27-29 September 2011, Paris, France

System-Level Thermal-Aware Design of 3D Multiprocessors with Inter-Tier Liquid Cooling

Arvind Sridhar Mohamed M. Sabry David Atienza

Embedded Systems Laboratory (ESL), École Polytechnique Fédérale de Lausanne (EPFL),
ELG 130 (Bâtiment ELG), Station 11, 1015 Lausanne, Switzerland.
{arvind.sridhar, mohamed.sabry, david.atienza} @epfl.ch

***Abstract*- Rising chip temperatures and aggravated thermal reliability issues have characterized the emergence of 3D multiprocessor system-on-chips (3D-MPSoCs), necessitating the development of advanced cooling technologies. Microchannel based inter-tier liquid cooling of ICs has been envisaged as the most promising solution to this problem. A system-level thermal-aware design of electronic systems becomes imperative with the advent of these new cooling technologies, in order to preserve the reliable functioning of these ICs and effective management of the rising energy budgets of high-performance computing systems.**

This paper reviews the recent advances in the area of system-level thermal modeling and management techniques for 3D multiprocessors with advanced liquid cooling. These concepts are combined to present a vision of a green data-center of the future which reduces the CO_2 emissions by reusing the heat it generates.

Keywords- 3D Integration, Liquid Cooling, System-Level Thermal Aware Design, Green Data-Centers.

I. Liquid Cooling of Microelectronic Systems

The economic and technological drivers pushing the trend of shrinking CMOS feature size and the increasing die size to incorporate larger functionality in Integrated Circuits (ICs) have slowed down in the recent years, but the demand for faster and more versatile electronic products remains insatiable [1]. Multiprocessor system-on-chips (MPSoCs) have helped meeting this demand via the integration of diverse functionalities on a single silicon die, and have revolutionized the electronics industry. However, increasing the size of the die in two dimensions for this purpose has depreciating returns in terms of performance enhancements due to increasing interconnect length and the resulting delays. Hence, even the MPSoCs are quickly approaching the limits of their computing throughput capacity.

In this context, 3D integration of multiprocessor ICs opens up a new dimension in design space for VLSI engineers. On one hand, 3D integration enables shorter interconnections, handling a larger IC design complexity and the possibility of heterogeneous integration [2]. On the other hand, it also brings compounded heat dissipation and larger thermal resistances to heat sinks, which results in chip temperatures well beyond the safe operating levels, thus severely undermining the already aggravated thermal reliability of MPSoC designs and lifetimes. As a result, conventional air-cooled heat sinking has become insufficient in the context of 3D-MPSoC integration [3,4].

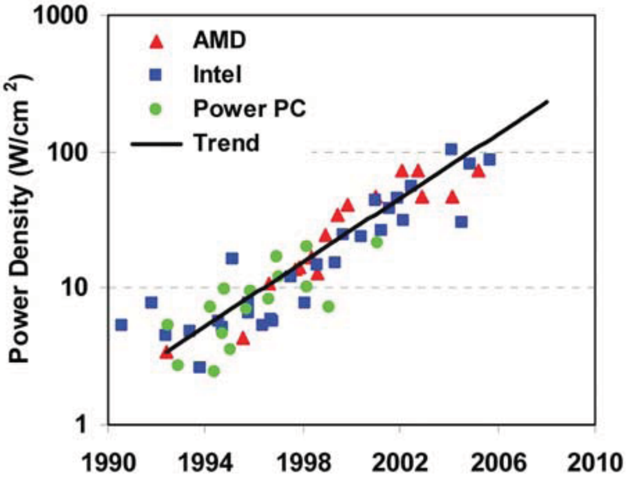


Fig. 1: On chip power density during the last two decades (Courtesy: [6], IEEE Special Issue on Thermal Engineering)

Thirty years ago, Tuckerman and Pease [5] published a seminal paper studying the merits of liquid cooling in ICs via microchannels etched directly on the substrate, and formulated the main determinants for the design of microchannel heat sinks that maximize cooling efficiency. With on-chip heat flux densities approaching 100 W/cm^2 in 2D processor chips- and twice or even thrice that value in 3D chips- in the recent years (Fig. 1), this work has inspired a renewed interest in the development of liquid cooled package design for ICs [4, 7-13]. The microchannel single-phase liquid cooling of 3D ICs is now envisaged to be the most promising short to medium solution to the problem of rising chip thermal reliability issues [4], while two-phase cooling is seen as the long term solution to meet the growing demands for energy in the high-performance computing installations and data-centers of the future [13].

A considerable amount of research effort has been invested in the last two decades towards the design of efficient microchannel liquid cooled heat sinks, especially single-phase cooling, of ICs. However, the electronics industry hasn't seen a large-scale acceptance of this technology. While this is partly due to the fact that the advent of CMOS circuits postponed the anticipated explosion of on-chip heat flux densities by almost two decades, the lack of knowledge-transfer from thermal engineers to the VLSI designers has also played a major role. For any new technology to be incorporated and exploited commercially by the electronics industry the formalization of three main aspects is required: 1) manufacturing methodologies, which can make the technology economically viable, 2) modeling and simulation methodologies, which helps designers to assess the performance metrics of the technology, and 3) design optimization methodologies, which help designers to effectively make use of the new technology and achieve the desired system-level objectives.

In the context of using liquid cooling for 3D-MPSoC ICs, these aspects entail, on one hand, the development of process and bonding techniques to create 3D IC stacks with microchannels etched on the back of each die, while maintaining the interconnection integrities of the TSVs that run through the channel walls. On the other hand, it is required to develop efficient modeling methodologies for electro-thermal co-simulation of MPSoCs that are cooled using microchannel heat sinks (especially during the early-stages of design), and then the development of “liquid cooling-aware” design- and run-time thermal management techniques, that maximize the electrical performance of the systems while maintaining safe operating conditions. These research topics have started gaining attention only recently and the goal of this paper is to review the latest results in these areas. Moreover, we will describe in detail, the new thermal analysis software called 3D-ICE, which is the first-ever compact and transient thermal simulator for 3D or 2D ICs with microchannel liquid cooling. We also describe our recent advances in thermal management policies developed for 3D multiprocessors. We briefly show an overview of two different approaches we undertake in the development of management policies. Moreover, we demonstrate the impact of the policies on a typical 3D multiprocessor performance.

Finally, we will present a vision of the culmination of these research efforts in the building of data-centers of the future. Data-centers, which form the backbone of the IT and the IT-enabled industries such as banking and telecommunications, consume 2% of the global electricity production. Hence, data-centers provide a great opportunity for the IT industry to contribute to the worldwide effort in combating rising carbon emissions and climate change. Data-centers typically consists of high performance computing servers, containing hundreds of processor cores stacked inside a chassis which are of the size of a small room. One or more of these server racks are stationed in an air-cooled facility to remove the heat generated by the electronic activity in these machines. Half of the energy consumed by data-centers is thus spent on cooling the facilities housing computing infrastructure.

3D-MPSoCs, with their reduced effective system size and increased computational throughput, have tremendous potential in making data-centers more compact and efficient. However, they bring with them aggravated thermal issues. But it is possible to see this as an opportunity in disguise: 3D-MPSoCs cooled using liquid coolants flowing inside microchannel heat sinks provide a valuable resource for extracting heat efficiently, which could be reused. Hence, in addition to cooling the servers much more cost effectively compared to conventional air-cooled heat sinks, energy reuse becomes practical, further reducing the effective carbon footprint of the data-centers. The vision presented in this paper focusses on the recent advances in the joint work between the Swiss Federal Institutes of Technologies (EPFL and ETHZ) and IBM Research Laboratory in Zurich towards building a zero-emission data-center that utilizes hot-water liquid cooling and enables the direct-reuse of the output heat for district heating in Europe.

II. Thermal Modeling for 3D-MPSoC ICs with Microchannel Heat Sinks

Thermal simulation of ICs with conventional heat sinks has a long history. There are many open source as well as commercial thermal/electro-thermal simulation tools and methods available for IC design [14, 19]. Most of these methods present simplified thermal models for steady state simulations and provide no information about the transient thermal behavior of the ICs. HotSpot [14] is an open source tool available for transient thermal simulation of 2D as well as 3D ICs.

Conventional compact modeling for thermal analysis in solid structures is based on the finite-difference method of dividing the IC structure into small cuboidal “thermal cells” and the construction of an equivalent electrical circuit based on the thermal-electrical analogy. That is, in each thermal cell, the thermal conduction in different directions are represented using electrical resistors, the volumetric storage of heat resulting in the rise of the material’s temperature is represented using an electrical capacitor, and the internal generation of heat due to chemical/electrical activity is represented using electrical current sources. Once such an electrical equivalent circuit for each thermal cell is constructed, these individual circuits can be connected through the interfaces of each thermal cell and its neighbors to create a compact RC circuit grid for the entire structure. The boundary conditions representing the escape of heat into the ambient in an air-cooled IC is represented using voltage sources at the exposed surfaces, and provides the ground, or the return path, for the equivalent circuit. This circuit mesh can be solved using conventional circuit simulators to obtain the temperatures in the IC. HotSpot, a tool based on this methodology, has been benchmarked against experiments conducted on industry grade ICs.

Forced convective liquid cooling in microchannel heat sinks has been extensively studied in the heat transfer literature [5, 8, 20-26]. Heat transfer is modeled using the Newton’s law of cooling, which states that the heat transferred from the wall of the heat exchange pipes/channels into the fluid is proportional to the temperature difference between them. The constant of proportionality is defined as the heat transfer coefficient, as follows:

$$q'' = h\left(T_{wall} - T_{\infty}\right) \quad \textbf{(1)}$$

Here q'' is the heat flux at the surface of the wall, *h* is the local heat transfer coefficient (in Watts/m^{2}K), T_{wall} and T_∞ are the temperatures of the surface of the wall and the liquid bulk respectively. The heat transfer coefficient is written as,

$$h = k\frac{Nu}{d_h}, \qquad (2)$$

where k is the thermal conductivity of the fluid, d_h is the hydraulic diameter of the channel, and Nu is the Nusselt number, which is traditionally calculated using empirical correlations based on experiments on channels of various crosssectional shapes and dimensions subject to different amounts of heat flux inputs.

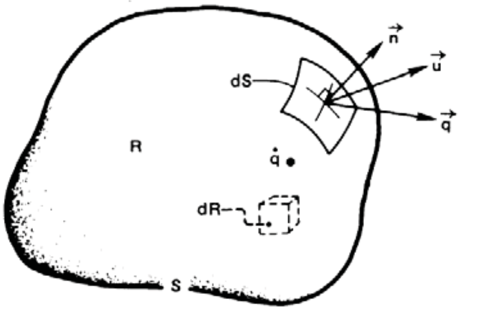


Fig. 2: Control volume of liquid

But all these models pertain to steady state analyses and study the forced convective cooling under idealized scenarios- constant and uniform heat fluxes along the channel walls with little or no conductive spreading of heat. This is in contrast to the reality of the thermal behavior of ICs- heat fluxes are highly non-uniform, and change with time at very high frequencies, and there is a considerable amount of heat spreading the silicon structures surrounding the microchannel heat sinks. This complex interplay between conduction and convection must be modeled for an accurate estimate of temperatures in the IC. Hence, we must find a compact model for the convective cooling in microchannel heat sinks, which can be interfaced with the conventional compact transient modeling methods for heat conduction in silicon, capturing this complex interplay of conduction and convection in ICs with liquid cooled heat sinks in the transient-domain.

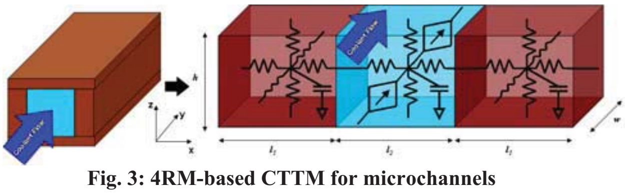


Fig. 3: 4RM-based CTTM for microchannels

In 2010, the 3D-Interlayer Cooling Emulator (3D-ICE), the first ever compact transient thermal model incorporating this complex interplay of heat diffusion, was presented [27-29]. This model advanced a new equivalent electrical representation of convective heat transport along the channel downstream, enabling the construction of a compact model for heat transfer in flowing liquids. This model has also been extended for the case of enhanced heat transfer geometries, such as pin fins, which are used for more efficient cooling of ICs. In the rest of this section, the theory of compact modeling behind 3D-ICE will be presented.

A. 3D-ICE

The development of the 4-resistor model-based compact transient thermal model, or the 4RM-based CTTM, which is the basis of 3D-ICE 1.0, starts with the energy balance equation in flowing liquids. The conservation of heat in a control volume of liquid (see Fig. 2) can be written as [20]:

$$\frac{d}{dt}\int_R \rho\,\hat{h}\,dR + \int_S (-k\nabla T)\cdot\vec{n}\,dS + \int_S (\rho\,\hat{h})\vec{u}\cdot\vec{n}\,dS = \int_R \dot{q}\,dR \qquad (3)$$

The first term in the LHS of the above equation represents the rate of increase of the internal energy of the fluid (capacitance), the second term represents the conductive heat transfer within the fluid in different directions (conductance), and the third term represents the convective heat transport due to the velocity of the fluid, $\vec{u}$. Taking the limit of the control volume to zero and applying Stoke's theorem we get,

$$C_v\frac{d}{dt}T + \nabla\cdot(-k\nabla T) + C_v\vec{u}\cdot\nabla T = \dot{q}\,. \qquad (4)$$

Here, C_v is the volumetric heat capacity of the coolant. If this equation is to be spatially discretized with the finite-difference approximation for the case of a microchannel as shown in Fig.3, where the entire cross-section of the microchannel forms the front and the rear faces of a single liquid “thermal cell”, then the heat transfer in this thermal cell is governed by the following equation (assuming the fluid is flowing in the direction $+y$):

$$\begin{aligned} &C_v\Delta V\frac{d}{dt}T + h\Delta A_x(T - T_{wall1}) + h\Delta A_x(T - T_{wall2}) \\ &+ h\Delta A_z(T - T_{wall3}) + h\Delta A_z(T - T_{wall4}) \\ &\qquad + C_v u_y \Delta A_y (T_{S2} - T_{S1}) = 0 \end{aligned} \qquad . \qquad (5)$$

Here, ΔV is the volume of the thermal cell, ΔA_x, ΔA_y and ΔA_z are the areas of the different faces of the thermal cell, $T_{wall1-4}$ are the wall temperatures at the different faces of this microchannel cell, T_{S1} and T_{S2} are the front and rear face temperatures for this cell, and h is the heat transfer coefficient at the walls of the microchannel, as calculated using (2). Hence, the rise in internal energy of the thermal cell is represented using an electrical capacitance, and the transfer of heat from the walls into the bulk of the fluid is represented using electrical resistance as before.

The new term for representing the convective heat transport along the downstream direction is a *voltage controlled current source* (or transconductance) element in this electrical analogy. The identification of this new analogy is the main innovation in this model and enables the compact transient analysis of heat transfer. The face temperatures T_{S2} and T_{S1} can be approximated as the average of temperatures of the current thermal cell, and the temperatures of its front and rear neighbors, respectively.

The above formulation results in the 4RM-based CTTM (“4RM” because there are four electrical resistances representing the convective cooling effect from the walls into the bulk of the fluid). This model was demonstrated to be flexible and accurate, given the availability of accurate estimation of convective resistances for Microchannels. In our experiments, the Nusselt number in (2) was estimated for fully developed flows using correlations provided by Shah & London [30]. Also, this model was validated against and was shown to be significantly faster than commercial CFD simulators like Ansys CFX, as shown in Fig. 4 (for detailed description of the 4RM-based CTTM, please refer to [27]).

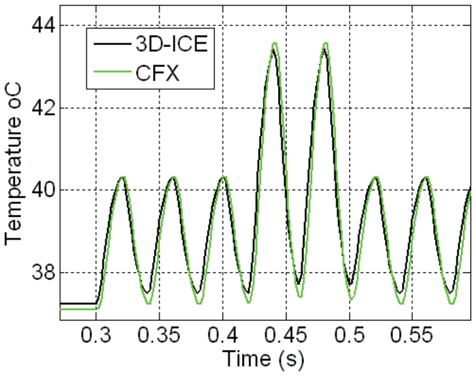
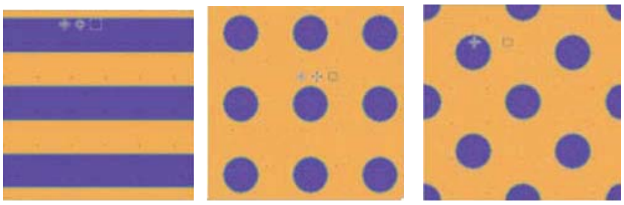


Fig. 4: Comparison of transient temperatures between for a 3D-IC test stack containing 3-dies and 4-microchannel cavities. 3D-ICE showed a 975x speed up compared to CFX [27].



(a) (b) (c)

Fig. 5: forced convective heat transfer geometries (dark: pin or fin)- (a) microchannel (b) pin fin inline (c) pin fin staggered.

3D-ICE 1.0 was released as an open source Thermal Simulator/Software Library for thermal analysis of 3D ICs in September 2010, and has since seen more than 50 subscriptions by universities and research groups all over the world [29].

B. 3D-ICE 2.0- The 2RM-based CTTM

One disadvantage of the 4RM-based CTTM is that the user is forced to use the channel width as the discretization length along the *x*-direction (i.e. transverse to the flow direction, *Δx*), since each fluid cell must encompass the entire cross section of the channel (Fig. 3). Given the typical microchannel width of 50-100μm in most interlayer cooling HTGs, this results in a significantly finer mesh for the thermal grid than what is necessary for the accuracy/resolution purposes of a VLSI designer. In addition, this model does not lend itself to the simulation of enhanced heat transfer geometries such as pin fins, illustrated in Fig. 5, which are envisioned as a solution for more efficient single-phase cooling of ICs [31].

In order to address this issue and in order to extend the scope of 3D-ICE to include enhanced heat transfer geometries (HTGs), we proposed the porous media based 2 resistor model (2RM) to replace the 4RM in the CTTM. For this, the porous medium approach advanced in [32] is incorporated.

Using the porous media based CTTM also allows the designer greater freedom to increase the discretization size, resulting in smaller problem sizes and faster simulations. This is because the porous media approach homogenizes the cavity layer into a porous medium, where the heat is transferred from the dies to the coolant via only 2 convective thermal resistances- one in the top and the other in the bottom. The heat transfer parameters for convection and conduction are modified based on the relative fraction of the volume of the cavity occupied by the fluid – called the porosity. Hence, the three dimensional heat transport from the solid domain to the fluid domain is reduced to a two dimensional circuit.

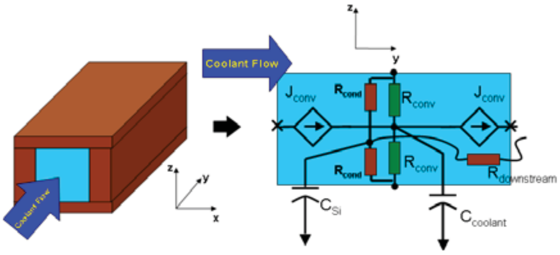


Fig. 6: 2RM-based CTTM for microchannels

The 2RM-based CTTM for microchannels is illustrated in Fig. 6. The basic idea here is to translate the convective heat transfer from silicon to the coolant from four directions (vertical from top and bottom walls, and lateral from the two side walls) into two directions (vertical), by projecting the heat transfer through the side walls onto the top and bottom surfaces, that is,

$$h_{eff,porous} = \frac{\int h \cdot dA_{wetted}}{A_{projected}}. \qquad (6)$$

Here, A_{wetted} represents the actual area that is wetted by the coolant, and $A_{projected}$ is the final area of projection of the heat transfer in the model. For example, if the vertical and the side heat transfer coefficients for the microchannel are equal (as in (5)), then the effective porous media heat transfer coefficient for the top and the bottom wall are given by:

$$h_{eff,porous} = h\frac{(w_c + t_c)}{p_c}. \qquad (7)$$

where, w_c is the width of the microchannel, t_c is the height of the cavity and p_c is the pitch of the channels.

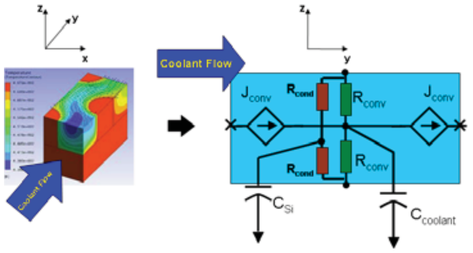


Fig. 7: 2RM-based CTTM for pin fins

In Fig. 6, R_{cond} represents the conductive resistance between the top wall and the bottom wall via the silicon walls separating the microchannels. $R_{downstream}$ represents the conductive resistance of these walls along the channel direction. All these parameters and the voltage controlled current source in this new model are scaled by the porosity factor, given by the expression $\varepsilon = w_c/p_c$ for the microchannel cavity [32].

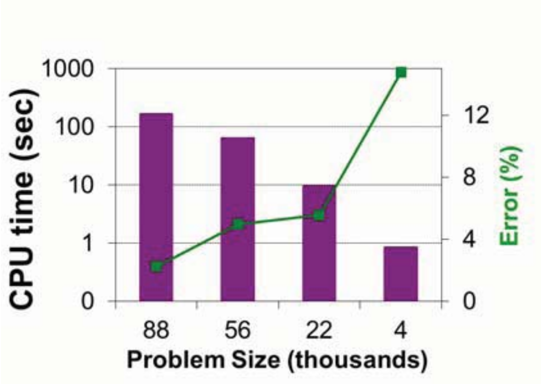


Fig. 8: Comparison of the CPU performance of 2RM-based CTTM, with the corresponding error incurred vs 4RM-based CTTM, as the cell size is increased

This 2RM-based CTTM can also be extended for the case of pin fins as shown in Fig. 7. Here, a pin-fin staggered HTG is considered. As can be seen, one main structural difference between this model and the 2RM-based CTTM for microchannels is the absence of the $R_{downstream}$. The model parameters are computed similar to the case of microchannels. The porosity for the pin-fin HTGs, with pin fin diameter d, is given by

$$\varepsilon_{\text{pinfin}} = \left(1 - \frac{\pi d^2}{4} \cdot \lambda\right), \cdot \qquad \textbf{(8)}$$

where λ is the pin density (number of pins per unit area) in the cavity. The effective heat transfer coefficient for the pin-fin HTG is obtained from the correlations presented in [32].

With this homogenization of the cavity layer, the user is free to use any discretization for the CTTM resulting in a reduced problem size, and in turn, reduced CPU time of simulation. This was demonstrated in [28] by conducting experiments on a real test 3D IC problem, comparing the 4RM-based CTTM with fixed discretization, with the 2RM-based CTTM with increasing cell sizes, for accuracy and simulation time, as shown in Fig. 8. The 2RM-based CTTM will be part of the next release of our thermal simulation software, 3D-ICE 2.0.

III. Dynamic Thermal Management of 3D-MPSoC ICs with Inter-tier Liquid Cooling

Temperature control, or dynamic thermal management (DTM), has been an important aspect in ameliorating the reliability and the lifetime of integrated circuits [33]. With the advent of 3D integration, system operating temperatures have escalated to alarming levels, implying the crucial need for temperature minimization and management [34].

Recently, research effort has been invested in the temperature management of 3D-MPSoCs using the conventional passive temperature control elements (e.g. dynamic voltage and frequency scaling) [35-37]. However, the reported temperatures in these works lie in the thermal runaway situations, where temperature exceeds 85°C. On the other hand, active cooling techniques, in particular inter-tier liquid cooling, manage to effectively reduce the operating temperature of multi-layer 3D-MPSoCs (e.g., 4 layers) to normal values [26, 38, 39].

Despite the significant impact of liquid cooling on temperature reduction, the thermal gradient within a single layer is aggravated [27]. As the liquid is passing through the microchannels, it is thermally developing from the inlet to the outlet. Thus, the amount of heat that can be transferred to the fluid is higher at the inlet than at the outlet.

In this paper, we briefly elaborate our recent advances in thermal management policies in order to achieve energy efficiency, temperature reduction and thermal balance through the interdisciplinary use of different control elements. First, we explain the applied control elements that are used in our policies. Next, we briefly explain two of our applied management policies, namely *rule-base fuzzy control* [40] and *hierarchical-based model predictive control* [41]. Finally, we explain the trade-offs in these two management policies with different simulation results.

A. Applied Control Elements

In our thermal management approaches, we deploy different control elements that have been used previously in various thermal management policies. These control elements are as follows:

- **Dynamic Frequency and Voltage Scaling (DVFS).** This technique has been used frequently for thermal management in 2D and 3D-MPSoCs. DVFS has a faster response time (μs range, 100-200μs) compared to the other thermal control techniques. However, DVFS typically implies a significant performance overhead [33].
- **Task scheduling and/or migration.** Job scheduling is an effective tool for reducing and balancing the temperature in MPSoCs [35-37]. This technique has a lower control decision frequency (fewer control actions per unit time) compared DVFS, as it relies on higher-level OS-based decisions, which are made at intervals on the order of tens of milliseconds.
- **Variant fluid flow rate.** Interlayer liquid cooling plays a major role in DTM of 3D stacked MPSoCs [38, 39]. Flow rate changes require hundreds of milliseconds as well as creating significant power overheads for high flow rates, as it relies on mechanical changes of the pumping network. On the other hand, this technique does not create any performance penalties, unlike DVFS or task migration, as it does not directly affect the workload of the processor.

B. Rule-Base Fuzzy Control

In this thermal management policy, we primarily rely on a combined design-time and run-time management policy. A schematic diagram of this design-time run-time policy is shown in Fig. 9. At design-time, we perform a thorough analysis of the aforementioned control elements to study their impact on the system temperature, thermal gradient, and energy consumption. We examine the impact of each control element value on individual aspects and on the overall system. An example of this analysis can be found in our previous work in [42].

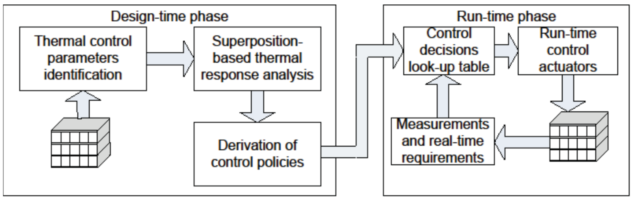

Design-time phase
Thermal control parameters identification
Superposition-based thermal response analysis
Derivation of control policies
Run-time phase
Control decisions look-up table
Run-time control actuators
Measurements and real-time requirements

Fig. 9: Schematic diagram of the design-time/run-time thermal management policy

We use this analysis results to derive the rule-base we use in the run-time fuzzy-logic controller. A general schematic diagram of this controller is shown in Fig. 10. This figure shows the construction modules in our proposed fuzzy controller.

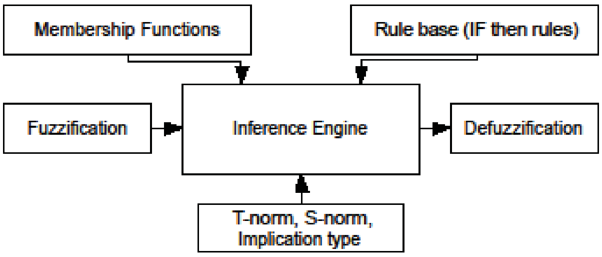


Fig. 10: Schematic diagram of the fuzzy-logic thermal controller

The controlled variables in this controller are the module temperatures, their physical locations, and the expected workload conditions. These variables are transformed from their numerical values to fuzzy-logic values using the fuzzification module. This transformation is crucial as the outputs from fuzzification are matched against the rule-base derived at design-time, to deduce the appropriate control action. More details on this controller implementation details are in our previous work [40].

One significant advantage of the rule-base fuzzy control is its lightweight, flexibility, and efficiency, as we show later.

C. Hierarchical-Based Model Predictive Control

In this management approach, we tackle the problem of the centralized controller scalability in systems with increasing complexity. Instead of using a single centralized controller with significant complexity, or the use of a complete distributed control with substantial communication overhead, we apply hierarchical control to our target 3D-MPSoCs. The structure of the proposed hierarchical thermal management system is shown in Fig. 11: the 3D-MPSoC architecture is partitioned into *p* tiers (or layers) where, without loss of generality, each tier is a subsystem of the 3D-MPSoC.

A tier consists of several units. These units could be cores or custom hardware blocks. Then, the units inside each tier, say tier *i*, are partitioned into *q*(*i*) frequency islands, and a local thermal controller manages the *q*(*i*) islands. The objectives of local controllers include preventing hot-spots and minimizing undone workload. Specific requirements (e.g. workload) come from a centralized unit (i.e., the *global thermal controller* in Fig. 11), which is responsible for the holistic coordination of the *p* local thermal controllers, and which regulates the heat extraction of the cooling system by setting the pressure of the coolant liquid (by controlling the cooling pump and/or the controlling valve).

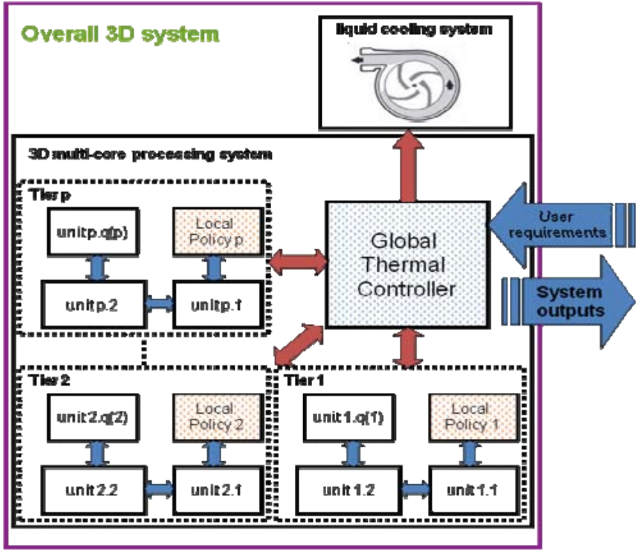


Fig. 11: Structure of the hierarchical thermal management system

This thermal management policy performs the control actions as follows: the global controller receives a workload requirement from the scheduler as well as a data vector containing the corresponding workload fulfillment status in each tier from all the *p* local controllers. This data vector contains two pieces of information: i) the maximum temperature measured online in the corresponding tier and ii) the already executed workload. The global unit splits the overall workload into *p* components. Hence, for each local controller, the global unit sets the amount of workload it has to execute. It is important to notice that the controller does not perform detailed task assignment, but just sets individual targets for each tier to satisfy the overall workload.

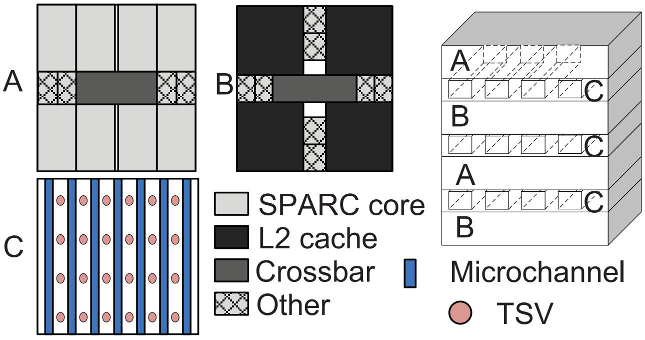


Fig. 12: Layout of the 3D-MPSoC used in our simulations

D. Simulation results

To evaluate the trade-offs in the two thermal management policies, we have applied both polices on a typical 3D-MPSoC shown in Fig. 12. The 3D-MPSoC we use in our experiments is based on the 90nm UltraSPARC T1 (Niagara-1) processor [43]. We have examined with power traces of actual workloads applied on the T1 processor. For brevity, we refer the reader to our work in [40] for more details on the power values and workload characteristics.

Our simulations show that both policies manage to maintain the temperature within the safe bounds (<85°C).

However, the *Hierarchical-Based Model Predictive Control* policy manages to maintain the thermal gradient within a single tier below 10°C, while *Rule-Base Fuzzy Control* has a higher thermal gradient within a single tier (15°). On the other hand, the overall thermal gradient in the whole 3D-MPSoC is lower when *Rule-Base Fuzzy Control* is applied compared to Hierarchical-*Based Model Predictive Control*. Moreover, when *Rule-Base Fuzzy Control* is applied, the system experiences a negligible computation overhead, which is about 0.1%. In contrast, when *Hierarchical-Based Model Predictive Control* is deployed, a significant computation overhead is experienced by the 3D-MPSoC. This overhead is due to model predictive control usage to deduce the appropriate control actions.

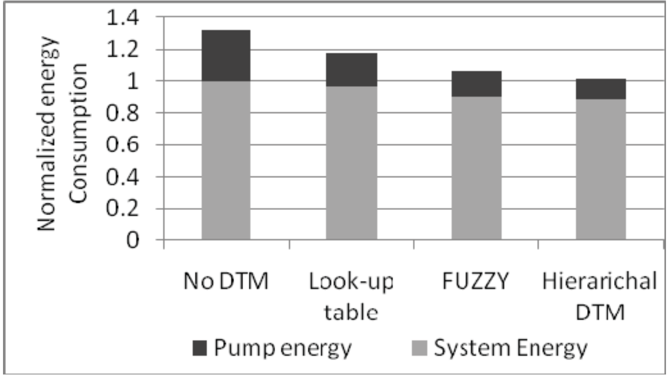


Fig. 13: System and pump energy consumption of the simulated 3D-MPSoC with different management policies with average workload. The results are normalized with respect to no DTM active case

In our simulations, we have also investigated the energy efficiency of our proposed DTM policies. Fig. 13 shows the 3D-MPSoC system energy consumption, as well as the liquid pump energy consumption. The values shown in Fig. 13 are normalized to the case when no DTM is applied to the targeted 3D-MPSoC. We also compare the energy efficiency with respect to the state-of-the-art look-up table-based DTM [38]. This figure shows that our both applied techniques manage to reduce the energy consumption by up to 60% compared to the worst-case design. Moreover, the results indicate that the *Hierarchical-Based* management policy reduces the pump energy consumption by an additional 40% when compared to *Rule-Base Fuzzy* policy.

From the above results, there is a clear tread-off between these two policies, which opens new directions in finding new policies that could utilize the benefits of the different management schemes towards energy-efficient 3D-MPSoCs that will be used in future data-centers.

IV. Towards Zero-Emission Data-centers

As discussed in Section I, the development of 3D-MPSoCs has profound implications for the future of data-centers, which represent high performance computing installations supporting the continuous progress of IT services. In particular, 3D-MPSoCs have the potential to make data-centers more compact and efficient in terms of computational throughput. However, their incorporation in future data-centers requires a comprehensive development of the innovative solutions presented in the previous sections on a large-scale to tackle the economic and environmental issues they bring to the already over-stressed cooling infrastructures of today's data-centers.

With increased global efforts towards reducing carbon emissions for combating climate change and dependence on fossil fuels [44], data-centers have become a major focus of these efforts. This is because the energy consumption of data-centers has soared to about 2% of the global electricity consumption and contributing to CO_2 emission-levels that are comparable to those of the aviation industry [45]. With 50% of this energy consumption going into the cooling infrastructure, innovative solutions, both hardware and software, are needed to reduce the emissions of data-centers.

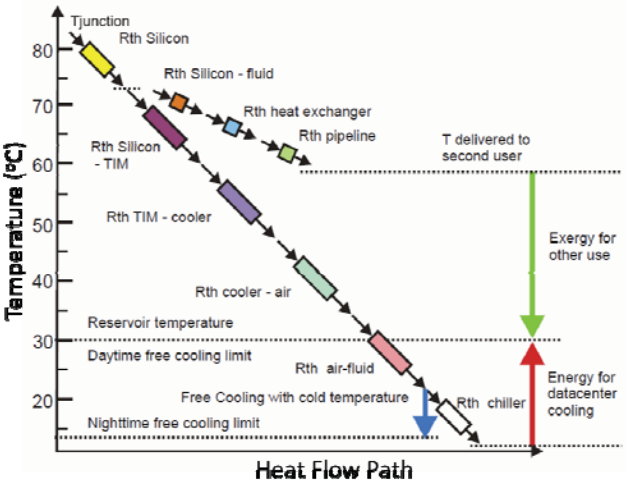


Fig. 14: Thermal resistances from the transistor junction to the ambient for a normal air-cooled datacenter ($R_{th,\ silicon} + R_{th,\ silicon\text{-}TIM} + R_{th,\ TI\ \text{-}cooler} + R_{th,\ cooler\text{-}air} + R_{th,\ air\text{-}fluid} + R_{th,\ chiller}$) and for an optimized liquid-cooled datacenter where all resistances have been minimized and the number of components is smaller ($R_{th,\ silicon} + R_{th,\ silicon\text{–}fluid} + R_{th,\ heat\ exchanger} + R_{th,\ pipeline}$) [47]

Since microelectronic heating due to processor activity is the biggest culprit in the generation of heat in data-centers, the presented chip-level liquid cooling of 3D-MPSoCs with system-level DTM to control the use of this cooling(cf. Section 2 and 3), is envisioned to transform the thermal reliability issues of data-centers into an opportunity for reducing their carbon-footprint. In fact, this approach considerably lowers the overall temperature gradient in the system (of the order of 10°C) compared to air cooling solutions, enabling the operation of transistors at the maximum allowable temperatures (~85°C) while using coolants at temperatures as high as 60-70°C. Hence, the direct reuse of this heat becomes possible, as illustrated in Fig. 14.

The economic viability of such direct reuse of heat by distributing the output heat of the data-centers into the local district heating systems in cold countries has been demonstrated by researchers at IBM Zurich [46], in conjunction with EPFL and ETHZ, using computer models and data from various countries in the European Union. The proposed solutions can reduce the carbon footprint of data-centers to zero without resorting to financial carbon-offset instruments or relying on emission-free electricity sources. In hot climates, these innovative "hot-water" cooling systems reduce the energy consumption by removing the necessity to provide high performance chillers to cool down the coolant to very low temperatures. Moreover, the latest results have outlined that these hardware solutions need to be used at system-level, i.e., in conjunction with the proposed adaptive software task migration and DTM techniques to adapt to the variable demand of computing resources, enabling large reductions in the energy consumption (and emissions) of data-centers.

A. *Aquasar*

A significant first step at realizing the proposed zero-emission data-center has been the Aquasar project, which is a new type of hot-water cooled supercomputer jointly built by the Swiss Federal Institutes of Technologies (EPFL and ETHZ) and the IBM Research Laboratory in Zurich [47]. This supercomputer consists of 33 IBM BladeCenter® QS22 (5 TFlops) and 9 IBM BladeCenter® HS22 and has an efficiency of more than 400 MFlops per Watt. The blades are equipped with high-performance micro-channel liquid coolers mounted directly on the MPSoCs (Fig. 15).

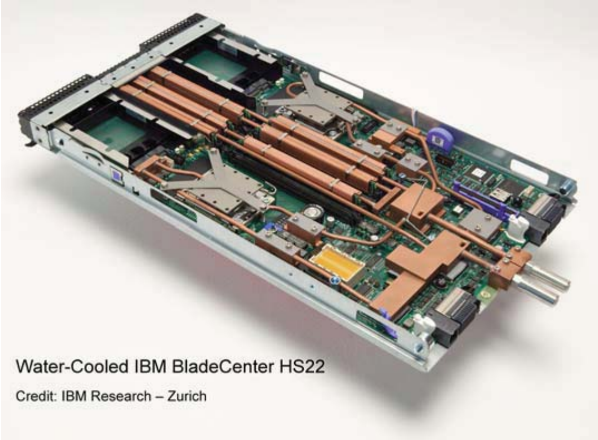

Water-Cooled IBM BladeCenter HS22
Credit: IBM Research – Zurich

Fig.15: Water-cooled IBM BladeCenter HS22 with two high-performance micro-channel liquid coolers that remove heat from the microprocessors and various further heat sinks that collect heat from other components [Courtesy: IBM RL, Zurich]

This innovative cooling system reduced the energy consumption of the supercomputer by up to 40% and its carbon footprint by up to 85% compared to similar systems using conventional air-cooling technology. The low carbon footprint is possible because the excess heat is used to heat the university buildings, as illustrated in Fig. 16. The fluid loops of the individual blades link to the larger network of the server rack, which in turn is connected to the main water transportation network. A heat exchanger transfers the excess heat from the coolant and feeds it directly into the heating system of ETHZ.

V. CONCLUSIONS

In this article we have presented a thorough review of the state-of-the-art system-level thermal modeling and management techniques for 3D stacked MPSoCs cooled using microchannel liquid-cooled heat sinks. In particular, we have summarized the research efforts currently undertaken by the Embedded Systems Laboratory (ESL) at EPFL in these directions, including 3D-ICE and the different system-level dynamic thermal management policies. Finally, a vision of zero-emission data-centers of the future, as the culmination of these joint research efforts by EPFL, IBM Zurich and ETHZ, has been presented, including the recently built and fully functioning Aquasar hot-water cooled supercomputer.

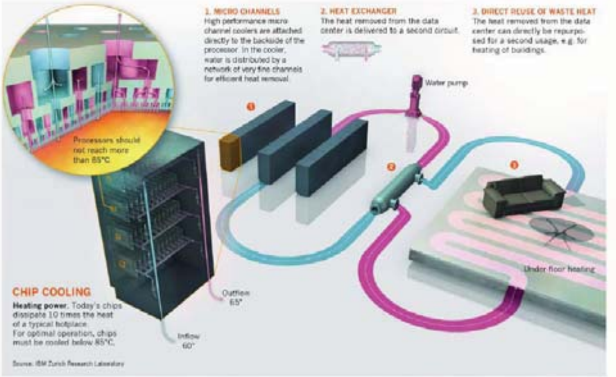

1. MICRO CHANNELS
High performance micro channel coolers are attached directly to the backside of the processor. In the cooler, water is distributed by a network of very fine channels for efficient heat removal.
2. HEAT EXCHANGER
The heat removed from the data center is delivered to a second circuit.
3. DIRECT REUSE OF WASTE HEAT
The heat removed from the data center can directly be repurposed for a second usage, e.g. for heating of buildings.
Water pump
Processors should not reach more than 85°C
Under floor heating
CHIP COOLING
Heating power. Today's chips dissipate 10 times the heat of a typical hotplace.
For optimal operation, chips must be cooled below 85°C.
Outflow 65°
Inflow 60°
Source: IBM Zurich Research Laboratory

Fig. 16: Schematic concept of the zero-emission data-center. Heat is collected from the individual microelectronic components and transferred via a heat exchanger to a district heating system to be used for space heating [Courtesy: IBM RL, Zurich]

ACKNOWLEDGMENT

We would like to thank Prof. Ayse Coskun of Boston University, MA, USA, and Thomas Brunschwiler of IBM Research Laboratory, Zurich, for providing us with their valuable inputs in the preparation of this manuscript. The research efforts discussed in this paper have been partially funded by the Nano-Tera RTD project CMOSAIC (ref.123618), which is financed by the Swiss Confederation and scientifically evaluated by SNSF, and the PRO3D project (ref. FP7-ICT-248776) financed by the European Community within its 7th Framework Programme.

REFERENCES

[1] C. Mack, "Fifty Years of Moore's Law," *IEEE Transactions on Semiconductor Manufacturing* , vol.24, no.2, pp.202-207, May 2011.

[2] K. Nomura, K. Abe, S. Fujita, Y. Kurosawa and A. Kageshima, "Performance analysis of 3D-IC for multi-core processors in sub-65nm CMOS technologies," *Proc. International Symposium on Circuits and Systems* (ISCAS 2010), pp.2876-2879, May 2010.

[3] F. Li, C. Nicopoulos, T. Richardson, X. Yuan, V. Narayanan, M. Kandemir, "Design and Management of 3D Chip Multiprocessors Using Network-in-Memory", *Proc. International Symposium on Computer Architecture* (ISCA 2006), pp.130-141, 2006.

[4] T. Brunschwiler, B. Michel, H. Rothuizen, U. Kloter, B. Wunderle, H. Oppermann and H. Reichl, "Forced convective interlayer cooling in vertically integrated packages", *Proc. Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems* (ITHERM 2008), pp.1114-1125, May 2008.

[5] D. Tuckerman, R. Pease, "High-performance heat sinking for VLSI" *IEEE Electron Device Letters*, vol.2, no.5, pp. 126- 129, May 1981.

[6] E. Pop, S. Sinha, and K. Goodson, "Heat Generation and Transport in Nanometer-Scale Transistors", *Proc. of the IEEE (Special Issue on Thermal Engineering)*, vol.94, no.8, pp.1587-1601, Aug. 2006.

[7] M. Vogel, "Liquid cooling performance for a 3-dimensional multichip module and miniature heat sink", *Proc. Semiconductor Thermal Measurement and Management Symposium* (SEMI-THERM 1994), pp.73-77, Feb 1994.

[8] F. Incropera, *Liquid Cooling of Electronic Devices by Single-Phase Convection*, John Wiley & Sons Inc., New York, 1999.

[9] K. Moores, Y. Joshi, and G. Schiroky, “Thermal Characterization of a Liquid Cooled AlSiC Base Plate with Integral Pin Fins”, *IEEE Trans. on Components and Packaging Technologies*, Vol. 24, No. 2, 213-219, 2001.

[10] X. Wei and Y. Joshi, “Stacked Microchannel Heat Sinks for Liquid Cooling of Microelectronic Components”, *ASME Transactions Journal of Electronic Packaging*, Vol. 126 No. 1, 60-66, March 2004.

[11] F. Alfieri, M. Tiwari, I. Zinovik, D. Poulikakos, T. Brunschwiler and B. Michel, “3D Integrated Water Cooling of a Composite Multilayer Stack of Chips”, *Journal of Heat Transfer*, Vol. 132, Issue 12, Dec 2010.

[12] B. Dang, M. Bakir, D. Sekar, C. King and J. Meindl, "Integrated Microfluidic Cooling and Interconnects for 2D and 3D Chips", *IEEE Trans. Advanced Packaging*, vol.33, no.1, pp.79-87, Feb. 2010.

[13] B. Agostini, M. Fabbri, J. Park, L. Wojtan, J. Thome and B. Michel, “State-of-the-Art of High Heat Flux Cooling Technologies”, *Heat Transfer Engineering*, vol. 28 no. 4, pp. 258–281, 2007.

[14] W. Huang, S. Ghosh, S. Velusamy, K. Sankaranarayanan, K. Skadron and M. Stan, "HotSpot: a compact thermal modeling methodology for early-stage VLSI design", *IEEE Trans. Very Large Scale Integration (VLSI) Systems*, vol.14, no.5, pp.501-513, May 2006.

[15] T. Wang and C. Chen, “3-D thermal-ADI: a linear-time chip level transient thermal simulator,” *IEEE Trans. Computer-Aided Design*, vol.21, pp.1434–1445, December 2002.

[16] S. Im and K. Banerjee, “Full-chip thermal analysis of planar (2D) and vertically integrated (3D) high performance ICs,” *International Electron Devices Meeting (IEDM 2000) Technical Digest*, pp.727–730, 2000.

[17] P. Li, L. Pileggi, M. Asheghi and R. Chandra, “Efficient full-chip thermal modeling and analysis”, *Proc. International Conference on Computer-Aided Design (ICCAD 2004)*, pp.319–326, November 2004.

[18] Y. Cheng, P. Raha, C. Teng, E. Rosenbaum and S. Kang, “ILLIADS-T: An electrothermal timing simulator for temperature-sensitive reliability diagnosis of CMOS VLSI chips”, *IEEE Trans. Computer-Aided Design for ICs and Systems*, vol.17, pp.668–681, August 1998.

[19] FlowTherm. URL: http://www.mentor.com/products/mechanical/products flotherm

[20] J.Lienhard-IV and J.Lienhard-V, *A Heat Transfer Textbook*, Cambridge, Massachusetts: Phlogiston Press, 2006.

[21] F. Incropera, D. Dewitt, T. Bergman and A. Lavine, *Fundamentals of Heat and Mass Transfer*, New York: John Wiley and Sons, 2007.

[22] W. Qu and I. Mudawar, “Thermal design methodology for high-heat flux single-phase and two-phase microchannel heat sinks” *IEEE Trans. Components and Packaging Technology*, vol.26, pp.598–609, 2003.

[23] X. Wei and Y.Joshi, “Optimization study of stacked micro-channel heat sinks for micro-electronics cooling” *IEEE Trans. Components and Packaging Technology*, vol.26, no.1, pp.55–61, 2003.

[24] Lei, N., A. Ortega, R. Vaidyanathan, “Modeling and optimization of multilayer mini-channel heat sinks in single phase flow,” *Proceedings of InterPACK 2007*, Vancouver, B.C., Canada, 2007.

[25] J. Koo, S. Im, L, Joang and K. Goodson, “Integrated microchannel cooling for three-dimensional electronic circuit architectures,” *ASME Journal of Heat Transfer*, vol.127, pp.49–58, 2005.

[26] T. Brunschwiler, B. Michel, H. Rothuizen, U. Kloter, B. Wunderle, H. Oppermann and H. Reichl, "Interlayer cooling potential in vertically integrated packaes", *Journal of Microsystems Technology,* (ITHERM 2008), pp.1114-1125, May 2008.

[27] A. Sridhar, A. Vincenzi, M. Ruggiero, T. Brunschwiler and D. Atienza, “3D-ICE: Fast Compact Transient Thermal Model for 3D ICs with Inter-tier liquid cooling”, Proc. International Conference on Computer-Aided Design (ICCAD 2010), pp. 463-470, November 2010.

[28] A. Sridhar, A. Vincenzi, M. Ruggiero, T. Brunschwiler and D. Atienza, “Compact Transient Thermal Model for 3D ICs with liquid cooling via Enhanced Heat Transfer Geometries”, Proc. International Workshop on Thermal Investigations of ICs and Systems (Therminic 2010), pp. 1-6, October 2010.

[29] 3D-ICE. URL: http://esl.epfl.ch/3D-ICE

[30] R. Shah and A. London, *Laminar flow forced convection in ducts*, New York: Academic Press, 1978.

[31] T. Brunschwiler, S. Paredes, U. Drechsler, B. Michel, W. Cesar, Y. Leblebici, B. Wunderle, and H. Reichl, "Heat-removal performance scaling of interlayer cooled chip stacks", *Proc. IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm 2010)*, pp.1-12, June 2010.

[32] T. Brunschwiler, S. Paredes, U. Drechsler, B. Michel, W. Cesar, G. Toral, Y. Temiz and Y. Leblebici, "Validation of the porous-medium approach to model interlayer-cooled 3D-chip stacks", *Proc IEEE Conference on 3D System Integration (3DIC 2009)* , pp.1-10, Sept 2009.

[33] A. Coskun, T. Rosing, and K. Gross, “Utilizing predictors for efficient thermal management in multiprocessor SoCs”, *IEEE Transactions on Computer-Aided Design*, vol. 28 no.10, pp 1503–1516, 2009.

[34] K. Puttaswamy and G. H. Loh, “Thermal analysis of a 3D die-stacked high-performance microprocessor”, *Proc. Great Lakes Symposium on VLSI (GLSVLSI 2006)*, pp. 19–24, Apr 2006.

[35] A. Coskun, J. Ayala, D. Atienza, T. Rosing, and Y. Leblebici, “Dynamic thermal management in 3D multicore architectures”, *Proc. Design Automation and Test in Europe Conference (DATE 2009)*, pp. 1410–1415, Apr.. 2009.

[36] X. Zhou, Y. Xu, Y. Du, Y. Zhang and J. Yang, "Thermal Management for 3D Processors via Task Scheduling”, Proc. International Conference on Parallel Processing (ICPP 2008), pp.115-122, Sept. 2008.

[37] C. Zhu, Z. Gu, L. Shang, R. Dick and R. Joseph, "Three-Dimensional Chip-Multiprocessor Run-Time Thermal Management", *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, vol.27, no.8, pp. 1479-1492, Aug. 2008

[38] A. Coskun, D. Atienza, T. Rosing, T. Brunschwiler and B. Michel, "Energy-efficient variable-flow liquid cooling in 3D stacked architectures," *Proc. Design, Automation & Test in Europe Conference & Exhibition (DATE 2010)*, pp.111-116, March 2010

[39] K. Matsumoto, S. Ibaraki, M. Sato, K. Sakuma, Y. Orii and F. Yamada, "Investigations of cooling solutions for three-dimensional (3D) chip stacks", *Proc. Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM 2010)*, pp.25-32, Feb. 2010.

[40] M. M. Sabry, A. Coskun, and D. Atienza, “Fuzzy control for enforcing energy efficiency in high-performance 3D systems”, *Proc. International Conference on Computer-Aided Design (ICCAD 2010)*, pp. 642–648, Nov. 2010.

[41] F. Zanini, M. M. Sabry, D. Atienza, and G. De Micheli “Hierarchical Thermal Management Policy for High-Performance 3D Systems with Liquid Cooling”, *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 1, no. 2, 2011.

[42] M. M. Sabry, D. Atienza, and A. Coskun “Thermal Analysis and Active Cooling Management for 3D-MPSoCs”, *Proc. IEEE Symposium on Circuits and Systems (ISCAS, 2011)*, pp. 2237-2240, May 2011.

[43] A. Leon, J. Shin, K. Tam, W. Bryg, F. Schumacher, P. Kongetira, D. Weisner, and A. Strong, "A Power-Efficient High-Throughput 32-Thread SPARC Processor", *Proc. International Solid-State Circuits Conference (ISSCC 2006)*, pp. 295-304, Feb 2006.

[44] International Panel on Climate Change, *Climate Change 2007-Mitigation of Climate Change: Contribution of Working Group III to the Fourth Assessment Report of the IPCC*, Cambridge University Press, Cambridge, ISBN 978 0521 70598-1.

[45] G. Meijer, “Cooling Energy-Hungry Data-centers”, *Science Magazine*, Vol. 328 no. 5976 pp. 318-319, April 2010.

[46] T. Brunschwiler, B. Smith, E. Ruetsche and B. Michel, “Toward zero-emission data-centers through direct reuse of thermal energy”, *IBM Journal of Research and Development*, vol. 53, no. 3, Nov 2009.

[47] T. Brunschwiler, G. Meijer, S. Paredes, W. Escher and B. Michel, “Direct Waste Heat utilization from liquid-cooled Supercomputers”, *Proc. International Heat Transfer Conference (IHTC 2010)*, pp. 1-12, Aug 2010.