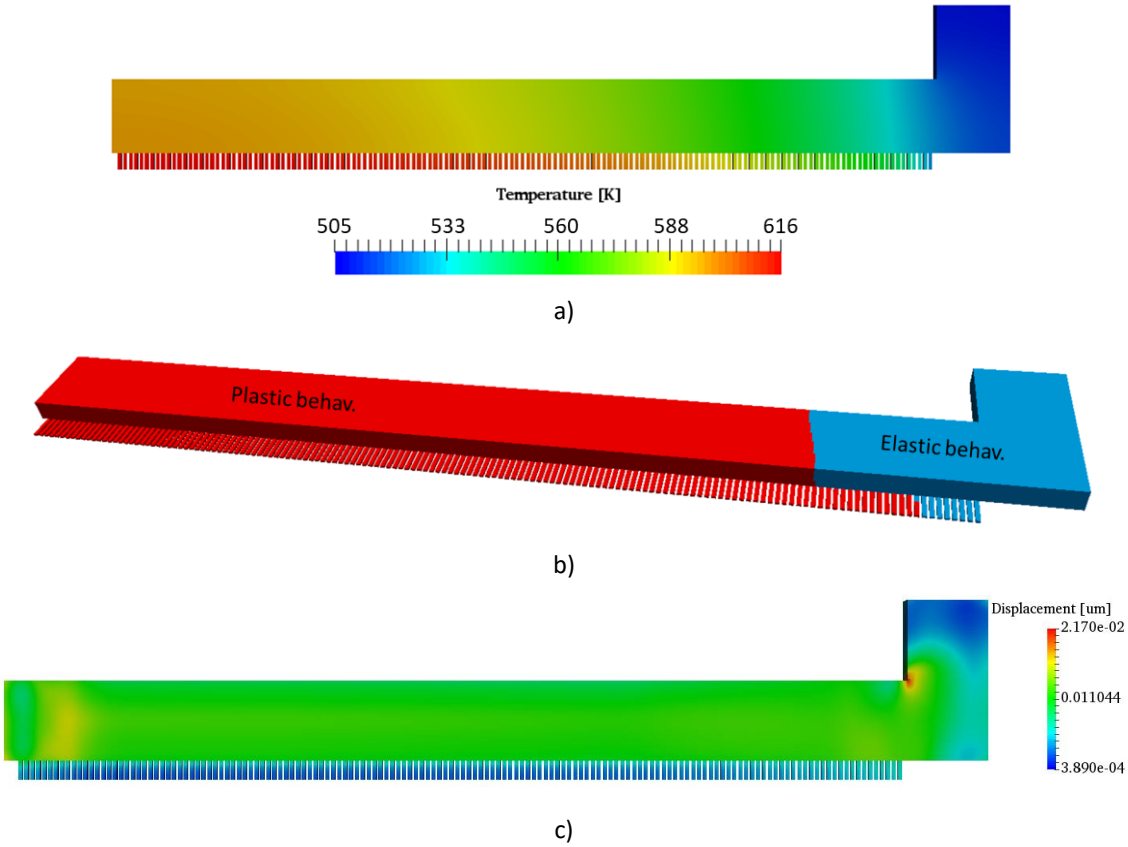


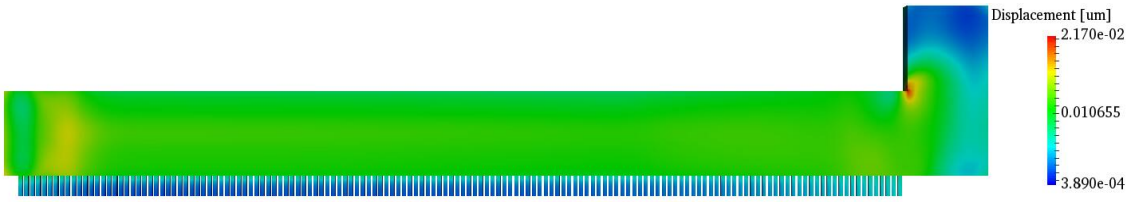

Figure 13: a) temperature distribution in the test model, b) determination of plastic or elastic behaviour depending on the temperature and c) results of mechanical deformations in the model.

Further, the methodology described above is used to simulate the IC structure presented in Figure 6 and Figure 5:. Thus, based on the temperature distribution inside the metallization system, see Figure 14: a), plastic or elastic material properties are applied to each mesh element, see Figure 14: b). The comparison of mechanical displacements, between cases where the plastic material properties are applied to metal regions entirely and the case where the metal region is divided in plastic and elastic subregions, is presented in Figure 14: c) and Figure 14: d). Both mechanical displacements are identical.

The CPU time reduced with c.a. 8% when c.a. 5% of the mesh elements were set to elastic material properties. The disadvantage is that for each time step, a check of plasticity/ elasticity domain behaviors shall be performed, hence, leading to a potential reduction of the CPU gain.

Further investigations on the computational efficiency are required and therefore will performed and presented in the following research reports.





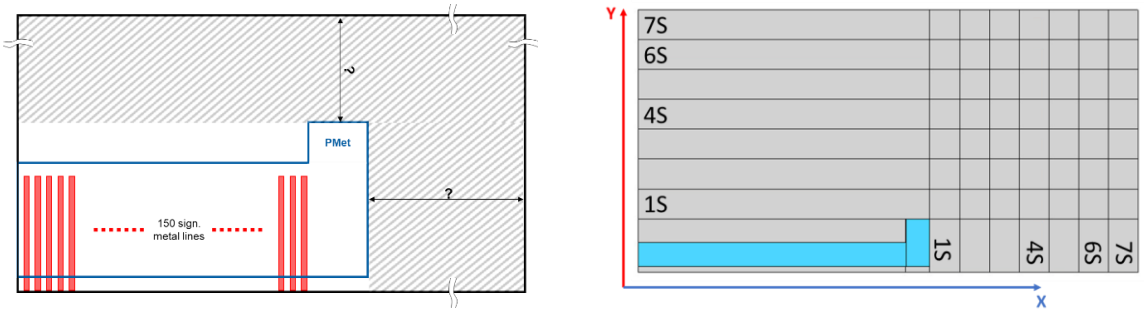

d)

Figure 14: a) Temperature distribution inside the metallization system b) subdivision in plastic and elastic material properties regions, c) displacement distribution for plastic material properties applied to the complete metallization system and d) displacement distribution for model subdivided in plastic and elastic material properties regions.

## 2.5 Si and $SiO_2$ computational model

Most of the smart power IC devices contain extra functionality integrated in the Si die in addition to the power switching area, like control, measurement, protection unit, etc. Nevertheless, device failures emerge only inside the power switching area. Thus, the extraction of this area is required in order to study the device reliability. Frequently, the simplification of the computational model is performed by symmetry, cyclic or far field boundary conditions. Therefore, the influence Si and $SiO_2$ domain size around the area of interest in order to find the optimal computational domain size for an efficient but still accurate simulation will be analyzed.

The temperature distribution and temperature gradient simulated during the electro-thermal analysis steep will final finally influence the choice of Si and $SiO_2$ substrates around the region of interest for an efficient thermo-mechanical simulation. For a minimum increase of Si and $SiO_2$ substrate dimensions, a balance between the dimensions of the added substrate and accuracy of the thermo-mechanical solution shall be established.

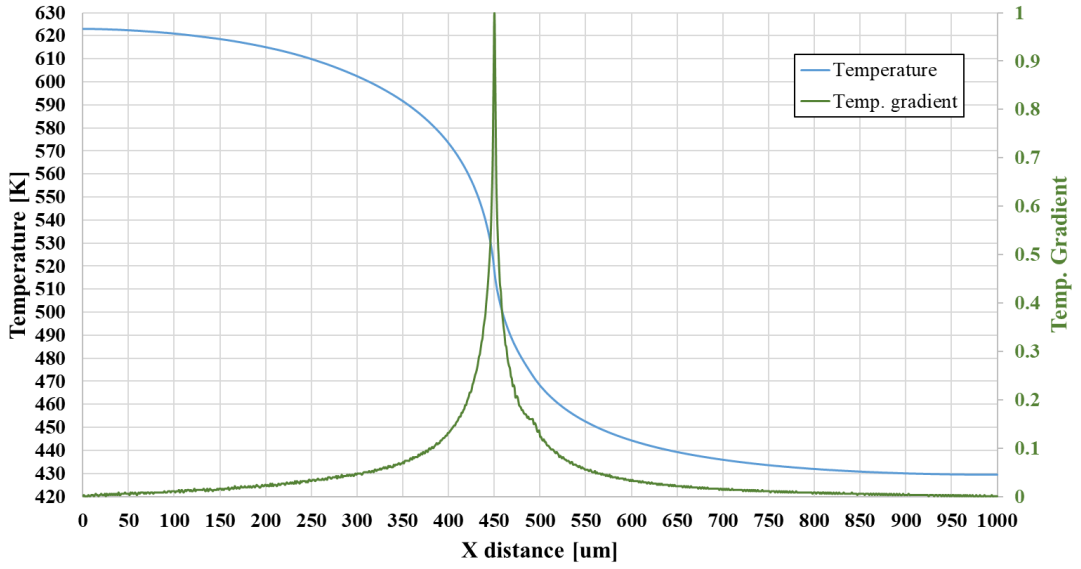



a) b)

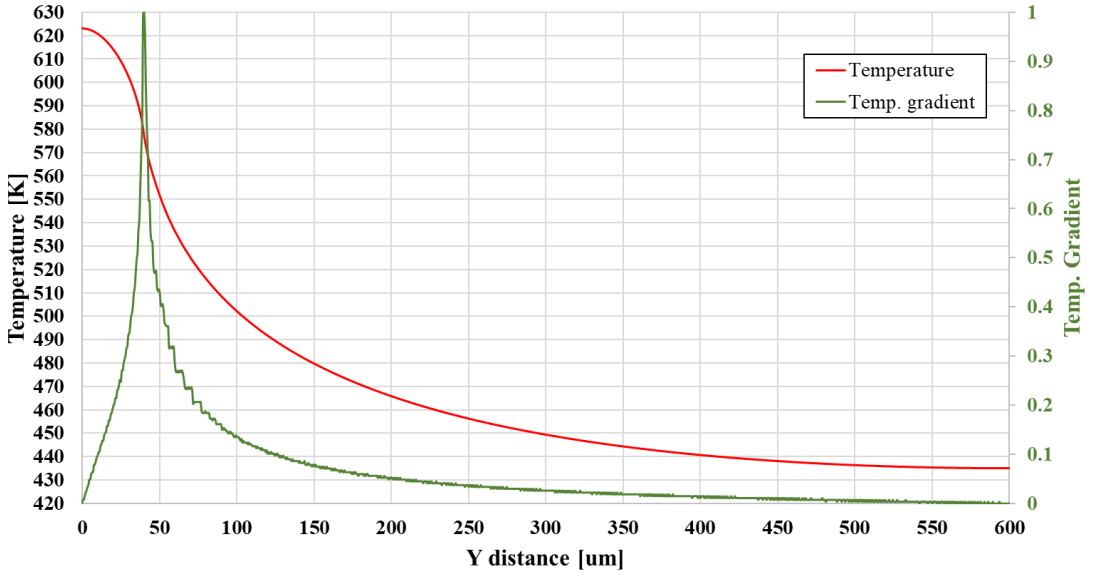
Figure 15: a) Schematic representation of computational model (dimensions are not scaled) and b) Si and $SiO_2$ size influence matrix test.

The interest region of this structure is represented by the area under $Pmet$. Starting from the IC structure presented in Figure 6 b), for fixed thickness of Si die, a set of 7 computational structures are generated, as described in Figure 15 a). The substrate of Si and respectively $SiO_2$ are progressively increased in strips of 50 μm width in both X and Y direction, Figure 15 b).

The temperature field (described in subchapter 2.2.2) produces a gradient distribution along the X and Y axes as presented in Figure 16 a) and b) respectively. The maximum temperature, 623K, is reached inside the model at the end of the power pulse. The temperature gradient decreases asymptotically to the reference temperature 780μm and 400μm along X and Y axes, respectively.




a)

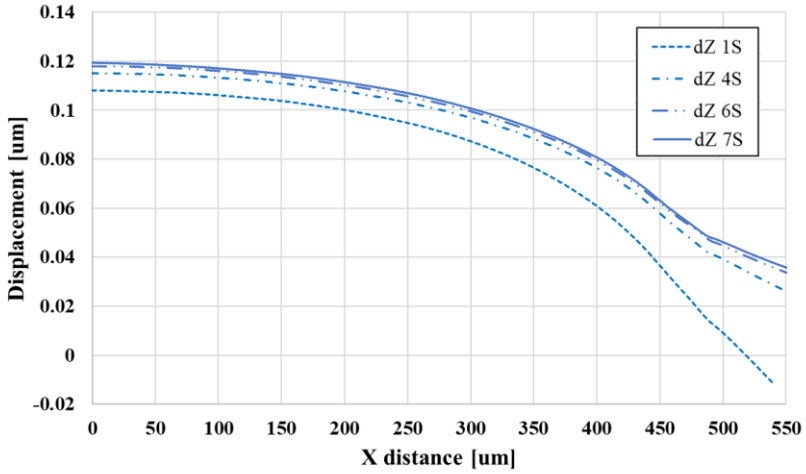



b)

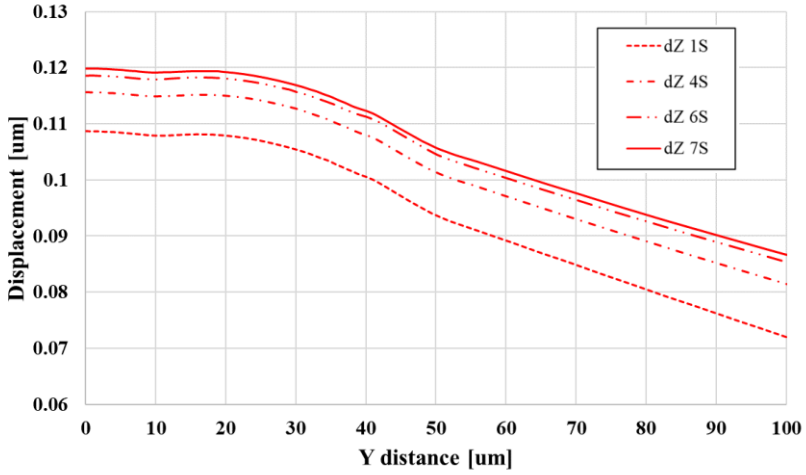
Figure 16: Temperature distribution and temperature gradient inside the model a) along X, b) along Y.

At the end of the power pulse the mechanical deformations converge for a 6-7 bands enlarged model, see Figure 17. This correspond to an area where the temperature gradient has a slope smaller than 2%.

The $Si/SiO_2$ computational model, requires for efficient, but accurate thermo-mechanical simulation, a bounding box of at least 350μm around the active (DMOS) device which cause the thermal excitation of the system.




a)




b)

Figure 17: Z displacement component at the Si/$SiO_2$ interface along X (a) and Y (b) axis, respectively.

## 2.6 Non-conformal mesh generation for modeling simplification

In case of IC's complex models, e.g. small and high aspect ratios between the topological features, complex geometrical features, etc., further optimized simulation approaches, e.g. domain decomposition and non-conforming mesh are investigated in order to increase the computational efficiency and accuracy.

Partition of the computational model is required first, in order to control de mesh density and accuracy function of zone of interest. The computational model is divided in regions of interest and the others, as flow chart presented in Figure 18: describes. This technique leads to a smaller number of mesh elements and nodes, but also to nonconforming meshes (non-matching nodes) along the multi-domain interfaces. The advantage of this method is that domains of complex topology can be independently meshed, without the failure risk and the mesh generation process can be easily parallelized. The disadvantage consists in a higher complexity of the mathematical model that shall overcome the nonconformity at the interface between non-matching domains.

An example of nonconforming mesh between two computational subdomains is presented in Figure 19 a), for a simple 3D computational domain $\Omega$ divided in two non-overlapping subdomains $\Omega_1$ and $\Omega_2$. The common shared nodes at the interface $\Gamma_{12}$ occur only at the corners and are highlighted with red dots. The face $\Gamma_1$ of $\Omega_1$ (meshed with triangular elements) is in contact with the face $\Gamma_2$ of $\Omega_2$, (meshed with quadrilateral elements), see Figure 19 b).

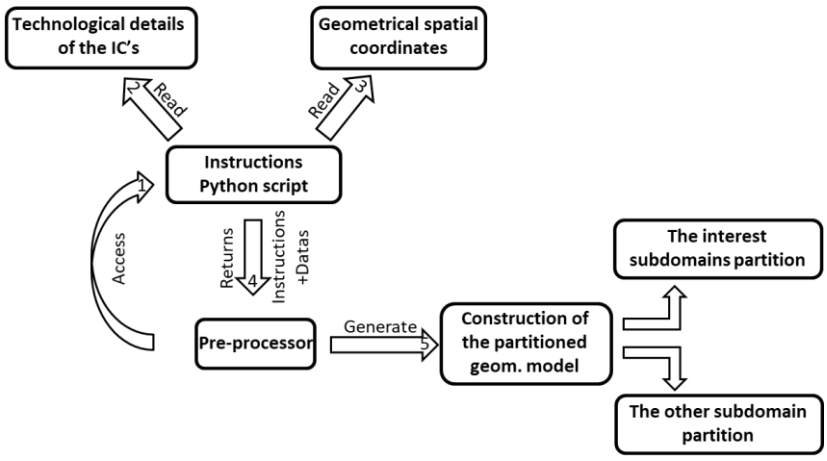
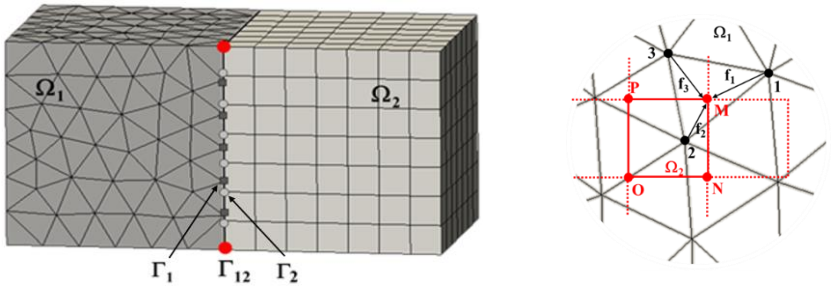



Figure 18: Automated partition computational domain process.




a) b)

Figure 19: Example of a) nonconforming mesh between the subdomains $\Omega_1$ and $\Omega_2$ and b) solution transfer at $\Gamma_{12}$ interface, from $\Gamma_1(\boldsymbol{\Omega_1})$ to $\Gamma_2(\boldsymbol{\Omega_2})$, based on the element shape functions $f_i$, where $\boldsymbol{i = \overline{1,3}}$

In case of thermo-mechanical simulation, the solution of the thermal induced displacement inside each subdomain is computed from the following equations (Fritz 2004):

$$-div\,\sigma(u) = f \qquad \text{in } \Omega \tag{16}$$

$$u(x) = 0 \qquad x \in \partial\Omega \tag{17}$$

$$\sigma\big(u(x)\big)\bar{n} = 0 \qquad x \in \partial\Omega \tag{18}$$

where, *f* is the displacement inside $\Omega$ domain, *x* is the computational node coordinate on $\Omega$ domain's boundary $\partial\Omega$ and $\bar{n}$ is the unit normal vector. The system of equations holds for a prescribed displacement *u* and given stresses *σ(u)* imposed on the outer boundary ($\partial\Omega$) of the computational domain.

At the non-conforming interface $\Gamma_{12}$, a linear relation is imposed for transferring the field values between $\Omega_1$ and $\Omega_2$ (Dureisseix and Bavestrello 2006):

$$\sigma(u_1)\overline{n_1} = \sigma(u_2)\overline{n_2} = \lambda \quad \text{on } \Gamma_{12} \tag{19}$$

where $\overline{n_i}$ for $i = \overline{1,2}$, is the unit outward normal vector and $\lambda$ is the Lagrange multipliers vector that ensures the equality of a specific field between $\Omega_1$ and $\Omega_2$.

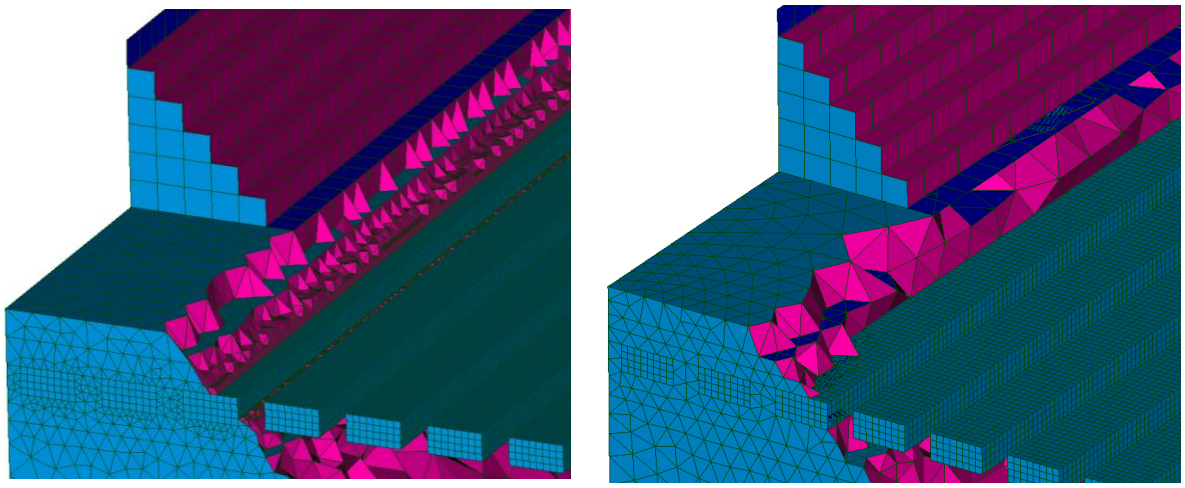
The displacement field u1 is transferred from $\Omega_1$ to $\Omega_2$ by linear interpolation, see Figure 19 b). Firstly, for an arbitrary node $M \in \Gamma_2 \subset \partial\Omega_2$ the value of the displacement field $u_1(M)$ is searched in the adjacent mesh element of $\Gamma_1 \subset \partial\Omega_1$. Secondly, the value of $u_1(M)$ is computed from the finite element shape functions with eq. (18) (Dureisseix and Bavestrello 2006):

$$u_2(M) = u_1(M) = N_1(M)U_1 \tag{20}$$

where $N_1$ is the vector of weights of the mesh element shape functions which fulfil the partition of unity. The term $U_1$ is the global nodal vector of the displacement field.

Meanwhile, on the non-conforming interface $\Gamma_{12}$ each node of $\Omega_2$ is locked by its projection on the adjacent domain $\Omega_1$. Although the couplings at the $\Gamma_{12}$ interface, defined with Lagrange multipliers, introduce additional lines in the system matrix, it still remains symmetric and no longer positive definite (Babuska, 1973).

Further, the non-conformal mesh between different domains technique is used to generate the mesh inside the IC simulation structure presented in Figure 6 and Figure 5:.The signal metal lines are set as the region of interest. In these conditions, the mesh is easily controlled and refined in the interest region, using different topology of mesh elements (e.g. hexahedrons) and higher density than the surrounding regions, as presented in Figure 20: b). Thus, the number of computational nodes is reduced compared to the conformal discretization, 1.55e6 nodes for the non-conformal mesh compared to 2.48e6 for the conformal mesh, and the time used to generate the spatial discretization inside the model is also significantly reduced.



a) b)

Figure 20: Conformal and non-conformal discretization of the computational structure in a), respectively b).

Although the non-conformity at the domain interfaces involves advanced interpolation schemes, the new proposed meshing approach offers higher flexibility by independent generation of the mesh on each model region, thus the refinement can be performed only inside the domains of interest.

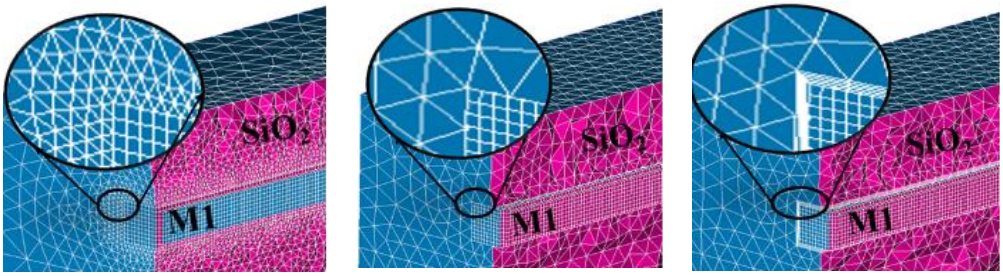
Further, this technique is applied in the DMOS computational structure, where the signal metal layers represent the region of interest.

## 2.7 Use of intermediate viscous layers

The non-conformity nodes at the domain interfaces introduce discontinuities inside the system matrix of the studied problem. Advanced interpolation schemes based on the Lagrange multipliers (Babuska, 1973) are used to handle the discontinuities. The results interpolation between multi-domain non-conformal interfaces can be improved by adding Viscous Layer (VL) mesh elements along the non-conforming interface. In addition the VL approach can be used to simplify the modeling of very thin intermediate layers within the metal stack.

The 3D VL elements are mesh elements with constant height and usually are represented by pentahedron for triangular mesh and hexahedron for the quadrangle mesh. The constant height of the layers around the non-conformal interface improve the interpolation at the interface and the progressive growth of the layers height leads to a continuous result on $\Gamma_{12}$, consistent with literature (Erwin 2011).

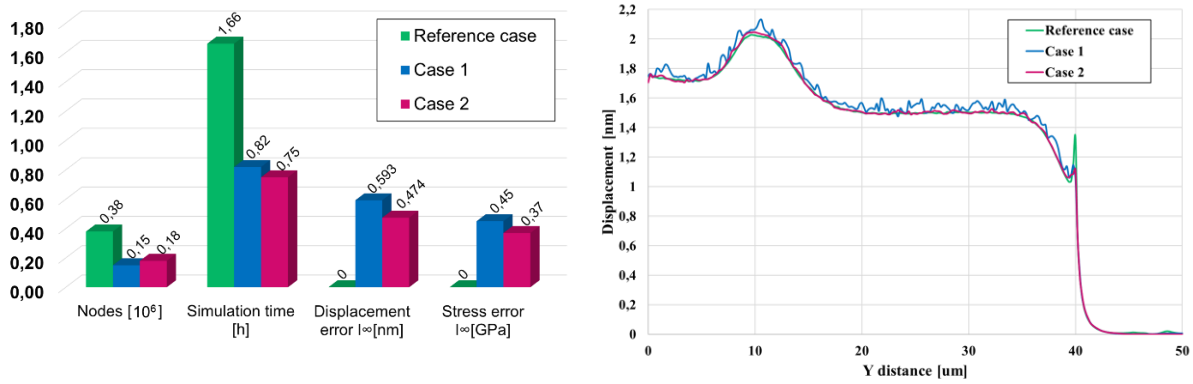
For the results comparison, the computational structure is further discretized with a fully conformal mesh (Reference case), non-conformal mesh without any VL (Case 1) and VL mesh elements (Case 2) in the $SiO_2$ region at the interface with metal lines, see Figure 21:.




Reference case Case 1 Case 2

Figure 21: The Reference case with conformal mesh distribution and Case 1 and 2 of nonconforming mesh between M1 and $SiO_2$ block without and respectively with VL.

The $dX$ component of the displacement field extracted along the signal metal line follows the same displacement distribution in all cases. The noisy results of displacement in Figure 22: b), obtained for Case 1, are smoothed out by adding 3 VL inside the surrounding $SiO_2$ region of the $\Gamma_{12}$ interface. Case 2 improves the accuracy of the results (fits the Reference case) and reduces the simulation time compared to the Case 1, see Figure 22: a).




a) b)

Figure 22: a ) Comparison between number of computational nodes, simulation time and maximum errors for the studied cases, b) x displacement component along the routing metal length.

Further, the methodology is extended to a full chip complex computational model, with full stack of signal metallization lines, extracted from a power DMOS structure. The signal metallization system consists in several thin metal layers. The model symmetry allows considering a quarter of DMOS active area further in simulations, Figure 23:. Three VL at the non-conforming interface are used for the mesh refinement. The non-conformal mesh consists of 795,770 computational nodes and the computational time of the mechanical simulation is 1h and 22min. The conformal mesh consisted of 996,542 computational nodes and computational time was 2h18min. Computational time reduced more than 40%.

The cumulative plastic strain is extracted in order to compare the results accuracy between conformal and non-conformal mesh approaches. The results are extracted within the interpolation Gauss point for more accuracy. The cumulative plastic strain obtained from simulations performed on the complex DMOS structure predict the failure spot obtained by experiments, as showed in Figure 24. Both simulation approaches, with conformal and nonconforming mesh capture the same peak plastic strain values at the edges of thin signal metal lines.

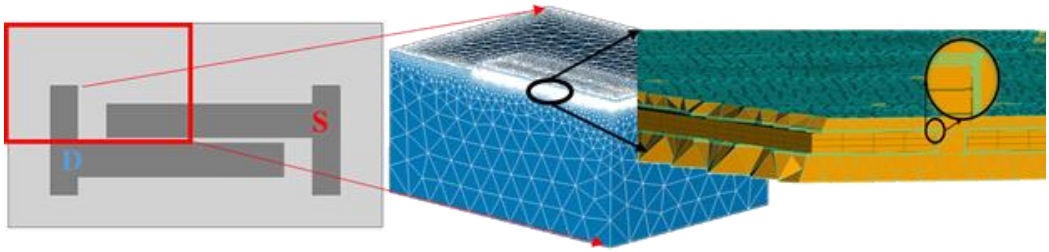



Figure 23: Schematic view of DMOS computational model (left) and mesh distribution inside the extracted model, with the nonconforming discretization of signal metal lines and surrounding $SiO_2$ (right).

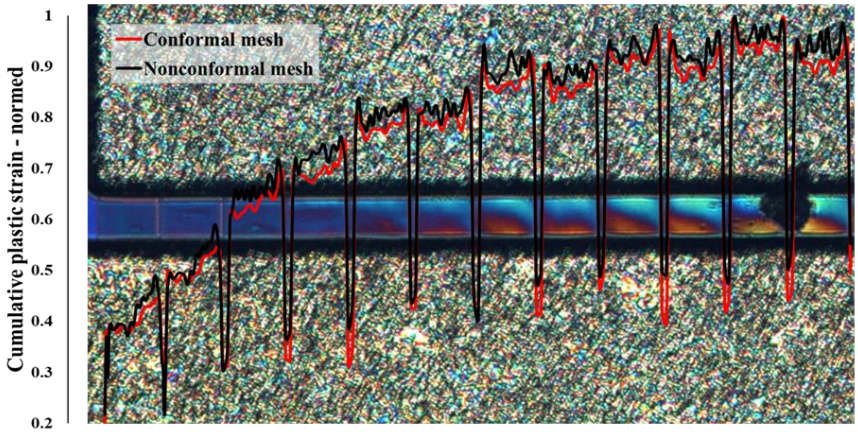



Figure 24: Image of differential interference contrast (Nomarski) from a failed device superimposed with cumulative plastic strain simulated results (normed values).

The kinematic coupling between the nonconformal mesh regions leads to good approximation of displacement and stress results compared to fully conformal mesh regions. Introducing of VL at the non-matching interfaces leads to more accurate results and reduced simulation time. This technique yields accurate results without mesh generation limitations for complex geometries. The proposed technique provides very good accuracy of results and smaller computational time compared to the conventional conformal case.

## 2.8 Conclusions

This report briefly presents the main methods and models for efficient simulation of electro-thermo-mechanical stress during the fast operation of power IC.

Computational domain simplification methods in chapter 2.1 are presented for more efficient thermo-mechanical analysis in versus of computing resources and CPU time.

Chapter 2.2 describes the method and models used to generate the thermo-mechanical stress conditions, which are then used for the initial assessment of the domain simplification methods.

Chapter 2.3 presents homogenization methods for thermal and mechanical properties. For mechanical properties it shows that if the system remains in the elastic domain, the method can be applied successfully, with a significant reduction in the simulation time without any loss in accuracy.

Chapter 2.4 presents the association of the plastic or elastic material properties to each mesh element in the metallization system depending on the temperature distribution. The technique is firstly tested on a simple computational model and extended to IC computational model described in chapter 2.1 resulting a computational time decreased with 10% compared to classical approach.

Connection between the temperature profile and the substrate dimensions is described in chapter 2.5. This is then used to determine the size of the Si and $SiO_2$ domains in order to reduce the simulation domain and thus improve simulation time.

Chapter 2.6 presents the methodology used to generate efficient computational model based on non-conformal meshing at the interface between the computational domains. The advantage of this method is that domains of complex topology can be independently meshed, without the failure risk and the mesh generation process can be easily parallelized. The disadvantage consists in a higher complexity of the mathematical model that shall overcome the nonconformity at the interface between non-matching domains.

Chapter 2.7 presents VL meshing at the nonconformal mesh interface between the model domains for improving both results accuracy simulation time. An example on real DMOS structure and comparison with measurements is presented.

The results of the first stage evaluation and the speed accuracy trade-off analysis for some of the methods presented in this chapter will be included in deliverable D7.8.

# 3 Test structure for validation of electro-thermo-mechanical simulation flow

## 3.1 Overview

In order to validate the simulation results for the targeted physics (i.e. electro-thermal and thermo-mechanical) and failure modes (i.e. IMD cracking), measurement results are needed. The studied DMOS power devices are repeatedly operated at high power levels, for $10^4$-$10^9$ on-off switching cycles. These devices act as heat sources, dissipating energy in their metallization system. Due to the cyclic heating/cooling, during which the constitutive materials expand/contract, the metallization system will suffer a progressive degradation, which will ultimately lead to chip (and) system failure.

## 3.2 Description of test structure

In order to validate any simulation results they need to be compared to real data extracted from measurements performed on a real device. A test chip containing two power DMOS devices was specially designed for this purpose. Though the structure of the two transistors is almost identical, there are some minor differences, because each transistor serves a different purpose. One of the transistors is equipped with a temperature sensor and serves as a calibration device while the other transistor is equipped with mechanical sensors used for the detection of some specific failure mechanisms during repetitive thermal cycling tests. The calibration device is subjected to a relatively small number of predetermined power pulses (in which various powers, pulse widths, and frequencies are tried) and the temperature will be captured during each test. The power-temperature relation will serve both as a mean to calibrate the electro-thermal simulation setup and to agree upon an adequate testing scenario for the other device. The transistor with embedded mechanical sensors will be operated under repetitive high energy pulses so that its metallization will be subjected to repetitive thermal cycling which will lead to the its degradation due to the repeated expansion-contraction of the materials in the metal stack, resulting in device failure after a number of cycles ranging from 104 to 109.

The detailed description of the test chip will start with the characteristics which are common to both transistors and the differences will be specified afterwards. The power transistors are composed of thousands of basic DMOS cells connected in parallel. All sources and drains respectively are shorted together to their corresponding terminals to which we will refer from now on as SF (source-force) and DF (drain-force). The device under test (DUT) floorplan is presented in Figure 25. The gate connections are grouped in two terminals, G1 and G2 respectively:

G1 – the gate terminal corresponding to the rest of the DMOS cells (the green region in Figure 25).

G2 – the gate terminal corresponding the DMOS cells in the middle of the device (the blue region in Figure 25).

In this way it is possible to control the region where energy is dissipated through activating or deactivating the G2 terminal by connecting it to the source terminal, SF (G1 will be always active during testing).

The maximum temperature is obtained in the center. Here is where the metallization is expected exhibit the most pronounced degradation. By deactivating G2 terminal, the temperature and the degradation in the center of the device are reduced. The power metal plates are represented by the pink rectangles and they are connected to the aforementioned terminals, SF and DF. The vertical yellow lines represent the sensors used to detect the inter-metal-dielectric (IMD) cracking and the red dots represent the power metal delamination sensors.

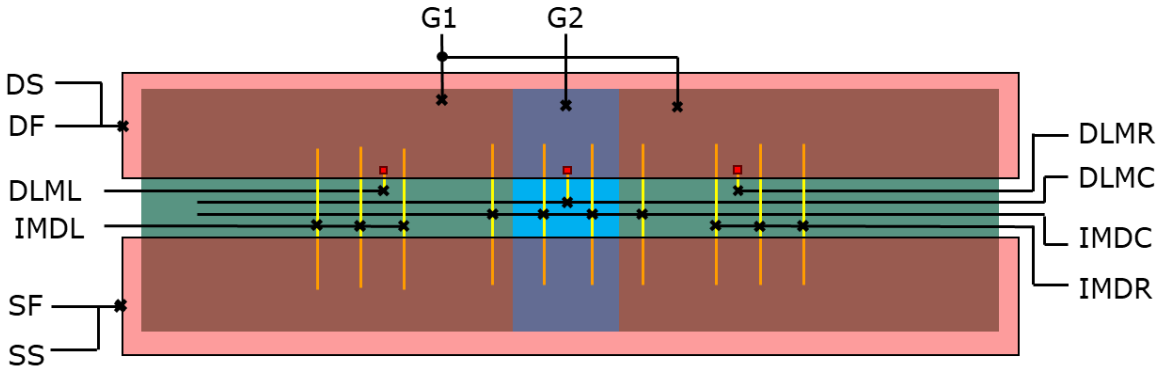



Figure 25: The DUT floorplan: G1 – gate terminal, green region; G2 –gate terminal, blue region; sensors to detect IMD cracking – yellow lines.

The device under test is equipped with integrated sensors and sensing terminals which will provide information in the form of electrical signals as follows:

1. The sensing terminals (found only on the calibration device) are used for accurate monitoring of the source and drain potentials and are referred to from now on as SS and DS (source-sense and drain-sense). The electrical signals coming from these sensing terminals will be used for accurate measurement and control of the drain-source voltage - $V_{DS}$, and the drain current - $I_D$. More details will be given in the sections 3.3 and 3.4.
2. The temperature sensors (found only on the calibration device) track the maximum temperature of the device. One such sensor is placed in the center of the device, as shown in Figure 26 and it consist in a thin meandered metal line. The electrical resistivity of this thin metal line varies with temperature. Therefore, the temperature is monitored by measuring the resistance of the meander. The resistance is measured in Kelvin connection (force and sense terminals) for better accuracy (more on section 3.4).

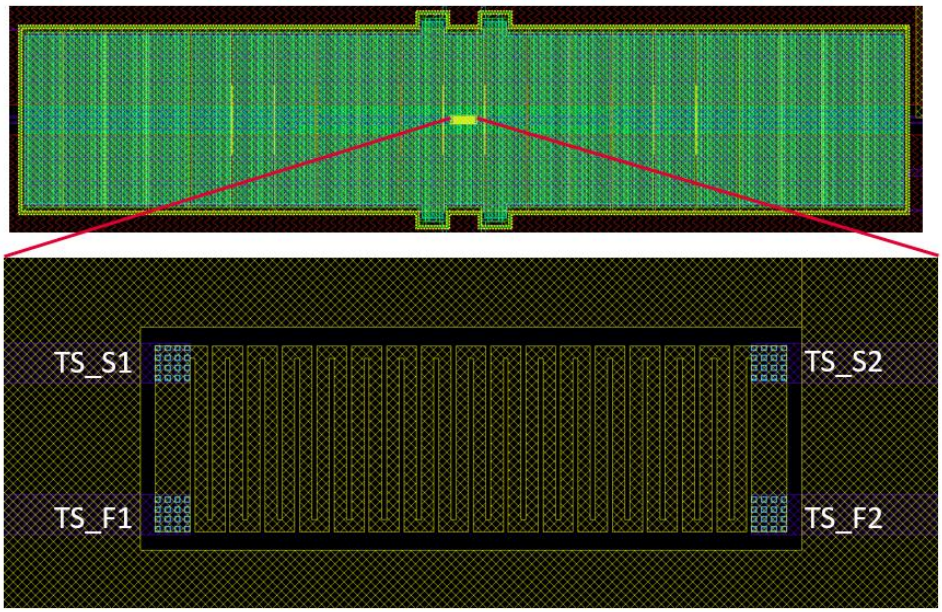



Figure 26: Top: the temperature sensor position on the device layout. Bottom: detailed layout of the temperature sensor consisting in thin metal meandered line.

3. The mechanical sensors (routed only on the second device) provide information regarding the stage of mechanical degradation of the metallization structure of the device. In order to describe the functionality of these sensors it is necessary to briefly explain the failure mechanisms emerging in modern power devices metallization that can be detected with these types of sensors.

a. Due to the repeated heating-cooling cycles through which the device undergoes, the materials composing the metallization will expand-contract differently, given their mismatch in thermal expansion coefficient. The thick metal layer on top (to which we will refer from now on as power metal) being for about 10 times thicker than the underlying routing metal lines, will expand up to 100 times more than the underlying insulator layer and up to 10 times more than the silicon substrate. With the time, through repeated plastic deformation, due to the shear forces that act upon it, the power metal will delaminate starting from the hottest edges.

   Therefore, the DUT will be equipped with a delamination sensor which consists in a via that connects the hottest margin of the power metal and the underlying signal metal line (the KVIA zone in Figure 27 and the red dots in Figure 25). The electrical resistance between the two metals is monitored (It is actually a series resistance composed of: the bond wires, the power metal plate resistance, the via resistance and the underlying metal lines resistance, as shown in Figure 27- left). When delamination takes place, the contact surface between the power metal and the underlying metal reduces, therefore an increase of the measured resistance will be observed.

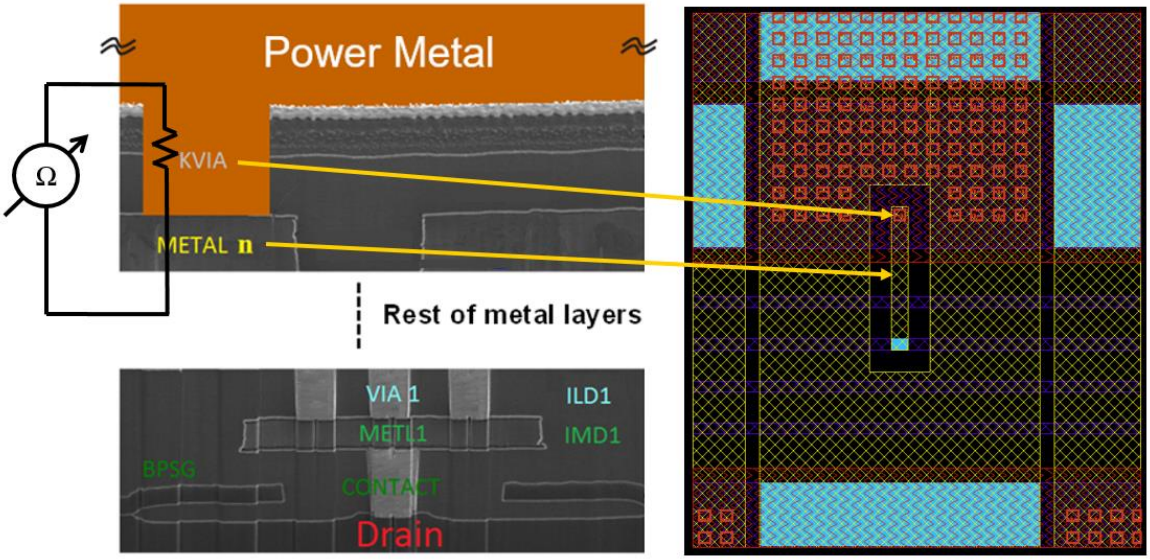



Figure 27: Left: cross-section through a modern power device metallization; the metallization connects the source, drain and gate of DMOS devices manufactured in high-power-high-integration BCD technology; power metal in orange; KVIA - power metal delamination sensor for monitoring series resistance composed of bond wires, power metal plate resistance, KVIA resistance and underlying routing metallization resistance. Right: delamination sensor layout.

b. Also due to the repeated heating-cooling cycles and due to the mismatch in thermal expansion coefficients of the power metal, the silicon die and the metal-insulator assembly also known as IMD – inter-metal-dielectric, which lies between the silicon surface (“drain” label in Figure 27-left) and the power metal, another thermo-mechanical phenomenon can cause mechanical degradation that leads to device critical failure. This failure mechanism manifests in the cracking of the insulator (e.g. silicon dioxide) (Smorodin, et al., 2008).

   This way, through cumulative plastic deformation of the metals in IMD, the mechanical force exerted on the oxide exceeds a critical point and a crack appears in the insulator that separates two adjacent metal lines. The metal fills up the crack and a short-circuit appears between the metal lines.

   In order to detect this phenomenon a crack sensor was designed. The sensor consists in a narrow metal line placed between two wider adjacent metal lines with functional role (lines that carry the source and drain signals). The sensor and the routing metal lines are

situated on the same metallization layer. While the device operates under repetitive thermal cycling, the wider metal lines will suffer from cumulative plastic deformation. When the pressure exerted on the oxide exceeds its yield limit, the oxide will crack (Figure 28).

In the absence of the crack sensor (the thin metal line), the crack will propagate directly from one line to the other (source and drain), resulting in a short-circuit that is destructive. When this sensor is present (Figure 29), however, the crack will propagate from the drain or source metal line to the sensor and the short-circuit is not destructive. If the potential measured on the sensor terminal is the source potential, then it means that the crack appeared between the source and the sensor. Otherwise if the potential measured on the terminal is the drain potential, it means that the crack appeared between the drain and the sensor.

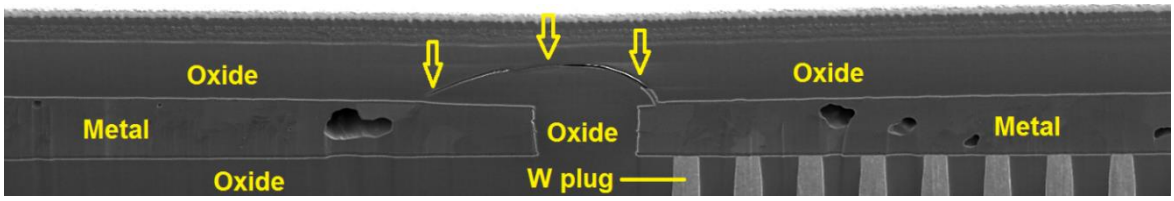



Figure 28: Cross section through the metallization of a power transistor. Only the layer right underneath the power metal is visible. One can observe that between the two metal lines a crack has formed due to the cumulative plastic strain in the metal. When the insulator cracks, the adjacent metal lines carrying drain and source signals will be shorted which will result in metal meltdown in the vicinity.

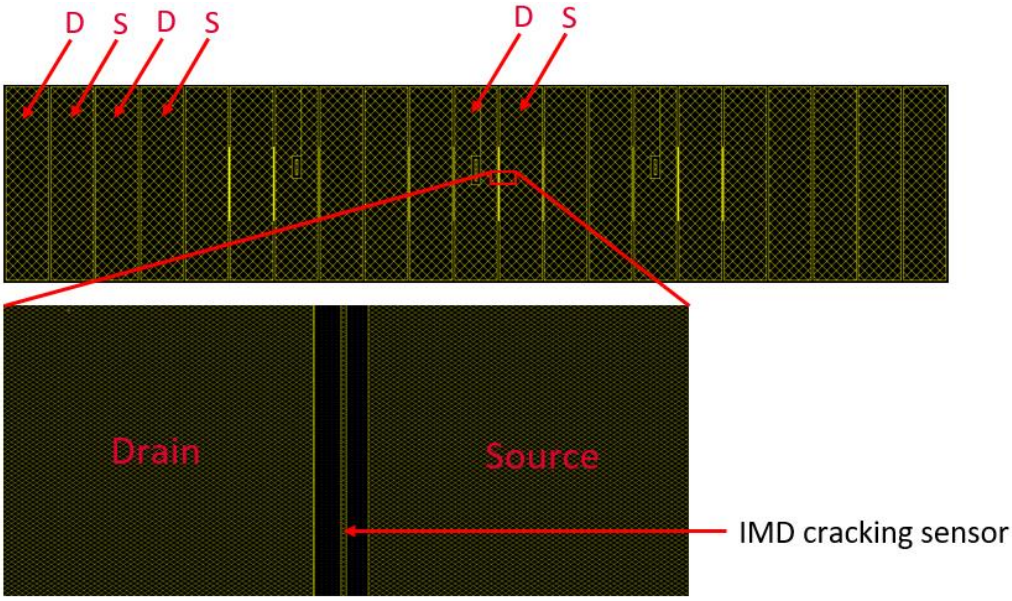



Figure 29: Top view of the metallization layer right underneath the power metal; metallization layer is defined as a succession of the wide metal lines carrying the source and source and drain signals, separated by a narrow region of oxide. A crack sensor (see the very narrow metal line in the bottom image) is placed between the two wider lines, having the role to prevent the crack propagation from source to drain. Hence, the crack appears between the sensor and one of the two lines and the short-circuit is not destructive. The early degradation stage is detected by monitoring the sensor terminal potential.

In order to decide where it was most suitable to place the mechanical sensors across the transistor area, some electro-thermal simulations were performed on a simplified model of the test chip which contains the silicon die, a simplified version of the transistor routing metallization and the power metal. Two cases were simulated: one where the transistor is operated at 100% active area and one where the central region is deactivated (G2 shorted to SF). The high-risk areas are the hotspots. The temperature distribution in power metal is shown in Figure 30. This serves as an indicator for areas

with high risk of delamination. The locations for delamination sensors should be near the hotspots, as close as possible to the power metal line edges.

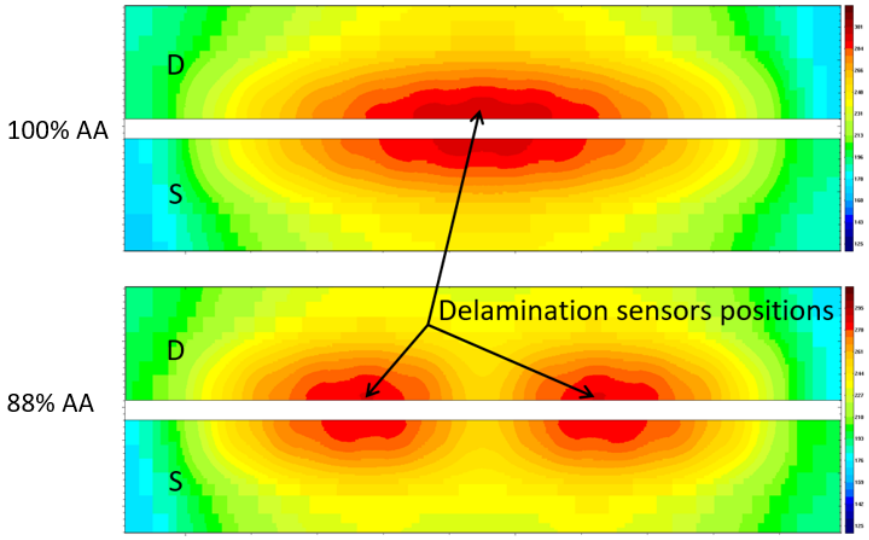



Figure 30: The temperature distribution in the power metal for 100% active area and for 88% active area respectively. The hotspots indicate areas which are prone to delamination and those are the locations where the delamination sensors should pe placed.

The temperature distribution in the thin metal layer right underneath the power metal is shown in Figure 31. and offers information regarding the high-risk areas for IMD cracking. The temperature is higher, and its distribution is more uneven than in the power metal.

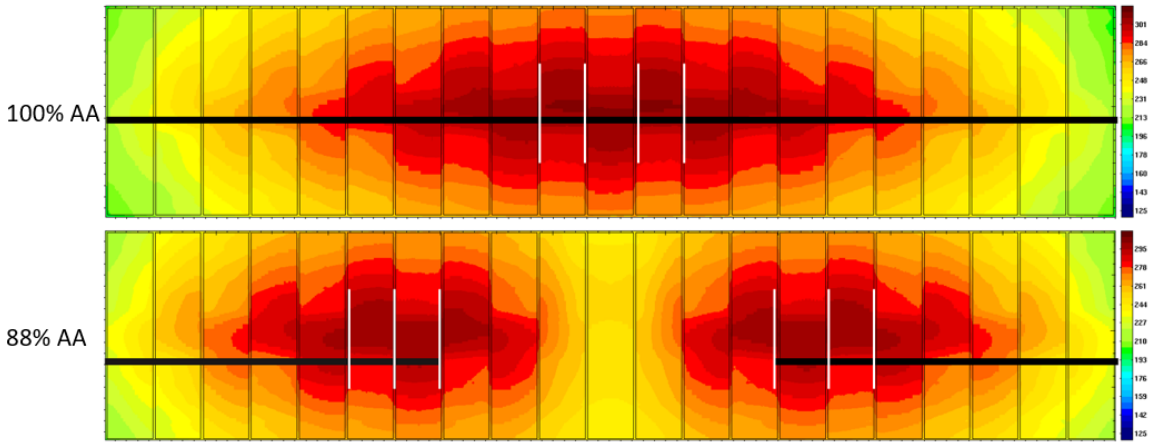



Figure 31: The temperature distribution in the metal layer right underneath the power metal for 100% active area and for 88% active area respectively. The hotspots indicate areas which are prone to IMD dracking, and the sensors must be

The areas prone to IMD cracking are those with the most pronounced temperature gradients. Another observation is that the temperature difference between two adjacent metal lines experiences a shift along the vertical direction when passing through the canter. This is due to the way this layer of thin metal line is connected to the power metal. This observation indicates that each of the two metal lines is equally probable of initiating the crack, therefore, the thin metal lines used as crack detectors must be long enough and disposed symmetrically to the horizontal axis that crosses the center of the device, in order to capture the early stage of crack formation. In addition, multiple such sensors must be grouped together in such a manner that they “contain” the high-risk areas and will be more likely to detect the defect. A suggestion of grouping such detectors is overlayed with white lines on the plot in Figure 31. This is the adopted solution for the test chip.

The layout of the whole test chip is presented in Figure 32. There are two DMOS transistors on this chip. The one on the bottom is the calibration structure which contains the temperature sensor and the one on the top is the transistor equipped with mechanical sensors, used in the repetitive thermal cycling tests. The source, drain, gate and other terminals are routed to the metal pads numbered from 1 to 24, as the chip is encapsulated in a CDIP24 package. The description of each terminal/pad is listed in Table 8.

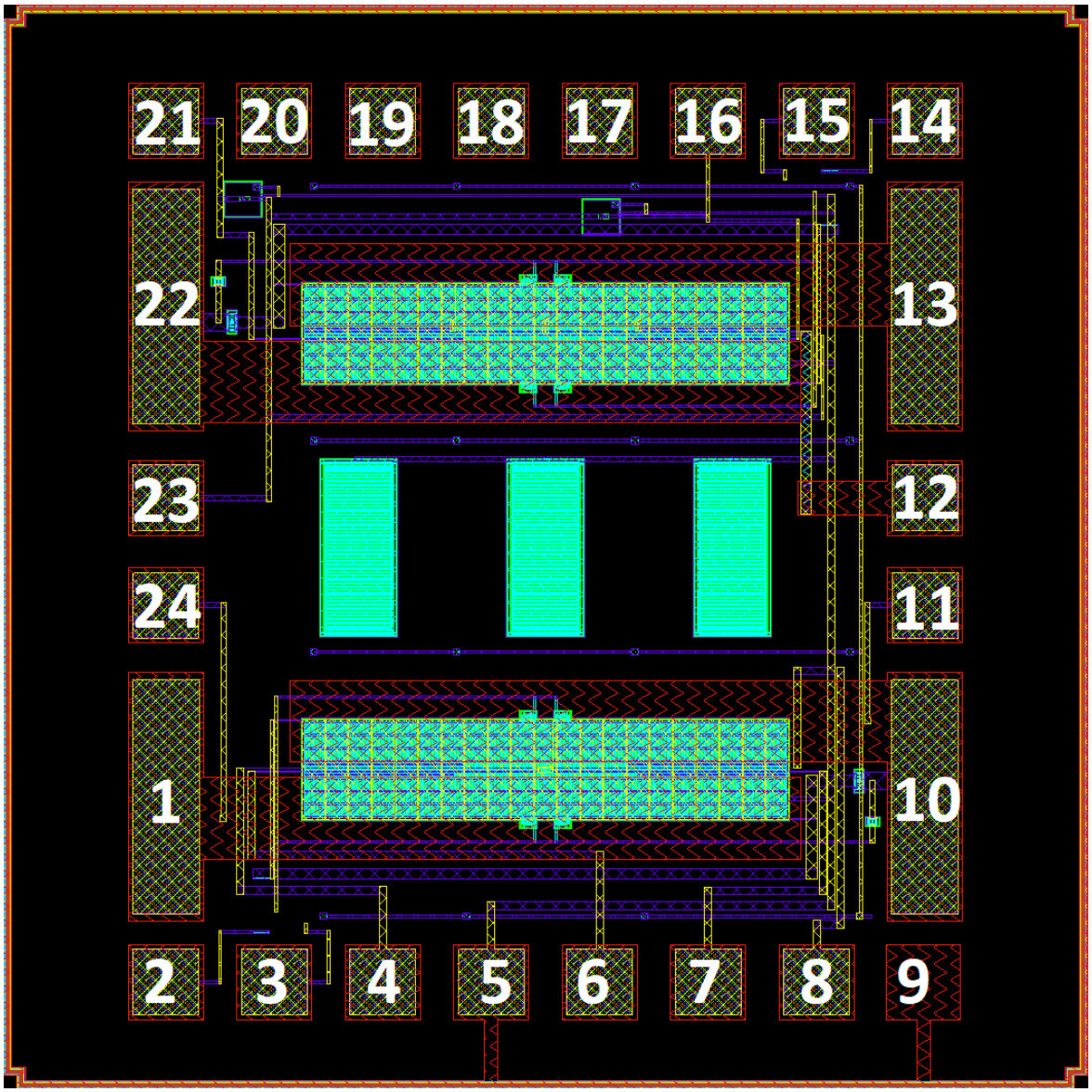



Figure 32: Top view of the whole test chip (layout). There are two DMOS transistors on this chip. The one on the bottom is the calibration structure which contains the temperature sensor and the one on the bottom is the transistor equipped with mechanical sensors, used in the repetitive thermal cycling tests. The aforementioned terminals (SF, SG, G1, G2 etc) are routed to the metal pads numbered from 1 to 24, as the chip is encapsulated in a CDIP24 package. The description of each terminal/pad is listed in Table 8.

Table 8 The description of each terminal/pad.

| Pad | Terminal Description | Transistor purpose |
|---|---|---|
| 1 | Drain Force, DF | calibration |
| 2 | The main gate that is always on, G1 | calibration |
| 3 | The central gate used for deactivation, G2 | calibration |
| 4 | Temperature sensor sense terminal no. 2, TS_S2 | calibration |
| 5 | Substrate, common to both transistors | both |
| 6 | Temperature sensor force terminal no. 2, TS_F2 | calibration |
| 7 | Temperature sensor force terminal no. 1, TS_F1 | calibration |
| 8 | Temperature sensor sense terminal no. 1, TS_S1 | calibration |
| 9 | Not connected, NC | N/A |
| 10 | Source Force, SF | calibration |
| 11 | Source Sense, SS | calibration |
| 12 | Delamination sensor | repetitive thermal cycling |
| 13 | Drain Force, DF | repetitive thermal cycling |
| 14 | The main gate that is always on, G1 | repetitive thermal cycling |
| 15 | The central gate used for deactivation, G2 | repetitive thermal cycling |
| 16 | Crack sensor, right side | repetitive thermal cycling |
| 17 | Not connected, NC | N/A |
| 18 | Not connected, NC | N/A |
| 19 | Not connected, NC | N/A |
| 20 | Not connected, NC | N/A |
| 21 | Crack sensor, left side | repetitive thermal cycling |
| 22 | Source Force, SF | repetitive thermal cycling |
| 23 | Crack sensor, center | repetitive thermal cycling |
| 24 | Drain Sense, DS | calibration |

## 3.3 Test System Design for Fast Thermal Cycling

It is not plausible to obtain measurement data by replicating the actual mission profiles undergone by the chips. This requires very long test times, possibly as long as several months or years (automotive grade chips are specified for 10,000 hours of operation). Instead, a test system for accelerated lifetime tests, which emulates the high-power conditions undergone by DMOS switches, is needed. Such a test system should be able to allow:

- to force the targeted degradation mechanisms
- to reliably extract the time to failure data (in number of cycles or number of hours to failure)
- testing as many devices (DUT) as possible in parallel
- easy control of the stress level
- to apply the same level of stress for all devices
- monitoring end-of-life criteria
- monitoring other quantities of interest (current through the device, temperature etc.)
- the protection of the test system and other DUTs when a DUT fails by, e.g., short-circuit

In order to force the targeted degradation mechanisms, the stimulus (high power pulses with low duty cycle) should be in the suitable range. This means, that the power should cause a significant junction temperature swing, during a short period of time: several hundreds of microseconds up to a few milliseconds. The period of the power pulses should be low enough for the device to cool down. The current level should be low enough to avoid other degradation mechanisms like electro-migration. The voltage level should be low enough to avoid degradation mechanisms like hot carrier stress. All these observations will place constraints on the electrical stimuli used to obtain the targeted destruction mechanism.

A microcontroller is added to the test system, so it will be possible to generate trigger signals to stress each device, in sequence. The microcontroller can *easily count the number of cycles to failure of each*

*DUT* and store them in its memory. The microcontroller, in case of a power outage of the electrical grid, can also easily save the state of the test system. Moreover, the circuits that monitor the sensor signals can interact with the microcontroller. Then, based on the input from these signals, it will make a decision: for example, if a region from a DUT is about to fail, the sensor responsible for monitoring that region will raise a flag and the microcontroller can cut-off power for that region (as detailed in Section 3.2).

A test system capable of testing *8 devices in parallel* is designed. This number was chosen from complexity and cost reasons. Moreover, the DUTs are assembled in ceramic packages with 24 pins, which are bulky and occupy a lot of space on a PCB. The DUTs need to be placed in a small oven, because they need to operate at high environment temperatures. Therefore, the DUTs are placed on a special PCB (burn-in board) and is connected to the test system through special cables, which can sustain high temperatures (Figure 33 presents the layout of this PCB).

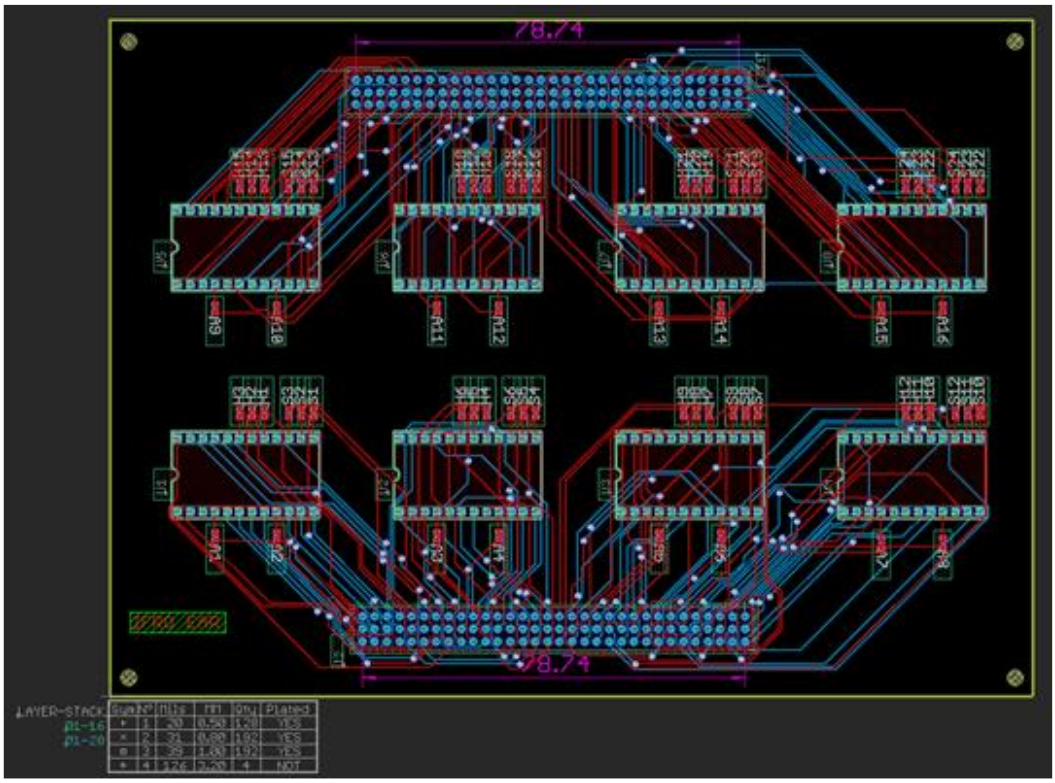



Figure 33: Layout of the burn-in PCB is used for mounting the devices in ceramic C-DIP-24 packages. The PCB contains 8 sockets for 8 ceramic chips and 2 connectors with 92 pins each. The connectors are used to establish the connection between the DUTs and the external control and monitoring test system. All PCB components can sustain up to 125°C.

***The stress level is easily controlled*** by a MOSFET gate driver, which automatically adjusts the current through the DUT to an arbitrary shape specified by the user. The drain-source voltage will be fixed during the power pulse; hence, the shape of the power pulse will be given by the shape of the drain current. The measurements were done by a prototype of the designed test system and are shown in Figure 34:.

***The same level of stress will be applied to all devices*** because only one gate driver is used and its output is de-multiplexed to each DUT. The gate driver will automatically adjust the gate potential of the DUT based on the reading from an electrical sensor, which provides the voltage drop on a precision shunt resistor (placed in the source of the DUTs). Therefore, the variation of the control circuit (gate driver offset, sense resistors etc.) is eliminated, which means that the same power level will be applied to all devices. Moreover, the stress conditions should not drift during the long test periods, in order to assure the same stress levels every time. That means that the minimum junction temperature ($T_{MIN}$) and maximum junction temperature ($T_{MAX}$), temperature swing $\Delta T$ should not change during the test.

Figure 34: shows the dissipated power, the above test parameters for a DUT, the temperature of the test system and the temperature of the oven. It can be observed that the minimum chip temperature rises from 125°C (oven temperature) and settles at 150°C after 5000 cycles. The temperature swing/rise during a power cycle is ΔT=200°C and it is constant during the entire 8 hours of the test period.

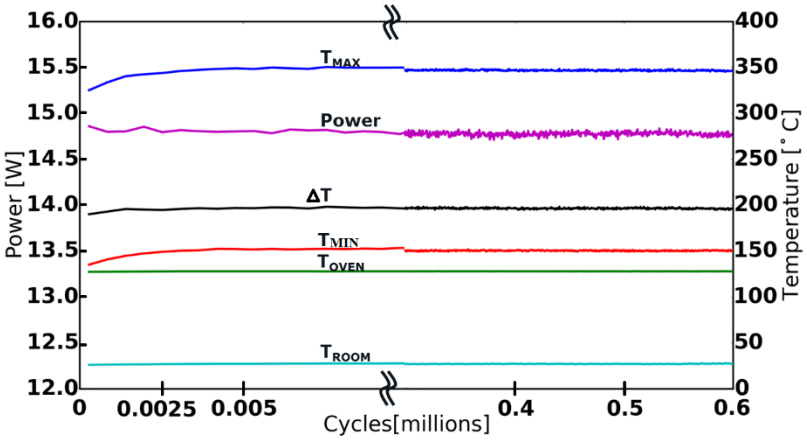



Figure 34: Power dissipated by a DUT, measured during 600k power cycles (c.a. 8.3h testing with a pulse period of 50ms – 20Hz).

Bias and monitoring circuitry are implemented in the test system to allow measurement of sensor signals. The sensors provide *information on the end-of-life criteria*. The mechanical sensors described in Section 4.2 provide the information for end-of-life criteria. A separate PCB (called extension PCB) is designed to bias and collect information from these sensors. In addition to this, a circuit implementing a feedback loop is designed, which allows to deactivate a device region (for example the active region controlled by the gate G1) when a mechanical sensor detects a defect. The schematic of this PCB is presented in Figure 35.

***Other quantities of interest are monitored by electrical sensors:*** drain-source voltage and drain current as presented in Figure 36. The junction temperature can also be monitored because the system has bias circuits (adjustable current sources) for integrated temperature sensors of two types: metal meanders and diodes.

***Protection and detection of DUT failure*** by either short-circuit or open circuit is necessary to both protect the test system for electrical over stresses and to record the cycles to failure of the device. Moreover, by implementing fast reacting protection circuits, the DUT will not be heavily damaged and failure analysis (i.e. optical inspection, focused ion beams etc.) will be possible afterwards:

- DUT short-circuit between drain and source can happen, as explained in paragraph 3.b) from Section 3.2.
- If the gate of the DUT cannot be controlled anymore by the gate driver (e.g. due to gate-source short-circuit) a circuit which detects the absence of current through the device will raise a flag and the microcontroller will interpret that the respective DUT has failed.
- The test system also includes an overheating protection circuit, in case the main board PCB gets too hot due to malfunction (the tests are supposed to run unattended, overnight).

The microcontroller is at the heart of the test system, but a PC is used to set-up the laboratory instruments (power sources, multimeters, oscilloscope) and to initialize the test. The microcontroller and the oscilloscope send to PC the data acquired on a periodic basis and the PC logs the data to text files.

Finally, the short-circuit protection circuit block, open-circuit detection, over-temperature protection, gate driver, temperature and electrical sensor bias and monitoring circuits can all be observed in the

schematic from Figure 36 of the main board. For clarity reasons, the connectors to external laboratory instruments and the extension PCB are not shown.

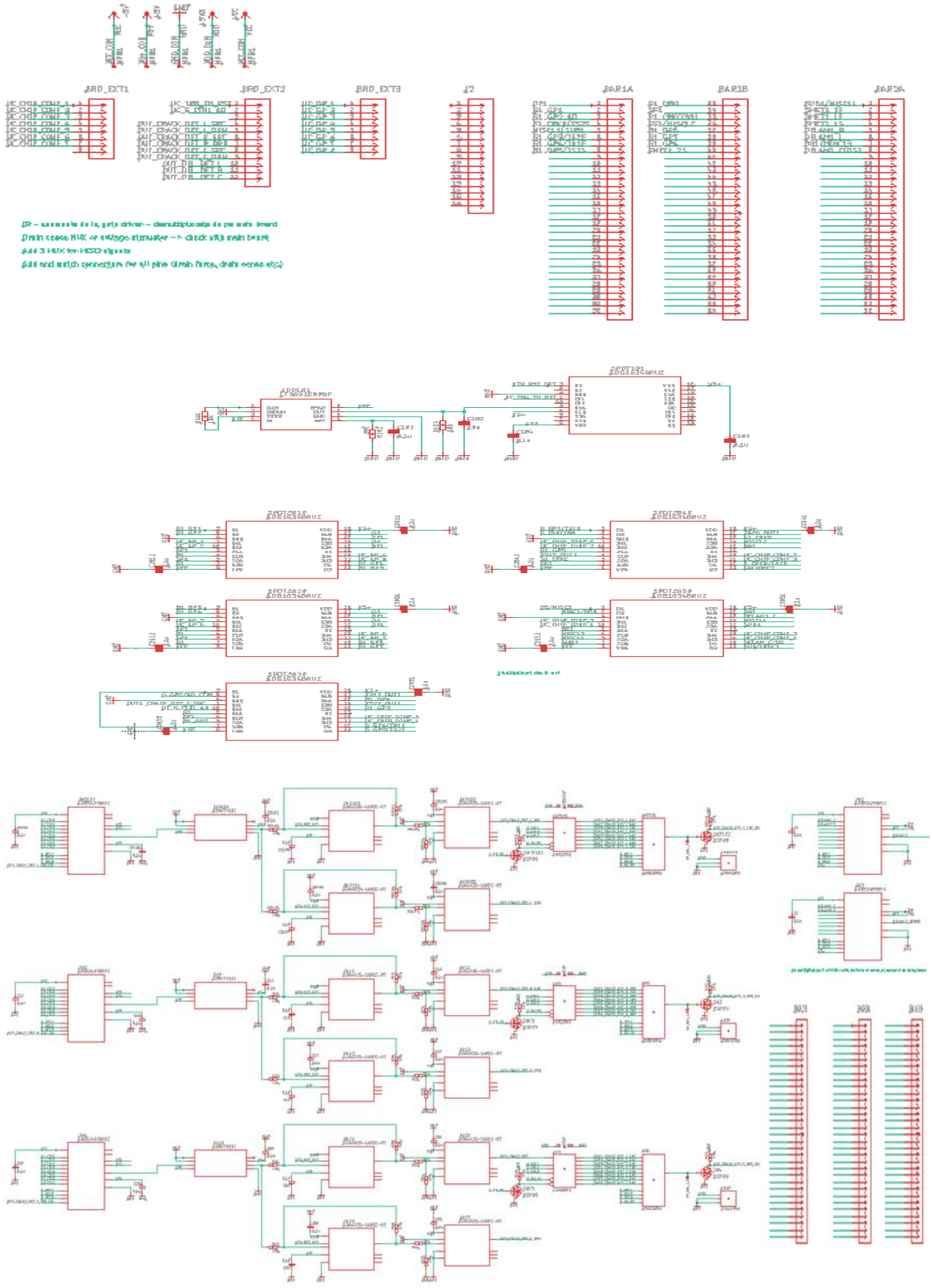


Figure 35: Schematic design of PCB extension, used to bias and monitor the mechanical sensors based on a feedback loop that automatically shuts-down the device region where destruction is about to occur.

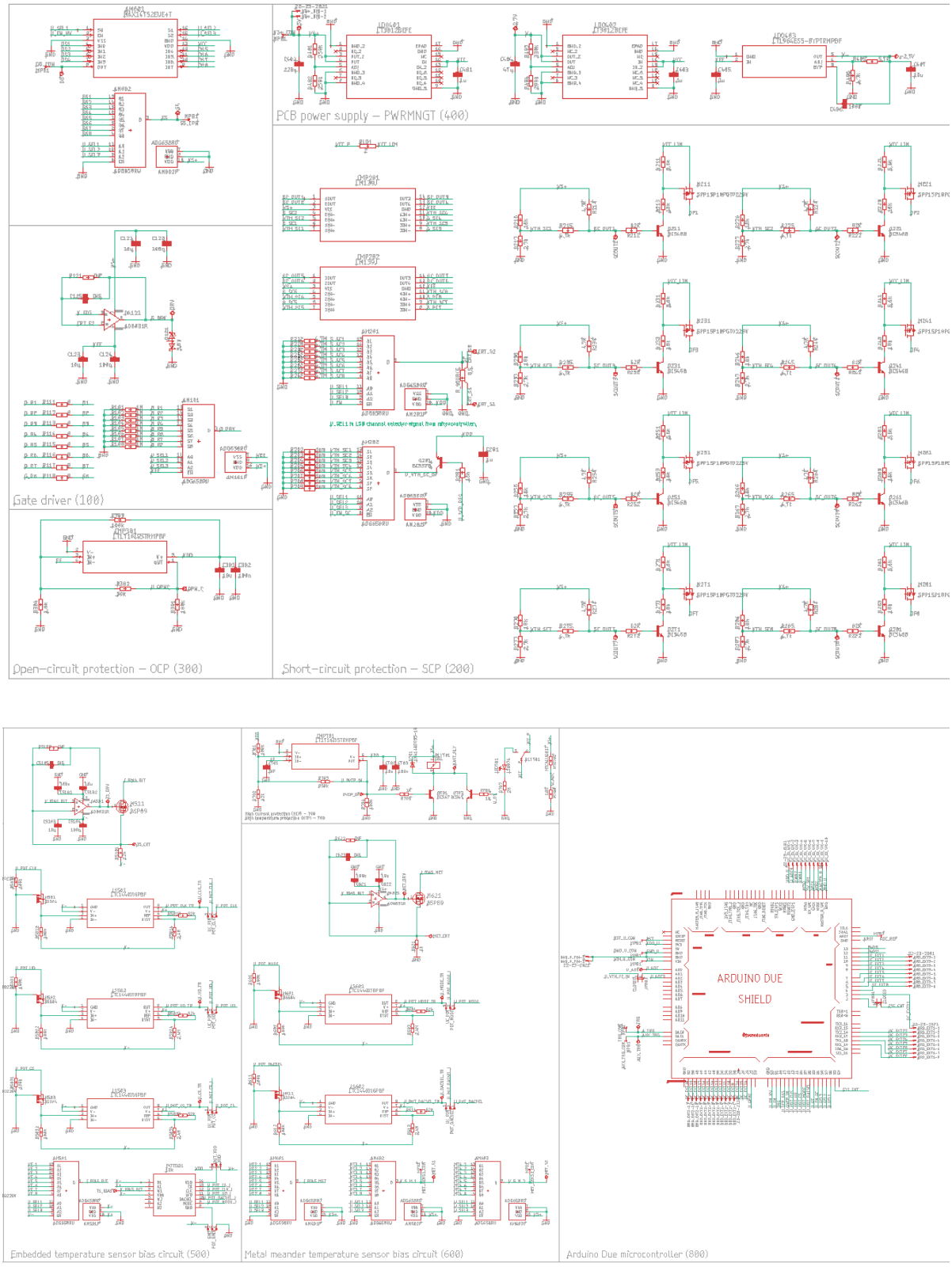


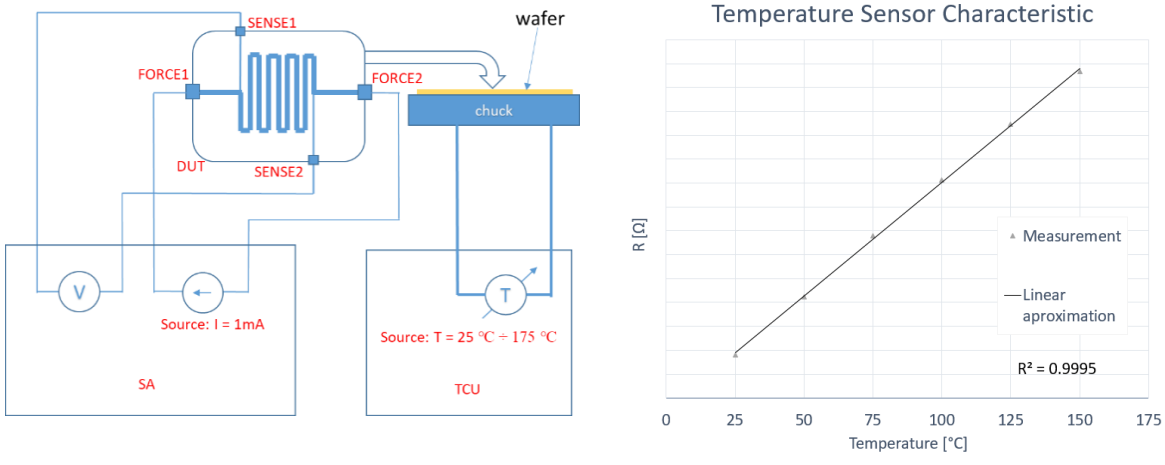

Figure 36: Schematic design of the main board PCB that controls and monitors the eight DUTs. The design contains circuit blocks for: short-circuit protection, open-circuit detection, bias for electrical and temperature sensors, multiplexers and de-multiplexers that allow control and monitoring of each DUT. It also contains the microcontroller which controls and synchronizes the test system.

## 3.4 Initial test results

Once the test chips were manufactured some preliminary test were necessary in order to verify the basic functionality of the transistors. First measurements were performed on-wafer. The silicon wafer containing the test chips was tested using a wafer prober and a semiconductor analyzer.

The first measurement was the characterization of the temperature sensor found on the calibration structure (Figure 26). The temperature sensor is a temperature-dependent resistance with a Kelvin connection (force and sense terminals) for accurate measurement. The wafer is fixed on a metal plate (chuck) whose temperature is controlled by a temperature control unit (TCU). A schematic representation of the setup is shown in Figure 37 a). A constant current of 1mA is forced through the FORCE terminals (TS_F1, TS_F2 in Figure 26) and the voltage is monitored at the SENSE terminals (TS_S1, TS_S2 in Figure 26). The temperature is swept from 25°C to 150°C with a step of 25°C. The obtained temperature characteristic is shown in Figure 37 b). One can observe that the temperature characteristic is linear, which is very helpful for measuring high on-chip temperatures.

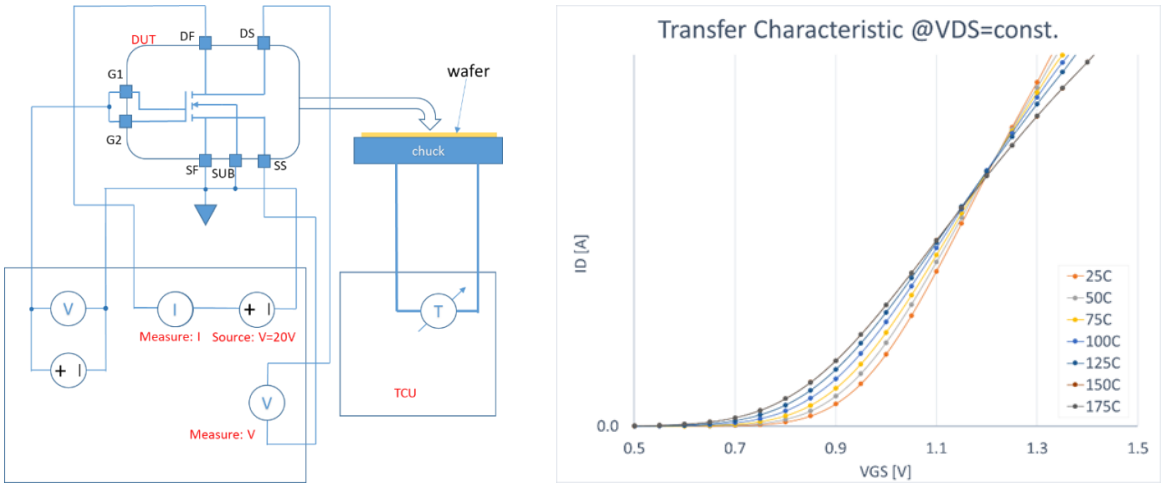
The next measurement was to trace the transfer characteristic of the DMOS transistor used for calibration, ID(VGS) for multiple temperature. The experimental setup is shown in Figure 38 a). The SF and SUB terminals are connected to the ground (V=0V). The DF terminal is kept at a DC potential V=20V. VDS is measured between DS and SS and is 20V. The temperature is swept from 25°C to 175°C with a step of 25°C. The two gates G1 and G2 are connected together. At each temperature step, the gate potential is swept from 0.5V to 1.5V with 0.05V step and the gate to source voltage VGS and the drain current ID are measured. The measurements results are shown in Figure 38 b). The transfer characteristics for different temperatures intersect in one point, known as thermal compensation point or TCP. VGS(TCP)=1.2V. TCP is very important in power electronics because DMOS transistors operating at VGS <VGS(TCP) (ID rising with temperature due to VTH decreasing with temperature) are likely to fail due to thermal runaway or due to activation of the intrinsic parasitic bipolar of the DMOS.




a) b)

Figure 37: On-wafer temperature sensor characterization: a) schematic representation of the experimental setup; b) the measured characteristic.

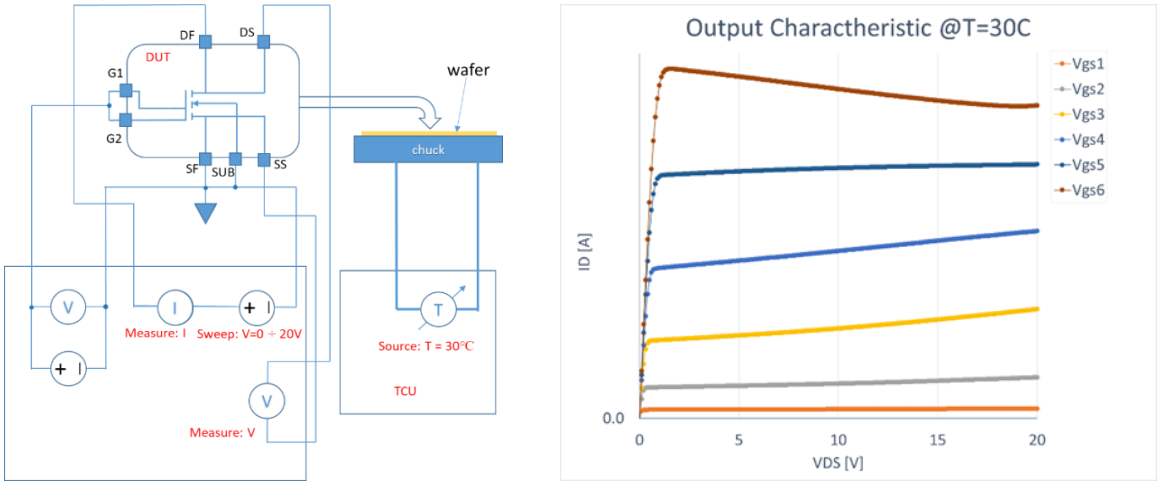
The next measurement was the tracing of the output characteristic of the transistor used for calibration at fixed temperature, for multiple gate-to-source voltages, ID(VDS). The schematic representation is shown in Figure 39 a). The SF and SUB terminals were connected to ground (V=0V). The temperature of the chuck is kept constant at 30°C. The who gates G1 and G2 are connecte together. The gate potential is stepped from 0.8V to 1.3V with 0.1V step and the gate-to-source voltage (VGS) is measured. At each VGS step the DF potential is swept from 0V to 20V with 0.1V step. The drain-to-source voltage VDS is measured between DS and SS terminals. Drain current ID is also measured. The experimental results are shown in Figure 39 b). From the plot one can observe that for high $V_{GS}$, $I_D$ tends to drop with increasing VDS. In DMOS transistors operating at $V_{GS}$>$V_{GS}$(TCP) the carrier mobility drops as a temperature rises, causing ID to drop.




a) b)

Figure 38: The transfer characteristic of the DMOS transistor for multiple temperatures: a) schematic representation of the experimental setup; b) the measured characteristic

Following the packaging, some additional sanity checks and measurements were needed to be performed on the test chips in order to identify chips with defects/malfunctions or outliers/out of specs. The chips that have defects or have parameters that are way out of the expected tolerance are not fit for fast thermal cycling testing because the experimental results would be then misleading and hard to interpret, therefore the chips that do not pass the sanity check will be further eliminated from the statistic.




a) b)

Figure 39: The output characteristic of the DMOS transistor for multiple VGS: a) schematic representation of the experimental setup; b) the measured characteristic.

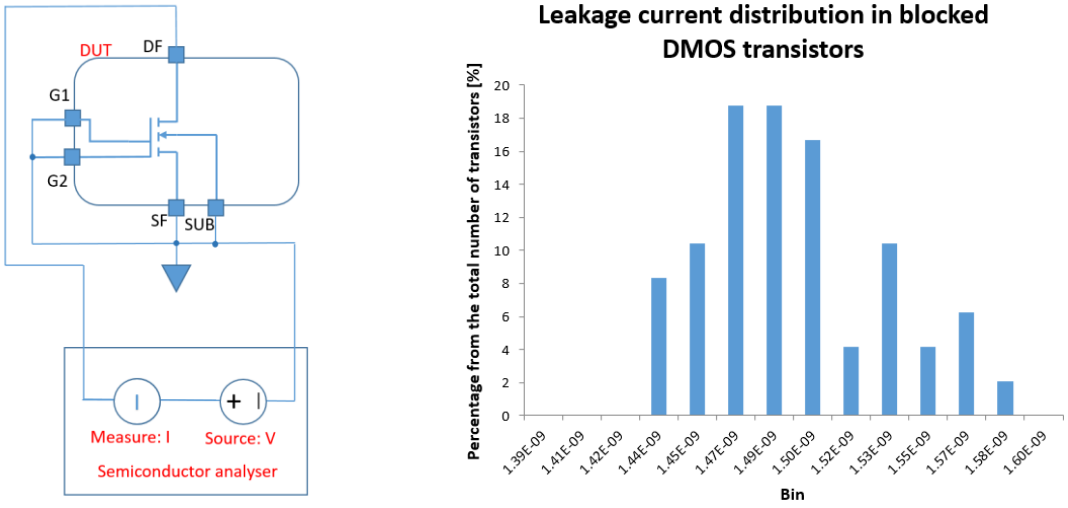
After the basic functionalities were tested, the wafer was diced, and the chips were packaged in ceramic C-DIP-24 packages to be suitable for fast thermal cycling reliability tests. Ceramic packages were chosen so that in the absence of the mold compound:

1. higher on-chip temperatures will be obtained during testing. This will cause a more pronounced/accelerated degradation of the metallization which will be easier to capture with the proposed integrated sensors.
2. the electro-thermal simulation setups are easier to calibrate because there is no need to accurately model the mold compound anymore.

The first set of measurements are performed at room temperature and consist in measuring leakage currents. Leakage currents are some residual currents that pass through a region which is poorly

conductive when subjected to high voltages such as a blocked junction, a blocked transistor or an insulator. A small current [pA to nA] is an indicator of that region's capability of remaining blocked or providing good insulation while withstanding high voltages. First, the leakage current in the blocked transistors (the ones used for calibration as well as the ones used for fast thermal cycling) was measured. The experimental setup is presented in Figure 40 a). All terminals except for the drain (DF) are connected to ground (V=0V). The drain potential is fixed at a value at about 60% higher than that of the bias value that will be used in the fast cycling tests and at about 8% higher than the maximum absolute voltage rating for this manufacturing technology. The measurement is performed on all transistors and the statistic is shown in Figure 40 b). The leakage current is extremely low (nA), which means that the devices can withstand high blocking voltages. The measured values are averaged and are distributed in bins obtained by equally splitting the interval (μ-3σ, μ+3σ). All measured values fall within the 6σ interval indicating a relatively low spread and no outliers or defect devices.

Another leakage measurement was performed in order to test the quality of the insulator. For this, only the transistors with embedded cracking sensors were used. The experimental setup is presented in Figure 41 a). All terminals except the central crack sensor are connected to ground (V=0V). The central crack sensor is connected to a high potential, because in order to test the quality of the insulator, the voltage used must be much higher than in the previous case. The insulator that separates the crack sensor and the rest of the metallization is subjected to a fixed voltage about 150% higher than the drain bias voltage in normal operation and about 70% higher than the maximum absolute voltage rating for devices manufactured in this technology. The results are shown Figure 41 b). The leakage in this case is also in nA range, which indicates that there are no defects in the oxide so it will provide good insulation at high drain-to-source voltages. Here as well, all values fall in the 6σ interval, no outliers are present, and the spread is low here as well.

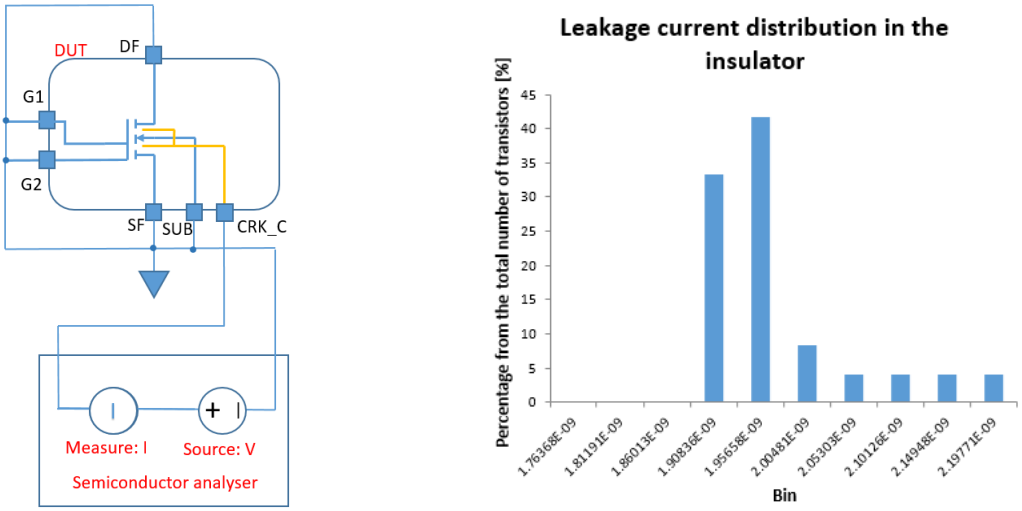



a) b)

Figure 40: The leakage current measurement on blocked DMOS transistors: a) schematic representation of the experimental setup; b) the leakage current distribution.

The second set measurements were electrical resistance measurements: 1. Statistical analysis at room temperature for the delamination sensor and at 25°C for the temperature sensor. 2. The temperature characteristic for the temperature sensors representing mean and the ends of the distribution.

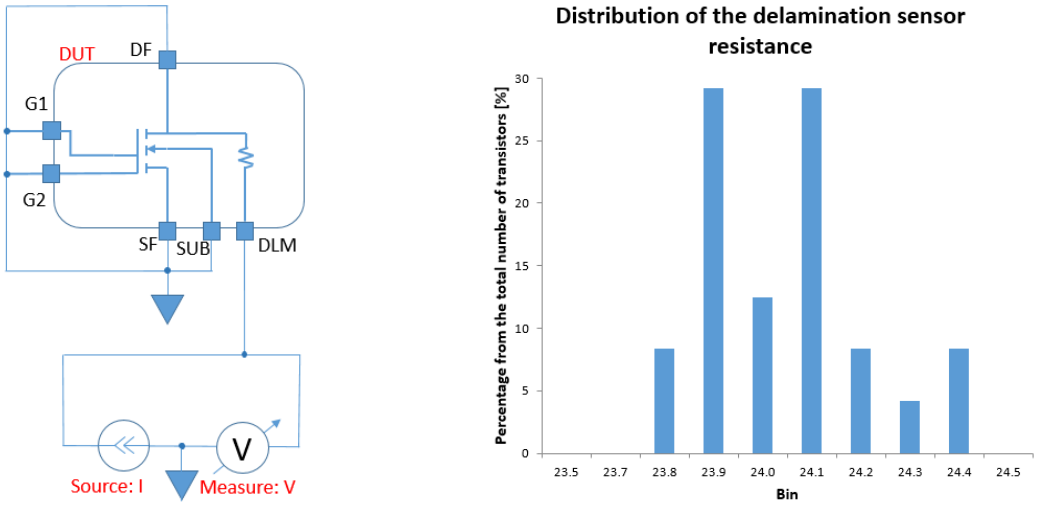
Next, all the temperature sensors (found only on transistors used for calibration) were measured at 25°C. The setup is the same as the one described at the on-wafer temperature sensor characterization, shown in Figure 37, except for the fact that the test structures are no longer on wafer. The packaged test chips are now placed in an oven that controls the ambient temperature. The resistance distribution is shown in Figure 43. The resistance follows a normal distribution and all values fall in the 6σ interval. No outliers are present, and the spread is less than 5% which is very desirable for sensors.




a) b)

Figure 41: The leakage current measurement in the insulator: a) schematic representation of the experimental setup; b) the leakage current distribution.

The delamination sensors are measured (found only on the transistor used for fast thermal cycling). The resistance is a series resistance composed of the resistance of the bond wires, the resistance of the power metal plate, the resistance of the via and the resistance of the underlying metallization, as described in Section 3.2 and shown in Figure 27. The schematic representation of the measurement setup is presented in Figure 42 a). All terminals except for the delamination sensor terminal are connected to ground (V=0V). A current of 1mA is forced though the delamination sensor terminal and the voltage is measured between the delamination sensor terminal and the ground (DF is connected to the ground as well). The resistance is obtained by dividing the measured voltage to the forced current. The resistance distribution is shown in Figure 42 b). All values fall in the 6σ interval, no outliers are present and the spread is low.




a) b)

Figure 42: The resistance of the delamination sensor at room temperature: a) schematic representation of the experimental setup; b) the resistance distribution.

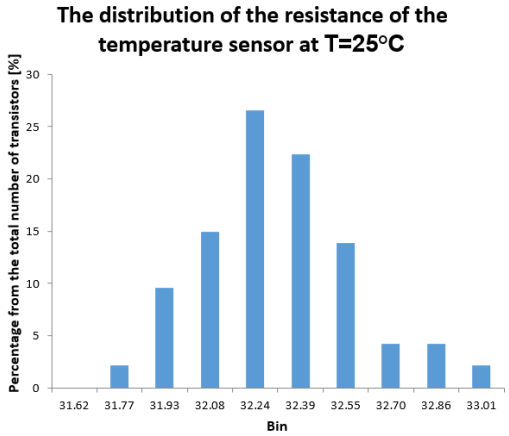



Figure 43: The distribution of the resistance of the temperature sensor at T=25°C.

From the distribution, three chips are selected for further temperature characterization: the chips corresponding to the mean and the ends of the distribution. The characterization results are shown in Figure 44. The characteristic obtained from the measurements performed on wafer was also plotted for comparison. The sensors measured in package are very linear and have the same slope, so only the correction with R(25°C) is necessary when measuring absolute temperatures using these sensors. No correction is necessary when measuring temperature differences. The characteristic obtained on wafer has a slope that is visibly smaller. This might indicate a possible outlier, or a decalibration of the thermal control of the chuck found on the wafer prober.

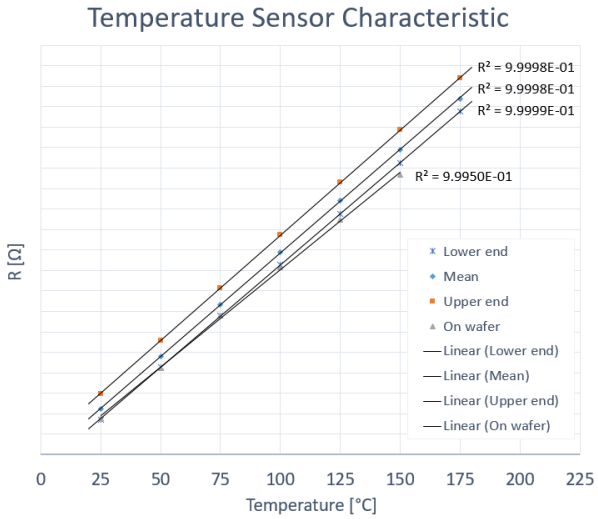



Figure 44: The resistance distribution of the delamination sensor at room temperature.

The third set of measurements were performed on the calibration transistors with the purpose of:

1. quantifying the spread in drain current and gate-to-source voltage for one specified bias point ($I_D$, $V_{GS}$, $V_{DS}$) = const
2. Verifying that all devices can be partially deactivated by shorting the central gate G2 to the source SF and quantifying the effective deactivation as a percentage of the total area.
3. Tracing a new transferred characteristic for packaged devices.

The setup used for this set of measurements is similar to the one used for tracing the transfer characteristic on wafer, shown in Figure 38 except for the fact that the test structures are no longer on wafer. The packaged test chips are now placed in an oven that controls the ambient temperature. The VDS is set to 20V, and the VGS is set so that ID is 100mA. The temperature inside the oven is set at 25°C. ID and VGS are measured and their distribution is calculated. For this specific bias point the transistor is operating below TCP, but in order to avoid significant self-heating the measurements were using short pulses to open the gate for short periods of time (ms range). These testing conditions are

like the device operation in real applications. The results are shown in Figure 45. All devices function properly, the distributions of ID and VGS are quite narrow, all values fall within the 6σ interval, no outliers or defect devices are present and the spread of ID and VGS is relatively low, below 5%. A low spread is very useful for future measurements and for easy electro-thermal simulation because there are virtually no corrections necessary.

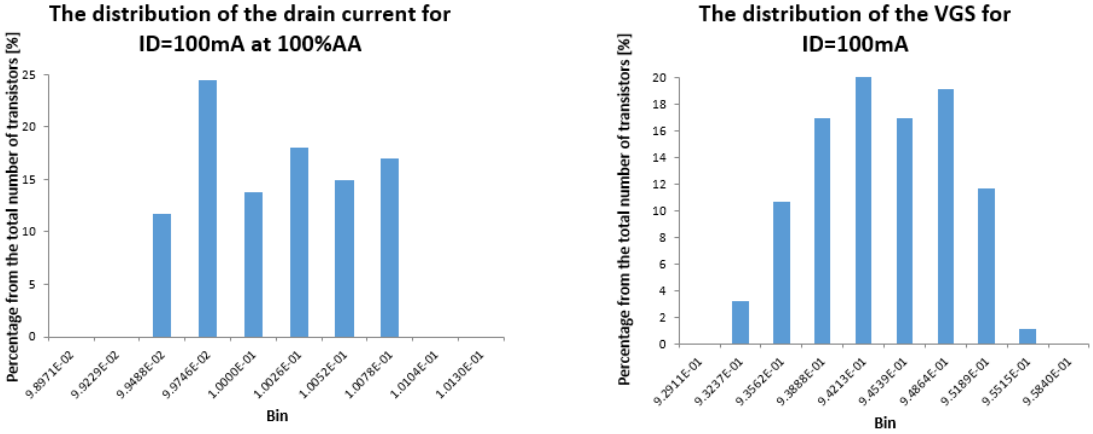



Figure 45: The drain current and gate-to-source voltage distributions for $I_D$=100mA, $V_{DS}$=20V.

For the same VGS, ID is measured once again, with the central gate G2 shorted to the source, SF to obtain the effective area deactivation. The area deactivation is calculated as a percentual difference between the currents measured in both cases:

$$Area\ deactivation = \frac{I_{D\ 100\%AA} - I_{D\ G2-SF}}{I_{D\ 100\%AA}} \times 100\ [\%] \tag{21}$$

The results are shown in Figure 46. One can observe that the actual deactivation when shorting G2 to SF is about 12% with a tolerance of less than 5%. This low tolerance is also very useful in the electro-thermal simulation of the test chip.

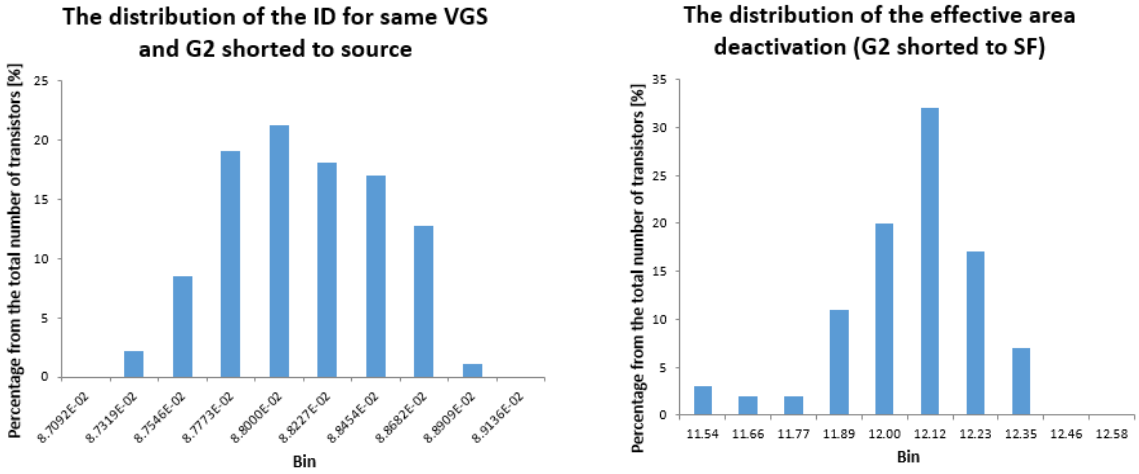



Figure 46: The drain current and area deactivation distributions G2 shorted to SF.

A new transfer characteristic was traced for a packaged device to see whether there is a difference compared to the on-wafer measurement. The transfer characteristic is plotted in Figure 47. TCP shifted from 1.2V to 1.3V, which clearly has no relation to any sort of spread or packaging artefacts. The difference comes from the testing conditions. During on-wafer characterization the transistor was always dissipating power, the DC sources were always on. This was possible due to the fact that the great silicon mass (the wafer) underneath the device could dissipate heat without running the risk of thermal runaway. However, this was no longer possible after packaging. So in order to prevent destruction, VGS was delivered in short pulses, in the millisecond range. This operating mode is also

close to one in the upcoming testing scenarios, so this characteristic will be taken as reference from now on.

The tests provide valuable information regarding the sanity of the test structure and its readiness for new thermal cycling tests.

1. The transistors function properly, their device characteristics are the expected ones. Deactivation of the central region is possible.
2. The temperature sensors are very linear and can be easily used to measure large ranges of temperatures
3. There are virtually no leakage currents; the blocked transistors and the insulator can withstand high voltages without experiencing breakdown.
4. The very low spread obtained in different measured characteristics is very helpful in future investigations (measurements or simulation) because it reduces the amount of necessary corrections.

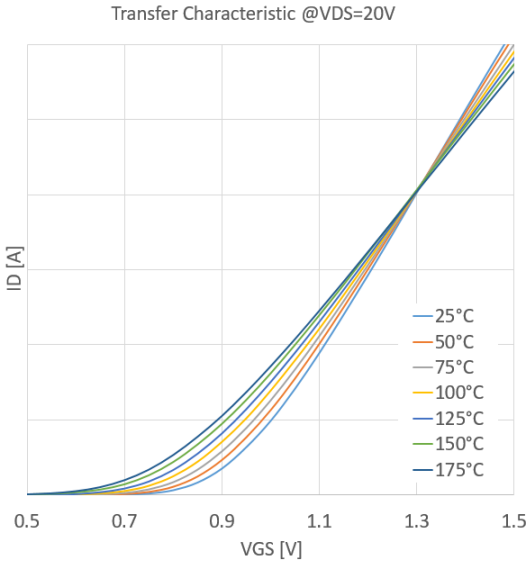



Figure 47: The drain current and area deactivation distributions G2 shorted to SF.

## 3.5 Conclusions

A test system was designed in order to validate the simulation results for the targeted failure modes: IMD cracking and power metal delamination, in a reasonable amount of time. The test system allows fully automated, long time tests, with minimal human intervention. The test bench is flexible because it allows bias and monitoring of various electrical, thermal and mechanical sensors. It consists of a main board and an extension board. The latter can be replaced with another one, if the concept of the mechanical sensors is changed.

A test vehicle (DMOS transistor) with mechanical sensors for detecting the failure mode and the location of defect was manufactured. The transistor is divided in two regions, which act as independent heat sources and can be automatically controlled by interacting with the test system. In this way the stress in the system can be controlled not only in terms of temperature variation, but also with a very course control, in terms of spatial variation. Initial, test results to check the basic functionality of the system: heat-source control, temperature and delamination sensors, voltage stress, was performed with very good results, showing the test structure can be used for fast thermal cycling operation. Only the functionality the ILD crack sensors is unproven as this can only be measured during direct failure conditions.

The next step is to stress to calibrate the power profiles over the DMOS heat-sources to get the desired temperature loading conditions, which then will be applied on the test IC in the fast-thermal cycling test to generate the cycles to failure and failure mode profiles. These profiles will be then compared

with the simulated data to verify the overall accuracy of the simulation flow and the accuracy-speed tradeoff for some of the domain reduction methods presented in chapter 2. These comparisons will be presented in deliverables D2.19 and D7.8.

# 4 Modelling of mechanicals stress influence on device matching

## 4.1 Overview

To enable the usability of the new FEM tool, the impact of solid power Cu plates has been studied which acts primarily on the mobility of charge carriers of integrated devices. The difference to verification approaches described earlier is the DUT or area of interest in the model. Here, the DUT is a group of two or more devices arranged in a matching configuration. In such a configuration, the difference in electrical performance of two or several integrated devices is of main interest for the behavior of the integrated circuit. Having identical devices in type, geometry and orientation, the difference is caused by local process fluctuations only.

However, in combination of a solid power Cu plate placed in the proximity or on top of the matching group, the mechanical restraints induced in the system can act non-symmetrically onto the individual group partners and lead to unwanted systematic shift in the circuit performance.

It is mandatory to know the safety distance between the power Cu plate and the matching group or the overlap length, in case the matching group is fully covered by the power Cu plate. This can be evaluated experimentally by measurements. However, the effort in time, layout and area on the IC is significant. Here, a well calibrated FEM tool can be of great support.

## 4.2 Models and vehicle for simulation (analysis)

Technically, there are two options how to deposit power Cu on top of an integrated circuit. On top of the last $SiO_2$, as a massive power Cu plate having a thickness of several 10th of microns. The drawback of such a deposition is the extremely non-planar surface on top that limits the placement or assembly of further structures there. On the contrary, the benefit is to drive large currents directly out of the IC into the external application. Such a structure has been chosen as a model for the first verification step of the new tool. The simplified model consists of three layers only (Figure 48), each with different physical properties: Silicon as the bulk material, followed by the IMDL as $SiO_2$ and the power Cu plate on top. The goal is to simulate the mechanical stress induced by all the power Cu plate down into the Si bulk material, where the integrated devices are placed. The desired result is to find the lateral area where the stress state tends to become low and homogenous. In Figure 48, these areas can be found on the Si surface with changing the stress level marking color from green to blue.

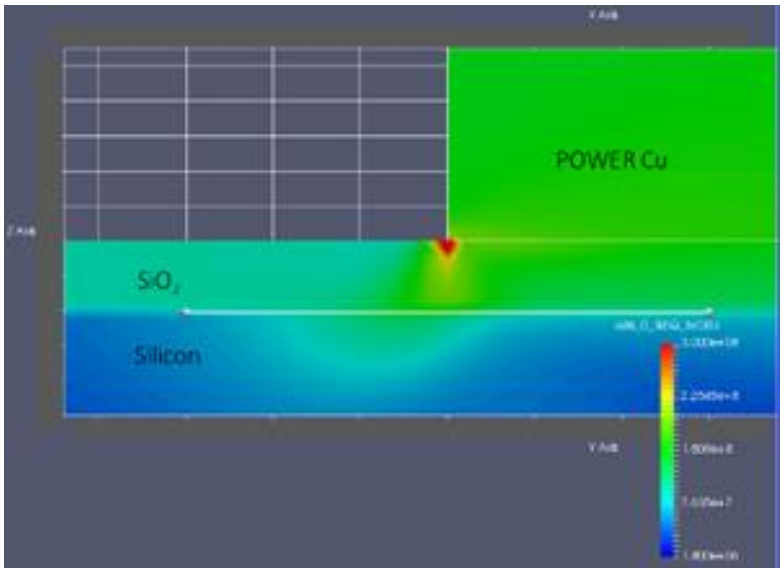



Figure 48: Power Cu plate on top.

Another known option for the power Cu deposition is to directly embed the thick power plate into the $SiO_2$. This type of power Cu in integrated systems is also known as damascene Cu. The advantage of

this damascene deposition is the high uniformity on top of the IC. This allows a reliable placement of further discrete structures on top of the IC. The drawback of this type of deposition is the strongly limited thickness of the power CU layer to a few microns only. If the thermal requested thickness exceeds this limit, several damascene Cu layers are necessary. In the model portfolio for the verification of the new FEM tool, several combinations of damascene Cu thicknesses and number of available layers has been chosen. Figure 49 represents one of these models consisting of two layers of damascene Cu. Also, here, the area of highest interest is the transition area (the Si surface) of the stress level with changing the color from green to blue.

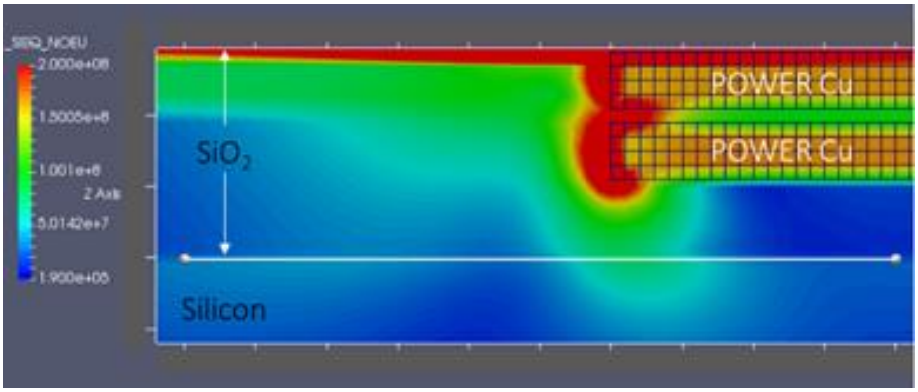



Figure 49: Two POWER Cu plates embedded into the IMDL layer.

## 4.3 Simulation flow

The thermo-mechanical simulation flow described in Section 2.1 was adapted to enable more facile computation of intrinsic stresses. The models have been built of three layers: Si bulk, $SiO_2$ and power Cu on top. The temperature, at which the entire system is in thermal equilibrium is around 400°C. By cooling down the system to ambient temperature, stress is introduced into the individual layers due to their different levels of volume change. The resulting stress values has been compared with benchmarking data. This data has been obtained from another FEM simulator that used a well-accepted model. This additionally was calibrated with data from measurements.

Facing discrepancies in the simulation results, a first enhancement of the simulation flow was necessary by introducing a stress liner. This stress liner consists of TiW (Titan-Tungsten). This liner has a physical background and is deposited shortly bevor the deposition of the power Cu. With this additional layer stress can be simulated that can be used as a stress pre-condition for the later stress simulation considering the power Cu assembly.

Having the possibility to use an initial stress state, the corresponding simulation results matched nicely the benchmarking data in terms of magnitude of the stress level and the shape of the stress gradient. Another important modification was necessary to use different densities of the mesh in areas of interest. In order to keep the simulation time at a most efficient level, areas containing devices of interest, here matching groups, needed a significantly higher mashing density compared to areas without them. Final important aspect of the new simulation flow is the generic data output. The raw data allows access to stress states for all known stress components and stress types, and this, for every desired position of the model. Thus, enhanced data post processing is mandatory to get the results for areas and orientations of interest.

The modification and enhancement to the thermo-mechanical tool, with respect to the main simulation flow are summarized:

- The most important change is the addition of a feature that allows the user to map stress and strain results from a previous simulation to a new simulation. The user can specify in *input file 2* the name of the previous simulation setup. During the solving stage, results are read from the out file of the previous setup and data is interpolated from the old mesh to the new mesh. If the new

model contains additional mesh regions with respect to the old one, the respective regions will be assigned with null stresses. The imported results will be used as initial conditions for the new mechanical simulation. The whole process is automated and transparent to the user.

- Additional mesh controls are added to the *input file 1*. The user can have more control on the quality of the mesh generated at the Surface of the Silicon, the region of interest.
- Since there is no need of repetitive and non-uniform temperature distribution, the data is no longer imported from the electro-thermal simulation. The whole model will start from the same reference temperature and will be heated/cooled to a uniform temperature specified by the user. The value of the uniform temperature field can be typed in the field with the name of the electro-thermal setup (see Section 2.1.B, paragraph ***Mesh controls and region definition file (Input file 2)***, for more details). Therefore, the ***Temperature distribution transfer to mechanical mesh*** pre-processing step described in Section 2.1.B is skipped, making the overall analysis much faster and simpler.
- Additional result types and quantities are enabled: e.g. stress component $\sigma_{xx}$ (this is necessary because the mismatch between electrical characteristics of devices is given by e.g., carrier mobility, which is affected by certain stress components).

This state of the development of the simulation flow allows to vary constructional parameters, allows the study of the stress states and to answer important questions of practical importance in the area of IC design.

In order to verify the correctness of the simulation results obtained in the FEM simulations described in the previous subchapters, a special test chip has been developed. It provides access to various constellations of POWER Cu placements combined with selected matching groups. The setup includes variations of the distance between the Cu edge and selected members of the matching group, deposition of one or more Cu layers and many more. First silicon will be available by mid of 2020. The readout and calibration of the simulation flow will follow in autumn 2020.

## 4.4 Next steps

In order to verify the correctness of the simulation results obtained in the FEM simulations described in the previous subchapters, a special test chip has been developed. It provides access to various constellations of power Cu placements combined with selected matching groups. The setup includes variations of the distance between the Cu edge and selected members of the matching group, deposition of one or more Cu layers and many more. First silicon will be available by mid of 2020. The readout and calibration of the simulation flow will follow in autumn 2020.

## 4.5 Conclusion

The verification of the newly developed FEM tool and its simulation flow passed first steps by comparing the simulated results with available benchmarking data. Several models have been developed in order to see the capabilities of the tool in terms of model creation, simulation time, and data acquisition. The results itself required an update of the simulation flow by introducing the option of an initial stress, induced by a stress liner. With this enhancement the result matched nicely the stress values which were used as benchmarking data for the tool and the selected model set-up. Up to now, as a stress origin, only the volume change has been considered due to thermal expansion. Discussions with experts revealed another potential origin for mechanical stress due to volume change due to internal crystal restructuring of special layers. In case of the simulation flow is not able to be calibrated with measurement data, this option needs to be considered for further system and/or simulation model enhancements. The complete results for this task will be presented in deliverable D2.19.

# 5 List of Abbreviations

| Abbreviation | Meaning |
|---|---|
| DFM | Design for Manufacturability |
| IC | Integrated circuit |
| ASICs | Application-specific integrated circuits |
| SPI | Service Provider Interface |
| DMOS | Double-Diffused Metal-Oxide-Semiconductor |
| BCD | Bipolar-CMOS-DMOS |
| SOA | The safe-operating-area |
| IMD / ILD | Inter-metal / inter-layer dielectric |
| FEM | Finite Element Method |
| $\alpha$ or CTE | Thermal expansion coefficient |
| PMet | Power metal |
| RVE | Representative Volume Element |
| VL | Viscous layer |
| DUT | Device under test |
| KVIA | Kelvin VIA delamination sensor |
| PCB | Printed Circuit Board |
| MOSFET | Metal–oxide–semiconductor field-effect transistor |
| TiW | Titan-Tungsten |

# 6 References


Fritz, A., Hüeber, S. and Wohlmuth, B. I. (2004): A comparison of mortar and nitsche techniques for linear elasticity. Calcolo 41, pp. 115-137.

Babuska, I. (1973): The finite element method with Lagrangian multipliers. Numerische Mathematik, Issue 20, pp. 179-192.

Bojita, A. I., Boianceanu, C., Prucar, M., Florea, C., Simon, D., and Plesa, C.-S. (2018): A simple metal-semiconductor substructure for the advanced thermo-mechanical numerical modeling of the power integrated circuits. *Journal of Microelectronics Reliability,* Volume 87, pp. 142-150.

Bojita, A. I. et al., 2019. A Simple Metal-Semiconductor Substructure Model for the Thermal Induced Fatigue Simulation in Power Integrated Circuits. In: Springer, ed. *Numerical Modelling in Mechanical and Materials.* Ghent: Springer, pp. 21-23.

Bosc, J.-M.et al., 2000. *Reliability characterization of LDMOS transistors submitted to multiple energy discharges.* Toulouse, 12th International Symposium on Power Semiconductor Devices & ICs, pp.165-168.

de Filippis, S., et al., (2012): Modelling of highly anisotropic microstructures for electro-thermal simulations of power semiconductor devices. *Miroelectronics Reliability,* 52(9), pp. 2374-2379.

Dureisseix, D. and Bavestrello, H. (2012): Information transfer between incompatible finite element meshes: application to coupled thermo-viscoelasticity. *Computer Methods in Applied Mechanics and Engineering, Elsevier*, pp. 44-47.

Li, C., Chen, L., Qiao, L. (2012): RVE based numerical evaluation on effective mechanical properties of composite with randomly distributed multi-phase inclusions. *Advanced Materials Research,* Volume Vols. 383-390, pp. 931-934.

Chaboche, J.-L. (2008): A review of some plasticity and viscoplsticity constitutive theories. *International Journal of Plasticity,* 24(10), pp. 1642-1693.

Dibra, D., Stecher, M., Decker, S., Lindemann, A., Lutz, J., Kadow, C. (2011): On the origin of thermal runaway in a trench power MOSFET. *IEEE Transactions on Electron Devices,* 58(10), pp. 3477 - 3484.

Erwin, J. T., Anderson, W. K., Kapadia, S., Wang, L. (2011): hree-Dimensional Stabilized Finite Elements for Compressible Navier–Stokes. *Computational Fluid Dynamics Conference*, 27–30 June.

Kravchenko, G., Karunamurthy, B., Nelhiebel, M. & Pettermann, H. E., 2013. *Finite element analysis of fatigue cracks formation in power metallisation of a semiconductor device subjected to active cycling.* Wroclaw, IEEE International Conference on thermal, Mechanical and Multi-Physics Simulation and Experiments in Microelectronics and Microsystems (EuroSimE).

Kurukuri, S., (Aug. 20, 2005):. *Homogenization of damaged concrete meso-structures using representative volume elements – implementation and application to Slang,* Bauhaus–University Weimar Germany: Master Thesis.

Lassnig, A. et al., (2012): *Accelerated lifetime measurements of Cu-Al ball bonded interconnects.* Cascais, IEEE International Conference on thermal, Mechanical and Multi-Physics Simulation and Experiments in Microelectronics and Microsystems (EuroSimE).

Nanakorn, K. T. a. P., (2013): Periodic boundary conditions for unit cells of periodic cellular solids in the determination of effective properties using beam elements. *The 2013 World Congress on Advances in Structural Engineering*, Sep., pp. 3738-3748.

Nguyen, H. V. et al. (2002): *Test Chip for Detecting Thin Film Cracking Induced by Fast Temperature Cycling and Electromigration in Multilevel Interconnect Systems.* Singapore, s.n., pp. 135-139.

Nguyen, V. H. & et al. (2003): *A Reliability Model for Interlayer Dielectric Cracking During Fast Thermal Cycling.* Montreal,Tokyo, s.n., pp. 295-299.

Nguyen, V.-D., et al. (2012): Imposing periodic boundary conditionon arbitrary meshes by polynomial interpolation. *Computational Materials Science,* pp. 390-406.

Pfost, M., Boianceanu, C., Lohmeyer, H. & Stecher, M. (2013): Electrothermal Simulation of Self-Heating in DMOS Transistors up to Thermal Runaway. *Transactions on Electron Devices,* 60(2), pp. 699-707.

Pfost, M., Boianceanu, C. M., Lascau, I. & Simon, D. I. (2014): Influence of the On-Chip Metallization on Self-Heating in Integrated Power Technologies. *IEEE Transactions on Semiconductor Manufacturing,* 27(2), pp. 169-177.

Pfost, M. & Joachim, J., (2008): *Measurement and simulation of self-heating in DMOS transistors up to very high temperatures.* Orlando, International Symposium on Power Semiconductor Devices and IC's.

Pham, G., Ritter, M. & Pfost, M., (2016): *Influence of metallization layout on aging detector lifetime under cyclic thermo-mechanical stress.* Pasadena, 2016 IEEE International Reliability Physics Symposium (IRPS), pp. 5B-5-1-5B-5-6.

Rudolf, R. et al., (2011): *Automotive 130 nm smart-power-technology including embedded flash functionality.* San Diego, s.n., pp. 20-23.

Simon, D. et al., (2016): Reliability Analysis for Power Devices Which Undergo Fast Thermal Cycling. *IEEE Transactions on Device and Materials Reliability,* September, 16(3), pp. 336-344.

Smorodin, T., Wilde, J., Alpern, P. & Stecher, M., (2008): A Temperature-Gradient-Induced Failure Mechanism in Metallization Under Fast Thermal Cycling. *IEEE Transactions on Device and Materials Reliability,* 8(3), pp. 590 - 599.

Spirito, P., Breglio, G., d'Alessandro, V. & Rinaldi, D., (2002): *Analytical model for thermal instability of low voltage power MOS and S.O.A. in pulse operation.* Santa Fe, s.n., p. 269–272.

Wagner, C. & et al, (2011): *Automotive 130 nm smart-power-technology including embedded flash functionality.* San Diego, s.n.

Wang, F. Y. a. H., n.d. A simple Python code for computing effective properties of 2D and 3D representative volume element under periodic boundary conditions.

Würkner, M., Berger, H., Gabbert, U. (2014): Numerical investigations of effective properties of fiber reinforced composites with parallelogram arrangements and imperfect interface. *Composite Structures*, Sep. , pp. 388-394.

Xia, Z., et al. (2006): On selection of repeated unit cell model and application of unified periodic boundary conditions in micro-mechanical analysis of composites. *International Journal of Solids and Structures,* Volume 43, pp. 266-278.