Towards Thermally-Aware Design of 3D MPSoCs with Inter-Tier Cooling

Mohamed M. Sabry[†], Arvind Sridhar[†], David Atienza[†], Yuksel Temiz[‡], Yusuf Leblebici[‡], Sylwia Szczukiewicz[§], Navid Borhani[§], John R. Thome[§], Thomas Brunschwiler[¶], and Bruno Michel[¶]
[†]Embedded Systems Laboratory (ESL), Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland.
[‡]Microelectronic Systems Laboratory (LSM), Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland.
[§]Laboratory of Heat and Mass Transfer (LTCM), Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland.
[¶]Avanced Thermal Packaging, IBM Zurich, Switzerland.

***Abstract*— New tendencies envisage 3D Multi-Processor System-On-Chip (MPSoC) design as a promising solution to keep increasing the performance of the next-generation high-performance computing (HPC) systems. However, as the power density of HPC systems increases with the arrival of 3D MPSoCs, supplying electrical power to the computing equipment and constantly removing the generated heat is rapidly becoming the dominant cost in any HPC facility. Thus, both power and thermal/cooling implications play a major role in the design of new HPC systems, given the energy constraints in our society. Therefore, EPFL, IBM and ETHZ have been working within the CMOSAIC Nano-Tera.ch program project in the last three years on the development of a holistic thermally-aware design.**

This paper presents the exploration in CMOSAIC of novel cooling technologies, as well as suitable thermal modeling and system-level design methods, which are all necessary to develop 3D MPSoCs with inter-tier liquid cooling systems. As a result, we develop energy-efficient run-time thermal control strategies to achieve energy-efficient cooling mechanisms to compress almost 1 Tera nano-sized functional units into one cubic centimeter with a 10 to 100 fold higher connectivity than otherwise possible. The proposed thermally-aware design paradigm includes exploring the synergies of hardware-, software- and mechanical-based thermal control techniques as a fundamental step to design 3D MPSoCs for HPC systems. More precisely, we target the use of inter-tier coolants ranging from liquid water and two-phase refrigerants to novel engineered environmentally friendly nano-fluids, as well as using specifically designed micro-channel arrangements, in combination with the use of dynamic thermal management at system-level to tune the flow rate of the coolant in each micro-channel to achieve thermally-balanced 3D-ICs. Our management strategy prevents the system from surpassing the given threshold temperature while achieving up to 67% reduction in cooling energy and up to 30% reduction in system-level energy in comparison to setting the flow rate at the maximum value to handle the worst-case temperature.

I. Introduction

The power density of high performance systems continues to increase with every process technology generation [11]. Thus, it is not possible to design the next-generation of supercomputers using the existing solutions for high-performance computing (HPC) systems. As the density of supercomputers increases with the arrival of chip multi-core and multi-threaded processors, supplying electrical power to the computing equipment while simultaneously removing the generated heat is rapidly becoming the dominant cost in any HPC facility, even further than the new server deployment costs. Thus, both power and thermal/cooling implications are increasingly playing a major role in the design of new HPC systems, especially given the current energy constraints in our society.

In addition, 3D integration [3] (i.e., multiple layers of processors, memories, etc. in the same chip stack) is a recently proposed design method for overcoming the limitations with respect to delay, bandwidth, and power consumption of the interconnects in large multi-processor system-on-chip (MPSoC) chips, while reducing the chip footprint and improving the fabrication yield. However, one of the main challenges for designing 3D circuits is the elevated temperatures resulting from higher thermal resistivity [12], [14], which irregularly spread in the 3D chip stack. Hence, it is more difficult to remove the heat from 3D systems with respect to conventional 2D MPSoCs. Temperature-induced problems are exacerbated in 3D stacking and are a major concern to be addressed and managed as early as possible in 3D MPSoC design. Therefore, innovative cooling strategies must be developed for 3D-based HPC systems.

Indeed, conventional back-side heat removal strategies, such as, air-cooled heat sinks and micro-channel cold-plates only scale with the die size and are insufficient to cool 3D MPSoC with hot spot heat fluxes up to $250W/cm^2$, as expected in forthcoming 3D MPSoC stacks [6]. On the contrary, inter-tier single- and two-phase liquid cooling is a potential solution to address the high temperatures in 3D MPSoCs, due to the higher heat removal capability of liquids in comparison to air [4]. However, no consistent design methodology has been proposed until now to develop 3D MPSoC stacks with the necessary inter-tier liquid cooling technology being integrated.

In this paper, we present the exploration of novel cooling technologies, as well suitable thermal modeling and system-level design methods, which are all necessary to develop 3D MPSoCs with inter-tier liquid cooling. On the one hand, the scalable liquid cooling technology being developed by the CMOSAIC Nano-Tera.ch RTD project involves injecting

The authors would like to thank Prof. Ayse K. Coskun for her support on this work, and acknowledge CMI staff, in particular Dr. Cyrille Hibert and Dr. Jean-Baptiste Bureau, for their valuable support in microfabrication processes. This research is partially funded by the Nano-Tera RTD project CMOSAIC (ref.123618), financed by the Swiss Confederation and scientifically evaluated by SNSF, and the PRO3D EU FP7-ICT-248776 project. 978-3-9810801-7-9/DATE11/ ©2011 EDAA

water, or an evaporating refrigerant, through micro-channels between the tiers of a 3D stack. On the other hand, the inclusion of such cooling technology that includes coolants ranging from liquid water and two-phase to novel engineered environmentally friendly nano-fluids is not sufficient to deploy energy-efficient HPC systems/architectures (i.e., maintain a balanced thermal profile of the HPC system below a certain threshold, while minimizing the energy consumption and performance degradation). In addition, an efficient design of HPC must include a thermal modeling tool, which is capable of modeling the HPC systems with inter-tier single- and two-phase cooling, as a fundamental step to design 3D ICs. Moreover, it is necessary to use a design-time specified inter-tier cavity, in combination with the use of run-time thermal management at the system-level to tune the flow rate of the coolant to achieve energy-efficiency in 3D MPSoCs. In particular, our experimental results on 2- and 4-tier 3D MPSoC designs show that, exploiting the features of the proposed new single- and two-phase liquid cooling technologies, our proposed system-level thermal management strategy prevents the system from surpassing the given threshold temperature while achieving up to 67% reduction in cooling energy and up to 30% reduction in system-level energy in comparison to setting the flow rate at the maximum value to handle the worst-case temperature.

The rest of the paper is organized as follows. First, in Section II, we describe the target 3D MPSoC architecture and the developed manufacturing technology to integrate different single-phase liquid cooling geometries developed in this project. Next, in Section III, we describe the baseline two-phase cooling technology that we have developed for future 3D MPSoCs and its advantages with respect to single-phase cooling are summarized. In Section IV we present the experimental results obtained in 3D MPSoCs after combining the different proposed strategies at the system level. Finally, Section V summarizes the main conclusions of this work.

II. 3D MPSoC Stacks with Inter-tier Liquid Cooling

A. Target 3D MPSoC Architectures

A typical structure of 3D MPSoCs consists of two or more more silicon tiers, which contain the processing and storage elements of the system. In particular, the 3D MPSoCs we use in this paper are based on the UltraSPARC T1 (i.e., Niagara-1) processor manufactured at 90nm node. The power consumption, area, and the floorplan of UltraSPARC T1 are available in [13]. UltraSPARC T1 has 8 multi-threaded cores, and a shared L2-cache for every two cores.

The communication between these tiers is realized with through-silicon vias (TSVs) that are etched in the residual silicon slab (cf. Subsection II-B). To account for inter-tier liquid cooling, the porous cavity is realized by etching porous structures of different form and shapes. However, both TSVs and porous cavity etching must be performed as a single integrated etching process, as shown in Subsection II-B.

In this paper, we target 2- and 4-tiers 3D MPSoC architectures. We place cores and L2 caches of the UltraSPARC T1 on

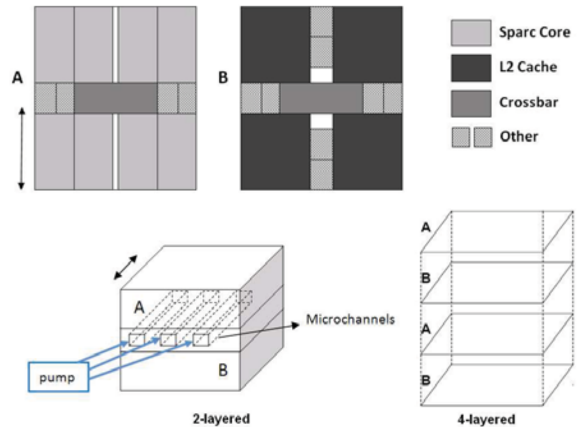


Fig. 1. Layouts of the 3D multicore systems.

separate tiers (see Fig. 1). Separating logic and memory layers is a preferred design scenario for shortening interconnections between the cores and their caches and achieving higher performance in 3D processing architecture [8]. The micro-channels are then built, and distributed uniformly, in between the vertical layers for liquid flow. The fluid flows through each channel at the same flow rate, but the liquid flow rate provided by the pump can be dynamically altered at runtime.

B. Manufacturing of 3D Stacks with Inter-Tier Liquid Cooling

The manufacturing of 3D CMOS stacks with TSV interconnections and micro-channels requires a series of microfabrication processes, namely (1) deep-reactive-ion-etching (DRIE) process for anisotropic silicon etching of both TSV openings and backside micro-channels; (2) conformal thin film deposition for TSV sidewall insulation; (3) electroplating for conductive layer formation; (4) grinding for chip thinning, and finally (5) wafer- or die-level bonding for the stacking. A simplified illustration of a 3D stack with inter-tier liquid cooling is shown in Fig. 2.

The compatibility and stability of the process steps play a crucial role in the reliability of the final 3D MPSoC. For instance, the thermal budget, the etching profile of the via sidewall, the quality of the dielectric layers, the aspect-ratio limitations of the thin-film deposition techniques, the stress induced during the grinding and the bonding steps are among the critical issues being investigated in the framework of the CMOSAIC project. Thus, we have been developing demonstrator chips to characterize the fabrication processes employed.

Our first generation TSV demonstrator chips involve SiO_2-insulated and fully-filled Cu TSVs having diameters ranging from 40 μm to 100 μm, fabricated in a 380 μm-thick Si wafer. The TSVs are connected in daisy-chain patterns for the electrical characterization tests. The compatibility of the TSV process to the micro-bump and thin-film bonding technologies, as well as the performance of the inter-tier cooling in the presence of TSVs is being tested.

The microfabrication of the test vehicle starts with a 380 μm-thick double-side polished silicon wafer having 200nm-thick thermally-grown SiO_2 (oxide) film on both sides.

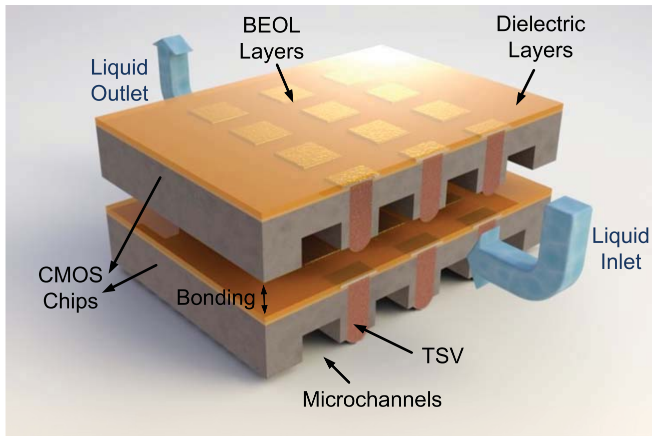
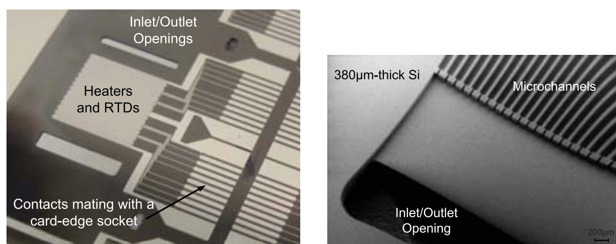


Fig. 2. Simplified illustration of 3D stack with inter-tier liquid cooling.



(a) Front-side with the Al micro-heaters (b) Back-side showing the micro-channels

Fig. 3. SEM photos the wafer with the inlet-outlet openings.

Microheaters and temperature sensors are fabricated by sputtering 50nm/1500nm thick Ti/Al layer and patterning the metal layers by RIE in Cl_2/BCl_3. A 4 μm thick photoresist is spin-coated on the front-side of the wafer and then inlet-outlet openings are etched by oxide RIE followed by 280 μm of Si DRIE. Figure 3(a) shows the front-side of the wafer with the Al microheaters and the inlet-outlet openings. Without removing the photoresist on the front-side, the oxide layer on the back-side is etched by wet oxide to be prepared for anodic bonding. After the oxide etching step, the photoresist on the front-side is stripped, and micro-channels are etched by DRIE on the back-side, as shown in Fig. 3(b). Finally, silicon-pyrex anodic bonding is performed to seal the channels.

C. Single-Phase Liquid Cooling Technologies for 3D MPSoCs

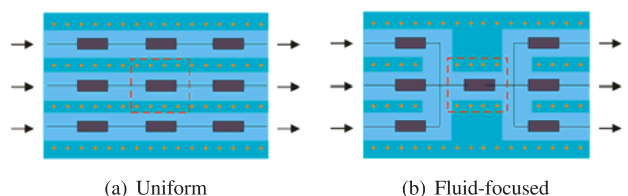
Inter-tier cooling is a solution to remove the heat with forced convection near the junction that scales with the number of tiers [6]. In the design phase of the porous media, the hydraulic diameter for the mass transfer of the coolant must take into consideration the dimension and placement of the manufactured TSVs. Due to the hydraulic diameter limitations that limits the maximum injected flow rate, the fluid temperature increase from inlet to outlet in single-phase cooling is significant (e.g. 40^oK in case of water as coolant at $130W$ power dissipation per tier [6]). Therefore, coolants with lower volumetric heat capacity and higher viscosity compared to water, such as dielectric fluids, are not acceptable, since they would degrade the inter-tier performance substantially. Using water, in contrast, demands for advanced sealing technology as described in Subsection II-B, to prevent electrical shorts, and electrochemical corrosion.

In addition, to use the water heat capacity more efficiently, the non-uniform nature of spatial power dissipation and local communication needs on a tier and between tiers should be considered. Such complex, hot-spot-aware heat transfer cavities can make use of the following building blocks/properties:

Heat transfer unit cell geometry: The shape of the heat transfer structure can be chosen freely in-plane, but is extruded normal to the surface. The only geometrical constraints are the implemented TSVs, which need to be embedded into the heat transfer structure we consider. Two fundamental geometries, i.e., channels and pin fins (circular, square, drop shape) are considered. We have investigated different pin arrangements (in-line, staggered) with respect to their heat removal performance. Our exploration has shown that, circular in-line pins result in low pressure drop at acceptable convective heat transfer, compared to staggered arrangement. In general, we conclude that low pressure drop structures should be targeted for 3D MPSoCs.

Heat transfer structure modulation: The effective convective resistance of heat transfer geometries can be adjusted spatially, by width or density modulation, in case of micro-channels or pin fin arrays respectively. The smaller the hydraulic diameter at a given mass flow rate, the higher the heat transfer and the associated pressure gradient. Accordingly, the maximal channel width, given by the TSV spacing, should only be reduced at locations where the maximal junction temperature would be exceeded. Thus, we have been able to report pressure drop and pumping power improvements by a factor of 2 and 5 [5].

Fluid focusing: The local flow rate on a hot spot location can be further increased with micro-channel networks or pin fin arrays in combination with guiding structures. Resulting super structures reduce the flow resistance from inlet to the hot spot and from the hot spot towards the outlet (Fig. 4). However, we only consider this option for 3D MPSoCs at a high heat flux contrast on the tiers, since the aggregate flow rate is reduced.



(a) Uniform (b) Fluid-focused

Fig. 4. Heat removal of a hot spot.

Electro-thermal co-design is mandatory to define the optimal fluid cavity and corresponding floorplan to achieve highest computational performance at minimal chip and pumping power needs, for the given temperature constraints [9]. In fact, using our developed models for 3D stacks [17], the scalability of inter-tier cooling has been already demonstrated. We compare the maximal junction temperature rise in a chip stack with a $1cm^2$ foot print and aligned hot spots of $250W/cm^2$ on three active tiers. Thus, we obtain an acceptable 55^oK in case of inter-tier cooling with four fluid cavities, compared to

the catastrophic 223^oK with back-side cooling [7].

D. System-Level Thermal Modeling and Management of 3D MPSoCs

In this very complex 3D MPSoC architectures, using detailed numerical analysis methods, such as finite-element methods (FEM) for thermal and cooling exploration is a too time-consuming process [6], [17]. This is not suitable for design-time architecture exploration and run-time thermal management of such complex 3D MPSoCs. Therefore, we need to use multi-scale thermal modeling concepts at system-level to improve computational efficiency of cooling requirements. Indeed, accurate thermal modeling is critical in the design and evaluation of temperature-aware systems and policies [18], [9]. To this end, the latest versions of HotSpot [16] (an RC-based thermal modeling tool for 3D MPSoCs) include 3D modeling capabilities for 3D MPSoCs, but with no inter-tier liquid cooling modeling capabilities. Hence, we have developed 3D-ICE [17] (3D Interlayer Cooling Emulator), which is a compact transient thermal model library (written in C) for the thermal simulation of 3D ICs with multiple inter-tier liquid cooling micro-channels. 3D-ICE is compatible with existing CAD tools for MPSoC designs, and offers significant speed-ups (up to 975x) over typical commercial computational fluid dynamics and thermal simulation tools while preserving accuracy (i.e., maximum temperature error of 3.4%) for 3D MPSoCs.

Moreover, even if inter-tier liquid cooling is a potentially effective cooling technology for future 3D MPSoCs, due to the limited diameter of the inter-tier micro-channels (channel cross-section less than $100 \times 50\mu m^2$), the energy spent in the pump that injects the coolant can be very significant. In fact, in an HPC cluster, the maximum pumping network energy required to inject the fluid to all stacks in this cluster is a significant overhead to the whole system, because it represents about 70 Watts (indeed similar to the overall energy consumption of a 2-tier 3D MPSoC). Thus, it is necessary to explore different cooling strategies and control the required cooling energy in the liquid injection architecture at system-level to achieve energy-efficient cooling infrastructures for 3D MPSoC-based HPC systems.

In our recent work, we have proposed a methodology to model liquid-cooled systems, and we have shown that dynamic flow rate control is able to reduce cooling energy consumption [9]. Moreover, we have developed, in this work, different thermal management strategies that use run-time varying flow rate in conjunction with other electronic-based thermal management options, i.e., task scheduling and dynamic voltage and frequency scaling (DVFS). In particular, we have developed a run-time fuzzy-logic thermal controller that uses run-time varying flow rate and DVFS to minimize the consumed energy while keeping the systems temperature below the thermal threshold (85^oC) for 3D MPSoCs [15].

III. Application of Two-Phase Liquid Cooling Technologies for 3D MPSoCs

Flow boiling heat transfer in micro-channels is also an excellent choice to consider for inter-tier cooling of 3D MPSoCs stacks, having been proven to be able to dissipate very high heat fluxes with similarly high heat transfer coefficients and a high uniformity in temperature in 2D tests so far [1]. Basically, flow boiling involves evaporating a refrigerant (dielectric fluids utilized as the working fluid of air-conditioning and refrigeration systems) within the micro-channels to remove the heat in the form of latent heat absorbed by the fluid as it changes from liquid into vapor as it flows along the channel. Since the latent heat of vaporization of most common refrigerants is large compared to the specific heat of water, e.g., about 150 kJ/kg of R-134a compared to 4.2 kJ/kg K of water. The flow rate of the two-phase coolant can be as little as 1/5 to 1/10 that of water, depending on the operating conditions and particular refrigerant. Since the pumping power to push the coolant through the micro-channels is directly proportional to the flow rate, two-phase cooling enjoys a significant energy savings with respect to water (about 80-90% less energy consumption in the micro-channels) [2].

Another important distinguishing characteristic of flow boiling as opposed to water cooling is that during evaporation along the micro-channels the refrigerant's temperature falls rather than increases. This is because the local saturation temperature of the refrigerant follows that of the local saturation pressure, which falls due to the pressure drop along the channel. Hence, in flow boiling the exit temperature of the refrigerant is lower than at the inlet [1], [2]. Since the local flow boiling heat transfer coefficient tends to decrease along the channel length, it is possible to approximately match the falling in local saturation temperature to the local rise in thermal resistance of the evaporating fluid to produce a uniform temperature along the micro-channel. Furthermore, since an evaporating refrigerant absorbs heat without an increase in its temperature, two-phase flow cooling has a transient flow thermal storage capacity, because simply more liquid evaporates into vapor, as long as dry-out of the annular liquid film evaporating on the channel walls is avoided, which is ideal for 3D MPSoC stacks. Furthermore, flow boiling in micro-channels is only a weak function of the flow rate, such that non-uniform flow distribution due to different heat dissipation tracks through the inter-tier does not create an imbalance in the local cooling capacity as long as dry-out is avoided. Hence, all of these characteristics of micro-channel flow boiling are of particular benefit to cooling of 3D MPSoCs.

On the other hand, the proper refrigerant must be chosen since its saturation pressure may be too high for 3D MPSoCs depending on the chip's operating temperature. In fact, Agostini et al. [1], [2] have tested several low pressure refrigerants in both once through flow (one inlet/one outlet) and for split flow (one inlet/two outlets) in silicon test sections with 134 parallel channels ($67/92/680\mu m$ channel width/fin thickness/channel height), where the split flow greatly reduced two-phase pressure drops. Heat fluxes up to $255W/cm^2$ were attained with pressure drops less than 0.9 bar. This geometry, except for its height, falls within the feasible range for cooling of 3D MPSoCs.

In order to validate our hypothesis of 2-phase flow benefits for 3d MPSoCs, we have developed test vehicles dedicated for two-phase liquid cooling experiments (Fig. 5). The chips

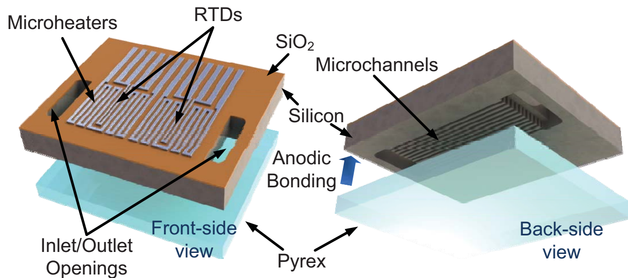


Fig. 5. Illustration of the front- and back-side of our test vehicles manufactured for two-phase liquid cooling experiments.

comprise microheaters emulating the power dissipated by active components in a CMOS chip, resistive-thermal-devices (RTD) as temperature sensors, backside micro-channels in various dimensions and configurations, and finally a pyrex cover for both channel sealing and visual inspection.

IV. Experimental Results

A. Single Phase Cooling and Run-Time Thermal Management

In this subsection we apply the proposed system-level control strategies [15] to our target 3D MPSoC architectures. In our experiments, we use workload traces collected from real applications running on an UltraSPARC T1. We record the utilization percentage for each hardware thread at every second for several minutes for each benchmark. We use various real-life benchmarks including web server, database management, and multimedia processing. Based on these utilization percentages, we calculate the power consumption of the 3D MPSoC. The peak power consumption of SPARC is close to its average value [13]. Thus, we assume that the instantaneous dynamic power consumption is equal to the average power at each state (active, idle). We compute the leakage power of processing cores as a function of their area and the temperature. For more details on power modeling, we refer to our work in [15].

We assume that each core has a temperature sensor, which is able to provide temperature readings at regular intervals (e.g., every 100ms). We implement various thermal management techniques to evaluate the thermal and energy efficiency of the proposed fuzzy-thermal management technique (LC_FUZZY). Dynamic load balancing (LB) balances the workload by moving threads from a core's queue to another if the difference in queue lengths is over a threshold. Temperature-triggered DVFS (AC_DVFS_LB) adjusts the VF settings of a core when the core's temperature exceeds 85^oC. In our implementation, as long as the temperature is above the threshold and there is a lower setting, we scale down the VF value at every scaling interval. When the temperature falls below another threshold value (82^oC), we scale up the VF values.

We use the parameters provided in Table I in thermal modeling. This table contains the thermal conductance and capacitance values of the various materials used in modeling the stack. In our experiments, we initialize the simulations with steady state temperature values. We compare air-cooled and liquid-cooled 2- and 4-tier 3D MPSoCs.

We experiment with both air-cooled (AC) and liquid-cooled (LC) systems for comparison purposes. In LC_LB, we apply the maximum flow rate (0.0323 l/min per cavity), while

TABLE I. Thermal and floorplan parameters deployed in the 3D MPSoC model

Parameter	Value
Silicon conductivity	$130W/(m \cdot K)$
Silicon capacitance	$1635660J/(m^3 \cdot K)$
Wiring layer conductivity	$2.25W/(m \cdot K)$
Wiring layer capacitance	$2174502J/(m^3 \cdot K)$
Water conductivity	$0.6W/(m \cdot K)$
Water capacitance	$4183J/(kg \cdot K)$
Heat sink conductivity (air cooling only)	$10W/K$
Heat sink capacitance (air cooling only)	$140J/K$
Die thickness (one stack)	$0.15mm$
Area per Core	$10mm^2$
Area per L2 Cache	$19mm^2$
Total area of each layer	$115mm^2$
Inter-tier material thickness	$0.1mm$
Channel width	$0.05mm$
Channel pitch	$0.15mm$
Flow rate range	$10 - 32.3ml/min$ per cavity
Pumping network power	$3.5 - 11.176W$

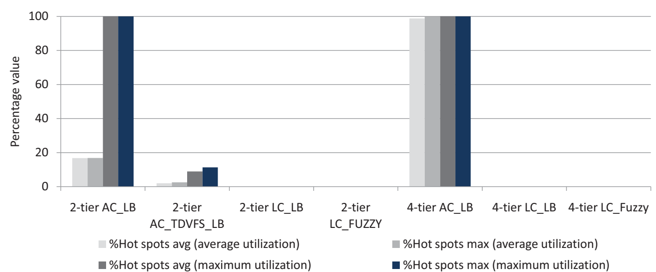


Fig. 6. Percentage of time we observe hot spots for all the policies, both for the average case across all workloads and for maximum utilization. The figure shows the % values averaged per core and the % of time hot spots are observed across the 2- and 4-tiers 3D MPSoCs.

the jobs are scheduled with LB. Thermal impact of all the policies is shown in Fig. 6. This figure compares the % of time spent above the threshold temperature for the average case across all the workloads (marked as hot spots avg) and also for the benchmark with maximum utilization rate. TDVFS help reduce the hot spots in air-cooled systems, while the integration of liquid-cooling removes all the hot spots. The peak temperature with LB and AC_DVFS_LB are 87^oC and 85^oC, respectively. However, in the 4-tier stack, due to increased stacking and limited cooling capabilities, the maximum temperature is much higher than 110^oC and reaching up to 178^oC, leaving little opportunity for any thermal management technique to successfully mitigate the hot-spots without severely degrading the performance.

On the contrary, the integration of liquid cooling removes all hot-spots in the tested 2- and 4-tiers 3D MPSoCs by reducing the temperature below the threshold, due to its ability of inter-tier heat removal. LC_LB reduces the 2-tier 3D MPSoC peak temperature to 56^oC, whereas the proposed fuzzy controller (LC_FUZZY) pushes the system into a higher peak of 68^oC, but still avoids any hot-spots. Moreover, the system temperature of a 4-tier 3D MPSoC is maintained even lower than the 2-tier 3D MPSoC in both techniques, due to the increased number of cooling tiers (cavities).

Fig. 7 shows the total energy consumed using the various policies on the 2-tier and 4-tier 3D MPSoCs for the average workload. Energy consumption values are normalized to the 2-tier AC_LB values. The proposed management strategy achieves major reduction in both the coolant and the overall

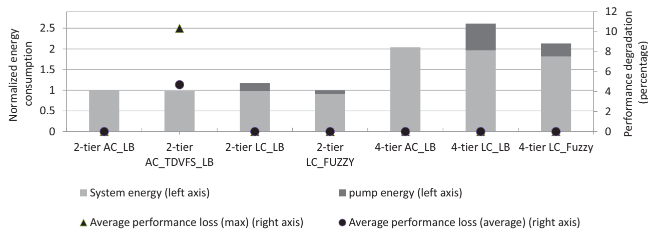


Fig. 7. Left-axis shows the energy consumption in the whole system (chip and cooling network) averaged per stack. The right-axis shows the % delay for each policy.

system energy consumption. LC_FUZZY reduces the 2- and 4-tier system energy by 14% and 18%, and cooling energy by 50% and 52% in comparison to LC_LB, respectively. The reason LC_FUZZY outperforms all other techniques in energy savings is due to the joint control of flow rate and DVFS at run-time based on each core thermal and utilization status. The proposed controller achieves up to 67% and 30% of coolant and overall system energy savings, respectively.

For our multicore 3D MPSoCs, the performance degradation of the average workload under a set of policies is shown in Fig. 7. Liquid cooling-based systems do not suffer from any performance degradation since the temperature of such systems does not rise to a value where another thermal management technique should be applied. Although our proposed fuzzy controller uses DVFS, as we apply DVFS based on the core utilization, the performance degradation results do not exceed 0.01%, which is negligible.

B. Two Phase Cooling

We examine two phase cooling with a 3D chip having 35 local heaters and 35 local temperature sensors on one face [10], and cooled by a two-phase refrigerant evaporating in 135 parallel micro-channels of $85\mu m$ width engraved in the opposite face of the silicon die. The 35 local heaters are organized in a 5×7 layout, where the first two and last two rows have a low heat flux ($2W/cm^2$) while the third row has a 15 times higher heat flux ($30.2W/cm^2$) applied. We show in Fig. 8 the thermal profile of this chip. In this figure, the refrigerant enters at a saturation temperature of 30^oC and leaves with a temperature of 29.5^oC. In particular, our results show the local heat transfer coefficient under the hot spot is 8 times higher so that the wall superheat (the wall temperature of the channel minus the local fluid saturation temperature) is only 2 times higher under the hot spot rather than 15 times with water cooling. Thus, the potential of two-phase micro-channel cooling is thus clear for inter-tier cooling of 3D MPSoCs; however, existing methods and experimental experience must be scaled down to the $50\mu m$ height of micro-channels permissible in between the TSVs.

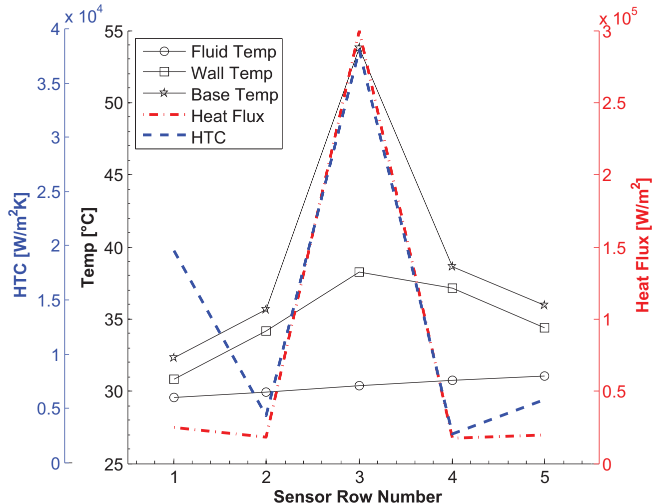


Fig. 8. Local hot spot test for a silicon micro-evaporator.

V. Conclusions

Inter-tier liquid and two phase cooling are promising cooling technology solutions to overcome the thermal challenges of 3D MPSoCs in HPC architectures. However, intelligent control of the coolant flow rate is needed to avoid wasted energy consumption for over-cooling the system when the system is under-utilized. In this paper we have presented the results of the CMOSAIC project on the development of novel system-level thermally-aware design methodologies as an effective and combined (mechanical-electrical) technology approach to achieve thermally-balanced 3D MPSoCs for high-performance computing systems. Our results with a 2- and 4-tier 3D MPSoC case studies using our novel system-level run-time management show that we are able to balance temperature across the 3D stack and to minimize system energy consumption while preventing thermal hot spots. Indeed, our controller maintains the temperature below the desired levels, while reducing cooling energy by up to 30% and achieving overall energy savings up to 67% with respect to setting the highest coolant flow rate to match the worst-case temperature.

References

[1] B. Agostini et al. High heat flux flow boiling in silicon multi-microchannels: Part I - Heat transfer characteristics of R-236fa. *International Journal of Heat Mass Transfer*, (51), 2008.

[2] B. Agostini et al. High heat flux two-phase cooling in silicon multi-microchannels. *IEEE TCPT*, 31, 2008.

[3] B. Black et al. Die stacking (3d) microarchitecture. In *MICRO 39*, 2006.

[4] T. Brunschwiler et al. Forced convective interlayer cooling potential in vertically integrated packages. In *11th ITHERM 2008*, 2008.

[5] T. Brunschwiler et al. Hotspot-optimized interlayer cooling in vertically integrated packages. In *MRS fall meeting*, 2008.

[6] T. Brunschwiler et al. Validation of porous-media prediction of interlayer cooled 3D-chip stacks. In *3DIC*, 2009.

[7] T. Brunschwiler et al. Heat-removal performance scaling of interlayer cooled chip stacks. In *ITHERM*, 2010.

[8] A. K. Coskun et al. Dynamic thermal management in 3D multicore architectures. In *DATE*, 2009.

[9] A. K. Coskun et al. Energy-efficient variable-flow liquid cooling in 3D stacked architectures. In *DATE*, 2010.

[10] E. Costa-Patry et al. Hot-spot effects on two-phase flow of R245fa in 85micron-wide multi-microchannels. In *THERMINIC*, 2010.

[11] M. Healy et al. Multiobjective microarchitectural floorplanning for 2-d and 3-d ICs. *IEEE Transactions on CAD*, 26(1), Jan 2007.

[12] W.-L. Hung et al. Interconnect and thermal-aware floorplanning for 3d microprocessors. In *ISQED*, pages 98–104, 2006.

[13] A. Leon et al. A power-efficient high-throughput 32-thread SPARC processor. *ISSCC*, 42(1):7 – 16, 2007.

[14] K. Puttaswamy et al. Thermal analysis of a 3D die-stacked high-performance microprocessor. In *GLSVLSI*, 2006.

[15] M. M. Sabry et al. Fuzzy Control for Enforcing Energy Efficiency in High-Performance 3D Systems. In *ICCAD*, 2010.

[16] K. Skadron et al. Temperature-aware microarchitecture. In *ISCA*, 2003.

[17] A. Sridhar et al. 3D-ICE: Fast compact transient thermal modeling for 3D-ICs with inter-tier liquid cooling. In *ICCAD*, 2010. http://esl.epfl.ch/3d-ice.html.

[18] C. Zhu et al. Three-dimensional chip-multiprocessor run-time thermal management. *IEEE Transactions on CAD*, 27(8), August 2008.