

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

Thermal analysis and interpolation techniques for a logic + WideIO stacked DRAM test chip

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:

Beneventi, F., Bartolini, A., Vivet, P., Benini, L. (2016). Thermal analysis and interpolation techniques for a logic + WideIO stacked DRAM test chip. IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN OF INTEGRATED CIRCUITS AND SYSTEMS, 35(4), 623-636 [10.1109/TCAD.2015.2474382].

Availability:

This version is available at: <https://hdl.handle.net/11585/545018> since: 2021-02-15

Published:

DOI: <http://doi.org/10.1109/TCAD.2015.2474382>

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).
When citing, please refer to the published version.

(Article begins on next page)

This is the post peer-review accepted manuscript of:

F. Beneventi, A. Bartolini, P. Vivet and L. Benini, "Thermal Analysis and Interpolation Techniques for a Logic + WideIO Stacked DRAM Test Chip," in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 35, no. 4, pp. 623-636, April 2016. doi: 10.1109/TCAD.2015.2474382

The published version is available online at:
<https://ieeexplore.ieee.org/abstract/document/7229268>

© 2015 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works

Thermal Analysis and Interpolation Techniques for a Logic + WideIO Stacked DRAM Test Chip

Francesco Beneventi, Andrea Bartolini, *Member, IEEE*, Pascal Vivet, *Member, IEEE*,
and Luca Benini, *Fellow, IEEE*

Abstract—Self-heating, high-operating temperature are major concerns in 3D-chip integration. In this paper we leverage a 3D test chip (WideIO DRAM on top of a logic die) equipped with temperature sensors and heaters to explore thermal effects and to develop advanced thermal modeling strategies suitable for complex 3D-stacked circuits. We correlate temperature measurements with the power dissipated by the heaters using model learning techniques. Moreover we defined a Thermal Basis Function obtained using power and thermal data available from the on-chip sensors. This function can be used to predict temperatures at chip locations far from the temperature sensors and to infer the power dissipation at any location of the chip. In addition the same Thermal Basic Function can be used jointly with formal interpolation frameworks like Radial Basis Function (RBF) methods to effectively estimate the full-chip thermal map. Results show that this methodology outperforms existing interpolation approaches for sparse integrated sensors.

Index Terms—Temperature sensor, Thermal interpolation, radial basis function, 3D Integration, Temperature estimation, Sensor Virtualization.

I. INTRODUCTION

3D-chip integration aims to overcome scalability and performance bottlenecks of planar ICs by stacking multiple silicon dies to augment the silicon active area (e.g. number of processing element, memory banks, etc.) accessible with low latency [1], [2]. In addition, the vertical dimension opens new integration opportunities making possible to stack different silicon technologies in the same package. Components with incompatible planar manufacturing processes, e.g. CMOS multicore processors and DRAM memories, can be efficiently coupled enabling new performance breakthrough. The capability of integrating heterogeneous technologies makes 3D staking a good candidate to alleviate the “memory wall” as shown by recent products [3]–[5].

The potential for three-dimensional DRAM-logic integration has been recognized by industrial standardization committees. As an example, the JEDEC WideIO interface specs were defined for SDRAM interface to deliver twice the bandwidth of the LPDDR3 specification while improving the power

efficiency [5]. WideIO is a 512-bit wide bus (four 128-bit channels) larger than previous solutions. The main concern in having such a large number of bus lines in off-chip connections is the routing at PCB level. Through Silicon Vias (TSVs) are formed through the dies along the vertical dimension and allow short path between adjacent layers. This technology has the potential to deliver efficient inter-layer interconnections [2] alleviating also routing problems.

The vertical integration of silicon dies has two drawbacks: it reduces the vertical heat conductivity of the bottom and central dies w.r.t. the ambient and it increases the peak power density. The former is primarily due to the lower thermal conductivity of the silicon with respect to copper and other materials used in the heat-spreader which is directly thermally-connected with the active silicon in planar chips [6]. This worsen the heat removal from the inner silicon layers as they cannot be directly connected to a heat sink. The latter is caused by the stacking of active dies, often characterized by a thinner bulk, for which even a small amount of power consumed can result in high power densities. Moreover replicating the same logic die in the 3D stacks can lead to the scenario where power consuming units in each layer are aligned along the vertical dimension; as consequence maximum temperature can easily reach worst case values causing dangerous hotspots and thermal gradients.

To overcome these issues technological strategies have been proposed to augment the heat dissipation capabilities. Some of them exploit a thermal-aware floorplan and additional TSVs placement to serve as heat pipes for reducing the resistance toward the ambient [7], [8] while other solutions adopt liquid cooling through micro-channels on the silicon die [9], [10]. Recent work from Santos et al. [11] verifies with real silicon and with accurate simulations that TSVs improves vertical heat transfer but due to the SiO_2 s insulation layer they causes lateral thermal blockage in the silicon substrates containing TSVs arrays which may lead to increased hotspot temperature.

Therefore practical evaluation of specific thermal design in manufactured 3D ICs is hard as classical thermal introspection methods used for 2D ICs are less effective. Indeed traditional approaches based on precise die temperature measurement by means of IR-cameras [12] are hardly applicable in a 3D stacking context. In addition, as in 2D chips, fine grain thermal measurements based on integrated thermal sensors causes high silicon area, power and design complexity penalties. These limitations also jeopardize the effectiveness of fast 3D-chips thermal simulators [13], [14] which cannot be carefully calibrated against real measurements.

With the goal of pushing the research toward new solutions

F. Beneventi, A. Bartolini and L. Benini are with the Department of Electrical, Electronic and Information Engineering (DEI), University of Bologna, Italy. E-mail: {francesco.beneventi, a.bartolini, luca.benini}@unibo.it

A. Bartolini and L. Benini are with the Integrated Systems Laboratory, ETH Zurich, Switzerland. E-mail: {barandre, lbenini}@iis.ee.ethz.ch

Pascal Vivet is with CEA-Leti, Grenoble, France. E-mail: pascal.vivet@cea.fr

Copyright (c) 2015 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending an email to pubs-permissions@ieee.org.

Manuscript received April XX, XXX; revised September XX, XXX.

for improving the thermal introspection of 3D ICs, test chips have been recently manufactured to evaluate 3D-integration thermal properties by embedding controllable heaters and thermal sensors [15], [16]. This allows to assess how the heat spreads in the silicon die. Unfortunately, since in practical applications only a limited number of sensors can be integrated due to cost/area reasons, they cannot provide a full detailed silicon temperature map which covers the entire chip area.

Thermal interpolation strategies allow to estimate the temperatures of the die at locations different than those corresponding to the thermal sensors. Common interpolation algorithms exploit existing measured data e.g. coming from on chip sensors to recover, at runtime, a full 2D thermal profile useful in DTM solutions for hotspots mitigation. Few works in literature tackle this problem, but they are validated on simulators or on planar devices [17], [18] and often rely on calibrated fine-grain thermal models of the underlying HW which can be obtained only by the manufacturer. In addition they do not account for environmental conditions as well as for real deployment scenarios.

On the other hand, classical curve interpolation algorithms are agnostic of the physics of thermal diffusion and of the physical properties of the specific device. This often leads to poor results in case of a limited number of thermal sensors. Kung et al [19] uses an approximation function to describe the chip thermal distribution as a function of the distance between two locations. Their approach approximates the Green's function (inside the analytical solution of the heat equation for the chip) to reduce the computational complexity of the thermal map reconstruction. As in previous techniques, this is based on simplified models of the underlining HW and it cannot capture real geometry and in-field environmental condition.

Radial Basis Function (RBF) methods [20] interpolate scattered data through "ad-hoc" univariate basis functions. This enable to bring an "a priori" knowledge of the problem physics in the interpolation process. In this work we exploit RBF in the interpolation of 3D ICs thermal field introducing as basis function the thermal resistance field directly-learned from the target physical device. This approach has the advantage of combining physical knowledge of the problem and in-field calibration with a robust formal identification methodology.

If assessing the temperature of an IC is a hard task, assessing its power consumption is even worse due to the difficulties in building accurate and low overhead power gauges [21]. HW thermal sensors have been exploited to estimate the power consumption of an IC by solving the temperature-to-power inverse problem. Unfortunately this procedure is well-known to be ill-conditioned and as a consequence it is strongly affected by measurement noise and models errors [22].

In this work we tackle the temperature-to-power problem by exploiting the thermal resistance field identified directly from the target device. We improve the robustness of the solution by taking advantage of partial power consumption knowledge by means of HW heaters and system level power gauges [21], [23].

We demonstrate the effectiveness of our techniques on a real 3D-chip (Mag3D) dealing with the quantization noise of the provided thermal sensors. Mag3D is a real 3D test

chip designed by CEA-LETI under the Wide IO Memory Interface Next Generation (Wioming) project, and features built-in heaters and thermal sensors [24], [25]. We use them to extract the static thermal model and to derive the thermal resistance field function. This function/model can be coupled with floorplan data to extrapolate thermal information at chip locations not covered by thermal sensors. Moreover, this model/function allows to estimate the power consumption of HW components by simply knowing their position. Finally to evaluate the performance of the RBF interpolation approach we built an accurate thermal model (FEM) of Mag3D to be used as a reference in the cross-validation step.

A. Related Work

Several approaches have been proposed for estimating the on-chip temperature map. Some authors deal with the statistical nature of the measured temperature data coming from the real silicon [18], [26], [27]. Liu et al. [26] propose a framework to build a spatial correlation model based on temperature measurements. A generalized least-square fitting approach and a Kriging technique are used to account for die variability and to maximize the accuracy of the model. Zhang et al. [18], [27] adopt a statistical method to obtain the thermal profile of the chip. Starting from the thermal sensor readings they found a maximum a posteriori (MAP) estimate of the power density distribution of the whole chip. Successively they used this information in a linear model to recover the corresponding temperatures. However the estimation based on the correlation model in [26] is strongly sensitive to noisy measurements as it is open loop. In [18], [27] the key elements of the proposed method are the prior knowledge of the cross-correlation among different chip power sources and a fine grain thermal model. Unfortunately authors demonstrate this approach for a simulated system composed of a single core only for which: (1) a strong cross-correlation holds in between functional units and (2) the thermal model is assumed to be exact. In a real multiprocessor SoC the cross-correlation in between the different power sources will decrease as the components activity will depend more on the specific program flow. In addition the thermal evolution of real devices diverges from thermal-models extracted at design-time which cannot account for ambient temperature and deployment conditions.

Computationally efficient methods leveraging fast convolution have also been investigated. Cochran et al. [17] proposed an alternative to Kriging, exploiting Fourier analysis techniques to fully recover the chip thermal map. In their work a convolution filter in the space domain is applied to the measurements vector of spatially distributed temperature sensors. Although this method benefits of the computational efficiency of the FFT algorithm, the final accuracy is limited by the non band-limited nature of the temperature spectrum and approximations are needed in case of non-uniform placement of the thermal sensors, which is common in heterogeneous SoC. Convolution is also at the base of the methods developed in [28]–[31]. These methods require the prior knowledge of the full chip thermal impulse response to each power source. A generic input power map then is convolved with

the impulse response of each power source to reconstruct the corresponding thermal map. Power Blurring [32], [33] is a recent techniques presented by Ziabari et al. which uses the discretized temperature response to a unity power impulse placed in the center of the silicon die to create 2D thermal mask. This thermal mask can then be convolved in the x and y domain to a generic spatial power stimulus to obtain the steady-state thermal map. In these works however, the thermal impulse response is obtained from FEM thermal models and this makes the methods difficult to be applied since these models are not always available and cannot account for different deployment conditions.

When dealing with noisy thermal sensors, linear predictive filtering techniques have been applied. In [34] Sharifi et al. use the Kalman filtering approach to estimate the temperature at different die locations starting from the available power and temperature sensors measurements. Zajo et al. [35] estimate the thermal field inside a 3D MPSoC by means of the unscented Kalman filter (UKF) and on-chip thermal sensors. The approach described in [34] gives good accuracy when dealing with noisy thermal and power sensors. However these method needs a prior knowledge of the chip thermal model and a model order reduction step to lower the model computational complexity. Moreover both [34] and [35] have been evaluated only on simulators with Gaussian noise assumption.

Design constraints and area costs are the main factors driving thermal sensors allocation in modern MPSoCs. Interpolation techniques have been shown to be very effective in managing spatially non-uniform scattered measurements. Recently Wang et al. [36] address the thermal map reconstruction problem by using thermal sensors interpolation. The approach is based on spline interpolation which can rely on mature algorithm implementation and shows high accuracy when compared with others not-standard approaches [17]. In a non-optimized numerical implementation the overhead of this method results in $O(N^3)$ considering N thermal sensors. However a pure interpolation approach using a generic basis function have limited accuracy in describing specific physical phenomenas as e.g. heat transfer. In cases where a very small number of thermal sensors are available, spline interpolation returns poor performance as we will show in our experimental results.

Several authors tackle the inverse problem: temperature-to-power estimation problem. This is of practical interest as integrated power-gauges have larger area overhead w.r.t. thermal sensors. However thermal model inversion is an ill-posed problem mainly due to its high sensitivity to the noise present in the measured thermal data. Nowroz et al. [37], have proposed to use AC stimuli instead of DC ones to reduce the impact of flicker noise and spatial heat diffusion in thermography and to improve the accuracy of the temperature to power inversion problem. Results show the efficacy of AC methodology in reducing the thermal noise enclosed in the final thermal map. Differently Peak et al. [38] have introduced transient analysis and probabilistic optimization to control the noise sensitivity of thermography approaches. Ranieri et al. [39], [40] have built a framework for optimal thermal sensor allocation oriented to maximize the overall accuracy in the full

thermal and power map recovery at runtime. Reda et al. [41] have formulated an NP-hard problem for the optimal thermal sensor allocation under design constraints. They shown how a temperature model obtained as a linear combinations of the measurements of real thermal sensors can improve the accuracy of an non optimal sensor allocation due to design limitations.

Although some of the presented algorithms [39], [40] have demonstrated accurate results, their performance are either tightly coupled with the optimality of the sensor spatial allocation which may not be feasible for real floorplans, or rely on IR-camera images [37], [38], [41] which are not applicable in 3D stacked chip. Wang et al. [42] present Power Trace which takes advantage of regularization extensions of the maximum likelihood used in image restoration and deblurring. This technique can bias the solution of the inverse problem to have sharp edges which are more realistic for power maps. This algorithm is shown to be more efficient than standard l_2 norm regularization algorithms. Like others, this approach relies on extensive simulation to obtain the power-to-temperature spatial transfer function which needs to be inverted. In our work we show a methodology to derive this function from the real device, exploiting built in heaters and thermal sensors. This allows to capture the non-idealities of real manufactured devices.

B. Contributions

This paper tackles the problem of augmenting the thermal and power introspection of 3D ICs. This is achieved by introducing a novel methodology which leverages built-in heaters and thermal sensors to directly extract from the target device a thermal basis function. This function, given the geometrical position of two points in the device, estimates the thermal resistance present in between these two points. We then show how this function can be combined at run-time with the available temperature sensors and heaters values to generate “virtual” sensors readings and power estimates for silicon regions not directly covered by the built-in HW sensors. We then show that the interpolation accuracy and computational complexity can be improved by exploiting formal radial basis function interpolation methods.

The main contributions of this paper are the following:

- 1) A novel interpolation approach that merges actual power and thermal information coming from the silicon to produce a detailed chip thermal map.
- 2) A new method to identify a Thermal Basis Function is presented. It is able to analytically describe the spatial distribution of the thermal resistance between two on-chip locations.
- 3) We exploit this learned function with on-line HW sensors measurements to create virtual sensors which estimates the temperature and the power consumption of unmonitored die regions.
- 4) We validate our approaches on Mag3D, a real 3D test chip implementing WideIO and TSV technologies.
- 5) We show the effectiveness of these novel methods in solving the “temperature-to-power” inversion problems.

- 6) We demonstrate how to include the thermal basis function in a well established methods like the Radial Basis Function (RBF) interpolation.
- 7) We compared the accuracy and complexity of the proposed methods w.r.t. state-of-the-art interpolation strategies for estimating the temperature of digital ICs. The proposed approach shows the highest temperature estimation accuracy with a practical number of thermal sensors and limited computational time.

Differently from [43] this paper introduces the thermal interpolation problem and postulates the usage of the identified thermal resistance function as a basis function for formal interpolation method based on RBF approach (Section V-A). To evaluate the performance of our approach we set-up a CFD thermal model of Mag3D and we extended the experimental results section with the comparison of the proposed thermal interpolation approach based on RBF and the thermal basis function against convolution, spline and the approach presented in the previous conference version (Section VI-B).

The paper starts with an overview of the Mag3D architecture in Section II. Section III provides the details of our thermal modeling approach while in Section IV we show how the modeling framework can be applied to practical cases to augment the thermal and power introspection in 3D chip. In Section V we introduce the Radial Basis Function interpolation method and describe how to integrate the custom basis thermal function in its flow. Finally in Section VI we illustrate the results of our analysis. Section VII concludes the paper.

II. MAG3D ARCHITECTURE

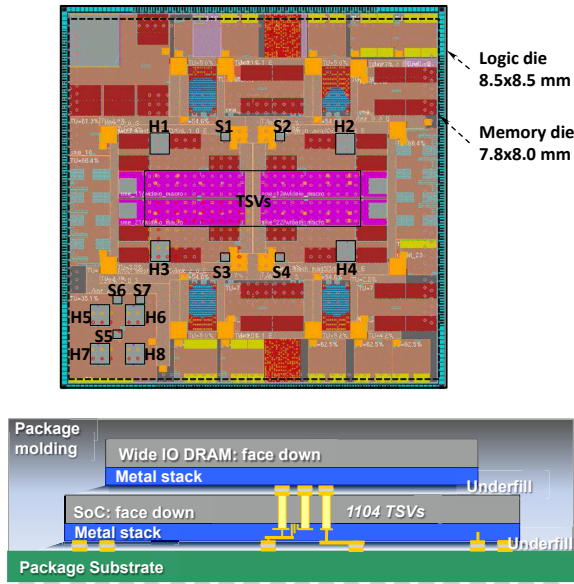


Fig. 1. Mag3D layout (top) and stack view (bottom) (Courtesy CEA-Leti).

Mag3D [5], [16], [25] is a WideIO memory-on-logic 3D circuit (65nm technology node) where the dies are stacked in a face-to-back configuration and are connected through TSVs and μ -bumps. The top die is a standard 1Gb mobile SDRAM provided by Samsung [4]. The bottom one is a SoC logic die

(flip-chip) containing a WIOMING circuit and is only $80\mu\text{m}$ thick to accommodate the integrated TSVs. The WIOMING circuit (Fig.1) implements an asynchronous NoC composed of 16 routers that connect 24 different units. Data managements units include SMEs (Smart Memory Engines) and MEPHISTO DSP units. MEPHISTO units are reconfigurable cores optimized for matrix computations [44].

The remaining units are dedicated IP cores with reconfigurable capabilities for DSP workloads. A generic application processor (ARM1176) is implemented to manage the overall system and acts as interface with the external world. The Mag3D chip is hosted on an application board designed around a Virtex5 FPGA. It implements an extension of the asynchronous NoC and some other units that provide debugging capabilities. Moreover it contains a NoC-ethernet bridge used to link the Mag3D chip to a host PC. The 3D-chip includes also HW sensors and programmable heaters to support thermal testing. These are distributed as follows:

- Each memory controller integrates one heater (H1,...,H4) and one thermal sensor (S1,...,S4).
- Four heaters (H5,...,H8) and three thermal sensors (S5,...,S7) are placed in the bottom left corner to emulate a quad-core processor (such as an ARM Cortex A9, for instance).

The heaters are made of poly-Si resistance and can dissipate up to 1W each. Each heater can be independently controlled from the board via embedded software, while integrated thermal sensors can be monitored in real time. Thermal sensor accuracy is $\pm 1^\circ\text{C}$ within the calibration temperature range (room temperature 27°C), which decreases at $\pm 4^\circ\text{C}$ at 100°C . Sensor resolution is 1°C in the entire range.

III. THERMAL BASIS FUNCTION

In this section we describe a methodology to exploit these built-in thermal sensors and heaters to extract a Thermal Basis Function which models the heat propagation in the device. We obtain it in two steps. First, from the sensors data and the heaters stress configurations we compute the static-thermal model. Second, we correlate the model's coefficients with the sensor geometrical position to fit the Thermal Basis Function. This will be used as basic building block in the next sections to estimate the power and temperature at unknown die locations.

A. Measurement Setup

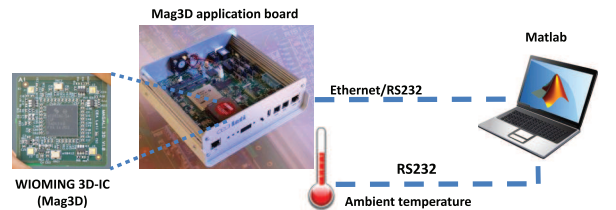


Fig. 2. Measurements setup

The training and validation data-sets are obtained with the measurement set-up depicted in Fig. 2. A Matlab [45] client

running on a lab workstation is directly connected to the on chip asynchronous NoC taking advantage of the ethernet-NoC interface provided by the board. We created a set of Matlab APIs to manage the interface with Mag3D (Fig. 3) which allows us to read the thermal sensors values and to set the heaters status. In addition our framework allows the measurements of the ambient temperature by means of an external temperature sensor connected to the Matlab environment through an Arduino board.

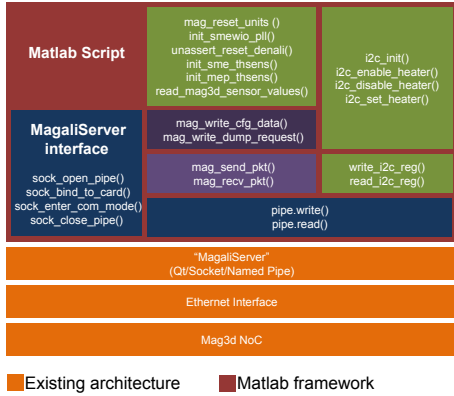


Fig. 3. Mag3D management interface

B. Static gain matrix

We extracted the steady-state thermal model of the Mag3D chip by applying a similar methodology as the one presented in [46] to the Mag3D Heater/Thermal sensor infrastructure. Given H heaters and S thermal sensors, setting the heater power stress $P = (p_1, \dots, p_H)$ the resulting temperatures $T = (t_1, \dots, t_S)$ measured at the sensor locations can be described by the following relation $T = T_{amb} + K \cdot P$ where K is the static-gain matrix and T_{amb} is the ambient temperature. As in [46], in this work we obtain the K matrix by a set of empirical measurements on the target chip.

The necessary data are obtained with the following procedure. We apply a set of power vectors, composed of binary power stress (heater on/off) to the chip considering all the possible 2^H combinations. At the same time we store the corresponding temperature sensors readings after waiting for the thermal response to get stabilized: the power vector is maintained active for 10 minutes which is enough to ensure that the transient effects are finished¹. We collect the heaters power measurements in the matrix P_m and the temperature measurements in the matrix T_m . The elements of the T_m matrix are relative to T_{amb} such as $T_m = T - T_{amb}$. Finally we apply a least squares optimization to compute the K matrix coefficients:

$$\hat{K} = \arg \min_K \|T_m - P_m \cdot K\|^2 \quad (1)$$

The static-gain matrix is a thermal resistance matrix since its coefficients have the units ($^{\circ}\text{C}/\text{W}$). We must note that the availability of programmable heaters enables accurate, low-noise and spatially well defined power stimuli which would

have been hard to obtain with active logic. Usually logic blocks spread among the entire die area and their power has higher fluctuation due to the program flow and software stacks interference (operating system and firmware run-time).

C. Thermal Basis Function

The thermal resistance is known to be a function of the distance between the heat source and the sensing point [47]. By knowing its analytical relation which links the coefficients of the K matrix to the geometrical distance of the heating sources to a given die location, it is possible to systematically compute extra coefficients related to thermal sensors and heating sources regions not covered by real one. This can be done by correlating the values of the K matrix with the reciprocal heater/sensor distance. The latter can be extracted from the geometrical coordinates of the heaters and the temperature sensors placed on the chip. Figure 4b displays the value of the coefficients of the K matrix versus the distance. As we notice, there is a strong correlation between these two quantities.

In this section we start from the K matrix and given the geometrical coordinates of the power source x_p and the thermal sensors x_s , we extract an approximation function that generates the coefficients of the K matrix as $K = f_K(x_p, x_s)$. To build the f_K function we use the function template described in [47].

$$f_K(x_p, x_s) = \frac{1}{((x_s - x_p)/a)^2 + 1} + \frac{b}{x_s - x_p} e^{\frac{(\ln \frac{x_s - x_p}{c})^2}{d}} \quad (2)$$

This function has four fitting parameters $\theta = \{a, b, c, d\}$. In this paper we use a non linear least squares algorithm to calculate them:

$$\hat{\theta} = \arg \min_{\theta} \|e_K\|_2^2 \quad (3)$$

where e_K is the error vector obtained by subtracting the coefficients of the original K matrix from those obtained using the f_K function.

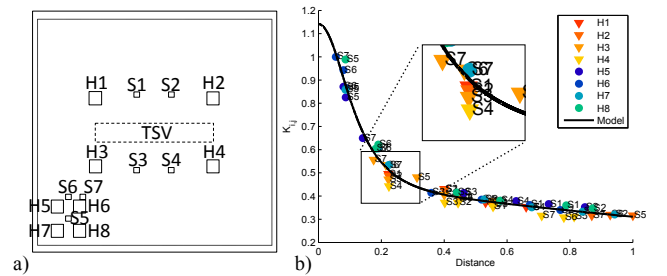


Fig. 4. (a) Mag3D floorplan; (b) Normalized Thermal Basis Function used to approximate the K matrix coefficients

By applying the procedure described above to the Mag3D chip we obtain the resulting approximation function shown in Fig.4b with label "Model" and solid line. This function represents an accurate analytical model able to approximate the real coefficients of the K matrix as function of the heater(source) to sensor(sink) distance. However, from the same figure, it is clear that this approximation function depends only on the distance and thus neglects information contained in the absolute position of the heater/sensor. For instance, by considering

¹This time period has been selected empirically.

the value of the coefficients describing the iteration between $H1, \dots, H4$ and the sensors $S1, \dots, S4$, highlighted in the Fig. 4b inset, we notice that different heater-sensor couples having the same reciprocal distance have different K coefficient values. Indeed in 3D-chips the floorplan is composed of spatial regions that involve different materials (i.e. TSVs). To account for this effect the basis function needs to be modified to consider also the absolute location at which the heater and sensors are located. Next section discusses our approach for introducing spatial adjustment in the template of the Thermal Basis Function and in the learning procedure. To distinguish the two approaches, the Thermal Basis Function presented in this section is named “Uniform” Thermal Basis Function while the one in the next section is named “Spatial Adjustment” Thermal Basis Function. It must be noted that due to current Mag3D sensor and heater placement limitations, which are all placed in the bottom die, the Thermal Basis Function has been trained for a planar heat dissipation. Consequently if it is used to estimate temperatures in the top die it will assume the same building materials composition as in the planar die. In future works we will extend this technique to handle the heat conductance heterogeneity and vertical temperature estimation [11].

D. Spatial Adjustment

The real values of the coefficients of the K matrix contain this spatial behavior as depicted by the Fig.4b inset. Therefore, to improve the accuracy of the approximation function we need an “adjusting” factor able to consider also the real position on the chip surface. The approach we propose is to weight the error k_{err} between the uniform approximation and the real coefficient. The weighting is done using a nearest-neighbor approach. This assumes that the properties of an unknown power source or thermal sensor are similar to the nearest existing one. Considering d_p as the distance of the power source and d_s as the distance of the sensing location from respectively the nearest known power source and temperature sensor, the adjusting factor can be defined as:

$$f_{ad} = K_{err} e^{-\frac{d_p + d_s}{\alpha}} \quad (4)$$

where α is a coefficient that regulates the spatial decaying of the adjusting term. Putting all together, the final function can be rewritten as $f_{K,ad} = f_K + f_{ad}$. In Section VI we will validate our modeling methodology with off-sample tests. In the following section we will present two main approaches to use this Thermal Basis Function to both create “virtual temperature and power sensors” and HW-aware interpolation.

IV. VIRTUAL SENSORS

In this section we discuss how these modeling strategies, when combined with heaters and thermal sensors infrastructure, can be applied to augment the introspection on real 3D devices in the modeling of the thermal properties, estimating the real temperature and power consumption of not physically monitored HW components by knowing only their position.

A. Thermal sensors virtualization

The Thermal Basic Function $f_{K,ad}$ can be exploited to infer the temperature at positions of the silicon surface not covered by any temperature sensors. Formally, in this case, the problem translates into the evaluation of the temperature T_x at the location x_{s_x} given K , the function $f_{K,ad}$ and the input power map P_m . To solve this problem first we evaluate the new row of the K matrix which correlates the existing power sources to the new virtual temperature sensor:

$$K_{S+1,h} = f_{K,ad}(x_{s_x}, x_h), \quad h = 1, \dots, H$$

Finally, after reconstructing the new matrix $K_s = \{K | K_{S+1,h}\}$ the temperature at the unknown location x_{s_x} of the chip is:

$$T_x = \sum_{h=1}^H K_{S+1,h} \cdot P_{m,h} \quad (5)$$

In the Section VI we will evaluate the performance of the proposed application directly on the Mag3D chip.

B. Temperature to power

The following procedure solves the orthogonal problem of identifying the power dissipated at unknown locations of the chip starting from the knowledge of the K matrix and the corresponding $f_{K,ad}$ function. We aim to calculate the power P_x of an “extra unknown” unit in the chip that generates the measurable (by the available thermal sensors) steady state thermal response T_m . Given the position x_{p_x} of the new power source we start generating a new column entry in the original K matrix

$$K_{s,H+1} = f_{K,ad}(x_{p_x}, x_s) \quad s = 1, \dots, S \quad (6)$$

obtaining the new matrix $K_N = \{K | K_{s,H+1}\}$. To calculate P_x we need to solve the problem $P = K^{-1} \cdot T$. We can use two approaches to obtain P_x .

1) *Model1*: The first approach “Model1” assumes that an external power gauges is available to monitor the full chip power consumption [48]. This translates in a constrained optimization problem where the complete chip power vector $P_T = \{P_m, P_x\}$ is the unknown. The optimization problem is:

$$\hat{P}_T = \arg \min_{P_T} \|K_N \cdot P_T - T\|^2 \quad s.t. \{P_T \geq 0, \sum_i p_i = P_{MAX}\}$$

2) *Model2*: A second approach “Model2” assumes that a limited number of power gauges are integrated on-chip to monitor the power consumption of a small subset of HW components, while others are not monitored [21]. This translates in the following least square optimization where the vector P_x is the only unknown:

$$\hat{P}_x = \arg \min_{P_x} \|K \cdot P_m + K_{S,H+1} \cdot P_x - T_m\|^2 \quad s.t. \{P_x \geq 0\}$$

V. 2D TEMPERATURE INTERPOLATION

Previous section described a methodology to obtain at run-time the temperature and power consumption of unknown die region by combining the obtained Thermal Basis Function with both the available heaters and thermal sensors measurements. In this section we exploit the Thermal Basis Function inside a formal interpolation framework that considers only the available thermal sensors to estimate a detailed temperature map of the entire die.

A. RBF Interpolation

In the \mathbb{R}^2 domain an interpolating surface $T(x)$ of scattered data can be expressed as a linear combination of basis functions

$$T(x) = \sum_{s=1}^S c_s \Phi_s(x), \quad x \in \mathbb{R}^2 \quad (7)$$

In the case of *Radial Basis Functions*, the basis are defined as $\Phi_s: \mathbb{R}^2 \rightarrow \mathbb{R}$ such that $\Phi_s(x) = \varphi(r)$ and $r = \|x - x_s\|$ is the Euclidean norm.

Assuming to know the *basis function* $\varphi(r)$, the only unknowns in the relation (7) are the coefficients c_s that are obtained solving the interpolation problem defined as follows. Given a set of measurements $T_s = (t_1, \dots, t_S)$ from sensors located respectively at the positions stored in the vector $x_s = (x_1, \dots, x_S)$ we want to find a continuous function $T(x)$ corresponding to our thermal map, such that

$$T(x_s) = T_s$$

From (7) this translates in finding the unknown coefficients $c_s = (c_1, \dots, c_S)$ by solving the following linear system

$$A_I c_s = T_s$$

where A_I is the interpolation matrix having the entries $a_{i,s} = \Phi_s(x_i)$. By knowing the basis function φ the entries of the interpolation matrix are easily calculated as $a_{i,s} = \varphi(\|x_i - x_s\|)$. The points $x_i \in \mathbb{R}^2$ are the *centers* where the basis functions are evaluated and they coincide with the spatial location $x_i = x_s$. As final step the thermal map $T(x)$ can be evaluated on an arbitrary grid defined by the set of points $x \in \mathbb{R}^2$ using (7). We consider the points defined by the $N \times M$ regular grid.

In this work we propose to use this approach together with the Thermal Basis Function described in Section III-C. In particular we consider $\varphi(r)$ to be the eq. (2) and $r = \|x_p - x_s\|$ the distance between two chip locations.

VI. RESULTS

In this section we evaluated the proposed methodologies on the Mag3D test chip. We split our analysis in two main parts.

In the first part we present the results obtained directly on the real Mag3D chip. This first family of tests evaluates the performance of the proposed learning algorithm for temperature prediction and temperature-to-power inversion by using real data and off-sample validation.

In the second part we focus on the evaluation of the accuracy and complexity of the proposed interpolation algorithm to predict the entire chip thermal map w.r.t. state-of-the-art

approaches [17], [36]. This evaluation cannot be performed on the real chip as it is impossible to directly measure a detailed thermal map due to the absence of fine grid of thermal sensors. We created a FEM model of the Mag3D chip to calculate the steady-state thermal profile of the chip surface and we use it as reference for the analysis.

A. Part 1: Real Mag3D chip results

1) *Sensitivity to Heaters Availability*: In the first test we evaluated the sensitivity of the learning procedure of the Thermal Basis Function and of the temperature estimation to the available number of heaters in the chip. Indeed in practical case due to the different HW costs, thermal sensors are more available than programmable heaters and power sensors. In our test chip there are $H = 8$ heaters and we emulate the case of reduced number of heaters by considering only a subset of the available ones in the test chip. We then compute a new K matrix using (1), named K_{red} matrix to distinguish it from the one constructed using all available heaters named K . Both of them are learned by considering all the available temperatures sensors ($S = 7$). The missing coefficients are thus recovered using the $f_{K,ad}$ function. Finally we validate the accuracy of this model by comparing the coefficients values with the full K ones obtained using all the heaters. More in details, as described in section III-B, we calculate the reduced K_{red} matrix using 2^{H-n} power vectors and temperature vectors; n indicates the number of removed heaters. Successively, using the method presented in Section IV-B we calculate the missing parameters of the K_{red} matrix. In this case the approximation function used in (6) is obtained using the K_{red} matrix and so it has a lower accuracy. Finally we compare the obtained coefficients with the corresponding ones in the original K matrix.

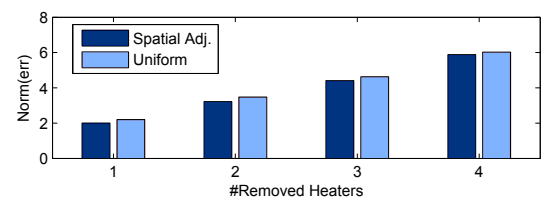


Fig. 5. Adjusted and uniform approximation function comparison

In Fig.5 we show the norm of the absolute error calculated for the cases with $n = 1, \dots, 4$ missing heaters. Since the performance of the approximation function are strictly correlated to which heater is removed, for each of the four configurations we calculated the average error considering all the n -choose- H combinations. Fig.6 reports this metric for the 3-choose-8 case. The results confirm that the adjusted approximation function behaves better in coefficient recovery than the uniform one since it can consider spatial variability. There is also a linear dependency between the error in the matrix coefficients and the number of removed heaters.

Moreover the higher QoS of the adjusted approximation function does not incur in a timing overhead when compared to the uniform one as the extra steps needed to compute the nearest-neighbour map and the adjustment term (eq.(4)) needs

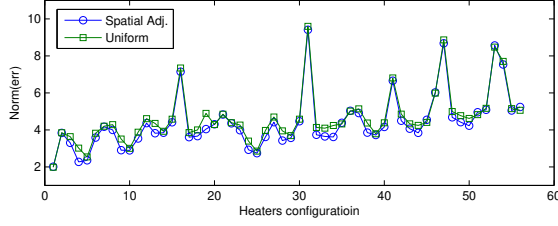


Fig. 6. Error comparison in the 3-choose-8 heater configuration test

to be computed offline only once. Then the runtime overhead for the uniform and the adjusted approximation function is the same.

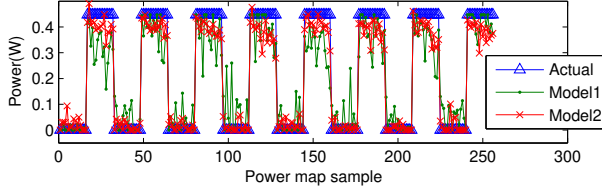


Fig. 7. Power traces recovered using two different methods to invert the K matrix

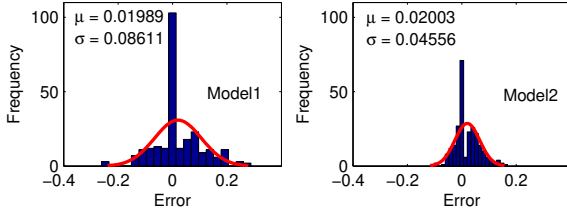


Fig. 8. Residuals of the Model1 and Model2 output obtained using the real heater power trace

2) *Temperature to Power Inversion Accuracy*: The second test aims to quantify the accuracy of the proposed Virtual Sensor method in estimating the power dissipated at unknown locations of the chip as described in Section IV-B. In particular we removed the heater H4 and we use its power trace as a reference for the validation. We then execute all the steps described for both the “Model1” and “Model2” approaches. The results of them are the two identified power vectors compared in Fig.7 against the real one. The real validation trace is indicated as “Actual”. “Model2” clearly has better performance due to the larger number of input information and less problem unknowns. Fig. 8 shows the residuals histograms. We notice that both the models has similar average residual but “Model2” shows significant smaller standard deviation. For both the average error is below 20mW.

3) *Virtual Sensor Accuracy*: In this test we evaluate the performance of the proposed virtual thermal sensors which exploits the Thermal Basis Function to compute the temperature at chip locations not covered by real temperature sensors. The procedure is explained in section IV-A. In this particular test we removed the sensor S6 and we used its real value as validation trace. We compute the error as the difference

between the real sensor output and the temperatures generated by the two models: the “Full Model” obtained using all the chip sensors (also the S6) and the recovered model (Rec. Model) obtained reconstructing the missing coefficients using the approximated function. Fig.9 shows the histogram of the

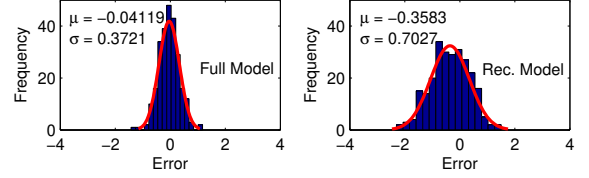


Fig. 9. Residuals of the Recovered model and the Full model output obtained using the real temperature trace

error for the two approaches. As expected the “Full Model” has a lower error than the “Rec. Model”. Both of them are unbiased as they have null error average. Moreover “Full model” has an accuracy of 1°C which is within the HW thermal sensor accuracy (1°C) and the “Rec. Model” has an accuracy loss of only (1°C) w.r.t. real HW thermal sensors.

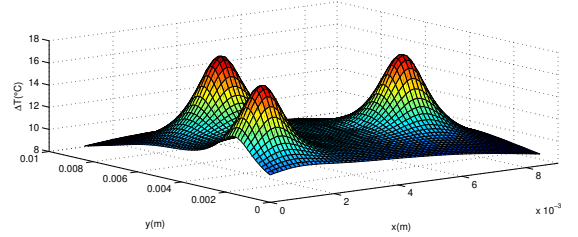


Fig. 10. Full chip thermal map

Using this procedure it is possible to generate a detailed thermal map of the whole chip surface. Assuming to subdivide the chip surface in a $M \times L$ grid where each point is a virtual thermal sensor point. Executing $M \times L$ times the procedure described in IV-A we can generate a larger K matrix K_l composed of $(M \times L) \times H$ coefficients, able to estimate the temperature at each point of the chip surface. Setting $M=L=64$ we reconstructed the steady-state thermal field of the chip corresponding to a generic power map (i.e. H1,H4 and H5 = on). The resulting thermal map is showed in Fig.10.

Unfortunately with this set-up it is difficult to evaluate the goodness of the fine grain thermal map estimation as we are missing the true reference thermal map, which is impossible to obtain with direct measurements in the Mag3D test chip. To overcome this limitation in the next section we built a finite element model (FEM) of the device on which we will evaluate the accuracy of thermal interpolation strategies.

B. Part 2: Mag3D FEM Model Results

The validity of the interpolation algorithms needs to be checked using an accurate and complete reference dataset. We obtained the reference full chip thermal map by creating a FEM model of the 3D stack. We use COMSOL [49] to build the Mag3D chip model. We built the FEM model using

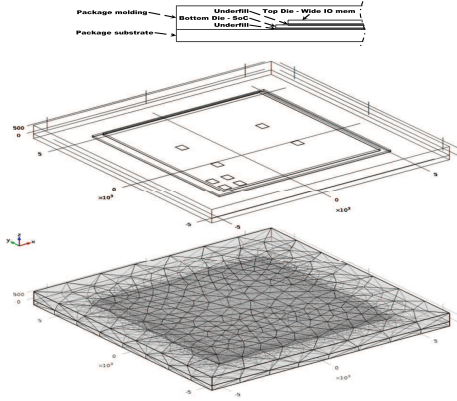


Fig. 11. FEM model of the Mag3D chip.

the chip size and materials coming from the manufacturer datasheet. We then validated it against real measurements done on the real Mag3D device. As shown in Fig.11, the model considers all the main layers of the 3D stack, up to the external chip package. The layers are considered homogeneous material blocks. However the heaters are obtained as poly-Si material blocks located inside the silicon layer. We have neglected the chip inhomogeneities like the metalization layers and the μ -bumps located in the underfill material layers. The size of each layer is obtained from the chip data-sheet as well as the materials properties which include the thermal conductivity and the heat capacity. We set up the simulation to get the steady-state heat equation solution at the chip boundaries considering the heaters as the unique heat sources. We consider also a convective heat flux as boundary conditions imposed on all the chip external faces. The heat transfer coefficient is empirically selected to minimize the mismatch in between the real thermal sensor readings and the thermal simulation results when the same power pattern is applied both to the real chip and the FEM model.

As a first step we recreated the entire chain of measurements done on the real device. This is done following the procedure described in Section III-B. This has been done by computing the COMSOL model as a batch simulation in a Matlab script. The script applies all the power vectors to the heaters in the model and stores the corresponding steady-state temperatures for all the silicon spatial coordinates that match the real thermal sensor locations. The resulting dataset obtained by the simulation is then used to compute the new K matrix following eq. (1) and computing the new basis thermal function as described in Section III-C.

We then used this set-up to evaluate the performance of the novel proposed thermal interpolation strategies which are based on formal RBF approach as presented in Section V. We evaluated the performance of the proposed method w.r.t. state-of-the-art methodologies and the proposed Virtual Thermal Sensor approach which uses in addition to the thermal sensors also per-component power information.

We carried out three main experiments to measure the performance of the novel interpolation method (RBF) introduced in Section V. Firstly we focused on the accuracy of

the temperature estimation at chip locations not covered by any thermal sensor. Secondly, we considered the ability of the algorithm in estimating the absolute temperature in the hot-spot location. Finally, we measured the computational overhead of the proposed algorithm to build the entire thermal map.

We compared the RBF method against two state-of-the-art approaches. More in details we have considered the spectral approach (Spectral) presented in [17] and the surface spline interpolation method (Spline) implemented in [36]. In addition we provide a comparisons w.r.t. the virtual thermal sensors approach discussed in Section.IV-A and presented in the conference version of the paper [43].

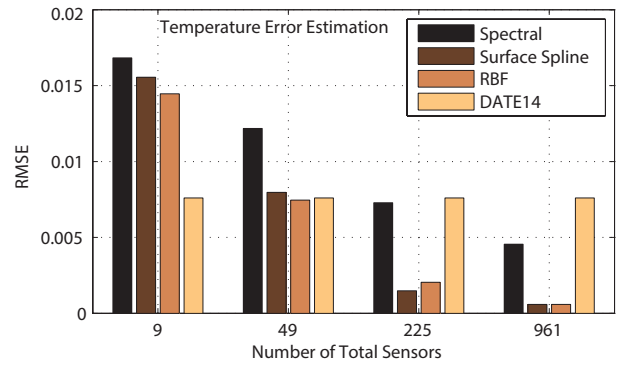


Fig. 12. Virtual sensor accuracy

In all the tests we assume the knowledge of the full chip thermal map which is obtained by simulating the FEM model. We selected as reference thermal profile the Mag3D temperature map shown in Fig.17 where only the heater H3 is turned on. This particular configuration emulates the presence of an hotspot. The reference thermal map is composed of 4K samples. The temperature field is sampled following a regular square sensors pattern and considering an increasing number of total thermal sensors as shown in the first column of Fig. 13. This emulates an increasing number of real temperature sensors and measurements available in the target device. We expect the estimation quality to increase as the number of input thermal measurements increases. In each test, we initially sampled the reference thermal map using a centred square grid of $S \times S$ points where $S = \{3, 7, 15, 31\}$. These are chosen to have the samples of the smaller grid included in the larger one and to minimize the noise in between the grids. These values are computed using the formula $S = \frac{\sqrt{4096}}{S_{step}} - 1$, where $S_{step} = \{2, 4, 8, 16\}$. Successively these values and the corresponding spatial coordinates are passed to the algorithms that compute the interpolated thermal map. For each map we used the RMS error as the interpolation performance index. We compute for each size of the reference temperature the same amount of output. These are chosen as the relative complement of the largest temperature reference input (31x31) in the total available reference samples (4K). This is done to consider always the same off-sample sensors in the computation of the error (RMSE).

1) *Virtual Sensor Accuracy*: The focus of this test is to measure the accuracy of the different algorithms in predicting

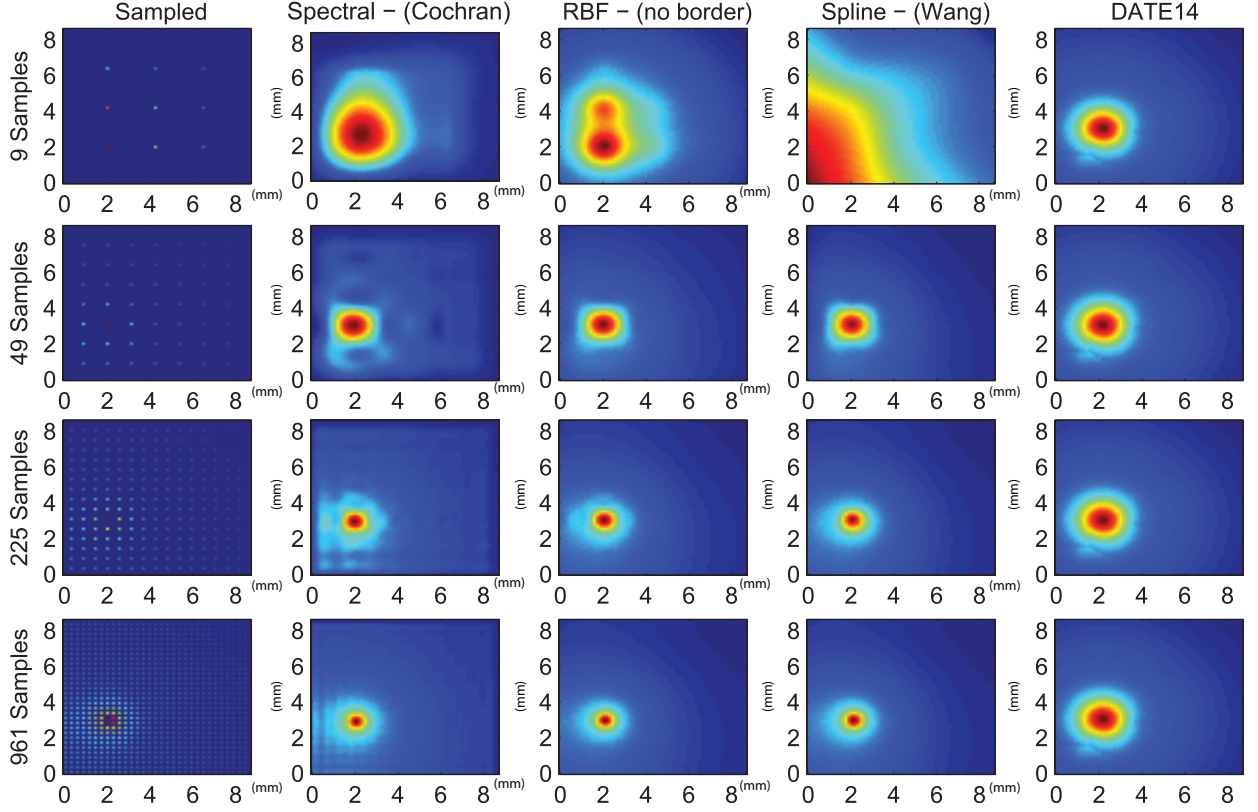


Fig. 13. Temperature maps: test (rows) and algorithm (columns). First column shows the input data obtained by sampling the reference map (Fig.17)

the temperature at unknown silicon locations. The temperature error (RMSE) is calculated only for the off-sample map points, namely the points of the map never considered as input samples in between all the tests.

Figure 12 shows the final results; the RBF algorithm shows the best accuracy for the different sensors numbers. In addition, we notice that as the number of sensors increases the accuracy of the RBF and Surface Spline algorithms increases faster than the Spectral algorithm that shows a relatively lower accuracy and higher error. As reported by the DATE14 bar, the thermal sensor virtualization strategy presented in Section IV-A has an accuracy which is constant with respect to the number of different input temperature sensor numbers. This is expected since the temperature estimation for this approach, once the thermal basis function is obtained, only depends on the power input estimation. We can notice that, if a good input power estimation is available on-line, this approach is beneficial for reduced numbers of input temperature sensors. If we focus the analysis to the pure thermal interpolation strategies (Spectral, Spline, RBF) we can observe a relatively higher performance of the RBF algorithm at low number of sampled points. Indeed both Spline and RBF methods use a basis function but what makes the difference is the specialized one used in the RBF approach. In the RBF case indeed, the basis function is learned using real silicon thermal data and thus it encloses more detailed thermal information. This property of the basis function is the key factor for a more accurate thermal map recovery as described in the paper. It must be noted that the relative higher accuracy of RBF w.r.t.

Spline decreases as the number of sensors increases. The reason of such artefact is that in RBF we do not re-train the basis function with the new thermal sensors, but we keep the original Mag3D number. This biases the basis function toward distant thermal sensors. Thus the reported results are conservative bounds for the RBF case. A qualitative overview of the various algorithms results can be found in Fig. 13. In the figure we can see the temperature estimated by the different approaches (along x-axis) for different input thermal sensors numbers (along y-axis).

Figure 14 shows the absolute error histogram for the corresponding interpolation algorithm and number of input temperature measurements. RBF achieves a full thermal map reconstruction with an accuracy of $\pm 2^\circ\text{C}$ starting from only 9 thermal sensors. With 49 thermal sensors RBF achieves an estimation accuracy similar of real HW ones ($\pm 1^\circ\text{C}$).

2) *Hot-Spot Estimation Error*: In this test we measured the capability of the algorithms in predicting the temperature of the hotspot location. This is done by considering the original temperature map and selecting the hottest point coordinate. Then, the absolute error in between the original temperature at this coordinate and the reconstructed ones is computed for each of the configurations in S . The plot shows that, despite the increasing number of total sensors, the absolute error has a non monotonic behavior. This can be explained considering that in the interpolation algorithm the relative position of the sampled points has a primary role in the surface reconstruction. The interesting result is that the spectral approach has better accuracy for the small sensor number case than the others al-

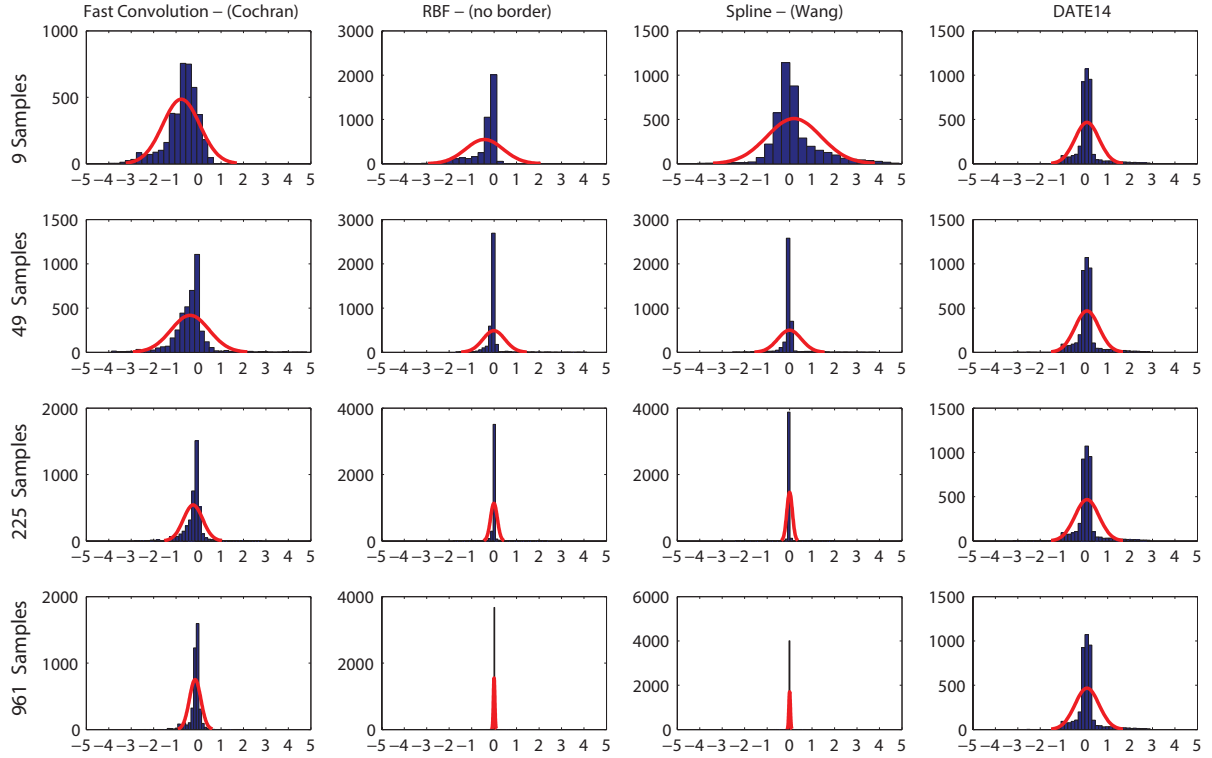


Fig. 14. Histogram of the absolute error: test (rows) and algorithm (columns). Errors in degrees celsius (x-axis). y-axis reports the frequency of a given error.

gorithms when considering a uniform sampling grid. However by increasing the sampling points the accuracy improves for the others algorithms and the RBF approach shows the best performance for a realistic and affordable number of points.

3) *Algorithms Overhead*: All the algorithms are implemented in Matlab and to have a comparison of their computational complexity in this test we measure their execution time. This is done for the different input temperature samples considered but for constant output interpolated temperature which is equal to 4K samples (64×64). More in detail we measure the time interval needed to calculate the final interpolated map starting from the measured sampling points. As expected the Spectral algorithm is the fastest approach since it exploits the FFT numerical efficiency. Indeed it must be considered that the basic implementation of the RBF and the Spline approaches have a computational complexity equal to $O(N^3)$ while for the Spectral method it is only $O(N \log N)$. In Fig. 16 are reported the normalized execution times of the three algorithms. We executed the algorithms on a dual core processor at 2.5GHz and 4GB of RAM. The maximum elapsed time is 5.42 seconds obtained for the RBF algorithm using 961 sampled points. Although it is a considerable high value for a run-time implementation it should be considered that it is computed for a large amount of sampling points (961 thermal sensors) that is an unpractical scenario since this sensor number would occupy the 53% of die area. The results highlight that the Spline method is relatively faster than the RBF. This can be caused by both the algorithm operation mode and the radial base function used. During the map interpolation

the radial base function is called repeatedly to compute its value for each evaluation point. The thin plate spline function used in the Spline algorithm is relatively simpler than the basis function in the RBF approach; consequently it requires a shorter evaluation time. However this performance difference can be alleviated optimizing the function evaluation, for example using LUTs.

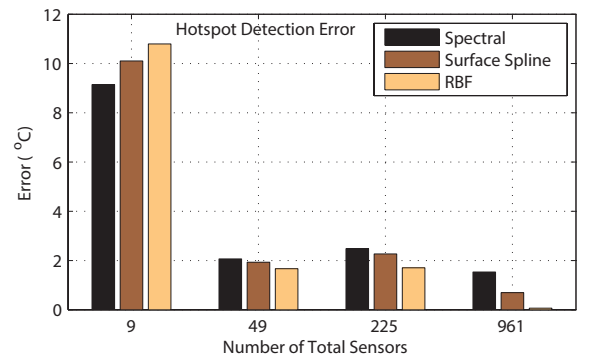


Fig. 15. Hot spot estimation error

C. Boundary Aware RBF

One of the main benefits of the RBF interpolation method introduced above is that it is agnostic of mesh constraints between the data points. This makes the algorithm extremely flexible in contexts where the data points follow an irregular spatial pattern. In this section we take advantage of this feature

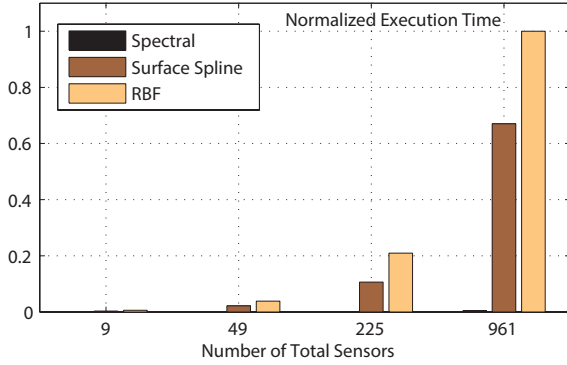


Fig. 16. Algorithm overhead

to improve the thermal map estimation when supplementary temperature measurements are available on top of the existing ones. As seen before the main data points are thermal information coming from the on chip sensors. In some cases however, additional chip thermal data are available directly from advanced compact thermal models (CTMs) [50] and thus temperatures at particular chip/package location can be easily inferred. This can be useful if a higher accuracy at certain chip location is needed. Indeed Fig. 17 shows that the real device (Comsol model) has a thermal gradient near the chip border. This information is not present in the radial basis function kernel by itself and it is partially included with the spatial adjustment as discussed in Section III-D. To test this assumption we assume to directly measure the temperature of the chip at SoC boundaries as described in Fig. 17. After adding these informations to the existing Mag3D data-set, following the same procedure early described in Section III-C we generated a new radial function namely RBF+boundary. Figure 18 shows the new thermal basis function. From this comparison we notice that the shape of the basis points gets modified to account for the newly added data points.

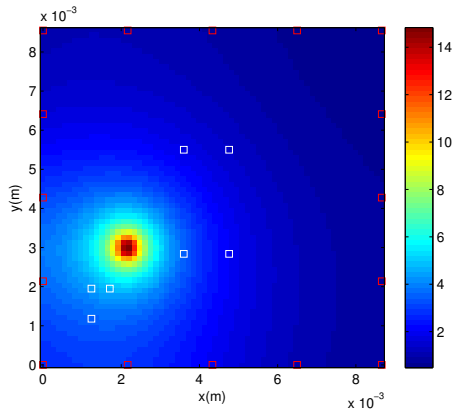


Fig. 17. FEM model reference temperature map and locations of thermal sensors (white squares) and points on the chip borders (red squares).

Starting from a coarse set of data (the power map, the thermal sensors and ambient temperature) we want to recover the fine-grain temperature field. We considered the scenario depicted in Fig. 17. Initially we identify the two radial

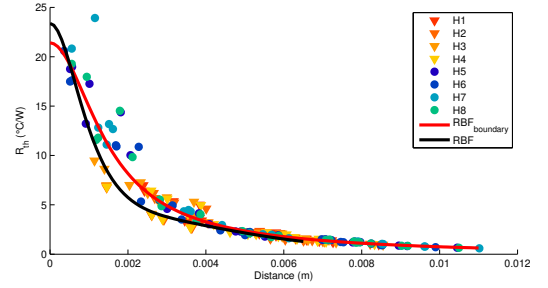


Fig. 18. RBF vs RBF+boundary function comparison

functions by fitting the parameters of the model (2). The first basis function (RBF) is obtained only using the thermal sensors data (white squares) while the second basis function (RBF+boundary) is computed using data points from both thermal sensors and borders temperature (red squares). We have then estimated the fine-grained thermal map using the proposed algorithm described in section V-A and using data points both from thermal sensors and borders temperature. In Figures 19 and 20 we evaluate the border-aware estimation when compared with the standard RBF estimation by means of two error metrics (RMS): e_{total} is the error computed considering all the map points while e_{bound} considers only the points of the map lying on its border. The results clearly show an improvement in accuracy near the points of interest (SoC borders) in the case where we used the RBF+boundary basis function (+35% local accuracy increase). However for the same case the global accuracy (e_{total}) is lower than the approach with the simpler basis function (RBF) (65% global accuracy loss). It follows that, when there is a particular interest in estimating the temperature at specific chip location, the accuracy can be significantly improved by considering supplementary data points in the identification of the thermal basis function. Moreover it is possible to design ad-hoc data points mesh to build a Thermal Basis Function which provides a given accuracy of the estimated thermal map.

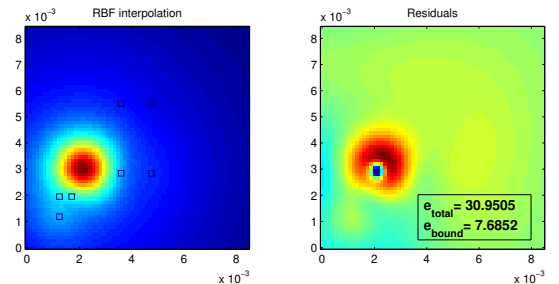


Fig. 19. Normal RBF: estimated thermal map (left) and residuals (right)

D. Final Remarks

From the proposed RBF approach results, the main benefit is introduced by the improved interpolation accuracy in cases with a lower number of available temperature sensors. Although in our tests we used a regular sampling grid, RBF

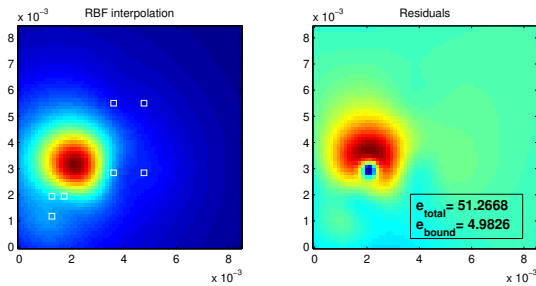


Fig. 20. RBF+boundary: estimated thermal map (left) and residuals (right)

(and Spline) methods do not have any particular restrictions for the position of the input sample points. This simplifies the placement of the thermal sensor in early design stages avoiding additional floorplan constraints. For this reason the RBF algorithm is indicated for optimal sensor allocation frameworks. The study of optimal sensor placement which provides the best accuracy in hotspot temperature detection goes beyond the scope of this work. It must be noted that to have the same flexibility in the Spectral approach it is needed a supplementary step for approximating the irregular sensor grid to a regular one before starting the interpolation procedure. However this step will jeopardize the overall accuracy of the interpolation method. Therefore the RBF approach can find the right trade-off between accuracy, area and design constraints at the cost of an additional post-silicon step needed for the calibration of the Thermal Basis Function for accurate temperature map estimation at run-time.

VII. CONCLUSION

In this paper we presented a novel thermal estimation framework suitable for the new forthcoming 3D-chips. It exploits on chip available thermal sensors and heaters to deliver a detailed full thermal map of the active silicon surface, in addition to model-based virtual thermal and power sensors. The focal point of this work is the Thermal Basis Function, an analytical representation of the thermal properties of the silicon derived empirically from a real 3D device. It correlates the silicon thermal resistance with the distance between two chip locations. Exploiting the on chip power and thermal sensors, the Thermal Basis Function is used directly to create virtual sensors and it is capable of estimating the temperature and power at chip positions not covered by any real sensor. In addition we showed that the same Thermal Basis Function can be used in classical RBF-based interpolation methods. We compared the proposed solution with state-of-the-art approaches. Since the Thermal Basis Function is tightly coupled to the target HW, it gives more accurate results when compared to a more generic basis function (i.e. Spline). Experimental results show that the Thermal Basis Function provides the best trade-off in between accuracy and interpolation cost for limited number of thermal sensors, which is the case of real HW devices.

ACKNOWLEDGMENTS

This work was partially supported by the FP7 ERC Advance project MULTITHERMAN (g.a. 291125), the H2020 FETHPC ExaNoDe (g.a. 671578),

the H2020 FETHPC ANTAREX (g.a. 671623) and by the YINS RTD project (no. 20NA21 150939), evaluated by the Swiss NSF and funded by Nano-Tera.ch with Swiss Confederation financing.

REFERENCES

- [1] A. W. Topol *et al.*, “Three-dimensional integrated circuits,” *IBM Journal of Research and Development*, vol. 50, no. 4.5, p. 491506, 2006.
- [2] E. Beyne, “The rise of the 3rd dimension for system intergration,” in *Interconnect Technology Conference, 2006 International*, 2006, pp. 1–5.
- [3] J. Jeddeloh and B. Keeth, “Hybrid memory cube new DRAM architecture increases density and performance,” in *VLSI Technology (VLSIT), 2012 Symposium on*, June 2012, pp. 87–88.
- [4] J.-S. *et al.*, “A 1.2 v 12.8 GB/s 2 gb mobile wide-I/O DRAM with 4x128 I/Os using TSV based stacking,” *IEEE Journal of Solid-State Circuits*, vol. 47, no. 1, pp. 107–116, Jan. 2012.
- [5] D. Dutoit *et al.*, “A 0.9 pJ/bit, 12.8 GByte/s WideIO memory interface in a 3D-IC NoC-based MPSoC,” in *2013 Symposium on VLSI Technology (VLSIT)*, 2013, pp. C22–C23.
- [6] F. Beneventi *et al.*, “An effective gray-box identification procedure for multicore thermal modeling,” *Computers, IEEE Transactions on*, vol. 63, no. 5, pp. 1097–1110, May 2014.
- [7] J. Burns, “TSV-based 3D integration,” in *Three Dimensional System Integration*, A. Papanikolaou, D. Soudris, and R. Radojcic, Eds. Springer US, 2011, pp. 13–32.
- [8] K. Ganeshpure and S. Kundu, “Reducing temperature variation in 3D integrated circuits using heat pipes,” in *VLSI (ISVLSI), 2012 IEEE Computer Society Annual Symposium on*, Aug 2012, pp. 45–50.
- [9] D. Sekar *et al.*, “A 3D-IC technology with integrated microchannel cooling,” in *Interconnect Technology Conference, 2008. IITC 2008. International*, June 2008, pp. 13–15.
- [10] M. Sabry *et al.*, “GreenCool: An energy-efficient liquid cooling design technique for 3-D MPSoCs via channel width modulation,” *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, vol. 32, no. 4, pp. 524–537, April 2013.
- [11] C. Santos *et al.*, “Using TSVs for thermal mitigation in 3D circuits: Wish and truth,” in *3D Systems Integration Conference (3DIC), 2014 IEEE International*, Dec 2014, pp. 1–8.
- [12] R. Cochran, A. N. Nowroz, and S. Reda, “Post-silicon power characterization using thermal infrared emissions,” in *Proceedings of the 16th ACM/IEEE international symposium on Low power electronics and design*, ser. ISLPED ’10. New York, NY, USA: ACM, 2010, p. 331336.
- [13] I. Arnaldo *et al.*, “Fast and scalable temperature-driven floorplan design in 3D MPSoCs,” in *Test Workshop (LATW), 2012 13th Latin American*, April 2012, pp. 1–6.
- [14] W. Huang *et al.*, “HotSpot: a compact thermal modeling methodology for early-stage VLSI design,” *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 14, no. 5, pp. 501–513, May 2006.
- [15] H. Oprins *et al.*, “Steady state and transient thermal analysis of hot spots in 3D stacked ICs using dedicated test chips,” in *2011 27th Annual IEEE Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM)*, 2011, pp. 131–137, 4.
- [16] C. Santos *et al.*, “Thermal performance of 3D ICs: Analysis and alternatives,” in *3D Systems Integration Conference (3DIC), 2014 IEEE International*, Dec 2014, pp. 1–7.
- [17] R. Cochran and S. Reda, “Spectral techniques for high-resolution thermal characterization with limited sensor data,” in *46th ACM/IEEE Design Automation Conference, 2009. DAC ’09*, 2009, pp. 478–483.
- [18] Y. Zhang, A. Srivastava, and M. Zahran, “Chip level thermal profile estimation using on-chip temperature sensors,” in *IEEE International Conference on Computer Design, 2008. ICCD 2008*, 2008, pp. 432–437, 1.
- [19] J. Kung *et al.*, “Thermal signature: A simple yet accurate thermal index for floorplan optimization,” in *2011 48th ACM/EDAC/IEEE Design Automation Conference (DAC)*, 2011, pp. 108–113, 2.
- [20] M. D. Buhmann, *Radial basis functions theory and implementations*. Cambridge; New York: Cambridge University Press, 2003.
- [21] H. David *et al.*, “RAPL: Memory power estimation and capping,” in *Low-Power Electronics and Design (ISLPED), 2010 ACM/IEEE International Symposium on*, Aug 2010, pp. 189–194.
- [22] Z. Qi *et al.*, “Temperature-to-power mapping,” in *2010 IEEE International Conference on Computer Design (ICCD)*, 2010, pp. 384–389.
- [23] M. Sadri, A. Bartolini, and L. Benini, “Single-chip cloud computer thermal model,” in *Thermal Investigations of ICs and Systems (THERMINIC), 2011 17th International Workshop on*, Sept 2011, pp. 1–6.

- [24] F. Clermidy *et al.*, "3D stacking for multi-core architectures: From WIDEIO to distributed caches," in *2013 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2013, pp. 537–540.
- [25] —, "3D embedded multi-core: Some perspectives," in *Design, Automation Test in Europe Conference Exhibition (DATE)*, 2011, 2011, pp. 1–6.
- [26] F. Liu, "A general framework for spatial correlation modeling in VLSI design," in *Proceedings of the 44th annual Design Automation Conference*, ser. DAC '07. New York, NY, USA: ACM, 2007, p. 817822.
- [27] Y. Zhang, B. Shi, and A. Srivastava, "Statistical framework for designing on-chip thermal sensing infrastructure in nanoscale systems," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 22, no. 2, pp. 270–279, Feb. 2014, 1.
- [28] T. Kemper *et al.*, "Ultrafast temperature profile calculation in ic chips," in *2006 12th International Workshop on Thermal Investigations of ICs and Systems (THERMINIC)*, Sep. 2006.
- [29] V. Heriz *et al.*, "Method of images for the fast calculation of temperature distributions in packaged VLSI chips," in *13th International Workshop on Thermal Investigation of ICs and Systems*, 2007. *THERMINIC 2007*, Sep. 2007, pp. 18–25.
- [30] S. Melamed *et al.*, "Junction-level thermal analysis of 3-D integrated circuits using high definition power blurring," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 31, no. 5, pp. 676–689, May 2012.
- [31] F. Maggioni *et al.*, "Convolution based compact thermal model for 3D-ICs: methodology and accuracy analysis," in *2013 19th International Workshop on Thermal Investigations of ICs and Systems (THERMINIC)*, Sep. 2013, pp. 152–157.
- [32] A. Ziabari and A. Shakouri, "Fast thermal simulations of vertically integrated circuits (3D ICs) including thermal vias," in *Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm)*, 2012 *13th IEEE Intersociety Conference on*, May 2012, pp. 588–596.
- [33] A. Ziabari *et al.*, "Power blurring: Fast static and transient thermal analysis method for packaged integrated circuits and power devices," *Very Large Scale Integration (VLSI) Systems*, *IEEE Transactions on*, vol. 22, no. 11, pp. 2366–2379, Nov 2014.
- [34] S. Sharifi and T. Rosing, "Accurate direct and indirect on-chip temperature sensing for efficient dynamic thermal management," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 29, no. 10, pp. 1586–1599, Oct. 2010, 1.
- [35] A. Zjajo, N. van der Meijs, and R. van Leuken, "Dynamic thermal estimation methodology for high-performance 3-D MPSoC," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. Early Access Online, 2013.
- [36] R.-l. Wang *et al.*, "Surface spline interpolation method for thermal reconstruction with limited sensor data of non-uniform placements," *Journal of Shanghai Jiaotong University (Science)*, vol. 19, no. 1, pp. 65–71, 2014.
- [37] A. Nowroz, G. Woods, and S. Reda, "Power mapping of integrated circuits using AC-Based thermography," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 21, no. 8, pp. 1398–1409, Aug. 2013.
- [38] S. Paek *et al.*, "PowerField: a probabilistic approach for temperature-to-power conversion based on markov random field theory," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 32, no. 10, pp. 1509–1519, 2013.
- [39] J. Ranieri *et al.*, "EigenMaps: algorithms for optimal thermal maps extraction and sensor placement on multicore processors," in *2012 49th ACM/EDAC/IEEE Design Automation Conference (DAC)*, Jun. 2012, pp. 636–641.
- [40] J. Ranieri, A. Chebira, and M. Vetterli, "Near-optimal sensor placement for linear inverse problems," *IEEE Transactions on Signal Processing*, vol. 62, no. 5, pp. 1135–1146, Mar. 2014.
- [41] S. Reda, R. Cochran, and A. Nowroz, "Improved thermal tracking for processors using hard and soft sensor allocation techniques," *IEEE Transactions on Computers*, vol. 60, no. 6, pp. 841–851, Jun. 2011, 1.
- [42] X. Wang *et al.*, "Power trace: An efficient method for extracting the power dissipation profile in an IC chip from its temperature map," *Components and Packaging Technologies*, *IEEE Transactions on*, vol. 32, no. 2, pp. 309–316, June 2009.
- [43] F. Beneventi *et al.*, "Thermal analysis and model identification techniques for a logic + WIDEIO stacked DRAM test chip," in *Design, Automation & Test in Europe Conference & Exhibition, DATE 2014, Dresden, Germany, March 24-28, 2014*, 2014, pp. 1–4.
- [44] C. Bernard and F. Clermidy, "A low-power VLIW processor for 3GPP-LTE complex numbers processing," in *Design, Automation Test in Europe Conference Exhibition (DATE)*, 2011, March 2011, pp. 1–6.
- [45] MATLAB. Natick, Massachusetts: The MathWorks Inc.
- [46] F. Beneventi, A. Bartolini, and L. Benini, "Static thermal model learning for high-performance multicore servers," in *2011 Proceedings of 20th International Conference on Computer Communications and Networks (ICCCN)*, 2011, pp. 1–6.
- [47] C. Torregiani *et al.*, "Compact thermal modeling of hot spots in advanced 3D-stacked ICs," in *Electronics Packaging Technology Conference, 2009. EPTC '09. 11th*, 2009, pp. 131–136.
- [48] J. H. et al., "A 48-core IA-32 message-passing processor with DVFS in 45nm CMOS," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, 2010 *IEEE International*, 2010, pp. 108–109.
- [49] C. AB, "Comsol multiphysics ver. 4.2a," 2011.
- [50] JESD15-4, "Delphi compact thermal model," 2008.

PLACE
PHOTO
HERE

Francesco Beneventi received the engineering degree in Electronic Engineering from the University of Bologna, Italy, in 2010. He is a Research Assistant in the Department of Electrical, Electronic and Information Engineering Guglielmo Marconi (DEI) at the University of Bologna since 2011. His research interests include energy-aware and high performance computing, system identification and thermal modelling of MPSoC.

PLACE
PHOTO
HERE

Andrea Bartolini received a Ph.D. degree in Electrical Engineering from the University of Bologna, Italy, in 2011. He is currently a postdoc researcher in the Integrated Systems Laboratory at ETH Zurich. He also holds a postdoc position in the Department of Electrical, Electronic and Information Engineering Guglielmo Marconi (DEI) at the University of Bologna. His research interests concern green computing and dynamic resource management ranging from embedded to large scale HPC systems with special emphasis on thermal and power-aware HW/SW co-design techniques.

PLACE
PHOTO
HERE

Pascal Vivet received the M.E. degree from UJF, Grenoble, France, in 1994, and the Ph.D. degree within France Telecom Lab, Grenoble, France, in 2001. After four years with STMicroelectronics, Pascal Vivet has joined CEA-Leti in 2003 in the advanced design department of the Center for Innovation in Micro and Nanotechnology (MINATEC), Grenoble, France. His topics of interests are covering wide aspects from system level design and modeling, to asynchronous design, Network-on-Chip architecture, low power design, many core architectures, 3D design, including interactions with related CAD aspects. Dr Pascal Vivet is the author or co-author of a couple of patents and of more than 50 papers.

PLACE
PHOTO
HERE

Luca Benini Full Professor at the University of Bologna and he is the chair of Digital Circuits and Systems at ETHZ. He has served as Chief Architect for the Platform2012/STHORM project in STMicroelectronics, Grenoble in the period 2009-2013. Dr. Benini's research interests are in energy-efficient system design and Multi-Core SoC design. He is also active in the area of energy-efficient smart sensors and sensor networks for biomedical and ambient intelligence applications. He has published more than 700 papers in peer-reviewed international journals and conferences, four books and several book chapters.