Thermal Effects of Silicon Thickness in 3-D ICs: Measurements and Simulations

Papa Momar Souare, Vincent Fiori, Alexis Farcy, François de Crécy, Haykel Ben Jamaa, András Borbély, Perceval Coudrain, Jean-Philippe Colonna, Sébastien Gallois-Garreignot, Bastien Giraud, et al.

Thermal Effects of Silicon Thickness in 3-D ICs: Measurements and Simulations

Papa Momar Souare, Vincent Fiori, Alexis Farcy, François de Crécy, Haykel Ben Jamaa, Andras Borbely, Perceval Coudrain, Jean-Philippe Colonna, Sebastien Gallois-Garreignot, Bastien Giraud, Severine Cheramy, Clément Tavernier, and Jean Michailos

***Abstract*— This paper presents the impact of silicon thickness on the temperature and the thermal resistance in a 3-D stack integrated circuits. This paper uses electrical measurements thanks to embedded *in situ* sensors and numerical design of experiments (DOEs). The primary objective is to provide the sensitivity of modeling factors by analyzing the variance on the basis of Sobol indices through DOE. The results show a strong influence of the silicon thickness and of the position of the hot spots with respect to the sensors on the maximum temperature and the thermal resistance of the total stack. The boundary conditions, in particular the heat-transfer coefficient of the bottom surface of the wafer, are also identified as significant factors. Therefore, simulation results and measurement approaches are compared. The measurements are carried out with embedded *in situ* sensors in the bottom die at wafer level. The results show a significant increase in temperature while decreasing the silicon thickness.**

***Index Terms*— 3-D integrated circuits (ICs), FEM simulation, self heating, sensor, thermal, thermoelectric measurement, through-silicon vias (TSV).**

I. Introduction

THE 3-D integration has attracted a common interest in the recent years as a mean to efficiently improve performance and miniaturization of integrated circuits (ICs). The integration is based on vertical die stacking, connected by through-silicon vias (TSV) and the bonding of active layers [1].

One of the most limiting factors in microelectronics, particularly in 3-D integration in terms of performance, is related to the generated heat. Merging several active silicon parts into a single package leads to denser heat fluxes in the final stack, which cause self-heating issues. The total power resulting from Joule effect in the transistors and interconnects may contribute to a strong increase of the global temperature of the chip [2]. These thermal issues directly impact the reliability of integrated circuits. Limitations in operating temperature are then fixed on products, according to the application, for example 85 °C for memory and 120 °C for processors [15].

In this paper, first the stack of our configuration of test chip are presented, and sensitivity analysis of the modeling parameters, which was carried out using the design of experiments (DOEs) is described. Then, the impact of silicon thickness on the temperature behavior under the dissipated power is investigated. This paper relies on measurements results through embedded sensors as well as numerical simulation results. As a conclusion, the temperature behavior depending to modeling factors is given and complemented with investigations of the impact of silicon thinning in 3-D ICs for self-heating challenge.

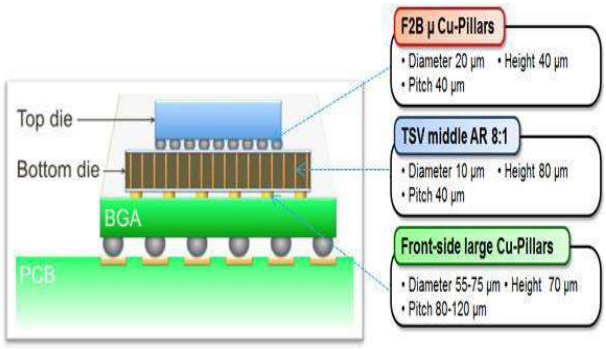


Fig. 1. 3-D stack of the thermal die.

II. Test Chip

To study this phenomenon, a dedicated thermal test chip has been designed and fabricated [3], [5]. It is formed by the stacking of a 7.5 mm^2 chip (top die) on a 31 mm^2 chip (bottom die) using μ-bumps and TSV-middle in a face-to-back configuration, as commonly used for wide I/O products [9] (Figs. 1 and 2). The peripheral part of the bottom die contains bumps to allow further mounting on a BGA substrate (Fig. 1). The connection path between the two active layers is achieved by TSVs in the bottom die and micro- and large copper

P. M. Souare is with STMicroelectronics, Crolles 38926, France, CEA-Leti Minatec, Grenoble Cedex 9 38054, France, and also with the Ecole des Mines de Saint-Étienne, Saint-Étienne 42023, France (e-mail: papa-momar.souare@st.com).

V. Fiori, A. Farcy, P. Coudrain, S. Gallois-Garreignot, C. Tavernier, and J. Michailos are with STMicroelectronics, Crolles 38926, France (e-mail: vincent.fiori@st.com; alexis.farcy@st.com; perceval.coudrain@cea.fr; sebastien.gallois-garreignot@st.com; clement.tavernier@st.com; jean.michailos@st.com).

F. de Crécy, H. B. Jamaa, J.-P. Colonna, B. Giraud, and S. Cheramy are with the CEA-Leti Minatec, Grenoble Cedex 9 38054, France (e-mail: francois.decrecy@cea.fr; haykel.ben-jamaa@cea.fr; jean-philippe.colonna@cea.fr; severine.cheramy@cea.fr).

A. Borbely is with the Ecole Nationale Supérieure des Mines de Saint-Étienne, Saint-Étienne 42023, France (e-mail: borbely@emse.fr).

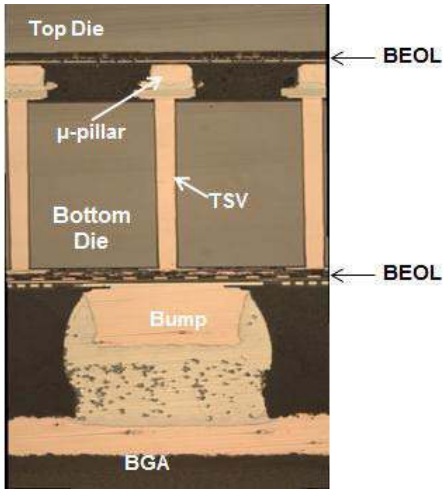


Fig. 2. 3-D stack optical cross section [10].

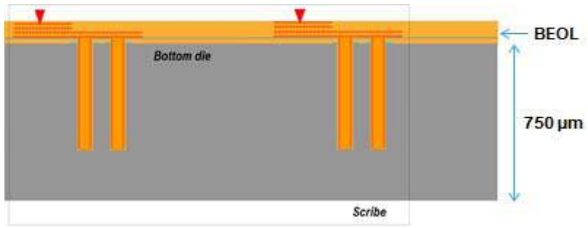


Fig. 3. Stack of bottom wafer unthinned: probed at the red triangles.

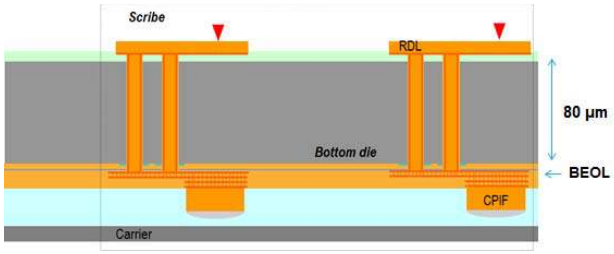


Fig. 4. Stack of bottom wafer thinned on temporary glass carrier: probed at the red triangles.

pillars (Fig. 2). Spaces between these layers are filled with an underfill material for mechanical and electrical reliability purposes.

All measurements presented in this paper were carried out at wafer level on the bottom die in thinned and unthinned configurations (Figs. 3 and 4). On the unthinned wafer, the measurement is performed directly on the aluminum pad of the back-end of line (BEOL), while on the thinned wafer the measurement is carried out on the backside Redistribution Layers, connected to the BEOL through the TSVs. The thinned wafer is on glass temporary carrier. Electrical measurements are carried out on a Semi-automatic prober tool.

The chips are designed as a combination of parallel test circuits: the bottom die has eight central (1B to 8B) deported on eight peripheral test circuits (1P to 8P). The central test circuits of the bottom die are connected to peripheral test ones to allow testing the bottom die after the stacking stage.

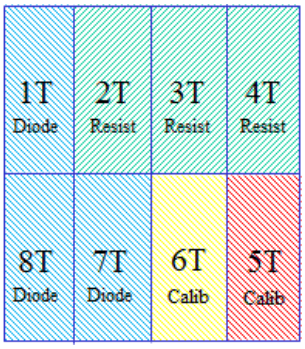


Fig. 5. Placement of the test circuits on the top die and type of sensor inside regions (top view).

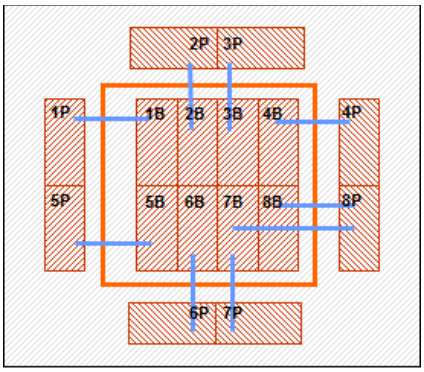


Fig. 6. Floorplan of test circuits on the bottom die (top view): test circuits of top die will be connected with central ones of bottom die.

This arrangement allows a flexible management of electrical measurements from die to wafer level. Among the eight central test circuits, six have heating elements and embedded temperature sensors (Figs. 5 and 6). The two remaining test circuits (named 5T and 6T) are used for sensors and heaters calibration.

Two types of sensors are used, depending on the test circuit (Fig. 5).

1) PN junction diodes use the linear dependence of voltage with temperature driven by a constant forward current [11], [13]

$$I = Is(T)\left(e^{\frac{qV}{kT}} - 1\right) \quad (1)$$

where Is is the saturation current, T is the junction temperature, V is the voltage across the diode, q is the electron charge and k is the Boltzmann constant such diodes are used to measure the temperature in silicon [10].

2) And the passive resistors based electrical resistance of copper change with temperature [11]

$$R = R0(1 + \mathrm{TCR} * \Delta T) \quad (2)$$

where R0 is the resistance at the reference temperature (25 °C), TCR is the temperature coefficient of resistance, and ΔT is the gap temperature relative to reference temperature such resistors are used to measure the temperature of copper at several BEOL levels [8].

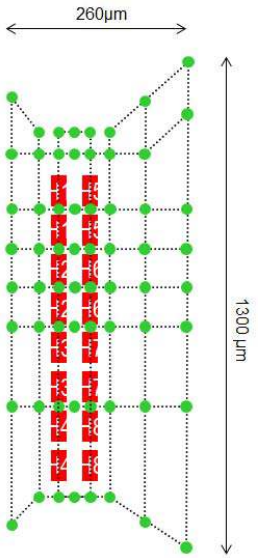


Fig. 7. Location of sensors (green dot) around heaters (red rectangle) on a test circuit.

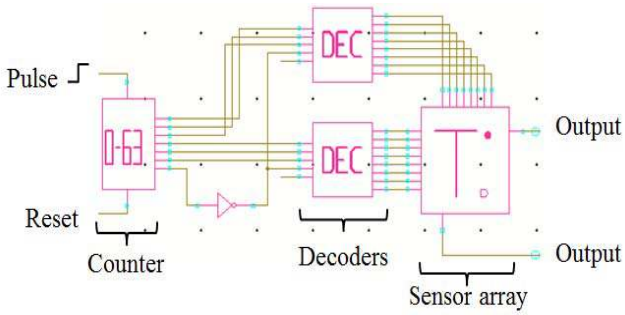


Fig. 8. Electrical schematic of the logic circuitry.

TABLE I
MATERIALS PROPERTIES

Materials		Units	XY Conductivity	Z Conductivity
Silicon		W/m*K	150	150
TSVs homogenized	TSV Ø10 µm SiO2 Thick 0.3 µm Si Pitch 40 µm		138.04	160.88
BEOL homogenized	M1 to M7		2.356	2.544
Glass			1.2	1.2

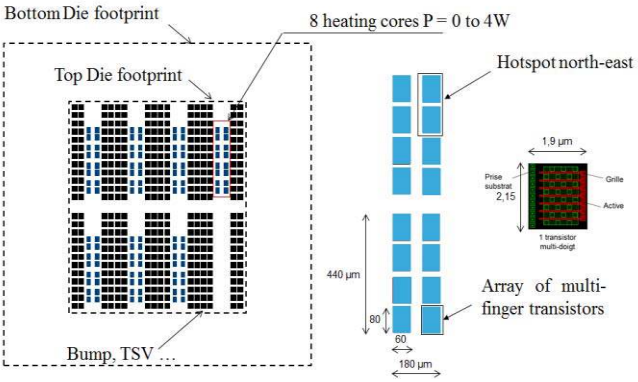


Fig. 9. Placement of heaters on the die.

Calibration structures are also embedded to compensate for ohmic losses and to evaluate nonlinear behavior of the sensors. Stacking test circuits leads to a total number of 128 sensors by test circuit: 64 on the top and 64 on the bottom die. Numerical addresses are assigned to generate the link of sensor locations with respect to the chip and make an automated temperature mapping around hotspots (Fig. 7). These functions are assumed by a logic circuitry; consisting in a counter and digital decoders (Fig. 8).

A counter of 128 bits used to address the 64 sensors of each chip (top and bottom dies) through two decoders. The addressing of the sensors array is done such that each input pulse selects a sensor and the 64 first pulses address those of the bottom die and the 64 others those of top die. For example, in this paper, the bottom dies being tested: the values of the 64 first pulses correspond to those across the sensors output and the 64 remaining to saturation values of sensors (Fig. 15). Reset allows to initialize the circuit.

Eight hotspots are integrated by test circuit, each of them being represented by two rectangles of 80×60 μm^2 (Fig. 8). The power of the heater core is generated by redundant multifinger metal–oxide semiconductor transistors placed in array [13]. Through Joule effect, this power will be the heat generator of each test circuit in the chip. They are individually controlled, with a generated power ranging from 0 to 4 W. In practice, this power shall be limited to a fraction of the maximum power owing to the weak ability of TSVs to bear a high current.

III. FEM SIMULATION

The objective of this paper is to quantify the sensitivity of modeling parameters and to provide for the temperature a surface response based on the keys factors. The simulation is done using ANSYS finite elements commercial software. The numerical model represents the test chip on a 300 mm wafer (Fig. 10) with the assumptions of homogeneous materials in the BEOL and TSVs [4]. The purpose of this homogenization of the materials is to simplify the complex subassemblies (such as TSVs, the copper pillar, etc.) in blocks of homogeneous but anisotropic equivalent material. This allows accounting structure details with limited CPU resources.

The model describes a silicon substrate thickness ranging from 80 to 750 μm with homogenized BEOL and TSV layers (Table I). The heating elements are represented by rectangle as described in blue in Fig. 9. One or two glass plates (thickness 750 μm each) are used for thermal insulation of silicon on the chuck (Fig. 10). The thermal insulation of the devices allows

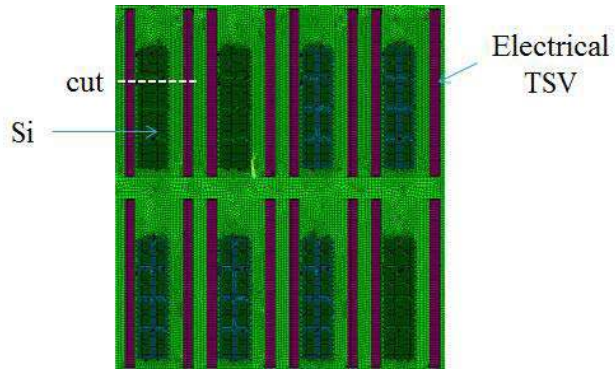


Fig. 10. Zoom in the DIE.

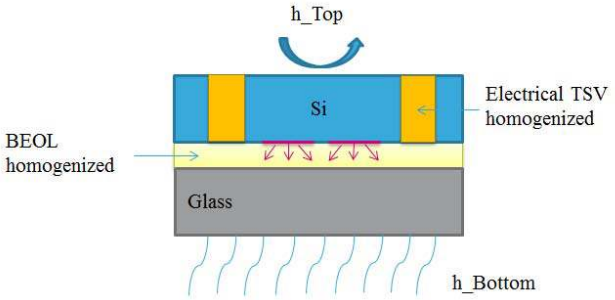


Fig. 11. Cut in one test circuit.

to prevent to the chuck to pump down all the heat, which would lead to measure temperature differences not significant.

The boundary conditions are applied as follows:

1) a convective heat transfer h_{bottom} varying from 262 to 2620 W/(m^2 · K) on the underside of the glass, equivalent to an air layer of 10–μ100 μm thickness between the glass and the chuck;
2) a convective heat transfer h_{top} equal to 8 W/(m^2 · K) on the top surface (Fig. 11).

The glass is placed on the chuck and thus h_{bottom} simulates the thermal contacts between glass and chuck. The use of a convective type transfert is chosen for simulation. The power dissipated by the heater ranges from 100 to 400 mW (Fig. 9), and the reference external temperature is set at 25 °C.

Because the simulation used a multiscale object (size TSV Ø10 μm versus Wafer Ø300 mm), a refinement mesh methodology is chosen to optimize the simulation time. The mesh is chosen coarse at the wafer level, with a maximum mesh size of $S_{\text{max}} = 5000$ μm and fine in the chip region with a minimum mesh size of $S_{\text{min}} = 5\mu$m (Fig. 12).

A. Example of Simulation Result

Fig. 13 shows a temperature map example with the excitation of a single hot spot: views are provided at the chip scale (a) and close to the hotspot (b).

The analysis shows that isotherms describe ellipses near the heaters, and tend toward circles away from them. The relationships driving the temperature between the distance from the sensors to the heaters and the technological parameters will

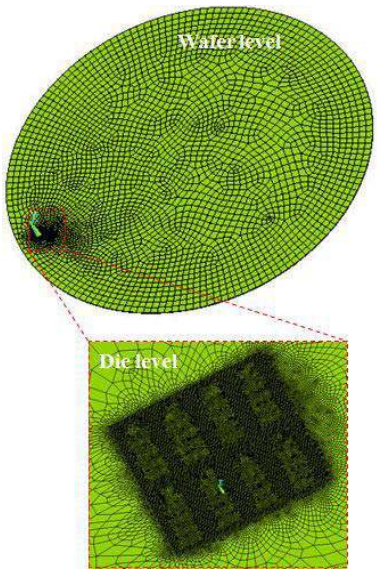


Fig. 12. FEM wafer level model and die level zoom + meshing.

TABLE II
RANGE OF INPUTS FACTORS OF DOE

Factor	Name	Units	Min	Max
A	Thermal insulation thickness	µm	750	1500
B	$1/h_{\text{bottom}}$	µm²*K/µW	381,6	3816,7
C	Eccentricity of the chip vs. wafer center	µm	0	138000
D	1/Silicon thickness	1/µm	0,00133	0,0125
E	Flux	µW/µm²	10,4	41,6
F	Ellipse parameter $a^{-0.5}$	$\mu m^{-0.5}$	0,02672	0,13482

then be studied using a DOE. Design Expert V8 software is used for this purpose.

B. Sensitivity Study

The kind of DOE used in this paper is the Optimal Design for Response Surface Modeling to minimize the number of runs. The primary objective of this DOE is to quantify the sensitivity of the aforementioned factors. The study includes six input factors, listed in Table II. The studied variable is $T - T_{\text{amb}}/\text{flux}$, which is proportional to the thermal resistance, where T is the temperature field in the die, T_{amb} is the reference external temperature, and Flux is the flux imposed in heaters.

The temperature in each of the 64 sensors is sought with 150 different simulations, that is, 9600 points.

The 150 runs of the DOE is the addition of an IV Optimal Design for 100 points, as proposed by Design Expert V8,

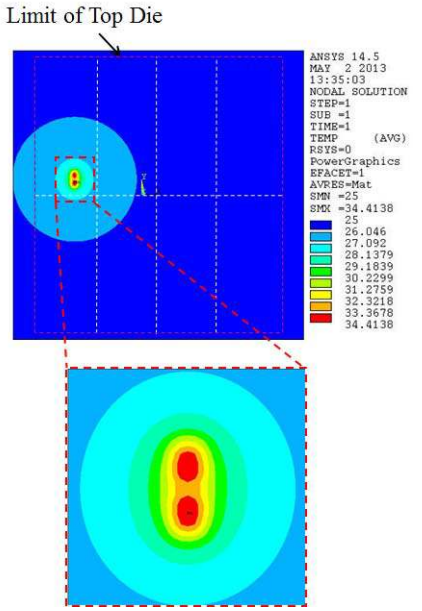
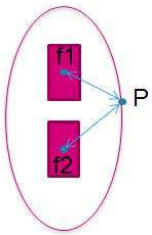


Fig. 13. Example of FEM mapping with excitation of one heater.

and a 50 points space filling design based on a Maximin Latin Hypercube Sampling. The IV optimal design offers a good tradeoff between a reasonable number of points in presence of a categorical factor (the heater location) and a low average variance across the simulated region. The addition of the space-filling design ensures that no part of the simulation region is too far from a simulation point.

1) The thermal insulation thickness is the thickness of the glass layer between the wafer and the chuck.
2) The high value of the bottom heat-transfer coefficient (h_{bottom}) particularly high simulates the thermal contacts between the glass and chuck.
3) The eccentricity of the chip versus wafer center is the position of the chip tested on the wafer.
4) The ellipse parameter (a) is the major semiaxis of ellipse, and the focal point is the center of hotspots.



$$\text{Ellipse parameters } (a) = \frac{pf1 + pf2}{2} \tag{3}$$

where $f1$ and $f2$ are the focal points of ellipse, and P is one point on the ellipse. Note that $f1$ and $f2$ correspond physically to the center of the transistor arrays.

In this paper, the top heat-transfer coefficient is fixed at 8 W/(m^2 · K), which models natural convection with air. Note that this quite low value limits the heat dissipation and does not have impact on trends [5].

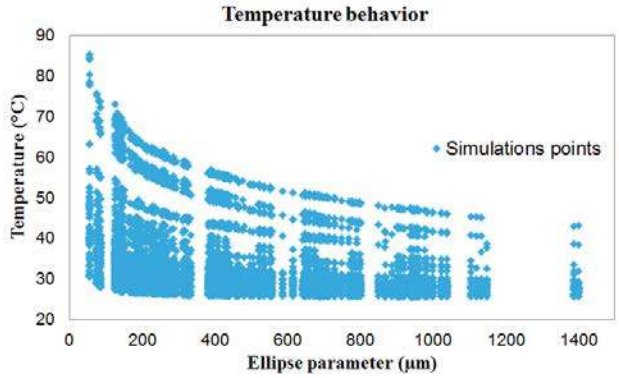


Fig. 14. Behavior of the temperature of the sensors in the silicon relative to the major semiaxis of ellipse (ellipse parameter) for each simulation point.

C. Result

As preliminarily mentioned, it is assumed that all isotherms are ellipses close to the hot spots.

From Fig. 14, the simulation points show a hyperbolic behavior of temperature relative to the ellipse parameter modeled by the following function:

$$T = f(a) \quad \text{with } f = \frac{1}{\alpha^{\beta}} \tag{4}$$

where T is the temperature, a is the ellipse parameter, and β is a parameter fixed to 0.5 [5], determined by optimizing the model. To have the most predictive numerical model possible in relation to 150 runs of DOE, the choice to model the behavior of the temperature with a hyperbolic function was done. This assumption allows to improve the coefficient of determination R^2 to 0.9972 and R^2 predicted to 0.9972.

The influence of each input factor on the responses is evaluated by an analysis of variance based on Sobol indices [14].

Sobol indices are a well-known method for sensitivity analysis, which is based on assigning a probability distribution (here a uniform density probability over the full range of the factors is assumed) to the input factors to model the uncertainty of each input parameter [6]. For example, Sobol indice of y such that $Y = f(X_1, \ldots, X_p)$ from X_i is given by the following formula:

$$\text{Si} = \frac{\text{Var}\mathbb{E}(Y|Xi)}{\text{Var } Y}. \tag{5}$$

These indices range from 0 to 1, and quantify the ratio of the variance of Y as a result of the variability of X_i alone to the total variance of Y.

The distribution of Sobol indices for all effects is shown in the pie graph in Fig. 15. It can be seen that the D factor (1/silicon thickness) strongly influences the output (i.e., to 52%): increasing the silicon thickness decreases the thermal resistance. The second important parameter is the ellipse major semiaxis (ellipse parameter) F (45%) and finally the boundary conditions (heat-transfer coefficient) (1%).

This result in the temperature being approximately expressed as a function of the input parameters based on variance analysis, which allows for the fitting between simulations and measurements.

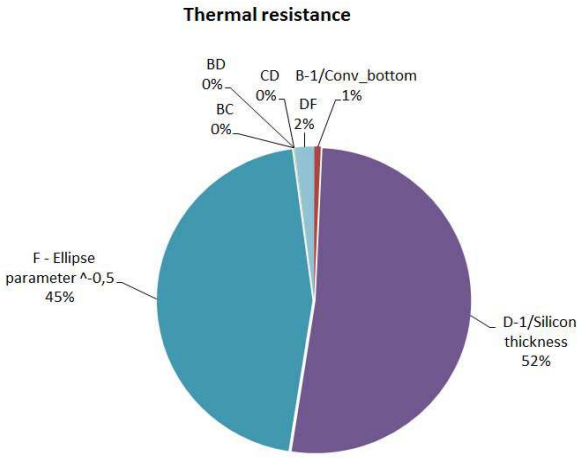


Fig. 15. Distribution of Sobol indices.

TABLE III
COEFFICIENT OF THE POLYNOMIAL WITH CODED FACTORS

$(T\text{-}T_{amb})$/Flux	Effect
0,590	Constant
5,697E-03	* A
0,0380	* B
2,739E-03	* C
0,370	* D
0,270	* F
-4,843E-03	* A * B
4,864E-03	* A * D
2,060E-03	* B * C
0,027	* B * D
-5,979E-04	* C * D
0,099	* D * F
-5,372E-03	* B^2
3,127E-03	* C^2
-6,699E-03	* D^2
0,0110	* F^2
0,0360	* D^2 * F
-0,0960	* D * F^2
-0,0110	* D^3
0,0470	* F^3

The values of coded factors (A to F) are the centered, reduced, and dimensionless factors, varying between -1 and 1 (Table III). The absolute value of these coefficients is also a measure of the relative importance of each effect. The name and actual range of these factors are explained in Table II.

IV. IMPACT OF SILICON THICKNESS

In 3-D ICs, thinning of silicon allows for shorter interconnects in stacked matrices, and thus a faster communication between chips and thus a higher performance [7]. From a technological point of view, thin silicon leads to have a reasonable aspect ratio for TVS process.

However, a strong influence of this parameter on the thermal resistance of the stack was observed. In this section, the impact of die thickness is investigated more closely with a combined approach of measurement and simulation in terms of thermal impact.

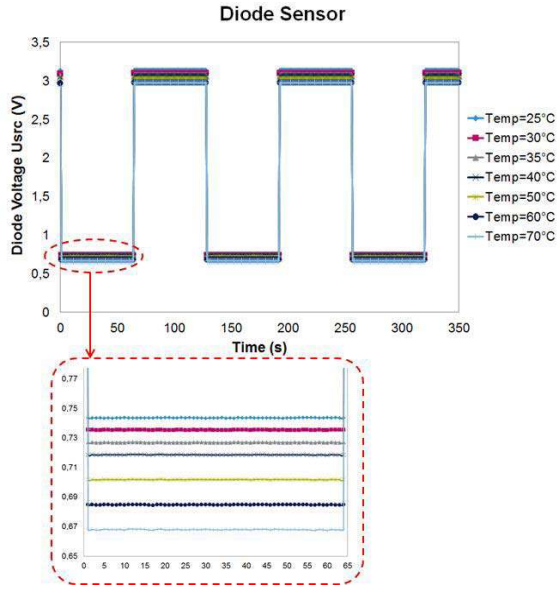


Fig. 16. Variation of the voltage across the 64 sensors according to the temperature over several measurement cycles.

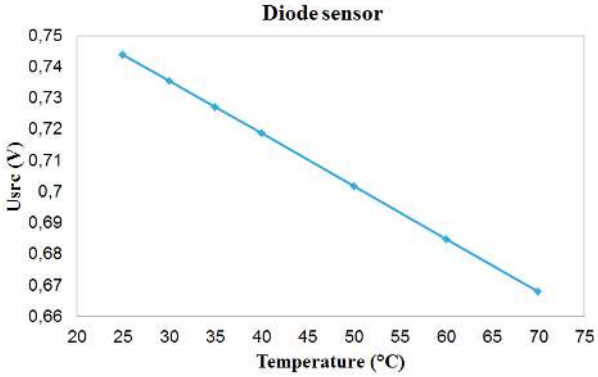


Fig. 17. Voltage variation across one sensor according the temperature.

A. Diode Sensor Calibration

First, the sensor calibration is performed to determine the sensitivity of each sensor with respect to temperature for a given current. To do so, the voltage of each sensor is measured at seven temperatures with a current fixed at 1 μA (Fig. 16). Then, for each sensor, the average variation of the measured voltage with respect to temperature ($\Delta V/\Delta T$) is determined.

The voltage has linear dependence temperature with across the diode; an example is provided on one of the 64 sensors shown in Fig. 17. The corresponding equation is

$$\Delta V/\Delta T = -1.68 \text{ mV/K}. \quad (6)$$

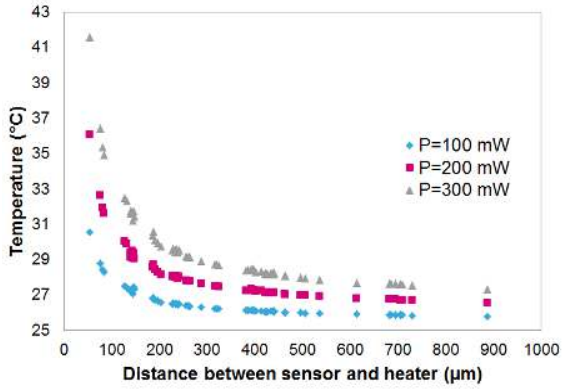


Fig. 18. Temperature profile at different heating power.

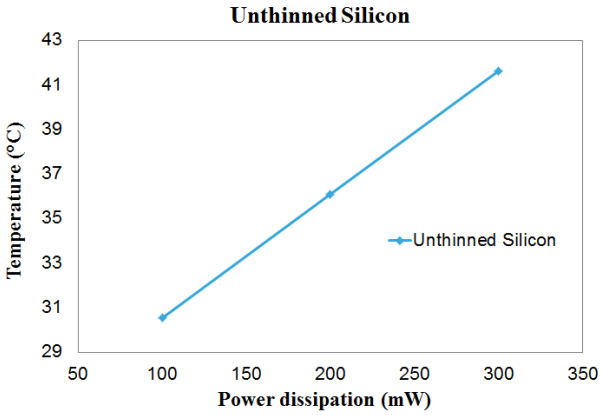


Fig. 19. Temperature at 50 μm distance from hotspot versus dissipated power in 750-μm unthinned silicon.

The same calibration procedure was applied for all the sensors.

B. Functionnal Measurement

After sensors calibration, the stack was tested at different value of the dissipated power while the output of temperature sensors was monitored. First, the influence of the power on the unthinned wafer was studied.

Figs. 18 and 19 show the temperature distribution at different heating powers on one hotspot (Fig. 18). A linear dependence was found between the temperatures at 50 μm distance from hotspot and the power (Fig. 19). Temperatures remain low at a silicon thickness of 750 μm since a large part of the heat is spread in the silicon.

Reducing the silicon thickness reduces heat spreading and hence the overall temperature is increased. In Fig. 20, the temperatures for silicon thickness of 80 and 750 μm are measured with several heating power. With a power of 200 mW, a strong increase of the maximum temperature by about 15 °C on the thinned silicon is observed. A slight difference in both temperature profiles is also observed, because the measurements are not carried out exactly in the same stacks and in the same areas: the one being on back side and the other on front side (refer to Figs. 3 and 4).

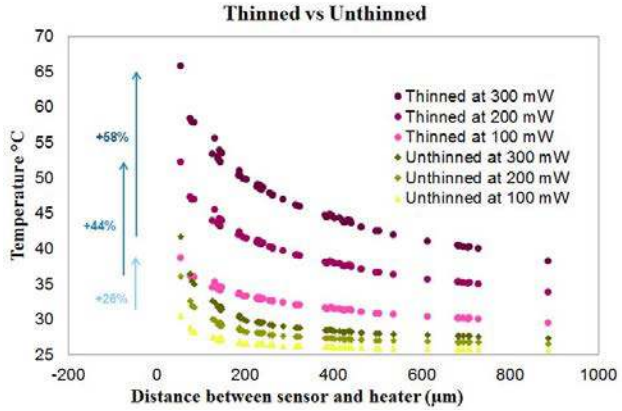


Fig. 20. Temperature profile in 80-μm thinned wafer and 750-μm unthinned wafer for different heating power.

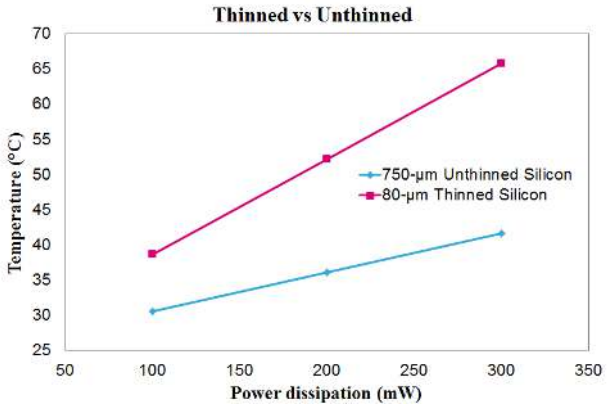


Fig. 21. Temperature at 50 μm distance from hotspot versus power dissipation in 80-μm thinned wafer and 750-μm unthinned wafer.

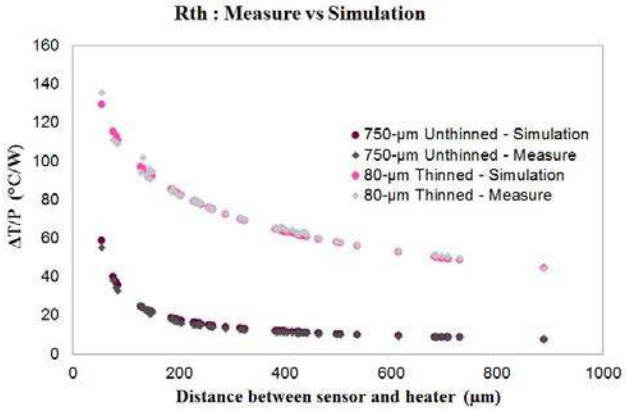


Fig. 22. Thermal resistance of both stacks silicon 750-μm unthinned and 80-μm thinned: measurements and simulations.

In these Figs. 21 and 22, the effect of silicon thinning on the temperature versus the power dissipation is depicted. The maximum temperature reaches more easily the admissible temperature limit, when the power dissipated increases on thinned silicon.

The impact of the silicon thinning on the temperature versus dissipated power can also be analysed by considering the

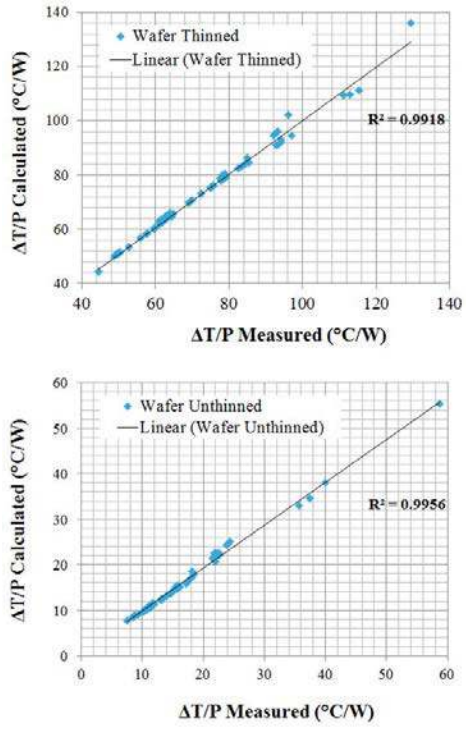


Fig. 23. Measures versus simulations.

total thermal resistance of stacks. Both the measurements and simulations are compared. The boundary conditions of the simulations are chosen such that on the top surface, a convection coefficient h is equal to 8 W/m^2 · K to model the natural convection and in the bottom surface, a convection coefficient h equal to 1230 W/m^2 · K is determined to fit FEM simulations with experimental results [5] by retro simulation. As discussed earlier, this high value physically corresponds to a conduction case.

Finally, a good agreement between simulations and measurements is found in the two stacks with a maximal error $\pm 1\%$ (Fig. 23). The silicon thinning leads to an increase of the thermal resistance of the stack, which means that the thinned silicon will be more favorable to hotspots generation.

V. Conclusion

In this paper, the impact of the silicon thickness on the temperature and on the thermal resistance for 3-D IC was studied by *in situ* measurement and simulations at wafer level. The temperature of each sensor has been studied. The sensitivity study shows a strong effect of the silicon thickness, and the distance between sensors and heaters representing the high gradient in the silicon: they both represent 99% of total variance of the thermal resistance. The boundary conditions, that is, the heat-transfer coefficients are also identified as a significant parameters. The study of the silicon thickness shows an increasing of thermal resistance, when the silicon thickness decreases. That leads to an increase of temperature while keeping power unchanged.

The further studies will address the full test of the 3-D chip in a package with electrical measurements and infrared thermal mapping. The comparison with simulation results will be done, and the impact of the stacking and the 3-D technology options on the thermal dissipation will be explored.

References

[1] C. C. Liu, *Circuit and System for Three Dimensional Integration Technology*. New York, NY, USA: Cornell Univ., May 2006, p. 470.

[2] J. H. Lau, "Evolution and Outlook of TSV and 3D IC/Si Integration," in *Proc. IEEE 12th EPTC*, Dec. 2010, pp. 560–570.

[3] P. M. Souaré, "Thermal behavior in stack-based 3D ICs," in *Proc. ESTC*, Sep. 2012, pp. 1–6.

[4] F. de Crécy, "A simple and approximate analytical model for the estimation of the thermal resistances in 3D stacks of integrated circuits," in *Proc. 18th THERMINIC*, Sep. 2012, pp. 1–6.

[5] P. M. Souaré, "Thermal correlation between measurements and FEM simulations in 3D ICs," in *Proc. IEEE 3DIC*, Oct. 2013, pp. 1–6.

[6] A. Saltelli, S. Tarantola, F. Campolongo, and M. Ratto, *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. New York, NY, USA: Wiley, 2004.

[7] C. Beelen-Hendrikx, "Trends in IC packaging," in *Proc. EMPC*, 2009, pp. 1–8.

[8] S. M. Sze and K. K. Ng, *Physics of Semiconductor Devices*. New York, NY, USA: Wiley, 1981.

[9] P. Coudrain *et al.*, "Towards efficient and reliable 300 mm 3D technology for wide I/O interconnects," in *Proc. 14th EPTC*, Dec. 2012, pp. 335–340.

[10] C. Gerard, M. Meijer, G. Wang, and F. Fruett, "Temperature sensors and voltage references implemented in CMOS technology," *IEEE Sensors J.*, vol. 1, no. 3, pp. 225–234, Oct. 2001.

[11] W. J. Kulesza, A. Hultgren, J. McGhee, M. Lennels, T. Bergander, and J. Wirandi, "An *in-situ* realtime impulse response auto-test of resistance temperature sensors," *IEEE Trans. Instrum. Meas.*, vol. 50, no. 6, pp. 1625–1629, Dec. 2001.

[12] A. G. McNamara, "Semiconductor diodes and transistor as electrical thermometers," *Rev. Sci. Instrum.*, vol. 33, no. 3, pp. 330–333, Mar. 1962.

[13] D. K. Sharma and K. V. Ramanathan, "Modeling thermal effects on MOS I-V characteristics," *IEEE Electron Device Lett.*, vol. 4, no. 10, pp. 362–364, Oct. 1983.

[14] G. E. B. Archer, A. Saltelli, and I. M. Sobol, "Sensitivity measures, anova-like techniques and the use of bootstrap," *J. Statist. Comput. Simul.*, vol. 58, no. 2, pp. 99–120, 1997.

[15] B. P. Dewan-Sandur *et al.*, "Thermal management of die stacking architecture that includes memory and logic processor," in *Proc. Electron. Compon. Technol. Conf.*, 2006, pp. 1–11.