A Design Framework for Thermal-Aware Power Delivery Network in 3D MPSoCs with Integrated Flow Cell Arrays

Halima Najibi, Alexandre Levisse, Marina Zapater
Embedded Systems Laboratory (ESL), EPFL, Lausanne, Switzerland

***Abstract*—Integrated Flow Cell Array (FCA) technology promises to address the power delivery and heat dissipation challenges in three-dimensional Multi-Processor Systems-on-Chips (3D MPSoCs) by providing combined inter-tier liquid cooling and power generation capabilities. In this paper, we present for the first time a design framework to accurately model the temperature-aware power delivery network in 3D MPSoCs, and quantify the effects of FCAs on the voltage drop (IR-drop). This framework estimates the power generation variation along FCAs due to voltage and temperature, in the case of uniform and non-uniform powermaps from several real processor traces. Furthermore, we explore different 3D MPSoC configurations to quantify their power delivery requirements. Our results show that FCAs improve the IR-drop with respect to state-of-the-art design methods up to 53% and 30% for dies with a power consumption of 60W and 190W, respectively, while maintaining their peak temperatures below 52°C, and at no additional Through Silicon Via (TSV) area overhead. In addition, as the presence of high power density regions (hotspots) can decrease the FCAs IR-drop reduction by up to 21% with respect to the average value, we present a scalable TSV placement optimization methodology using the proposed framework. This methodology minimizes the IR-drop at hotspots and guarantees an optimal and uniform exploitation of the IR-drop reduction benefits of FCAs.**

***Index Terms*—3D multi-processor systems-on-chip, integrated flow cell arrays, liquid cooling, on-chip power generation, 3D power delivery network design**

I. Introduction

In the recent years, 3D integration became a promising solution to overcome scaling limitations of Multi-Processor Systems-on-Chip (MPSoCs) [1], by enabling important improvements in chip area, power, heterogeneity in terms of technology and architecture, and communication bandwidth [2] [3]. However, 3D MPSoC design faces two major challenges: heat extraction and power delivery [2]. The heat removal difficulty fundamentally increases as heat traverses several layers of low thermal conductivity to reach the sink. Thus, the dies reach high temperatures that increase their leakage power [4]. 3D MPSoC power delivery is also problematic as the supply voltage (V_{DD}) decreases before reaching the gates, due to the resistivity of the delivery network. This phenomenon, referred to as IR-drop, manifests with local V_{DD} drops that result in slower transitions of the affected gates and lead to timing violations [5]. In this context, many techniques have been proposed to improve heat dissipation and minimize IR-drop in 3D MPSoCs, by increasing the number of thermal [6] and Power and Ground (P/G) TSVs [5], respectively. However, these techniques generally involve a high TSV area overhead, which makes their implementation impractical and costly.

This work has been supported by the ERC Consolidator Grant COMPUSAPIEN (GA No. 725657)

978-1-7281-2954-9/19/$31.00 ©2019 IEEE

Integrated Flow Cell Array (FCA) technology addresses the aforementioned challenges by providing combined on-chip liquid cooling and power generation in 3D MPSoCs [7]. The FCAs consist of micro-channels, filled with an electrolytic solution, and etched in the silicon substrates of the dies. Their capabilities were experimentally validated in [7], showing that they can generate up to $3.6W$ per cm^2 under optimal voltage and load conditions. However, previous analyses used a simple analytic model that did not cover the 3D MPSoC design aspects, such as the connectivity of the FCAs to 3D power delivery networks, or the power generation variations between different flow cells with respect to temperature and voltage levels across the dies.

In this paper, we propose a thermal-aware 3D power delivery network design framework to quantitatively analyze the effects of connecting FCAs to 3D MPSoC P/G grids. The framework uses real powermaps to extract the load sinks, estimate the power grid densities of the dies, and calculate the FCA temperature maps using a thermal analysis tool [8]. It then builds a fine-grained electrical model of the 3D P/G network, and calculates the voltages across the dies. The proposed framework can be used in early 3D MPSoC design stages to meet specific power and performance requirements. The contributions of the paper are listed in the following:

- We introduce for the first time a fine-grained 3D P/G network model to quantify the effects of FCAs on 3D MPSoC power delivery. We show that, compared to state-of-the-art methods, FCAs reduce the IR-drop by up to 53% for dies with uniform power consumption, and up to 30% for dies containing high power density regions (power hotspots), at no additional TSV area overhead.
- We analyze the variation of current generation along FCAs due to temperature and voltage, using multiple real processor power traces. We show that, while the average IR-drop reduction increases by 10%, at the power hotspots it remains 21% below the average value. Therefore, it is key to have an accurate P/G delivery network modeling approach for 3D MPSoC design, as opposed to qualitative profiling approaches.
- Finally, we propose an optimization algorithm that uses the proposed modeling methodology to find the best placement of P/G TSVs in 3D MPSoCs. Our algorithm ensures a minimal IR-drop and an optimal FCA IR-drop reduction across dies. Moreover, it runs in under 4 minutes per 3D MPSoC sub-grid and per TSVs placement, for our case study based on the POWER8 processor with 16 P/G sub-grids of 3362 nodes and 15 TSV placement options. This result proves the scalability of our optimization algorithm for large 3D MPSoCs.

II. Related Work

A. 3D integration trends and challenges

3D integration using TSV technology promises to overcome IC scaling challenges, by offering opportunities to increase the integration density, enable mixed technologies and architectures, and minimize the delays and power dissipation of interconnects. 3D integration was successfully adopted in the memory industry, with the High Bandwidth Memory (HBM) [9] and the Hybrid Memory Cube (HMC) [10]. Both achieve breakthrough bandwidths (over 256GB/s and 320GB/s, respectively), while consuming a lower energy-per-bit compared to 2D DRAMs. Following the emergence of 3D memories, efforts were directed towards memory-on-processor architectures to speed up memory-intensive applications and lower the communication power consumption [11] [3]. However, these 3D MPSoCs face two major challenges: heat extraction difficulty and voltage distribution complexity [12], particularly due to the non-uniformity of power consumption.

In this context, methodologies were proposed to manage the thermal aspect of 3D MPSoCs in early floorplanning stages [13]. In particular, the insertion of thermal TSVs improves the heat extraction in 3D MPSoCs [6]. Nonetheless, using thermal TSVs has a high area overhead of 47% for a maximal temperature drop of only 38%. Moreover, researchers proposed a technique to co-optimize the locations, sizes and number of P/G TSVs as well as the floorplan of major 3D MPSoC blocks, in order to prevent IR-drop constraints violations [5]. This method recovers up to 11.8% of IR-drop, but uses a high number of TSVs (up to 200 TSVs for a grid of 300 nodes) of large diameter (20μ m). In this work, we accurately model and exploit FCA technology to address the above challenges. We achieve up to 4% and 42% better results than state-of-the-art methods (for the temperature and IR-drop reduction, respectively), at no additional TSV cost.

B. Power delivery grid modeling

Power delivery analysis is a critical step in the early design and floorplanning of high-performance MPSoCs. As the interconnect dimensions shrink and the currents traversing them increase, the IR-drop becomes more critical [5]. Therefore, it is crucial to accurately model the P/G grid in order to ensure a correct chip operation. A typical P/G grid consists of intersecting horizontal and vertical metal lines dedicated to voltage delivery [14]. Then, the lines are interconnected with each other and to the gates using vertical vias. Each element has specific electrical resistivity and dimensions, which allows to estimate the resistances between the nodes and to the sources. This model is commonly used for P/G delivery network design and analysis [15] [16]. In this work, we extend it by integrating the FCAs as additional power supply components.

C. Integrated FCA technology

FCA technology consists of inter-tier micro-channels that are placed in the silicon substrate of 3D MPSoC dies, and where an electrolytic solution flows. They absorb the heat generated in the chip, thus serving as on-chip liquid cooling. At the same time, a continuous and stable electrical current

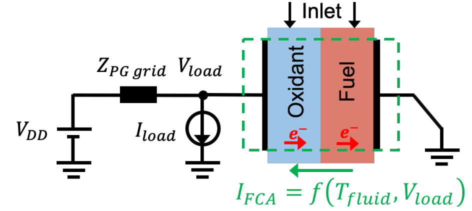

Fig. 1: A single flow cell connected to the load and P/G grid

flows when a load is inserted between FCA electrodes, due to electro-chemical reactions between the electrolytes (Figure 1). The rate of these reactions increases with temperature, hence generating more current, and allowing the FCAs to effectively transform heat into power to partially supply the logic gates.

The PowerCool tool was recently introduced in [7] to perform electro-thermal simulations of 3D MPSoCs with integrated FCAs. It is a combination of the original PowerCool simulator [17], which enables compact electro-chemical simulation of FCA power generation, and the 3D-ICE simulator [8], which analyzes the FCA cooling capabilities. The accuracies of PowerCool and 3D-ICE were validated against fine-grained physics models and real measurements, respectively. The authors of [7] showed that FCAs can generate up to $3.6W$ per cm^2, but considered optimal load and voltage conditions without taking into account the impact of connecting the FCAs to the 3D power delivery network. Moreover, they did not study the current generation variations along FCAs due to voltage and temperature, hence missed to evaluate the effects of powermap non-uniformities. We thereby accurately model in this work the 3D MPSoC P/G delivery networks with integrated FCAs, and evaluate the effects of hotspots on the FCAs IR-drop reduction capabilities.

III. Design framework description

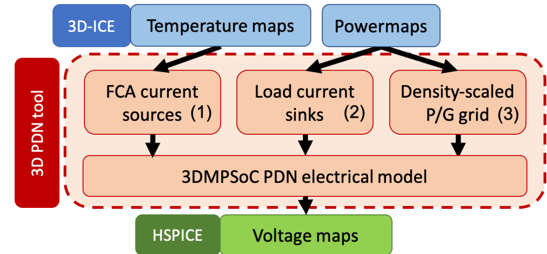

Fig. 2: Thermal-aware power delivery network design framework

To efficiently design 3D MPoCs with integrated FCAs, we propose a framework that comprehensively evaluates their power generation performance, and analyzes their effects on the voltage delivery across dies. The framework is summarized in Figure 2, where the proposed power delivery network (PDN) modeling tool is shown in red, its inputs in blue and its output in green. The framework builds a fine-grained electrical model of the 3D PDN using the modeling tool. The FCAs are included in this model as additional voltage and temperature-controlled current sources, and connected to the P/G grids due to their stability [7] [18]. Next, the P/G grids and TSVs are modelled as a mesh of resistances that are estimated based on the power densities at the nodes. Finally, a DC analysis is performed to evaluate the IR-drop across the 3D MPSoC. The following subsections describe the methodologies for

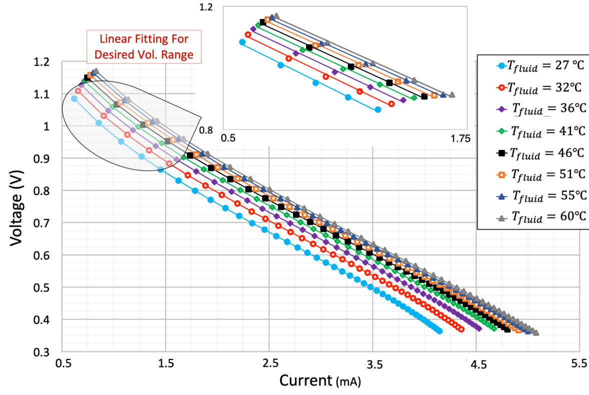

Fig. 3: IV curve for a single $\mathbf{200\mu m}$-long flow cell

the FCAs and the 3D PDN modeling, and for the network optimization.

A. flow cell electrical model

The framework models each flow cell as a voltage and temperature-controlled current source, and connects its electrodes to nearby V_{DD} and V_{GND} metal lines in the P/G grid. Then, the voltage between the metal lines generates a continuous and stable current between the flow cell electrodes, which is directly supplied to the gates.

We use PowerCool [7] to assess the current-voltage relation (IV curve) between the flow cell electrodes, and analyze the effect of the fluid temperature on it. Figure 3 shows the IV curve of a single flow cell of length $200\mu m$, for different fluid temperatures ranging from 27°C and 60°C. The selected length minimizes the temperature variation inside the flow cell (< 0.5°C) and the number of flow cells per channel (112 cells). The IV curve is not linear in the full analysis range, but it has a linear behaviour around 1V (typical V_{DD} value for sub-30nm technologies). Furthermore, the slope of the linear relationship in this range is independent of the temperature, which only affects the offset. Following this analysis, we model the flow cell at position (i, j) using equation (1), where $V_{FCA,i,j}$ and $I_{FCA,i,j}$ are the voltage and current between the flow cell electrodes, R_{FCA} is the FCA resistance and $V_{offset}(T_{fluid,i,j})$ is the offset corresponding to fluid temperature $T_{fluid,i,j}$ (extracted from the temperature map):

$$V_{FCA,i,j} = V_{offset}(T_{fluid,i,j}) - R_{FCA} I_{FCA,i,j} \qquad (1)$$

B. 3D power delivery network model

We model the 3D MPSoC power delivery network including P/G TSVs, 2D P/G grids, and connected FCAs (Section III-A). The 3D MPSoC layers are partitioned into nodes representing cells with uniform power-per-surface-units. These nodes are connected horizontally via the equivalent resistances of 2D P/G grids, and vertically by the equivalent resistances of TSVs. Each flow cell is connected to the node it supplies. Consequently, we construct a network for each layer, as shown in Figure 4, with the following elements:

- Current sources representing the flow cells (Box (1) in Figure 2), controlled by the voltage and temperature at the corresponding nodes, according to Section III-A.

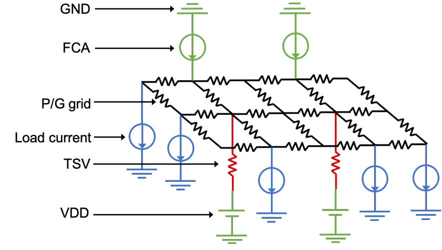

Fig. 4: Electrical model of a 3D MPSoC layer

- Load current sources, which are extracted from the powermaps and the operating voltages at the corresponding nodes (Box (2) in Figure 2).
- TSVs resistances, which are a function of the TSV length L_{TSV}, radius r_{TSV}, and the copper resistivity ρ_{Cu} [5]:

$$R_{TSV} = \frac{\rho_{Cu}\, L_{TSV}}{\pi\, {r_{TSV}}^2} \qquad (2)$$

- 2D grid resistances, estimated as the sum of the resistances of the wires and vias in the metal layers dedicated to voltage delivery [15]. The resistance of a wire between two nodes in a metal layer M_i is computed as follows:

$$R_{M_i} = p_{usage}\, R_{sM_i} \frac{L_{wire}}{W_{wire}} \qquad (3)$$

$$\text{with:} \begin{cases} p_{usage}\text{: power grid usage , } W_{wire}\text{: wire width} \\ R_{sM_i}\text{: } M_i \text{ sheet resistance, } L_{wire}\text{: wire length} \end{cases}$$

The power grid usage refers to the percentage of available P/G lines and vias. It is scaled to the power density, as the locations and paths between the gates are unknown at pre-layout design stages [15] (Box (3) in Figure 2).

The scalability of the 3D PDN model depends on several parameters: cell size, TSV number, FCA size, pitch, etc. However, it is dominated by the cell size which determines the number of unknowns (i.e the nodes), and thus impacts the size of the mesh of resistances. The complexity of the model is *O(n)*, where *n* is the number of nodes.

C. Power delivery network optimization

To control the effects of power hotspots on the voltage delivery in 3D MPSoCs with integrated FCAs, we use the 3D thermal-aware P/G network modeling and analysis methodology described in III-B to optimize the placement of TSVs, and fully exploit the FCAs IR-drop reduction capabilities. In this context, we propose the optimization Algorithm 1. This algorithm takes as input a set of 3D MPSoC nodes N, and a predefined set of TSVs positions P, as opposed to an exhaustive search, which is unfeasible for large MPSoCs. Our algorithm searches for the optimal position p_{opt} in P, considering a quality metric Q defined to evaluate the power delivery network (e.g., the maximum IR-drop variation with respect to the constraint value ΔV_{ref}). For each position p in P, we build the corresponding power delivery network (PDN) model (line 3 in Algorithm 1), considering the temperature dependency of FCA currents. Then, we perform a DC analysis to extract the voltages $[V_n]_{n \in N}$ (line 4 in Algorithm 1) and calculate the IR-drops $[\Delta V_n]_{n \in N}$ at the nodes, with respect to their operating voltages $[V_{op,n}]_{n \in N}$ (lines 5-7 in Algorithm 1). Finally, we calculate the quality score Q_p at position p (line 8 in Algorithm 1). We select the optimal position p_{opt} with the best quality score Q_{opt} (lines 9-12 in Algorithm 1).

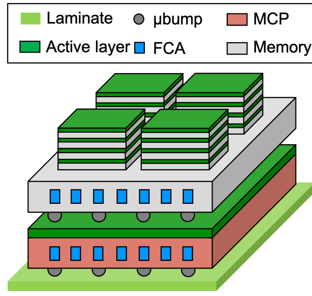

Fig. 5: 3D MPSoC (B6)

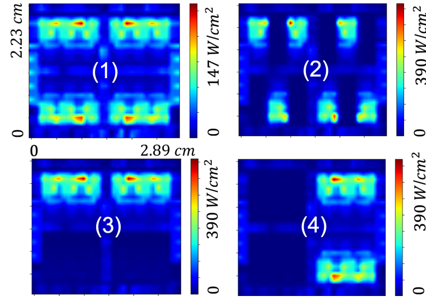

Fig. 6: Non-uniform POWER8 powermaps

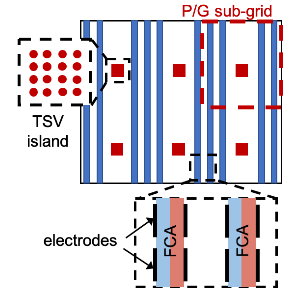

Fig. 7: TSVs and FCA placement

Algorithm 1 PDN optimization

1: Initialize Q_{opt} and p_{opt}
2: **for** p in P **do**
3: Build thermal-aware PDN model (cf. Section III-B)
4: Run HSPICE DC analysis and calculate the voltage V_n for each $n \in N$
5: **for** n in N **do**
6: Calculate the IR-drop $\Delta V_n = V_{op,n} - V_n$
7: **end for**
8: Calculate the quality score $Q_p = \frac{\max_{n\in N} \Delta V_n - \Delta V_{ref}}{\Delta V_{ref}}$
9: **if** $Q_p < Q_{opt}$ **then**
10: $Q_{opt} = Q_p$
11: $p_{opt} = p$
12: **end if**
13: **end for**

IV. Experimental Setup

A. 3D MPSoC and 3D power delivery network description

To test the proposed 3D MPSoC thermal-aware power delivery design framework, we used the stack composition from [7]. Figure 5 presents an overview of the 3D MPSoC. It contains a Multi-Core Processor (MCP) with the IBM POWER8 architecture and a layer with four DRAM memories, each with the profile of a second generation 4GB 4-layer HBM [9]. For each HBM, we use a uniform powermap with a total power consumption of $15W$ (i.e., assuming a uniform memory access pattern). For the MCP, we use four different non-uniform powermaps (cf. Figure 6). These powermaps correspond to real power traces of the POWER8 processor, and have a total consumption of $190W$. In powermap (1), all cores are fully loaded with the maximal nominal frequency, whereas in (2), (3) and (4), half of the cores are in Performance Boost mode while the other half are idle.

TABLE I 3D MPSoC dimensions

	Width (mm)	Length (mm)	Height (μm)	Pitch (μm)
Die	28.9	22.3	$6^{(1)}+2^{(2)}+120^{(3)}$	-
			$6^{(1)(*)}+2^{(2)(*)}+48^{(3)(*)}$	-
FCA	0.05	22.3	100	100
TSV	0.005		120, $48^{(*)}$	10

(1) BEOL (2) Active (3) Bulk (*) w/o FCA

The dies are partitioned into nodes representing areas down to $40\mu m \times 40\mu m$. Figure 7 shows the placement of the FCAs and P/G TSVs. Each die has regions, which we call "TSV islands", where the P/G TSVs are placed, including the TSVs powering the die itself and those traversing it to power the above die. There are 16 TSV islands in total, each of them delivers V_{DD} to the sub-grid that supplies its corresponding area. FCAs are etched in the remaining available area, considering that sufficient signal TSVs are placed to fulfill the communication bandwidth needs of the system. Table I shows the dimensions of the dies, TSVs and FCAs.

B. Design space exploration

We explore different 3D MPSoC configuration possibilities in terms of die and FCA placement. The same configurations were analyzed in [7] to qualitatively score their FCA heat-removal efficiency and power interconnect requirements, under optimal load conditions. Therefore, this analysis did not consider the FCAs connectivity to the P/G network, nor did it model their local power generation variation due to the non-uniformity and non-optimality of the voltage and temperature. Therefore, we propose to quantitatively analyze these 3D stacks using our framework.

Figure 8 presents the simulated stack configurations. Notations are the same as in [7] for categories A and B. In category A, FCAs are only etched under the memory. In category B, FCAs are present under both the processor and the memory. In category C, FCAs are only under the processor. For all the above categories, we explored different arrangements and orientations of the dies.

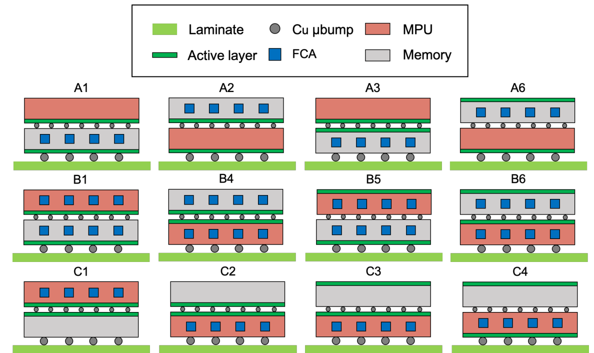

Fig. 8: 3D MPSoC stack configurations

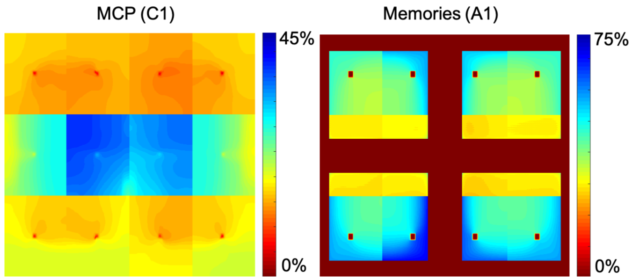

Fig. 9: IR-drop reduction maps for the MCP (non-uniform powermap) and memories (uniform powermaps)

C. Experimental flow

We perform a thermal analysis of the 3D MPSoCs using 3D-ICE [8] to extract the temperatures of the fluid inside the micro-channels, which drive the flow cells current generation (Section III-A). Then, we model the load currents and P/G grids using the dies powermaps (Section III-B). FCAs are added using a Verilog-A module, of which the behaviour description corresponds to Equation 1 in Section III-A. The temperature-dependent voltage offsets $V_{offset}(T_{fluid})$ are set by interpolating between simulated fluid temperatures, ranging from 27°C to 60°C. Then, we perform a DC analysis using Synopsis HSPICE, to extract the voltages at the 3D MPSoC nodes. And finally, we calculate the percentage of recovered IR-drop with respect to the case with no FCAs.

V. Experimental Results

A. Evaluation of IR-drop reduction using FCAs

Using the proposed framework in this work, we derive fine-grained temperature and voltage maps for 3D MPSoC dies. Figure 9 shows the IR-drop reduction percentage at the nodes of the MCP in configuration C1 and the memories in configuration A1, both correspond to powermap (1) in Figure 6. Figure 10 presents the average and standard deviation of the IR-drop reduction due to FCAs, calculated between the die nodes. Our results show that FCAs enable an IR-drop reduction on average from 50 to 53% for the memories, and from 23 to 30% for the MCP, with a standard deviation up to 14% and 18%, respectively. The IR-drop reduction is higher for the memories, as they consume in total 3× less power than the MCP, so the FCAs provide a more significant portion of the current needed to drive them. In general, FCAs enable up to 42% higher IR-drop reduction than the maximum achieved in [5]. They also keep temperatures under 52°C (versus 90°C without FCAs), which is 4% more than the maximum in [6]. Both the IR-drop reduction and cooling are achieved with no TSV area overhead compared to [5] and [6].

Furthermore, Table II presents a summary of the results for the stack categories A, B and C. In particular, it presents the percentage of total FCA-generated current with respect to the current consumed by the MCP and the memories. In general, the FCAs provide a significant amount of current to drive the memories and the MCP, up to 52% and 19%, respectively. The FCAs current generation is highest for powermap (3) where the MCP has the highest hotspots, thus generates more heat that increases the current between the FCAs electrodes (values in bold).

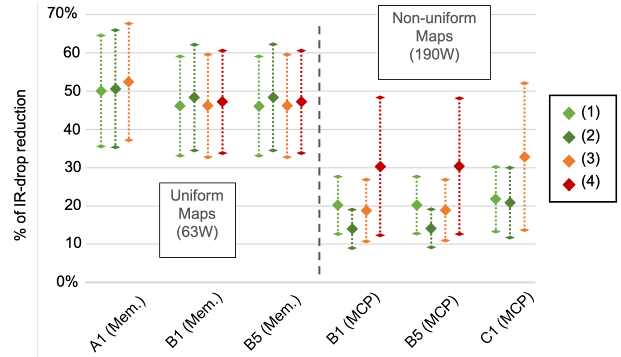

Fig. 10: Average and standard deviation of IR-drop reduction for powermaps (1), (2), (3) and (4)

In addition, Table II presents the percentage of IR-drop reduction at the most critical node (i.e., the node with the maximal IR-drop). This table shows that the IR-drop reduction is proportional to the current generation in the memories, which have a uniform power distribution, unlike the MCP that has hotspots. The percentage of IR-drop reduction in the most critical node of the MCP decreases up to 5% from powermap (1) to (4), which have increasing concentrations of hotspots (values in italic bold in Table II). Conversely, Figure 10 shows that the average IR-drop reduction in the MCP increases between powermaps (1) to (4) up to 10% with a difference of 21% from the most critical node. This observation indicates that for a non-uniform powermap, less critical nodes can better benefit from the FCAs in a non-optimized P/G grid, which highlights the necessity of the proposed analysis in 3D MPSoC power delivery network planning.

Finally, our results invalidate the quantitative profiling of some 3D MPSoC configurations presented in [7]. For example, configuration B1 was rated with a much higher score compared to A1 in terms of P/G interconnect requirements. However, our results show it is not necessarily the case, especially for powermaps with more critical hotspots (cf. Table II). A similar observation can be made between configuration B4 and A3 proposed in [7], which are not specifically shown in this paper due to space limitations. These observations prove that a qualitative analysis, considering an optimal FCA power generation, is not sufficient for 3D MPSoC design with FCAs.

B. TSV placement for optimal IR-drop reduction with FCAs

We tested the power delivery network optimization algorithm in Section III-C using the MCP in stack B1 with

TABLE II FCA Simulation results

		A	B		C
	Powermap	Mem.	Mem.	MCP	MCP
% FCA current	(1)	49	45	16-17	18
	(2)	50-52	46-47	16	17-18
	(3)	**52**	46-47	17	**18-19**
	(4)	-	45	16-17	-
% IR-drop reduction at critical node	(1)	35-36	34-35	12-13	13-14
	(2)	37	35	11-12	12
	(3)	37	35	***9***	***9***-10
	(4)	-	33-34	***9***	-

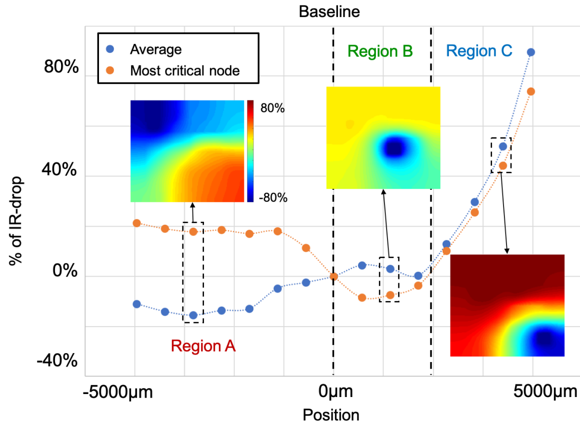

Fig. 11: IR-drop variation with respect to the baseline

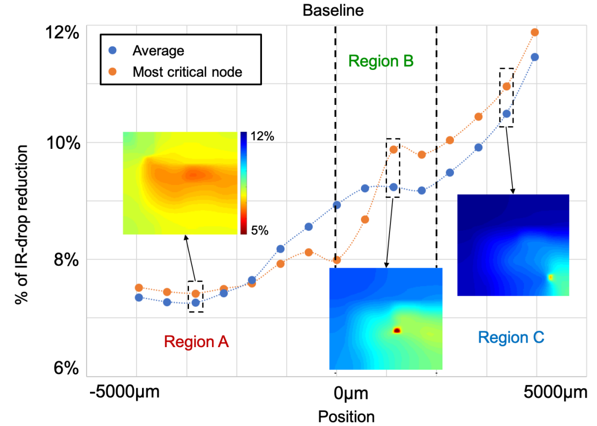

Fig. 12: IR-drop reduction with FCAs

powermap (4), which is the most critical case in the previous results. We selected among the 16 MCP sub-grids, the one with highest power hotspots. Then, we tested 15 different placements of the TSV island, along the diagonal axis from the least power-consuming to the most active region (hotspots). We ran the optimization algorithm considering the quality metric Q, which represents the IR-drop variation with respect to the baseline position (i.e., TSVs at the center). Additionally, we defined a new quality metric Q^*, which represents the FCAs IR-drop reduction with respect to the case with no FCAs. Our algorithm runs in under 4 minutes per position of TSVs, for a sub-grid of 3363 nodes, each representing an area of $40\mu m \times 40\mu m$. This result proves the scalability of our proposed algorithm for larger 3D MPSoCs.

Figure 11 shows the values of Q for different TSV island positions, and Figure 12 shows the values of Q^*. Our results show that in Region A, where TSVs are farthest from the hotspots, the IR-drop reduction is uniform but low on average. Moreover, the IR-drop at hotspots is highest (up to 21% more than the baseline). In Region B, the overall IR-drop decreases and the IR-drop reduction increases for all nodes. Finally in Region C, the IR-drop reduction is the highest and most uniform. However, while the IR-drop decreases at hotspots, it significantly increases at less power consuming areas (over 80% more than the baseline). This effect switches the location of critical nodes to parts of the die that are farther away from hotspots. Thus, the best placement of TSV islands, according to Algorithm 1, and considering both metrics Q and Q^*, is at the intersection between Regions B and C. This placement results in an optimal IR-drop for all the nodes of the MCP sub-grid, and an optimal use of FCA current generation benefits.

VI. Conclusion

In this paper, we have presented a new design framework for thermal-aware power delivery in 3D MPSoCs with integrated FCAs, which can be used at early design stages to achieve optimal power delivery. The framework constructs a fine-grained 3D network model including current generation variations along FCAs, and estimates the IR-drop across dies. We proved that FCA integration permits to reduce the IR-drop by up to 53%, while maintaining temperatures below 52°C. We also showed that hotspots in non-uniform powermaps increase the IR-drop on average up to 10%, but reduce the IR-drop at critical nodes up to 5%. Finally, we proposed a fast and scalable optimization methodology to find the placement of TSVs that best manages power hotspots, and guarantees full exploitation of FCA IR-drop reduction potentials. This work motivates the design and optimization of power delivery networks to enable a high-efficiency FCA power generation under different operation voltages, with a low FCA power supply switching noise.

References

[1] R. Brain. Interconnect scaling: Challenges and opportunities. *IEEE International Electron Devices Meeting*, 2016.

[2] S. Bokar. 3D Integration for Energy Efficient System Design. *Design Automation Conference (DAC)*, 2011.

[3] D.H. Woo et al. An Optimized 3D-Stacked Memory Architecture by Exploiting Excessive, High-Density TSV Bandwidth. *International Symposium on High-Performance Computer Architecture*, 2010.

[4] P. Zhou et al. 3D-STAF: scalable temperature and leakage aware floorplanning for three-dimensional integrated circuits. *International Conference On Computer Aided Design (ICCAD)*, 2007.

[5] S. Wang et al. P/G TSV Planning for IR-drop Reduction in 3D-ICs. *Design, Automation and Test in Europe (DATE)*, 2014.

[6] E. Wong et al. 3D Floorplanning with Thermal Vias. *DATE*, 2006.

[7] A.A. Andreev et al. PowerCool: Simulation of Cooling and Powering of 3D MPSoCs with Integrated Flow Cell Arrays. *IEEE Transactions on Computers*, 2018.

[8] A. Sridhar et al. 3D-ICE: a compact thermal model for early-stage design of liquid-cooled ICs. *IEEE Transactions on Computers*, 2014.

[9] D. Lee et al. A 1.2 V 8 Gb 8-Channel 128 GB/s High-Bandwidth Memory (HBM) Stacked DRAM With Effective I/O Test Circuits. *Journal of Solid-State Circuits*, 2015.

[10] J. Jeddeloh et al. Hybrid Memory Cube New DRAM Architecture Increases Density and Performance. *IEEE Symp. on VLSI Tech.*, 2012.

[11] M. Daneshtalab et al. Memory-Efficient Logic Layer Communication Platform for 3D-Stacked Memory-on-Processor Architectures. *International 3D Systems Integration Conference*, 2011.

[12] D.H. Kim et al. Physical Design and CAD Tools for 3-D Integrated Circuits: Challenges and Opportunities. *IEEE Design and Test*, 2015.

[13] J.H. Lau et al. Thermal Management of 3D IC Integration with TSV. *IEEE Electronic Components and Technology Conference*, 2009.

[14] M. Popovich et al. *Power Distribution Networks with On-Chip Decoupling Capacitors*. 2008.

[15] S. Nassif et al. Fast Power Grid Simulation. *DAC*, 2000.

[16] A. Zou et al. Efficient and Reliable Power Delivery in Voltage-Stacked Manycore System with Hybrid Charge-Recycling Regulators. *DAC*, 2018.

[17] A. Sridhar et al. PowerCool: Simulation of integrated microfluidic power generation in bright silicon MPSoCs. *ICCAD*, 2014.

[18] A. Andreev et al. Design Optimization of 3D Multi-Processor System-on-Chip with Integrated Flow Cell Arrays. *International Symposium of Low Power electronics and Design (ISLPED)*, 2018.