

Foundations and Trends® in Electronic Design
Automation
Vol. 8, No. 2 (2014) 117–197
© 2014 M. M. Sabry and D. Atienza
DOI: 10.1561/10000000032



Temperature-Aware Design and Management for 3D Multi-Core Architectures

Mohamed M. Sabry and David Atienza
Embedded Systems Lab (ESL)
Ecole Polytechnique Fédérale de Lausanne (EPFL)
1015, Lausanne, Switzerland
mohamed.sabry@epfl.ch, david.atienza@epfl.ch

Contents

1	Introduction	118
1.1	3D-ICs for Augmented Performance Per Unit Area	120
1.2	Thermal Issues in 3D MPSoCs	122
1.3	Thermal Impact on 3D MPSoC Reliability and Performance	124
1.4	Advanced Cooling Technologies for 3D MPSoCs	128
1.5	Survey Overview and Outline	131
2	Thermal Modeling for 3D MPSoCs	134
2.1	Heat Transfer Principles	134
2.2	Thermal Modeling Frameworks for 2D/3D ICs	137
2.3	Compact Thermal Modeling for 3D MPSoCs	138
2.4	Summary	142
3	Temperature-Aware Design Optimizations for 3D MPSoCs	144
3.1	On-Chip Design-Time Optimizations	146
3.2	Off-Chip Design-Time Optimizations	152
3.3	Summary	162
4	Temperature-Aware Runtime Management for 3D MPSoCs	163
4.1	Temperature-Affecting Control Knobs	165
4.2	Centralized Versus Decentralized Management Schemes . .	166
4.3	Reactive Versus Proactive Management Schemes	171

4.4	Heuristic Versus Algorithmic Management Schemes	176
4.5	The Global Perspective on Thermal Management	182
4.6	Summary	184
Conclusion		185
Acknowledgements		188
References		189

Abstract

Vertically-integrated 3D multiprocessors systems-on-chip (3D MP-SoCs) provide the means to continue integrating more functionality within a unit area while enhancing manufacturing yields and runtime performance. However, 3D MPSoCs incur amplified thermal challenges that undermine the corresponding reliability. To address these issues, several advanced cooling technologies, alongside temperature-aware design-time optimizations and run-time management schemes have been proposed. In this monograph, we provide an overall survey on the recent advances in temperature-aware 3D MPSoC considerations. We explore the recent advanced cooling strategies, thermal modeling frameworks, design-time optimizations and run-time thermal management schemes that are primarily targeted for 3D MPSoCs. Our aim of proposing this survey is to provide a global perspective, highlighting the advancements and drawbacks on the recent state-of-the-art.

Keywords: System-Level Design, Thermal Management, MPSoC Cooling, Temperature Optimization, Reliability, Vertical Integration.

1

Introduction

The last decades have seen a revolution in data gathering, processing, information storage and communication. This revolution has been caused by electronic computing systems, which nowadays are one of the key building blocks of the world's information technology (IT) infrastructure. In fact, computing systems and IT services are an essential pillar of the developed world, contributing up to 50% of its economy [1]. The IT and computing systems revolutions have been the result of the advancements in IC processing technology, where the number of components (transistors) on the same die area have been doubled every 18 months [2], which is also known as Moore's law. This has been the drive to generate more complicated computing systems with higher performance and computational functionality.

As feature sizes scale with advanced processing technologies, the performance of processing units has increased because of greater functionality and higher computational capabilities. This functionality augmentation was accompanied by an increase in the operating frequency of the processing unit. Micro-architects have conventionally used operating frequency as a measure for the processing unit performance, as higher frequency implies more instructions executed per unit time.

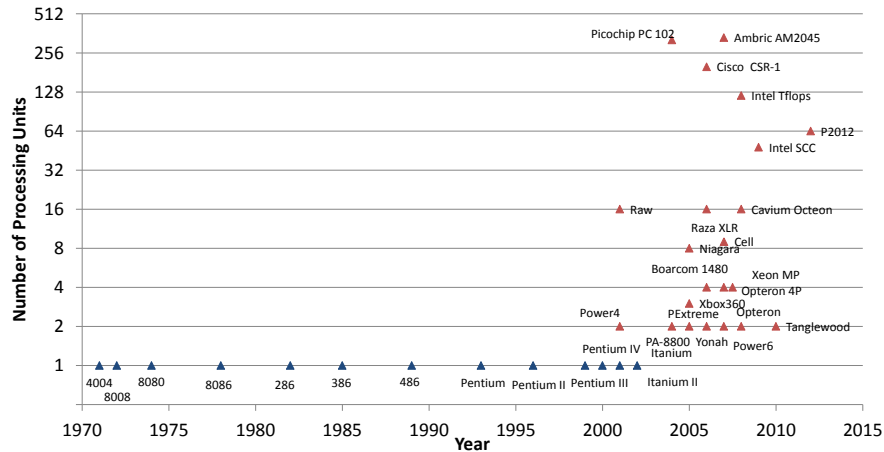


Figure 1.1: Number of processing units integrated in a single IC, as evolved with time. Blue-labeled ICs are single-core architectures, while red-labeled ICs are multi-core architectures.

However, at the sub-micron level, the circuit-level delays are getting dominated by wiring and interconnect delays, which led to frequency flattening. If the operating frequency would increase with this technology, significant additional power consumption is required by the processing unit, which results in an increased heat generation. For example, a 90 nm fabricated AMD processing unit would require 60% additional power consumption to increase the operating frequency by 400MHz [3].

Multi-core architectures have been proposed as an alternative design paradigm to frequency increase in single-core architectures, to continue performance improvement with technology scaling [4]. Multi-core architectures integrate two or more processing units, with shared or distributed memory modules, interconnected through an on-chip bus or a network-on-chip [5]. As feature sizes scale with advanced processing technologies, the number of processing units in multi-core architectures dramatically increases. Fig. 1.1¹ shows that the number of cores inte-

¹This figure is based on a similar figure found in the course slides given by Prof. S. Amarasinghe of MIT <http://groups.csail.mit.edu/cag/ps3/pdf/6.189-info-session.pdf>

grated in a single IC has started to ramp-up in the beginning of the 21st century. Recent multi-core architectures integrate a number of processing units, multi-level memory hierarchy, an interconnect module, and in the case of embedded domain, special peripherals such as analog-to-digital converters (ADC), co-processors, or wireless RF antennas. This architecture, which is known as *Multiprocessor Systems-on-Chip* (MPSoC), has been widely used in various domains. An example of recent MPSoC utilization at the high performance computing systems level is data-intensive computing systems, which is also known as the *fourth paradigm* [6]. Another example for MPSoC utilization is found at the embedded systems domain, with small- or tiny-size computing systems, such as on-body [7] or in-body [8] health monitoring systems.

1.1 3D-ICs for Augmented Performance Per Unit Area

While the performance of computing architectures has been enhanced by MPSoCs, MPSoCs' performance has been recently challenged by increased propagation delay, primarily due to longer interconnects [9]. This is mainly due to the increased wire-to-gate delay ratio. This delay would lead to degraded MPSoC performance or increased energy consumption.

This delay limitation, combined with the continued demands for increased integrated functionality while preserving the performance and area efficiency, have led to the development of vertically-integrated 3D ICs. 3D integration is viewed as an attractive solution to provide increased functionality with better yield, as well as a technique of combining several technologies in a single enclosure (package) [10]. From a design perspective, 3D integration can be split into the following categories:

- **Monolithic 3D integration** [11]. This integration technology fabricates the tiers serially at the transistor granularity, within a single fabrication process. From the bottom tier, the corresponding transistors are fabricated then a substrate layer is placed on top where another tier is fabricated. These layers are connected with vertical interlayer vias. Thus, this integration technology is

promising in terms of providing higher density and performance gains.

- **3D stacking** (also referred to as Parallel 3D integration) [10]. This integration technology stacks vertically 2D die layers to form a single 3D IC. To enable the communication and power delivery to these dies, there are several techniques adopted such as wire bonding, microbumps, and *through-silicon vias* (TSVs). TSVs are vertical wires that carry power and signals between different dies, which are etched in the silicon substrate between the 2D dies.

In this survey, we primarily target 3D stacked ICs with several digital logic dies and TSV-based interlayer communication. We refer to the targeted 3D ICs throughout this review by 3D multiprocessors systems-on-chip, or 3D MPSoCs.

3D MPSoCs are multi-layered stacked 3D ICs, where each die contains a number of processing units, memory modules, and other peripheral and interconnection units. Examples of typical 3D MPSoCs integrate a number of processing layers that contain all the processing units, and a number of memory layers. Another 3D MPSoC example is where the processing units and the memory modules are co-placed in each die of the 3D stacked layers. These examples have been shown in previous works [12, 13], and are shown in Fig. 1.2, which includes an UltraSPARC T1 [14] version of a 3D MPSoC. This vertical integration of logical modules brings several benefits to multi-core architectures, which are as follows:

- Vertical stacking shortens the wiring length between two modules. In this respect, the propagation delays, which are recently dominated by interconnect delays [15], are dramatically reduced leading to an increased performance of 3D MPSoCs. Thus, 3D MPSoCs would outperform 2D MPSoCs.
- 3D MPSoCs allow heterogeneous integration of different components, such as DRAM on multi-core architectures [13]. 3D MPSoCs enhance the memory access bandwidth and throughput, by bringing the memory modules (e.g. DRAM) to the top or bottom

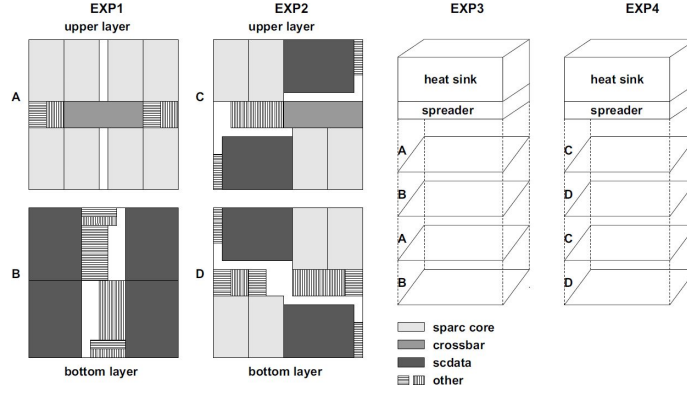


Figure 1.2: Schematic view of a 4-tier 3D MPSoC with different architectures [12].

of processing layers. This would enable high-speed, massively-parallel data access to these stacked DRAM layers.

1.2 Thermal Issues in 3D MPSoCs

Despite the performance and throughput enhancements that 3D MP-SoCs bring, 3D MPSoC designs face major challenges, particularly in the extreme elevated temperatures accompanied with high-performance designs.

While technology continues to scale-down the transistor features, MPSoCs voltage supply (V_{dd}) could not be scaled down accordingly [16]. Recent work [17] (Fig. 1.3(a)) has shown that the supply voltage scaling is saturating that, if combined with increased integration within a unit area due to reduced transistor features, leads to an increase in power consumption. In this respect, multi-core architectures design trends have taken the direction of increasing the power density by integrating more processing units on the chip (with a fixed chip area), as shown in Fig. 1.3(b). If the MPSoC power density keeps increasing, it will eventually reach the same magnitude of nuclear power plants [4, 18].

With such increased power densities, MPSoCs face a tremendous

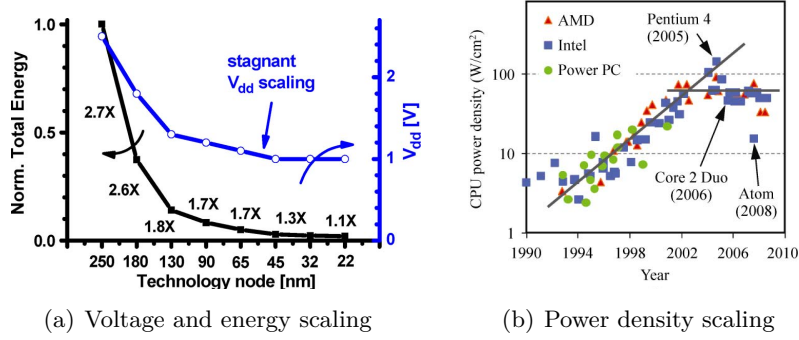


Figure 1.3: Voltage/Energy [17] and power density [18] scaling trends.

increase in heat generation that has a direct impact on the lifetime of MPSoCs. This increased heat generation leads to high temperature in the MPSoCs. While high-frequency single processing units have faced the similar case of high temperatures, the thermal profile of MPSoCs can have a more severe impact. This is mainly related to the localized heat generation of the several processing units of MPSoCs. Thus, the localized heat generation creates several localized high temperatures, which is known as thermal *hot spots*. The existence of several thermal *hot spots* would imply that there are other localized cold spots, which leads to the creation of the undesirable spatial *thermal gradients*. Moreover, the time-varying nature of workload processing requirements, or even when the processing goes to power-up and power-down cycles, leads to temporal *thermal cycles* [19] formation. To demonstrate the MPSoC thermal issues, an example of various hot spots location and thermal gradient is shown in Fig. 1.4(b). This figure shows a thermal response snapshot of the UltraSPARC T1 (Niagara) [20] MPSoC to a typical workload execution.

While high-density 2D MPSoCs face strong thermal challenges, these challenges are more prominent in the vertically-stacked 3D MPSoCs [21, 12]. Due to the vertical stacking of different dies, the thermal resistance of 3D MPSoCs is significantly increased to alarming values [12], compared to the increased temperature we demonstrate in the case of 2D MPSoCs. This is mainly due to the increased and non-

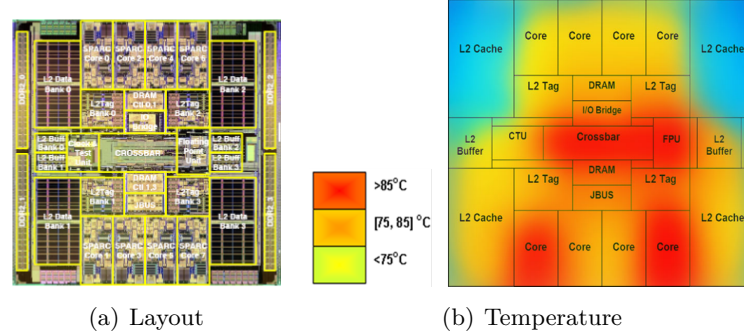


Figure 1.4: Floorplan layout and thermal response of the UltraSPARC T1 MP-SoC [20].

uniform thermal resistance at different stacked layers, based on their relative heat dissipation paths using conventional techniques such as placing heat sink on top of the top-most layer. For example, Fig. 1.5 shows the temperature of an emulated 3D MPSoC. This emulator is built by stacking 4 heat dissipation tiers on a substrate tier. In each of the heat dissipation tiers there is a number of controllable micro-heaters that are used to emulate the heat generation pattern of processing units similar to the actual pattern of each processing unit. In addition to these micro-heaters, there is a number of thermal sensors to capture the thermal profile of this emulator. This 3D thermal emulator has a heat sink placed at the top-most layer. The temperatures shown in this figure indeed confirm the expected high temperature and thermal gradient values of prospective high-performance 3D MPSoCs.

Thus, it is expected that high-density high-performance 3D MP-SoCs are more prone to hot spots and thermal gradients. The existence of hot spots, thermal gradients, and thermal cycles heavily affect the MPSoC (2D and 3D) operation and lifetime, as shown in the following section.

1.3 Thermal Impact on 3D MPSoC Reliability and Performance

High temperature is undesirable in 3D MPSoCs operation due to the different device and interconnect reliability and degraded performance

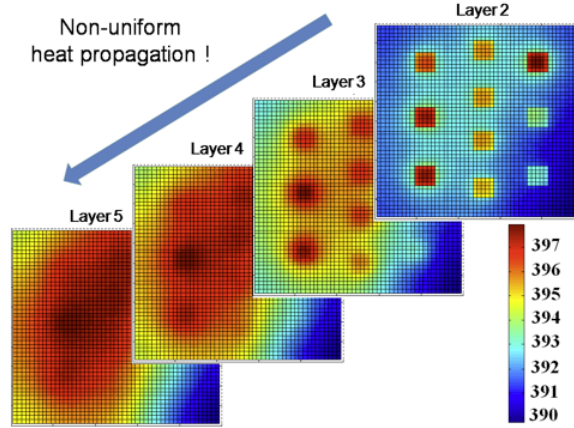


Figure 1.5: Thermal state of a 5 tier (4 thermal dissipating tiers + 1 substrate base tier) high-density MPSoC thermal emulator. Temperature values are in Kelvin [22].

sources that are highly affected, directly and indirectly, by this rise in temperature. These sources would affect the reliability of 3D MP-SoCs by accelerating the processor aging or the *Mean-Time to Failure* (MTTF) [23], which is the statistical average time for the MP-SoC to breakdown permanently, as well as creating irreversible functional failures in the computation modules (e.g., storage) that limit the full utilization of these modules. In addition to the impacted reliability, high 3D MPSoC temperature would eventually lead to degraded performance by reducing the operating frequency due to increased propagation delays or reduced energy-efficiency resulted from the increased leakage power consumption. Thus, it is important to identify the temperature-induced reliability and degraded performance sources and elaborate more on their corresponding impact. The following paragraphs give an overview of these sources:

Bias Temperature Instability (BTI) [24]. This factor causes instabilities in the device behavior, due to the stress applied on the bias (e.g., a negative bias on the gate source voltage of a PMOS transistor). BTI can be split to two types, namely Negative BTI, which is related to the PMOS device stress, and Positive BTI, which is related to the NMOS device stress. The main degraded parameter due to BTI is the

threshold voltage, as shown in prior work [24]. The change in threshold voltage during stress (increase in threshold voltage) and release times (decrease in threshold voltage) has a dependency on a number of factors, which includes temperature [25]. Higher operating temperatures indeed have a direct impact on the threshold voltage based on BTI, which results in longer circuit delays and increase in dynamic power consumption.

Hot-Carrier Injection [26]. Hot-carrier injection occurs when a carrier gains enough energy to tunnel from the transistor source or drain to the dielectric material. This even accompanied with a rise in the device temperature. Hot-carrier injection occurs at normal temperature range, but the injection rate is increased as the operating (or stress) temperature is increased [27]. Based on the above observations, hot-carrier injection could lead to thermal positive feedback loop situation, i.e., the injection leads to increase in temperature that may trigger an increase in the injection rate. Consequently, hot-carrier injection would lead to thermal run-away.

Time-Dependent Dielectric Breakdown (TDDB) [28, 29] in high-k device dielectric and low-k interconnect dielectric. This is modeled as trap generated that leads to a leakage path through the oxide layer of the transistor. This is also referred in literature as *gate oxide breakdown*. TDDB has an exponential dependency on temperature [28] that accelerates the failure of a transistor by breaking down the dielectric, hence forming a constant conducting path. As a consequence, the faulty transistor would be permanently in a conducting state.

ElectroMigration (EM) in metallic interconnect [30]. Electromigration is a phenomenon that occurs in the IC interconnects (metal layers) due to high current densities. EM leads to a shift in the conducting ions location, hence causing a breakdown in the interconnects. EM has a strong dependency on the temperature resulted from the utilization of the IC and the joule self-heating of the interconnects due to the high current density. In this respect, higher operating temperatures would

eventually lead to a breakdown of the metal layer, which may result in a complete failure of the IC.

Subthreshold Leakage Current [31]. Subthreshold leakage current is one of the sources of leakage current, and hence static power consumption in MPSoCs. It is the drain-source current of a transistor operating in the weak inversion region. Unlike the strong inversion region in which the drift current dominates, the subthreshold conduction is due to the diffusion current of the minority carriers in the channel. Subthreshold leakage is found to be the dominant component in the overall leakage current sources [32]. Subthreshold leakage current has a strong dependency on the operating temperatures with a sensitivity of $8 - 12x/100^{\circ}C$ [31].

The aforementioned sources can lead to system failure if no proper measures are taken. But these sources also degrade the MPSoCs (2D and 3D) operation from its original (also called time-zero) operating conditions. The following paragraphs highlight more on these affected parameters:

Mean-Time To Failure [23, 31]. The mean-time to failure would be heavily impacted due to the temperature impact on Time-Dependent Dielectric Breakdown (TDDB), Electromigration (EM), as well as stress migration and thermal cycling.

Temperature-Dependent Propagation Delays [33]. This change in the time required for a signal to travel between two modules is related to the thermally-induced delays in the logic gates (e.g. resulted from BTI), as well as the increase of the interconnect resistivity (e.g. resulted from Electro-migration). Another cause for the propagation delay is in the clock skew between different modules experiencing diverse thermal stress. In this respect, propagation delays have a strong dependency not just on the overall thermal state of the IC, but on the spatial thermal gradient as well. Indeed, previous work reports that a spatial gradient of $40^{\circ}C$ would create a 10% clock skew between different modules within a single IC.

Temperature-Dependent Leakage Power [31]. Leakage power is one of the sources of static power consumption in MPSoCs. The leakage power is a cause of various elements, such as the reverse-biased junction leakage current and the subthreshold leakage. These elements have a strong dependency on the operating temperatures. In fact, it has been shown by previous work that the leakage power has an exponential [34] dependency on temperature. Thus, it is crucial to prevent the operating MPSoCs from entering thermal runaway situations. Temperature-dependent leakage power may not be viewed as a failure mechanism, but high leakage power would cause a significant degradation in the power efficiency, as it would surpass the dynamic power consumption, where dynamic power is the effective power used in computations and is mainly workload dependent.

1.4 Advanced Cooling Technologies for 3D MPSoCs

To address the increasing thermal rise of 3D MPSoCs, several research initiatives have explored several advancing cooling strategies for the target architectures. For instance, there has been several research efforts to insert dummy thermal through-silicon vias (TTSVs) [35] to dissipate the heat generated from the layers further from the heat sink in a more efficient way.

Another advanced cooling technology uses injected fluids (single- or two-phase), between the different layers of the 3D MPSoCs. This cooling methodology, which is also known as interlayer liquid cooling [36], is achieved by manufacturing a cavity layer, for example rectangular microchannels or micro pin-fin structures [37], on the back-side of each silicon layer. A typical structure of 3D MPSoCs consists of two or more silicon tiers, which contain the processing and storage elements of the system. The communication between these tiers is realized with through-silicon vias (TSVs) that are etched in the residual silicon slab. To account for inter-tier liquid cooling, the porous cavity is realized by etching porous structures of different form and shapes (cf. Fig. 1.6). In the example shown in Fig. 1.6, the micro-channels are built, and distributed uniformly, in between the vertical layers for liquid flow. The

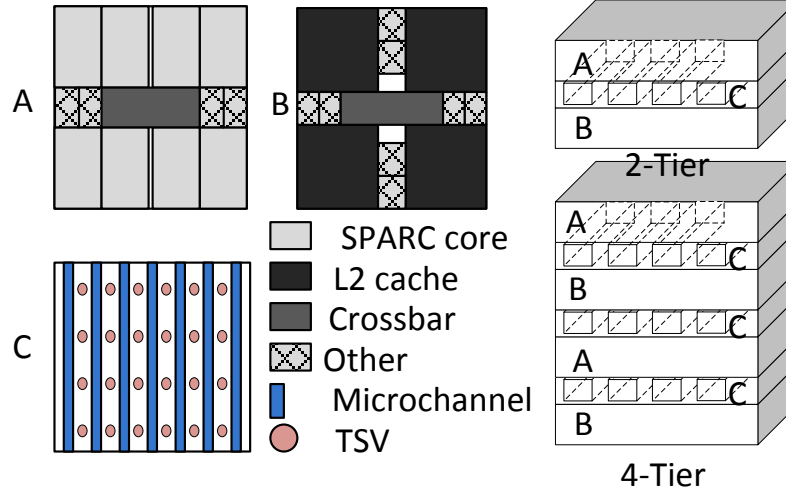


Figure 1.6: Layouts of the interlayer liquid-cooled 3D MPSoCs [38].

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.

Manufacturing 3D MPSoCs with TSV interconnections and microchannels requires a series of microfabrication processes, namely (1) deep-reactive-ion-etching (DRIE) process for anisotropic silicon etching of both TSV openings and backside microchannels (cf. Fig. 1.7); (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. 1.8.

Despite the benefits that liquid cooling brings in terms of significant thermal reduction, liquid cooling adds additional challenges to obtain a balanced thermal state with low spatial thermal gradients. As the coolant flows in microchannels, it experiences sensible heat absorption along the path [41]. This results in the coolant temperature increase from inlet to outlet, which causes a thermal gradient formation on the MPSoC surface even when the heat dissipation is uniform, as shown in Fig. 1.9(a). More commonly, in 3D MPSoCs with non uniform heat

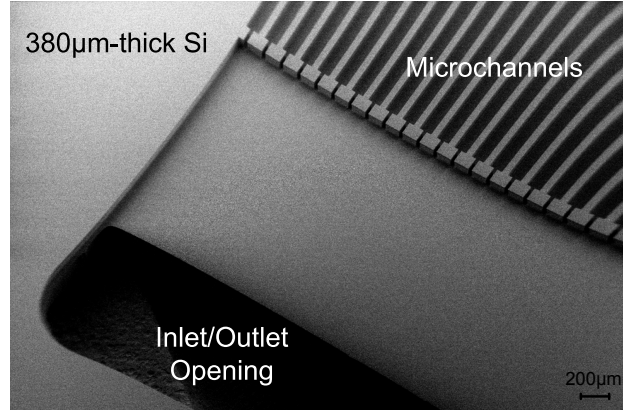


Figure 1.7: SEM photos the wafer Back-side with the inlet-outlet openings while showing the micro-channels [39].

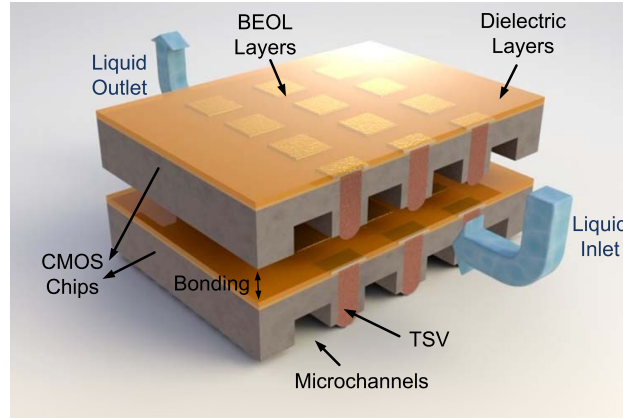


Figure 1.8: Simplified illustration of 3D stack with inter-tier liquid cooling [39].

dissipations, the existing thermal gradients and hot spots are aggravated by this characteristic of interlayer liquid cooling, as shown in Fig. 1.9(b). As a result, thermal gradients proliferate in 3D MPSoCs with liquid cooling. These gradients cause uneven thermally-induced stresses on different parts of the MPSoC, significantly undermining overall system reliability [31] (cf. Section 1.3).

From these observations we deduce that these new advanced cooling technologies bring both additional benefits and challenges. Thus,

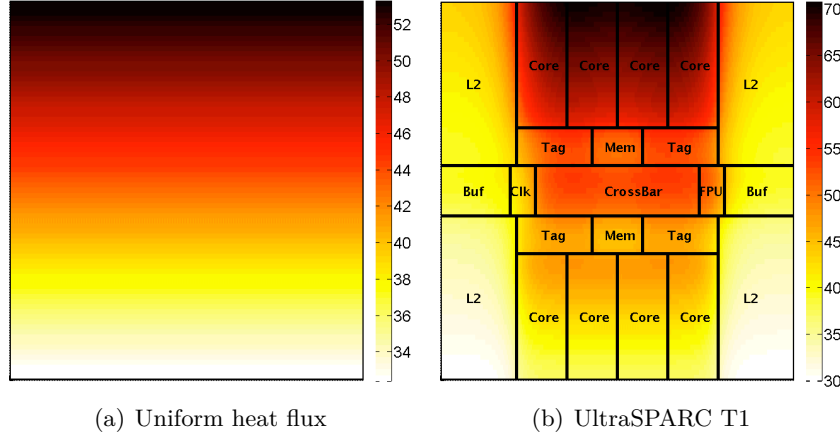


Figure 1.9: Steady-state temperature distribution of a 14mm x 15mm two-die 3D IC with (a) uniform (combined) heat flux density of $50\text{W}/\text{cm}^2$ and (b) the UltraSPARC T1 (Niagara-1) chip architecture [14]- the (combined) heat flux densities range from $8\text{W}/\text{cm}^2$ to about $64\text{W}/\text{cm}^2$. Direction of the coolant flow is from the bottom to the top of the figure [40].

it is crucial to find an optimized methodology to apply these technologies, alongside conventional 2D temperature balancing techniques in the context of resource-efficient temperature-aware 3D MPSoCs. Our interest in resource-efficiency varies from area to applied energy.

1.5 Survey Overview and Outline

In this survey, we provide an extensive survey that covers temperature-aware design optimizations and run-time management schemes for 3D MPSoCs, to avoid the rapid degradation of these architectures due to the thermal impact on reliability. The survey shows how the state-of-the-art tackles the thermal issues in the emerging 3D MPSoCs that include advanced cooling mechanisms, such as thermal-through-silicon-vias (TTSVs) and interlayer liquid cooling. We perform this survey exploration in a top-down manner to cover all the temperature-aware aspects in the target architecture. In particular, we provide two main research directions that tackle the thermal issues, namely:

1. **Design-time technological solutions and temperature-aware optimizations.** In this category we explore various techniques that address, at *design-time*, the means of generating and dissipating heat in 3D MPSoCs. This includes the optimization of new advanced cooling and heat dissipation mechanisms, temperature-aware floorplanning, and design-time optimization of different parameters that would affect the thermal behavior of 3D MPSoCs.
2. **Run-time thermal management mechanisms.** We show in this category the various temperature-affecting knobs in the target 3D MPSoC, and how the state-of-the-art utilizes these knobs in developing several *run-time* thermal management mechanisms.

1.5.1 Related Survey Works

Our proposed survey overlaps with several surveys that exist in literature. An initial survey in thermally-aware design [42] explores the various design-optimization mechanisms for MPSoCs, both planar 2D and vertical 3D. In particular, this previous survey explains the various modeling framework approaches and explores thermal-aware floorplanning, and the means to recover from temperature-induced parameters degradation such as run-time shifts. However, there is no exploration for advanced cooling mechanisms, other design-time optimization mechanisms, or run-time thermal management techniques, which are explored in our proposed survey.

There is a recent survey work that explores various thermal management mechanisms for processing architectures [43]. In this previous survey, several mechanisms for balancing temperatures in 2D and 3D MPSoCs are explored, namely thermal sensor placement, run-time thermal management, floorplanning, operating system/compilation techniques, and a brief exploration on liquid cooling. However, this survey does not provide a systematic classification of the temperature optimization research field. In our proposed survey, however, we provide our classification in a more systematic methodology that follows a top-down reasoning to cover most of the research directions in 3D MPSoC

temperature optimization.

Finally, another survey explores thoroughly vertical integration in 3D ICs [44]. This previous work explores the various electronic-design automation tools needed for 3D architectures. Then it explores the various architecture flavors in 3D ICs, namely 3D field programmable gate arrays (FPGA) and 3D MPSoC designs. Thus, this previous survey work is complementary to our proposal.

1.5.2 Survey Organization

This survey starts with providing an overview of the recent thermal modeling approaches developed for the target 3D MPSoC architectures in Section 2. In Section 3, we explore the design-time optimization schemes to minimize the temperature-oriented issues in 3D MPSoCs. Section 4 shows our exploration in run-time thermal management schemes for 3D MPSoCs. We first start by stating the various temperature-affecting control knobs, then we show the classification of different management schemes that use these knobs to balance the dissipated heat in the target architecture. Finally, we summarize our work in this survey in the conclusion.

2

Thermal Modeling for 3D MPSoCs

In this section, we explain the fundamentals of heat transfer of the targeted 3D MPSoCs, as it is a basic requirement for developing the appropriate thermally-aware optimization schemes. Based on these heat transfer fundamentals, we highlight on the state-of-the-art thermal modeling frameworks.

2.1 Heat Transfer Principles

Due to the hybrid solid-fluid material composition of the targeted 3D MPSoCs that may include interlayer liquid cooling (solids in the silicon dies, fluids in the used cooling material), the heat transfer in this target architecture is based on three factors, which are as follows:

1. Heat flow via **conduction in solids**, which occurs between the adjacent die layers or within the layer.
2. **Convective heat transfer** at the solid-liquid interface at the microchannel walls.
3. **Sensible heat transport** via mass flow of fluids from inlet to

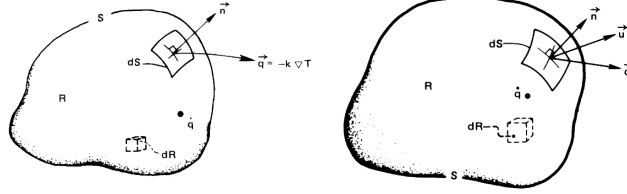


Figure 2.1: (a) Control volume of a solid with heat conduction. (b) Control volume of a liquid (coolant) with heat conduction and convection [45].

outlet inside the microchannels (or liquid-carrying cavity structures).

The modeling of heat flow via **conduction in solids** begins with Fourier's law [46, 45]. Consider a control volume of a solid R as shown in Fig. 2.1(a). The energy balance for this control volume can be written using the following integral equation:

$$\frac{\partial}{\partial t} \int_R \rho \hat{u} dR + \int_S (-k \nabla T) \cdot \vec{n} dS = \int_R \dot{q} dR, \quad (2.1)$$

where ρ is the density of the material, \hat{u} is the enthalpy, S is the surface area of the control volume, k is the thermal conductivity of the material, \vec{n} is the unit normal vector on the surface of the volume and \dot{q} is the volumetric heat generation rate inside the volume. In the above equation (Eq (2.1)), the first term on the left hand side is a volume integral representing the amount of heat energy stored in the volume. The second term is a surface integral representing the loss of heat from the volume due to heat conduction. The term on the right hand side is a volume integral representing the heat generation rate inside the volume due to conversion from another form of energy, which is the electric switching activity in the case of the target 3D MPSoC. Thus, taking the limit $R \rightarrow 0$, applying Stoke's theorem [45], and assuming the material has isotropic thermal conductivity (that is, the same value of thermal conductivity in all directions), the above integral equation reduces to the following diffusion equation:

$$c_v \frac{\partial T}{\partial t} + (-k \nabla^2 T) = \dot{q}, \quad (2.2)$$

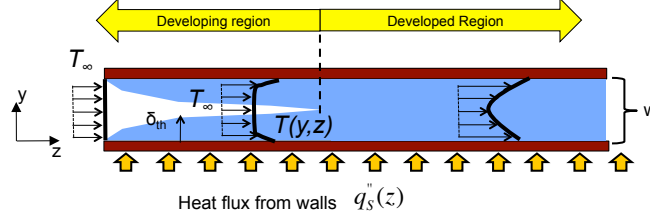


Figure 2.2: Convective heat transfer at the solid-liquid interface of a microchannel.

where c_v is the volumetric specific heat of the material and T is the temperature of the control volume.

Modeling the **convective heat transfer** at the interface between a microchannel wall and a flowing liquid (as seen in Fig. 2.2) traditionally begins with the Newton's law of cooling as described by the following equation [47]:

$$q_s'' = h (T_s - T_\infty), \quad (2.3)$$

where q_s'' is the heat flux at the surface entering/leaving the solid, T_s is the temperature of the solid surface, T_∞ is the temperature of the fluid bulk and h is the heat transfer coefficient at the surface. Combining the above equation with the Fourier's law at the interface in steady state assuming one dimensional heat transfer (Fig. 2.2):

$$q_s'' = -k_f \left. \frac{\partial T}{\partial y} \right|_{y=0}, \quad (2.4)$$

Then we can write the heat transfer coefficient as:

$$h = \frac{-k_f \partial T / \partial y|_{y=0}}{T_s - T_\infty}, \quad (2.5)$$

where k_f is the fluid thermal conductivity. The heat transfer coefficient in (2.5) is derived by studying the fluid flow nature, the hydrodynamic and thermodynamic developing layers and the thermal properties of the fluid to compute the terms on the right-hand side. It is then typically expressed in the form of a dimensionless parameter of the flow called the Nusselt number Nu [47, 45].

Once the parameters for convective heat transfer at solid-liquid interfaces are computed, they can be combined with Fourier's law to

solve for the temperatures in the entire structure. While inside the solid parts of the structure (2.1) are used for this purpose, inside the fluid parts an additional term must be included to account for the heat transfer associated with the liquid flow, in order to account for the rise in T_∞ as it flows from inlet to outlet absorbing heat (**sensible heat transport**), as shown below:

$$\frac{d}{dt} \int_R \rho \hat{u} dR + \int_S (-k \nabla T) \cdot \vec{n} dS + \int_S (\rho \hat{h}) \vec{u} \cdot \vec{n} dS = \int_R \dot{q} dR, \quad (2.6)$$

2.2 Thermal Modeling Frameworks for 2D/3D ICs

The mentioned factors in Section 2.1 have been used before in various modeling techniques. For heat conduction modeling, numerical techniques involve the discretization of the above partial differential equation (Eq. 2.2) first in space, creating ordinary differential equations. Next, the ordinary differential equations are numerically integrated in time to solve for the temperature at the relevant points in space and time in a given computational domain. The first step is usually accomplished by either finite-element [48] or using compact finite-difference frameworks such as HotSpot [46]. In the case of alternate direction implicit method (ADI) [49], the discretizations in space and time are done simultaneously and the solving in both domains are intertwined.

In the case of convective heat transfer modeling, finite-volume methods are used in traditional numerical fine grained simulators such as ANSYS CFX [50], where the microchannel and the surrounding structures are divided into very small cells, and the three-dimensional fluid velocity profile and the temperatures are simultaneously computed by solving the flow of fluids, governed by the Navier-Stokes equations, and the Fourier's law equations respectively in three dimensions. Hence, the computations of convective heat transfer at solid-liquid interfaces (i.e. terms in (2.5)) are implicit in this method. Such methods, while being accurate, are extremely slow and are not desirable in complex structures such as interlayer liquid-cooled 3D MPSoCs.

To model the thermal behavior of the target architectures, many other approaches have been proposed for this problem. For example, one dimensional methods such as [36] can be used to simplify the heat

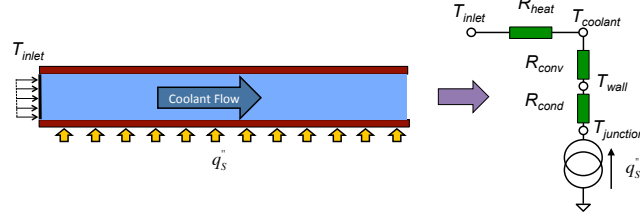


Figure 2.3: A simple one-dimensional model for forced convective cooling in microchannels [36].

transfer problem by assuming that heat enters vertically from the silicon surface into the fluid in the microchannel via a vertical thermal resistance that combines the conduction resistance and the convective resistance at the solid-liquid interface. This is illustrated in Fig. 2.3. In order to take into account the rise in fluid temperature as it accumulates heat from inlet to outlet, another thermal resistance (R_{heat}) is used to represent the sensible heat absorption connecting to the inlet of the microchannel where a constant temperature boundary condition is applied. This method, while being easy to build and solve for, oversimplifies the problem and does not account for non-uniform heat fluxes generated by the heterogeneous multi-core 3D MPSoC, and the three-dimensional spreading of heat in the structure surrounding the microchannel. Hence, the applicability of this model is limited.

2.3 Compact Thermal Modeling for 3D MPSoCs

Compact thermal models for MPSoCs have been widely used in design-space thermal explorations [52] and developing thermally-aware optimization schemes [12], as they provide superior speed-ups with acceptable accuracy compared to the finite-element methods. Initial work on faster modeling approach has been conducted by Koo et al. [51]. In this modeling approach the governing heat transfer equations are employed via the finite difference method and applied to a discretized volume of the 3D ICs. The main assumption in this work is that the heat is dissipated vertically from the silicon tiers to the microchannels. As shown

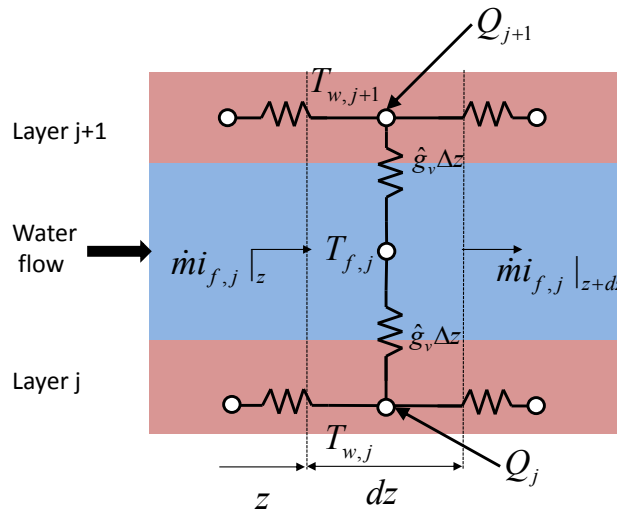


Figure 2.4: Schematic of microchannels and implemented in a 3D circuit and thermal modeling of microchannel cooling for a 3D circuit. Only one channel is analyzed in a cooling layer by geometric and thermal symmetries. Dotted lines indicate the discretized control volume used the model [51].

in Fig. 2.4, a resistive network is generated which is governed by the following equations:

$$\frac{d}{dz} \left(k_{w,z,j} A_{w,j} \frac{dT_{w,j}}{dz} \right) = q_j'' p + h_{conv,j} \eta_0 (w + d) (T_{w,j} - T_{f,j-1}) \quad (2.7)$$

$$\begin{aligned} \dot{m} \frac{d_{f,j}}{dz} = & h_{conv,j} \eta_0(w+d)(T_{w,j} - T_{f,j}) + (T_{w,j} \\ & - T_{w,j+1}/R_{th,j}) + (T_{w,j} - T_{w,j-1}/R_{th,j-1}) \\ & + h_{conv,j} \eta_0(w+d)(T_{w,j+1} - T_{f,j}) \end{aligned} \quad (2.8)$$

where q_j'' is the applied heat flux at the j th layer, $T_{w,j}$ and $T_{f,j}$ are the average local temperatures of the solid wall and the fluid, respectively. The subscript (j) indicates the property of the j th layer. The pitch of microchannels (distance between two adjacent microchannels) is denoted as (p). The depth and width of the microchannel are represented as (d) and (w), respectively. The forced convection coefficient

cient for heat transfer between the solid wall and the working fluid is $h_{conv,j}$. The fluid enthalpy per unit mass (i_f) is expressed in terms of local thermodynamic equilibrium fluid quality (x). The effective thermal conductivity of solid in the (z) direction is $k_{w,z,j}$ and $A_{w,j}$ is the effective solid cross-sectional area. $R_{th,j}$ is the conduction thermal resistance between control points on layers (j) and ($j+1$) and η_0 is the overall surface efficiency, which is employed to simplify the temperature variation in the (y) direction within the channel walls.

This work while seems promising at its publication time, has not provided any validation support to the models. Moreover, this model could only provide steady-state thermal profiles without any transient thermal behavior insights. Thus this model is indeed limited to be applied as-is for realistic 3D MPSoC thermal exploration and run-time thermal analysis and management.

To enhance the previous model, Kim et al. [53] have extended the previous model to account for lateral thermal dissipations and non-uniform heat flux distribution scenarios. This work explores several fluid inlet configurations, as well as the potential of two-phase cooling. This advancement however, is still applicable to the steady-state case.

The *3D-ICE* [41] compact thermal modeling framework has been proposed for transient thermal simulation of 3D MPSoCs with inter-layer liquid cooling. This model has been extensively validated for accuracy against fine-grained numerical simulations as well as measurements from real liquid-cooled emulated ICs [41, 54].

In this modeling framework, finite-difference approximations are applied to the aforementioned equations (2.2) and (2.6). When finite-difference approximation is applied to (2.2), the given volume of solid is discretized into “thermal cells” along the 3 cartesian coordinates to generate a thermal grid. Thus, electrical analogy is invoked here with the temperature represented by voltage, heat flow represented by electric current [46], the first term on the left hand side in (2.2) as a capacitor and the rest of the discretized represented as conductances, giving rise to an RC circuit [41, 45]. Then, for a given thermal cell of length l , width w and height h as shown in Fig. 2.5(a), the compact thermal model consists of six resistances representing the conduction of

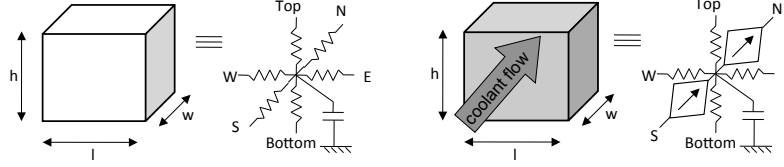


Figure 2.5: (a) Equivalent electrical circuit for heat conduction in a solid thermal cell. (b) The 3D-ICE model for a liquid thermal cell [41].

heat in all the six cartesian directions and a capacitance representing the heat storage inside the cell.

In the case of liquid cells, the microchannel layers in the 3D MPSoC are divided into thermal cells similar to the other solid layers as described before. Next, for each thermal cell inside the microchannel, an equivalent electrical representation similar to Fig. 2.5(a) is constructed. The only difference of this cell with respect to the solid cell is the average velocity of the fluid through the cell with fluid entering from the front end and exiting from the rear end of the fluid cell as indicated in Fig. 2.5(b). Based on the discretization of (2.6) and by invoking the electrical analogy, the velocity-related term can be translated into a *voltage-controlled current source* in the equivalent RC circuit. These *voltage controlled current sources* model the transport of heat from the inlet to the outlet of the microchannel and, hence, account for the rise in temperature of the coolant as it flows through the microchannel.

3D-ICE has been shown its accuracy when validated against the commercial computational fluid dynamics (CFD) simulation framework, i.e. Ansys CFX [50]. Simulation results show that 3D-ICE achieves peak temperature error of only 3.4% while providing 975x speed-ups when compared to CFD simulations [41] (cf. Fig. 2.6(a)). More advances show that 3D-ICE can even reach speed-ups up to 13478x with 10% peak temperature error [54]. Finally, the recent validations against a realistic test bed [55] show that 3D-ICE achieves an average error of 8.5% with respect to experimental thermal measurements of the test bed (cf. Fig. 2.6(b)).

Despite the accuracy of 3D-ICE, recent work [56] highlights the lim-

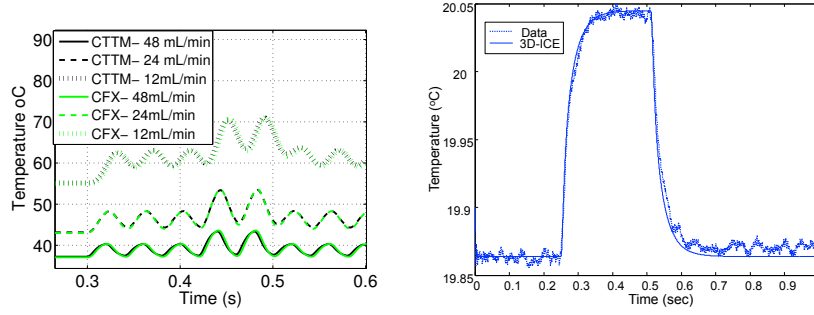


Figure 2.6: Validation of 3D-ICE simulations against (a) Ansys CFX CFD simulation [41]. (b) Experimental measurements of a 3D test bed [55].

itations of this tool. For instance, 3D-ICE ignores the fluid entrance and the accompanying thermo-hydrodynamic developing region. Thus, 3D-ICE tends to under-estimate the liquid-cooling thermal gradient. Moreover, 3D-ICE neglects the thermal effects of TSVs on the overall 3D-IC thermal profile. In this respect, Qian et al. propose 3D-ACME [56] a compact thermal modeling framework that takes into consideration the impact of TSVs and the microchannels entrance region. By taking these two factors, 3D-ACME manages to reduce the error with respect to finite-element methods to 0.2% which is one order of magnitude less than the error of 3D-ICE with respect to finite-element methods (4.1%). However, 3D-ACME is a steady-state modeling framework that has a higher runtime for large problem size compared to 3D-ICE. Thus, 3D-ACME cannot be used for runtime thermal monitoring of 3D MPSoCs with time-varying workload conditions.

2.4 Summary

In this section, we have shown the basic heat transfer principles that are used in various thermal modeling frameworks. We have shown the thermal conduction in solids, convection and sensible heat absorption in fluids. These factors are later used in the different modeling frameworks we have shown, which are either dedicated to 3D MPSoCs or

generalized to various modeled platforms. We have elaborated more on the compact thermal modeling concepts of 3D MPSoCs, due to their effective utilization in various temperature-aware explorations.

3

Temperature-Aware Design Optimizations for 3D MPSoCs

The continued scaling and integration in Multiprocessor Systems-on-Chip (MPSoCs) designs have caused an increase in the power density. This increase has augmented the heat generation in MPSoCs, hence higher operating temperatures. These temperatures lead to accelerated degradation rates in various thermally-induced parametric factors [31] (cf. Section 1.3), such as electro-migration (EM) [30], bias-temperature instability (BTI) [24], gate oxide breakdown [29], propagation delays [57, 58], leakage power [59], and power/ground integrity [31]. Thus, it is crucial to design these high-power density MPSoCs in a thermally-aware manner to abate the mentioned degradation factors, even before experiencing the non-homogenous time-varying workload characterizations. It is worth mentioning that these thermally-aware designs should be designed to handle two pivotal parameters, namely the overall thermal state of the system defined by peak temperatures (*hot spots*) and balancing the temperature distribution within the MP-SoC platform (*thermal gradients*).

While the mentioned thermal issues are alarmingly rising in high power-density MPSoCs, they are even more furious in the prospective vertically-stacked 3D MPSoCs [21, 12]. This is mainly due to the in-

creased and nonuniform thermal resistance at different stacked layers, based on their relative heat dissipation paths using conventional techniques such as placing heat sink on top of the top-most layer. With limited surface exposure to the heat sink, temperature in 3D MPSoCs exacerbates since the inner layers do not have a direct access to heat sink for effective heat removal.

The accompanied high temperatures of these 3D MPSoCs have triggered the research and development of design-time cooling solutions. Research directions include better heat dissipation paths such as thermal through-silicon vias (TSV) insertion [35, 60], temperature-aware floorplanning [61], and using active cooling techniques such as forced convective liquid cooling [36, 62, 63, 53] and thermoelectric cooling [64, 65] techniques.

Despite the benefits that new cooling technologies bring to 3D MPSoCs, these techniques can bring additional challenges if not optimized properly. Moreover, if these techniques are applied to arbitrary 3D MPSoC architectures, without any temperature-aware consideration, several resources such as cooling energy can be over-utilized. Thus, it is crucial to optimize, at design-time several temperature-related features of 3D MPSoCs, which has been already performed in literature through several directions and approaches.

In this Section we provide a detailed state-of-the-art review on the design-time optimization techniques of 3D MPSoCs that target thermal attenuation and balancing. At the outset, temperature-aware design optimization can be classified to mechanisms that optimize the heat generation and the flow in the 3D MPSoC, and techniques that optimize the heat dissipation paths from the 3D MPSoC to the surrounding environment. In this review, we refer to these mechanisms as *on-chip* and *off-chip* design-time optimizations, respectively. We start by the *on-chip* optimization mechanisms in Section 3.1. Then, we explore the *off-chip* optimization mechanisms in Section 3.2. Finally, we summarize the contents in Section 3.3.

3.1 On-Chip Design-Time Optimizations

The first category we elaborate in this survey is related to the heat generation and dissipation patterns inside the 3D MPSoC, hence the term *on-chip* optimization. It is important to mention that this exploration does not include how the heat is removed from the target system. In the case of targeted 3D MPSoCs, on-chip design optimization can be classified to the following categories. The first category is related to the *heat generation* of each computational module. This can be optimized either by the way the modules are designed or utilized. The second category elaborates the thermal impact different modules have on each other, which we refer to as *on-chip heat dissipation*. Please note that this category is different from *off-chip* heat dissipation as the former category is related to the intra-chip heat dissipation inside the 3D MP-SoC, while the latter category is concerned with the inter-chip heat dissipation between the 3D MPSoC and the surrounding environment. In the following subsections, we explore the state-of-the-art works that belong to the mentioned categories.

3.1.1 Heat Generation Optimization

Power, or heat, generation in computing systems can be adjusted at design-time to prevent reaching any thermally-based critical situation. Since computing systems design is defined by the target platform architecture and application characteristics, addressing the power (or heat) generation at design-time can be split between techniques that are applied at the *platform* level and techniques that are applied at the *application and mapping* level. These general trends can be followed in 2D and 3D MPSoC design optimizations. Thus, we demonstrate the state-of-the-art in these directions in the case of both planar 2D and 3D MPSoCs.

Platform Oriented At the platform level, different modules can be designed to reduce the overall power density, hence heat generation, while preserving the system functionality. This approach has been taken recently in low-power (hence low temperature) processor designs such as

ARM big.LITTLE processing architecture [66]. Another approach at the platform level is to reduce the operating power supply of the platform to near-threshold values [17]. Near-threshold computing allows the processing units to operate close to the voltage threshold value of the used transistor, hence reducing the overall power and thermal density.

In the case of 3D MPSoCs, recent work proposes multiple supply voltages utilization to optimize the voltage islands distribution in 3D MPSoCs [67]. In this work, a temperature-aware voltage island generation methodology is proposed that formulates this problem as a mixed-integer linear programming (MILP) problem. The main aim in this work is to minimize the thermal hot spots in 3D MPSoCs while keeping the performance and timing requirements satisfied. The interdependency between power and heat densities made it feasible to formulate this problem and achieve significant results.

Another work utilizes various microarchitectural techniques to control the thermal hot spots in 3D MPSoCs via thermal herding [68]. This technique explores different architectural disciplines by splitting several microarchitectural blocks between the different layers of 3D MPSoC to enhance the throughput while controlling the thermal hot spots such as register file and integer adder splitting, as shown in Fig. 3.1 and Fig. 3.2. This splitting is based on general application trends and the significance of particular instructions or data locations to the execution flow. When thermal herding is applied to a 2-tier 3D MPSoC, significant enhancement in the performance figures is obtained when compared to conventional 3D integration, by 51% mean performance improvements. This performance improvement is accompanied with 12% power savings, with 5K peak temperature reduction.

Application and Mapping Oriented In addition to platform level techniques, the operating applications can be designed to alleviate the heat generation of the processing unit. For example, recent work [69] distributes the idle time between the tasks running on the same processing unit to allow for a cool-down period between running tasks. This work applies a distribution time heuristic and non-linear program-

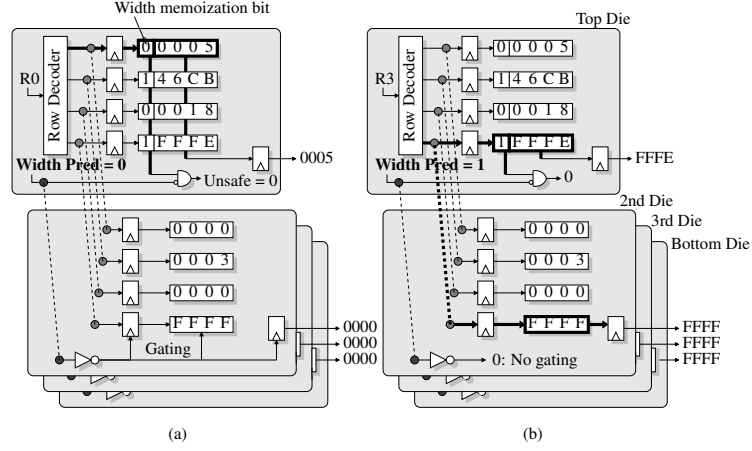


Figure 3.1: Thermal Herding in register files: examples where (a) a low-width value only requires circuit activity on the top die, and (b) a full-width value requiring reading state from all four die [68].

ming to optimize the idle time placement. Other work [70] proposes a pseudo instruction scheduling technique for VLIW processors that maps parallel instructions to the coolest functional units, and gives for each instruction a possible list of thermally-inactive functional units for temperature reduction.

3.1.2 On-Chip Heat Dissipation Optimization

As mentioned earlier, this subclass achieves thermal reduction via intra-MPSoC (or inter-module) heat dissipation adjustments. Such adjustments can be done at the **platform level**, where the microarchitecture of the MPSoC (2D and 3D) modules and the accompanying middleware have major impact on the heat flow and propagation of the target architecture. Additionally, the adjustment of heat flow can be managed by **application level** customizations, both algorithmic and mapping, such that the application impact on heat spreading is minimized.

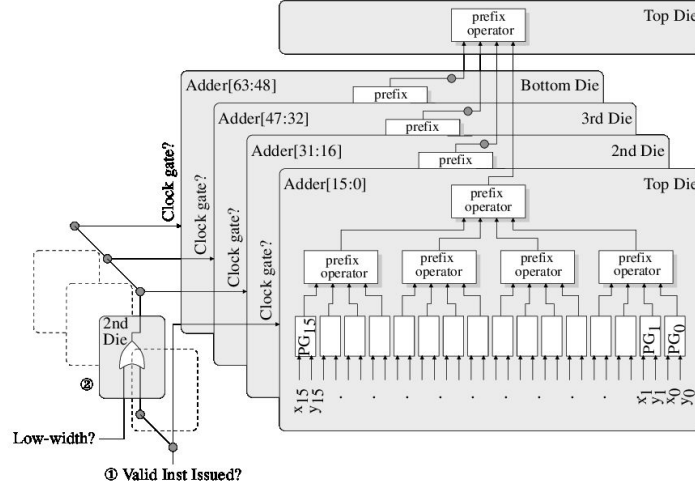


Figure 3.2: Thermal Herding in an integer adder with the most active least significant bits placed on the top die [68].

Platform Level . Previous works have investigated the rearrangement of various hardware modules within the MPSoC to minimize the global thermal impact, which is also known in literature as *temperature-aware floorplanning*. Initial work on temperature-aware floorplanning [71] has shown its significant impact on reducing the peak temperature. This work has defined a metric called *thermal diffusion* that resembles the lateral heat dissipation. This metric has been used in an optimization problem to maximize the gains of *thermal diffusion*. Other similar works have proposed simulated annealing utilization [72] or genetic algorithms [73] to achieve temperature-aware floorplanning.

In the context of 3D MPSoCs, temperature-aware floorplanning has also been extended by including the interlayer thermal dissipation and interconnect characteristics [74, 72, 75, 61]. For example, initial work has been proposed [76] for temperature-aware microarchitectural floorplanning. The main objective in this work is to place the processing submodules of a single processor in several layers such that the wire lengths and the temperatures are minimized. To achieve this, a mixed integer linear programming (MILP) problem is formulated to minimize the weighted sum of performance, area and thermal-related aspects.

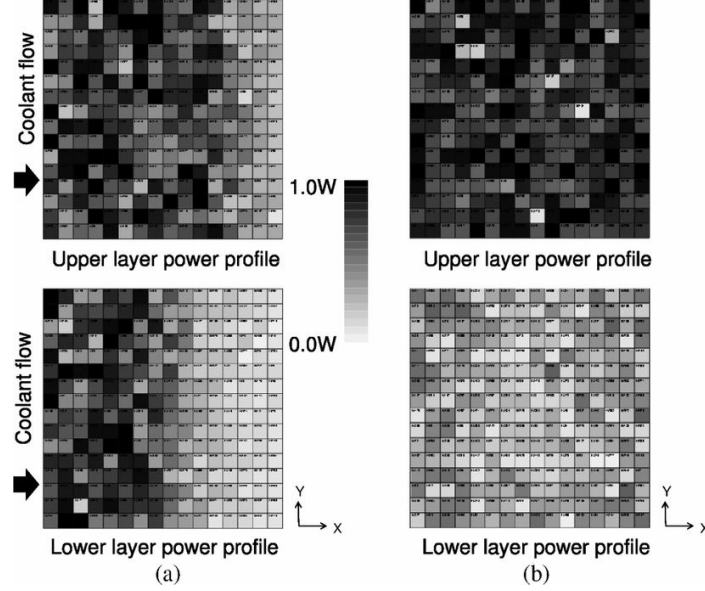


Figure 3.3: Power profile of two layered (a) microchannel cooled and (b) conventional air cooled 3-D ICs with thermal aware placement [52].

Results show that this floorplanning approach reduces the peak temperature by 24% when compared with performance-driven floorplanning approach. Another work uses simulated annealing to minimize the temperature of 3D MPSoC via floorplanning [75] by considering the additional power consumption of the interconnects.

As for liquid-cooled 3D MPSoCs, Mizunuma et al. use their thermal model to explore floorplanning solutions [72] to homogenize temperature distributions in this architecture [52]. In particular, this work explored the impact of the Polish expressions-based, simulated annealing (SA) thermal placer [72] on the allocation of modules in liquid-cooled 3D MPSoCs. The results in this work, which is further assisted by the observations in other work [38], show that in the case of liquid cooled 3D MPSoC, temperature-aware floorplanning follows the trend of placing more heat dissipating modules at the fluid inlet port, while lower heat dissipating modules at the outlet port, which is different than the

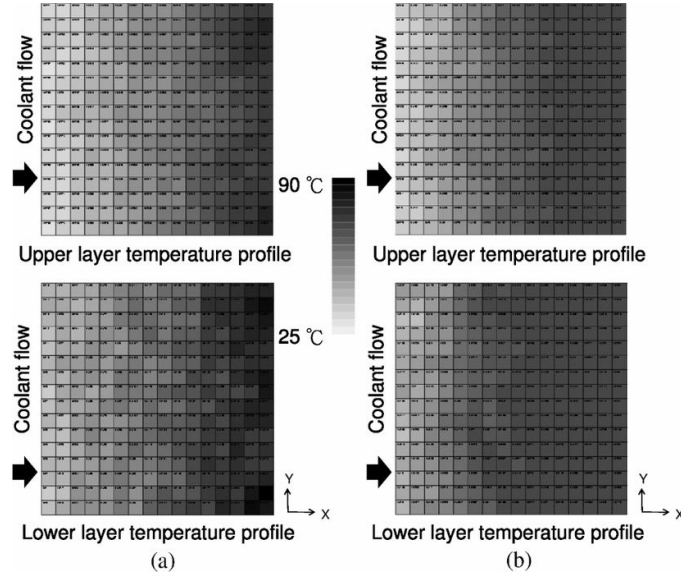


Figure 3.4: Temperature profile of two layered liquid-cooled 3-D ICs with (a) random and (b) thermal aware placement [52].

case of air cooling as shown in Fig. 3.3. In other words, the optimal heat dissipation pattern for temperature-aware floorplanning would be monotonically decreasing from the distance of the fluid inlet port, which would generate the optimal thermal map as shown in Fig. 3.4.

Application Level. The **application mapping** to a specific MPSoC has a significant impact on the corresponding thermal behavior. This mapping can be performed at design-time, where several temperature-aware compilation schemes have been proposed. Previous work [77] proposes register file reallocation technique for VLIW processors to homogenize the resulted temperature distribution within a VLIW processor. This work uses register renaming and variable live range splitting to achieve uniform thermal distribution. In addition, other work on thermally-aware register file utilization [78] targets the homogenous use of register window-based processing architectures. This work vir-

usually splits the application to a number of smaller sub-applications, where each sub-application is assigned to a different register window for better thermal dissipation. This work manages to reduce the peak temperature by 11% and the thermal gradient by 38%. While this research branch has shown its effectiveness in the case of planar 2D MPSoCs, it is yet to be explored for 3D MPSoCs.

3.2 Off-Chip Design-Time Optimizations

Off-chip optimizations primarily target the customization of heat dissipation paths of the different tiers in 3D MPSoCs to the heat sink layer(s). Significant research efforts have been invested to find techniques that aid in heat removal (also referred by vertical dissipation) from the 3D MPSoC. In the context of this survey, heat removal is defined as dissipating the local generated heat of a certain module to the surrounding environment outside the 3D MPSoC. These techniques aid in better heat removal by providing better means of thermal dissipation without any external effort (power) requirements, which is referred to in this text by *Passive* heat removal techniques. Alternatively, heat removal can be enforced by the use of *Active* techniques, where an external effort (power) and medium are required to enhance this off-chip heat dissipation. Typical examples of active cooling mechanisms involve liquid [36] and thermoelectric [65] cooling methods, while passive cooling examples include thermal through-silicon vias (TTSVs) [35].

3.2.1 Passive Heat Dissipation Optimization

Prior work related to this category has investigated using new materials, with better thermal properties than silicon, in the fabrication process. For example, Carbon nanotubes [79] and Graphene [80, 81] are recently proposed materials to be used in the transistor fabrication, which are characterized by high thermal conductivity values. Another approach introduces other passive elements in the ICs. Particularly, thermal through-silicon-vias (TTSVs) [60, 35, 82] have been recently proposed to alleviate the corresponding thermal issues for 3D MPSoCs. Thermal TSVs provide a good thermal conducting medium to transfer

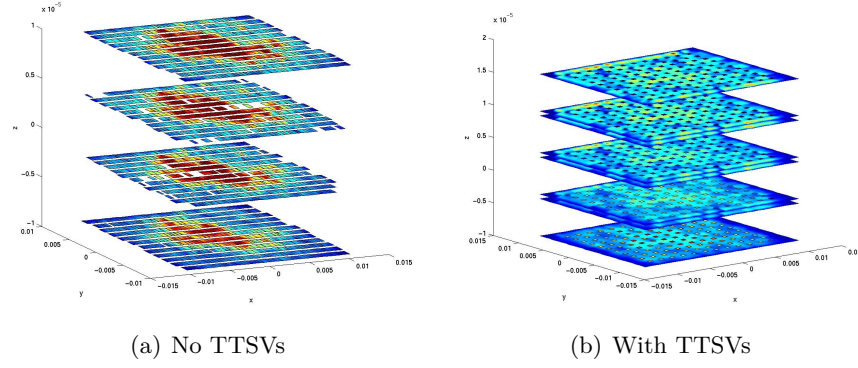


Figure 3.5: The thermal profile of multi-tier 3D IC with and without TTSVs [82].

the generated heat from the lower layers to the heat sink for an overall better thermal performance.

An early work on thermal vias insertion and optimization has been performed [82]. This work has used an iterative method of TTSVs insertion while deducing the resulted thermal state by finite-element analysis. The resulted thermal vias density determines the thermal conductivity of the surrounding region. Results have shown the significant temperature enhancement with the TTSV insertion. As shown in Fig. 3.5, this work manages to significantly reduce the 3D IC thermal profile by up to 47.1% while incurring 1.19% area overhead for TTSVs occupation.

Previous work [35] proposes the optimization of the thermal TSVs (TTSVs) locations both horizontally and vertically in 3D MPSoCs. This work formulates the horizontal placement of TTSVs as a constrained nonlinear programming problem, while providing a fast solving heuristic for this formulation. This formulated heuristic achieves near-optimal solution (within 1% deviation from the optimal one) while achieving 200x solution time speed-ups. In addition, this work formulates the vertical placement of TTSVs as a convex optimization problem, hence guaranteeing a global minimization of the vertical heat dissipation from the lower layers to the heat sink. This work manages to reduce the TTSV density demand to satisfy a certain thermal constraint by 68%

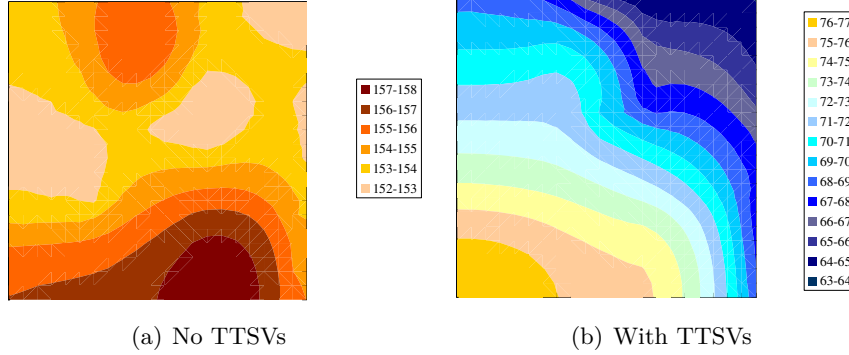


Figure 3.6: Temperature distribution of 3D-IC top layer [35].

compared to the state-of-the-art [74], with significant reduction in the thermal profile as demonstrated in Fig. 3.6.

Another work [83] uses the previous algorithm and combines it with another optimization problem that targets whitespace redistribution to optimize further the TTSV insertion, with better heat dissipation properties and reduced performance degradation. This work deduces the TTSV requirements for each block and uses this deduced information to formulate a linear programming problem to distribute the whitespace in the targeted 3D MPSoC to meet the TTSVs requirements. This optimization is also combined with performance requirements as another objective.

Temperature-aware routing of lateral thermal wires and vertical thermal vias has been proposed [84]. This work adds the utilization of lateral thermal wires, which are signal non-carrying wires thus have no electric functionality. These thermal wires utilize the residual routing tracks from signal wires routing to distribute the heat more evenly in 3D MPSoCs. Thermal wires provide direct connectivity to vertical thermal vias for more heat dissipation capabilities. The insertion of thermal vias and wires in this work has been formulated as a linear programming problem to minimize the temperature of thermal hot spots, while using the minimal thermal vias and wires.

Another recent work studies the impact and utilization of power

distribution networks (PDN) to remove the generated heat in 3D MP-SoCs [85]. This work shows that power distribution network populated with interlayer vias can aid significantly in reducing the temperatures of 3D MPSoCs and monolithic 3D ICs. Significant thermal reductions are observed that can enable various dies of a 3D MPSoC to operate at higher power density figures while keeping the thermal constraints.

3.2.2 Active Heat Dissipation Optimization

In some cases where passive techniques are not sufficient, active cooling methodologies provide the required support. A typical example in active cooling techniques is injected air cooling [86]. Forced air cooling injects air at different volumetric flow rates through the mounted heat sink to augment the convective heat transfer between the hot plate and the surrounding environment. Thermoelectric cooling is another active technique that removes the heat from the IC by using a series of actively-connected devices on top of the target IC [65]. When these devices are electrically active, they transfer the heat from the IC to the surrounding environment using the *Peltier effect*. Other active techniques involve cooling the targeted computing system by fluids, single or two-phase. For example, a recent cooling methodology proposes using hot water to cool-down server processing units [87], which is now deployed in the AQUASAR data-center. More recent research efforts propose interlayer liquid cooling for 3D MPSoCs [36], where a single phase coolant is injected in silicon-etched microchannels within the integrated layers. This injection causes the forced convective heat transfer from the processing elements to the coolant. This methodology is being extended to incorporate two-phase cooling [88], where the fluid phase change absorbs significant thermal energy values, hence cooling down the target 3D MPSoC. Additional research efforts have investigated micro pin-fin structures [89, 90] for enhanced single-phase fluid cooling potential. While these active cooling mechanisms are shown to be effective in heat removal, these techniques can be over utilized, hence resulting in resources (e.g. energy) inefficiency. In this respect, active cooling mechanisms require design optimization for their efficient application.

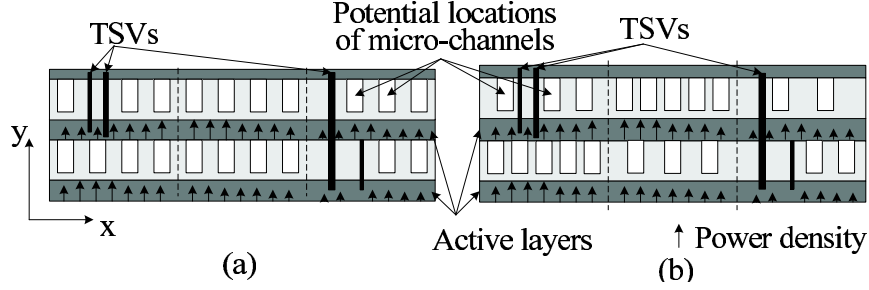


Figure 3.7: Potential locations of micro-channels: (a) uniform spreading, (b) non-uniform spreading [94].

For the case of liquid-cooled 3D MPSoC, previous work [91] proposes a channel clustering methodology where microchannels are grouped into clusters of channels. Within a single cluster, the channels have the same flow rate in order to customize the cooling effort based on the demands of computing elements. Recently, the same authors [92] extend their previous work by proposing an energy-efficient microchannel clustering technique where the primary aim is to minimize the pumping energy consumed to achieve a given peak temperature constraint.

Another work by Shi et al. [93] proposes a customized non-uniform channel allocation technique, where the density of the etched microchannels reflects the cooling demands of various regions of the MP-SoC, as shown in Fig. 3.7. However, the channel allocation technique primarily targets the improvement of energy efficiency instead of the thermal distribution. The same authors have extended this work and have proposed a methodology that includes the previously proposed non-uniform channel allocation mechanism alongside a run-time thermal management scheme [94] that both manages to reduce the cooling power consumption by up to 70%, when compared to uniform channel allocation.

The above mentioned methods are beneficial when the hot spots line up perpendicular to the coolant flow, but these works do not address the case where multiple different hot spots lie along the fluid flow direction in the microchannels.

Brunschwiler et al. investigate channel width modulation, four-port

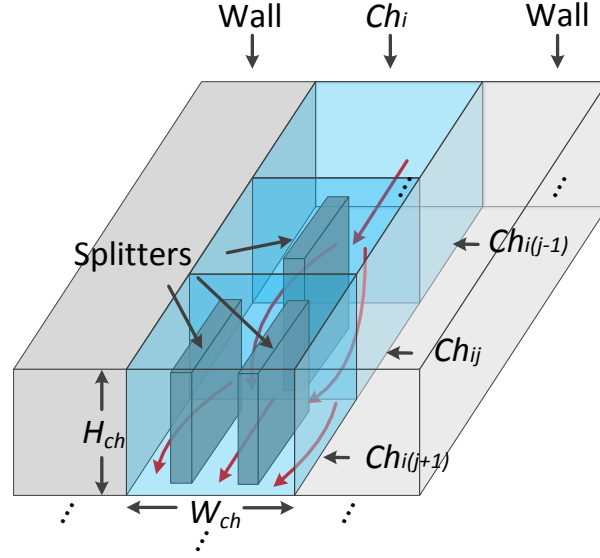


Figure 3.8: Illustration on splitter insertion for channel splitting [95].

fluid access, and the use of fluid guiding structures for thermal optimization [37, 89]. They show that by changing the channel width from inlet to outlet, customized cooling can be applied and thus achieve thermal balancing. However, their channel modulation scheme relies on heuristics without providing any optimality guarantee.

A recent work by Qian et al. [95] proposes placing a number of channel splitters, as illustrated in Fig. 3.8, in order to reduce the convective thermal resistance at the desired locations along the microchannel. The main aim is to reduce the peak temperature and balance the temperature of 3D MPSoCs. Results show that this work manages to reduce the peak and average temperatures by up to $13K$ compared to straight channels. However, this reduction may come with the cost of increased cooling power due to the introduced splitting mechanism.

The recent proposed framework, namely *GREENCOOL*, optimizes the active cooling path of microchannel-based interlayer liquid cooled 3D MPSoCs to balance the thermal profile of the target 3D MPSoC while significantly reducing the active cooling energy demands [96].

This design-time optimization methodology uses the concept of channel modulation, where we change the microchannel aspect ratio (channel width/channel height) to enhance the heat transfer capability from the target 3D MPSoC via changing the convective thermal resistance [97]. Using the conventional CMOS fabrication process for etching the channels, such as deep reactive ion etching [98], it is possible to modulate the width of the channel from inlet to outlet (and hence its aspect ratio) and create any kind of channel width profile, while keeping the height of the channels constant. Thus, channel width modulation requires only a change in the patterns on the masks used for etching channels amounting to minimal additional fabrication costs. To summarize, using careful design it is possible to modify the local channel aspect ratios so as to contain the pumping power while constraining the thermal gradients.

To understand how the channel width affects the change in temperature due to convection (ΔT_{conv}) in detail, an analysis is performed on a single microchannel shown in Fig. 3.9. We start by the following set of equations governing the Nusselt number (a dimensionless form of heat transfer coefficient), and the product of friction factor and Reynold's number for microchannels, under fully developed conditions [97]:

$$\begin{aligned} Nu &= 8.235 \cdot (1 - 2.0421AR + 3.0853AR^2 \\ &\quad - 2.4765AR^3 + 1.0578AR^4 - 0.1861AR^5) \\ fr \cdot Re &= 24 \cdot (1 - 1.3553AR + 1.9467AR^2 \\ &\quad - 1.7012AR^3 + 0.9564AR^4 - 0.2537AR^5), \end{aligned} \quad (3.1)$$

where AR is the aspect ratio reciprocal (height/width) of the channel. Using the Nusselt number, the heat transfer coefficient (a measure of the amount of heat transferred per unit area for one Kelvin difference in temperature between the fluid and the microchannel wall surface, expressed in W/m^2K) can be written as:

$$h = \frac{k_{coolant} \cdot Nu}{d_h} \quad (3.2)$$

where $k_{coolant}$ is the thermal conductivity of the coolant and d_h is the hydraulic diameter of channel. The effective heat transfer coefficient as

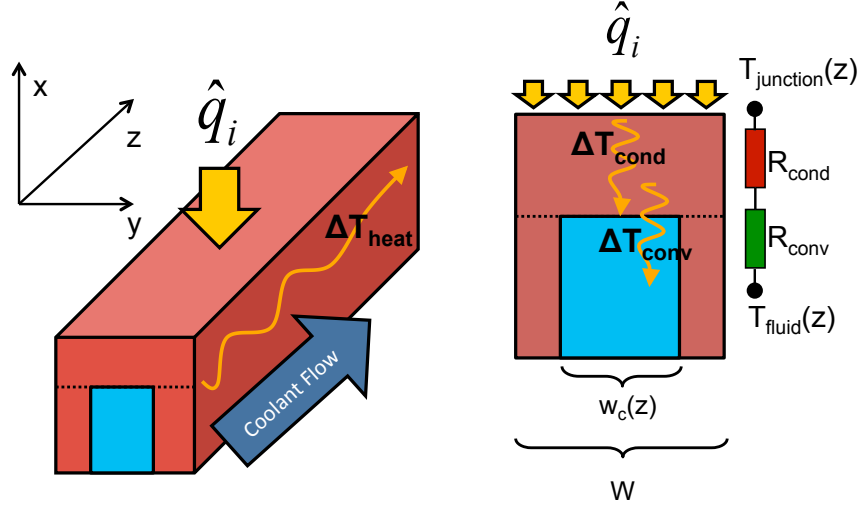


Figure 3.9: Test structure: a single microchannel cooling a strip of an IC with uniform heat flux distribution. The figure shows both the 3D and the cross-sectional views [96].

seen by the junction looking down the channel from the top can be written by projecting the heat transfer coefficient above from the side wall surfaces onto the top as follows:

$$h_{eff} = h \frac{2 * H_C + w_C}{W} \quad (3.3)$$

where H_C is the height and w_C is the width of the channel, and W is the total width of the structure as shown in Fig. 3.9. The convective resistance R_{conv} for this structure can be obtained as a reciprocal of this quantity. The R_{conv} for this structure is plotted as a function of w_C in Fig. 3.11, assuming water as the coolant, $H_C = 100\mu\text{m}$, $W = 100\mu\text{m}$ and varying w_C from $10\mu\text{m}$ to $50\mu\text{m}$.

Fig. 3.11 shows that the convective resistance (and also ΔT_{conv}) drops quickly as the channel width is reduced. Since the goal is to modify the convective resistance to compensate for ΔT_{heat} , it can be postulated that the channel width must no longer be a constant but instead should be a function of the distance along the channel $w_C(z)$. The width must be larger near the inlet where the fluid temperature is low and smaller near the outlet where the fluid temperature is high.

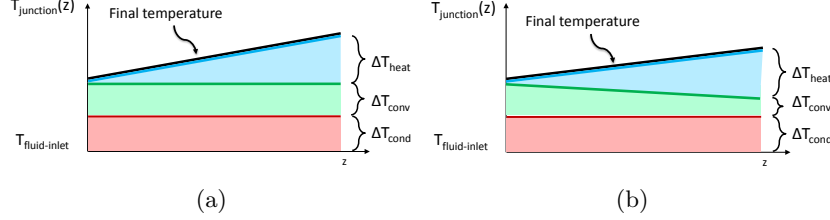


Figure 3.10: Junction temperature distribution for the structure in Fig. 3.9: (a) with uniform non modulated channel width, (b) with modulated channel width to compensate for sensible heat absorption [96].

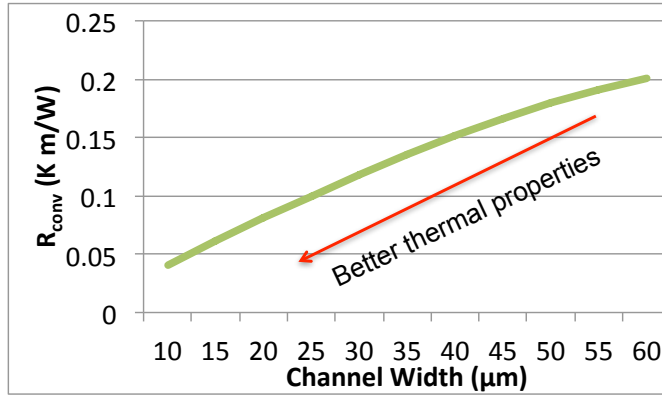


Figure 3.11: R_{conv} as a function of the channel width for the structure in Fig. 3.9.

Hence, theoretically, for the case of uniform heat flux, it is possible to lower the final thermal gradient by steadily modulating the channel width from inlet to outlet, as shown in Fig. 3.10(b).

GREENCOOL uses this principle in formulating an optimal control problem to find the optimal channel width profile for each microchannel, from the fluid inlet to outlet ports. The target of this optimization is to minimize the peak temperature and thermal gradients of the 3D MPSoC, as well as reducing the energy needed by cooling. When applied various 3D MPSoC architectures, significant thermal gradient reductions as well as cooling power savings, with respect to worst-case designs are achieved. For instance, when *GREENCOOL* is applied to different architectural layouts of the UltraSPARC T1 Niagara MP-SoC [14], a 31% thermal gradient reduction is achieved. Fig. 3.12 shows

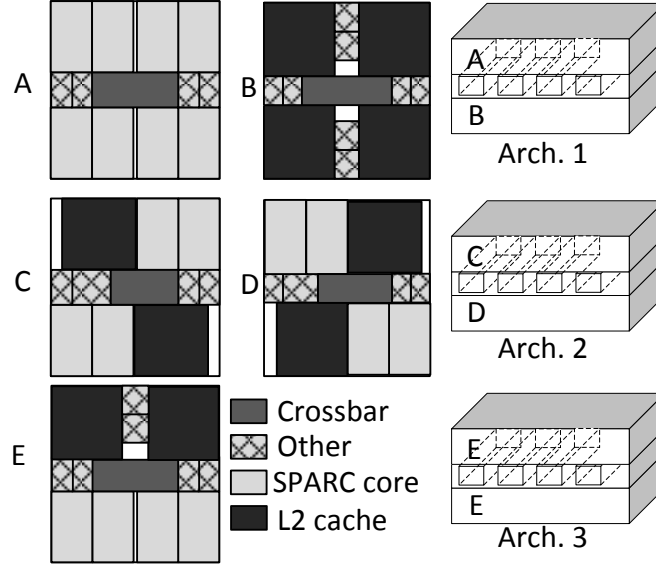


Figure 3.12: Layout of the 3D-MPSoCs used in channel modulation evaluation [40].

the layout of the different two-dies 3D-MPSoCs used in this experiment. The dies are of size $1\text{ cm} \times 1.1\text{ cm}$ and the heat flux densities range from 8 W/cm^2 to 64 W/cm^2 in the two dies. Further details about the floor-plan and power dissipations can be found in pervious works [99, 38, 14].

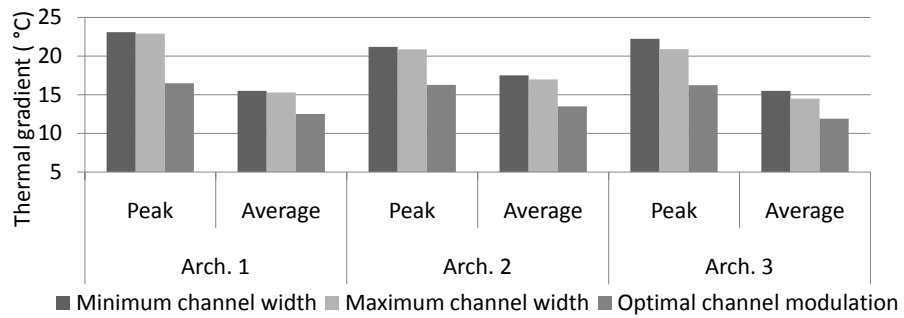


Figure 3.13: Thermal gradients observed in the different 3D-MPSoC architectures dissipating peak and average level heat fluxes, using maximum, minimum and optimally modulated channel widths [40].

In this experiment, the worst-case (peak) power dissipation of the 3D-MPSoC functional elements [99, 38, 14] (obtained using measurements) are used in the optimization process. *GREENCOOL* achieves a thermal gradient reduction of 31% (23°C to 16°C). When the peak heat flux levels were replaced by average values, this same optimal channel modulation configuration manages to reduce the thermal gradient by 21% compared to the uniform channel width case. The thermal gradients obtained for the different cases and for various channel types are plotted in Fig. 3.13.

In another set of experiments to demonstrate the energy-efficiency of *GREENCOOL*, significant cooling energy savings that reach up to 80% has been achieved [96]. Furthermore, *GREENCOOL* aids in developing efficient cooling layout in the cases where uniform cavity utilization is infeasible.

3.3 Summary

In this Section, we have explored the various temperature-aware design-time optimization mechanisms for 3D MPSoCs. We generally split these mechanisms to two research directions that, on the one hand, optimize the heat generation and the flow in the 3D MPSoC (*on-chip* optimization), and on the other hand optimize the heat dissipation paths from the 3D MPSoC to the surrounding environment (*off-chip* optimization). In each of these classes, we further classify them into subclasses to highlight the several research directions that all aim to optimize the thermal profile of 3D MPSoCs. We demonstrate these optimizations on different 3D MPSoC technologies, which include thermal through-silicon vias (TTSVs) and interlayer liquid cooling. In addition, we have shown our own contribution in design-time optimization of liquid-cooled 3D MPSoCs, namely *GREENCOOL*. *GREENCOOL* manages to minimize the cooling energy demands, as well as the peak temperatures and thermal gradients of the target architecture.

4

Temperature-Aware Runtime Management for 3D MPSoCs

The recent advances in thermally-aware design-time optimization schemes, that are explored in the previous section, manage to reduce the thermal dissipation, as well as the cooling costs, for 3D MPSoCs. Thus, 3D MPSoCs would be able to achieve maximal operating conditions without incurring significant degradation. However, thermally-aware design optimization techniques often target the worst-case or the average-case scenarios [92]. This may not be the usual case, as the thermal profile of a typical 3D MPSoC is mainly dependent on the workload conditions [100].

These workload conditions vary based on the processing and memory access activities [101], as well as the target processing domains. If the target domain is within the application-specific case where workload conditions are known in advance, then design-time optimization schemes are beneficial [100]. However, in the case of high-performance computing (HPC), or even in data-dependent domains, large variations in the processing demands occur that diminish the effectiveness of the design-time optimization schemes. For example, recent work [102] demonstrates the significant variation in a typical HPC server workload in time, as shown in Fig. 4.1. Fig. 4.1 shows the utilization rate of two

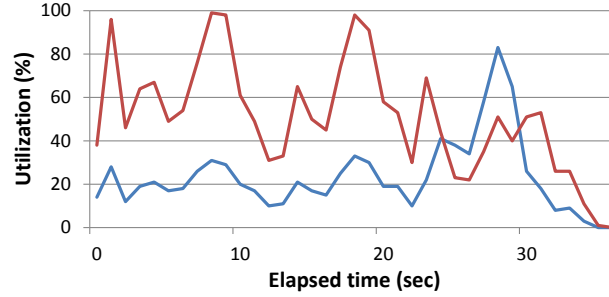


Figure 4.1: Workload-dependent server processing utilization rate variation [102]. Each curve represents the workload activity of a virtual machine that is mapped to a number of physical processing machines.

virtual machines that are mapped to a number of physical processing machines.

If the mentioned thermally-aware design optimization techniques are used as-is in the mentioned various operating conditions, several cases of overcooling or thermal run-away situations would occur [12]. Thus, it is essential to design run-time dynamic thermal management (DTM) schemes that adapt either the heat generation, heat flow, or removal for energy-efficiency, thermal balancing, and thermal hot spots diminishing. These management schemes exploit the utilization of the existing temperature-affecting control knobs in different layers of abstraction of the system, to aid in thermal reduction and balancing. In addition, a fundamental challenge of any developed thermal management scheme is to have minimal performance degradation. If thermal management mechanism has a significant impact on the processing performance, it interferes with the architectural characteristics, hence considered a degrading rather than a benefitting element. Being associated with generated power, implicit thermal balancing can be achieved by power management. For instance, Ogras et al. [103] propose the control of power usage in processing elements (PEs) and routers by using model predictive control at design time, and Bogdan et al. [104] elaborate further this approach by considering both PEs and routers in the control scheme for voltage and frequency. However, they only consider power management and do not explore thermal control aspects. In fact,

consolidating the power consumption in processing elements could undermine temperature issues while the power consumption is reduced. Thus, explicit thermal management schemes that include temperature as a key role in optimization or imposing temperature as a constraint are required for thermal balancing.

In this section, we explore various multi-level thermal management strategies for 3D MPSoCs. These strategies utilize different control knobs to achieve holistic, robust, and energy-efficient run-time temperature optimization. We start by showing a classification of the possible control knobs in 3D MPSoCs in Section 4.1. Afterwards, we explore the various management schemes based on several classifications. In particular, we show the trends differences between centralized and decentralized management schemes in Section 4.2. Then, we show the split between reactive versus proactive management techniques in Section 4.3. We show the different trends in developing heuristic-based or algorithmic-based thermal management schemes in Section 4.4. Finally, we provide a generalized perspective on the thermal management schemes in Section 4.5 and we summarize the main exploration aspects of run-time thermal management in Section 4.6.

4.1 Temperature-Affecting Control Knobs

As mentioned earlier, run-time management schemes utilize various control knobs that either reduce the causes of high heat generation, or increase the ability of the utilized cooling methodology. In the case of 3D MPSoCs, these control knobs are classified as follows.

Workload Activity Knobs At the software-level (application, system software, and OS), workloads can be altered and customized such that they can be thermally-aware. For example, task scheduling and task migration [105] have been extensively used to balance the workload on planar 2D MPSoCs [20]. Another example involves the intra-task instruction scheduling to prevent the processing element temperature from elevating to alarming values.

Circuit Switching Activity Knobs This class of control knobs affects the operating conditions of the processing element. These knobs may stall the processing element temporarily to reduce the heat generation, such as clock gating [19]. Alternatively, these knobs may reduce the operating speed (and/or voltage) of the processing element, which implies lower power consumption, hence lower heat generation, such as dynamic frequency scaling (DFS) or dynamic voltage and frequency scaling (DVFS) [19, 106].

Thermal Package Control Knobs The knobs at the thermal package level are responsible of changing the cooling capacity, which is related to the injected fluid in the case of 3D MPSoCs with liquid cooling, or the rotation speed of the cooling fan in the case of air-cooled 3D MPSoCs. For instance, the volumetric flow rate of the injected fluid can be varied by changing either the liquid pumping power [99], or varying the value of a flow-control valve [107].

4.2 Centralized Versus Decentralized Management Schemes

The first classification we explore is centralized versus decentralized thermal management schemes. This classification relates to *where the thermal management decision is taken*. The difference between centralized and decentralized management schemes can be briefly observed in Fig. 4.2. In the centralized category, a single control unit, algorithm, or a software thread is responsible for the management decision of all the used control knobs. Centralized thermal management schemes are preferably used in the following scenarios:

- The number of control variables is fairly low. This implies that the targeted problem is small and the overhead in creating a decentralized management scheme is significant with respect to the gains.
- The targeted problem has significant similarity cases, which means that several control knobs would follow the same control decision that a single thermal management unit is sufficient to deduce this decision.

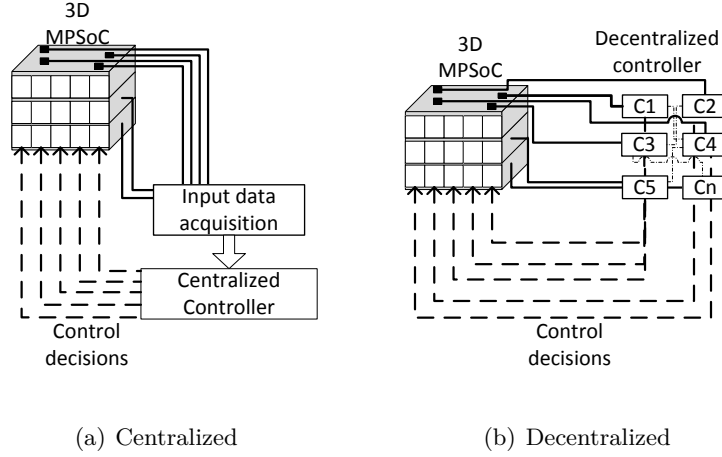


Figure 4.2: Schematic diagrams of (a) centralized and (b) decentralized thermal management schemes.

The decentralized category however, involves a number of small-scale (or micro) thermal managers that each controls a specific portion of the whole system, and interacting with each other to achieve a predefined holistic goal. This category of decentralized management schemes is beneficial and deployed in the following cases:

- The problem size is large with many control knobs that each requires individual management policy. If a centralized controller is used instead, it can be found to be impractical in some cases [108]. Furthermore, due to the increased complexity and communication overhead, a centralized scheme can be infeasible due to the significant imposed latency constraints, which cannot be met due to the significant delays between the centralized controller and the corresponding control knobs.
- The target problem is highly heterogeneous that a single online centralized management scheme would not generate an optimal policy, with low complexity, for all of the existing knobs. In fact, the increased problem complexity is the bottleneck of deploying a centralized management scheme.

Decentralized thermal management schemes can be classified to hierarchical and distributed schemes. The hierarchical subcategory includes a number of management levels, where in the top-most level a single manager decides the policies that the lower-level managers would follow to meet the holistic goal known to this top-most manager. Moreover, the inter-manager communication is always from one management level to the precedent or the succussing levels.

Distributed management is different from hierarchical management since all small-scale managers are of the same level and they can communicate with each other in an ad-hoc manner. In this respect, all small-scale managers contribute equally in achieving the holistic objective by several means such as consensus or majority voting.

In the following subsections we demonstrate several state-of-the-art approaches that use either centralized or decentralized management schemes.

4.2.1 Centralized Thermal Management Scheme

As mentioned earlier, the centralized thermal management scheme consists of a single management unit. Thus, a single thread-based manager implemented at the operating system (OS) level is a clear example of how a centralized thermal management scheme is realized. We briefly demonstrate several thermal management schemes that exist in literature.

A centralized management scheme applied at the OS-level for temperature-aware task scheduling has been previously proposed [109]. This work uses the non-uniform heat propagation pattern found in air-cooled 3D MPSoC to allocate the tasks, based on their workload requirements, to the processing units. In particular, this work groups the tasks into sets of super tasks, where each super task is a collection of a number of tasks. The total workload requirement of each super task is similar to the rest of the super tasks. Then processing cores are grouped to super cores, such that a super core contains all the cores that are vertically-aligned. Then, a one-to-one mapping occurs between the super tasks and super cores based on the workload requirements of the super tasks and the temperature of the super cores. Within a

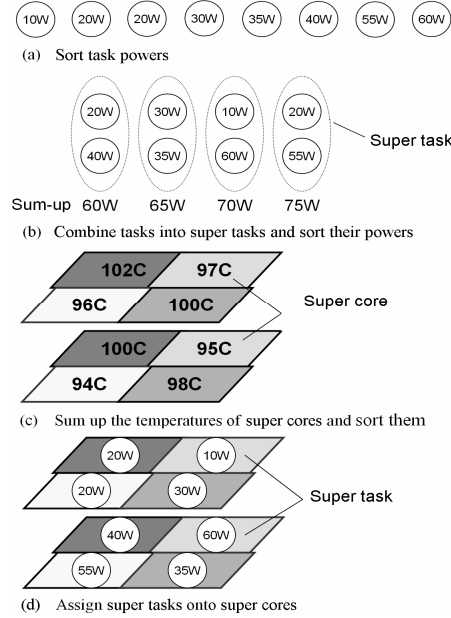


Figure 4.3: Illustration of the centralized management scheme proposed in [109].

single mapping, the tasks within a super task are mapped to the cores based on the relative distance from the heat sink. Fig. 4.3 illustrates the grouping of tasks and cores concept in this work. This technique also combines the task scheduler with DVFS on a 2-tier 3D MPSoC with 8 P4 Northwood cores [109]. When this technique is applied on the previously mentioned 3D MPSoC, it achieves the minimal spatial thermal gradient compared to several task scheduling techniques, in addition to significant peak temperature reduction that reaches 24°C when compared to no thermal management application.

Another centralized management scheme was developed for targeting flow rate control in interlayer liquid-cooled 3D MPSoCs [91]. In this work, a cyber-physical approach is developed to acquire the temperature-related data and apply the flow-rate control mechanism. The authors allow for non-uniform flow rate control in the 3D MPSoC via channel clustering, where in each cluster an arbitrary liquid

flow rate can be applied. Within each cluster, the management algorithm checks the peak temperature. If the peak temperature exceeds a predefined threshold, the flow rate is augmented based on the difference between the peak and threshold temperatures. Similarly, the flow rate is reduced if the peak temperature drops below another predefined threshold. With the fine grained non-uniform control of flow rate, this technique manages to keep the operating temperature of the 3D MP-SoC below the defined peak threshold (85°C) while saving up to 72.1% of cooling power consumption, when compared to uniform flow rate control.

4.2.2 Decentralized Thermal Management Scheme

The structure of a decentralized thermal manager consists of a group of intercommunicated controllers (cf. Fig. 4.2(b)), either through a hierarchical or complete distributed fashions. In the following paragraphs we demonstrate the state-of-the-art in decentralized management schemes.

Previous work proposed a hierarchical thermal management scheme to control task assignment, migration, and DVFS in air-cooled 3D MP-SoC [21]. This previous work explores thermal profiles of adjacent processing elements being on the same vertical column (inter-layer adjacent) or within the same layer (intra-layer). Based on this analysis, the authors derive a set of guidance rules to maximize the throughput locally while minimizing the global thermal profile of the target 3D MPSoC. These guidelines are later used to derive a combined DVFS and task migration policy, named *THERMOS*. *THERMOS* uses both OS and hardware layers in the applied schemes, as shown in Table 4.1. At the OS layer a global thermal-power budget and workload migration policy is implemented. At the hardware layer, DVFS and clock throttling control is applied locally to each processing unit to meet the transient thermal variations to meet the global power and thermal demands. When compared to a distributed approach for only DVFS control [19], *THERMOS* achieves 29% throughput enhancement, while guaranteeing 0% thermal violation.

Another approach uses agent-based thermal management [110] for 3D MPSoCs. This approach applies proactive temperature-aware task

Table 4.1: ThermOS Implementation [21]

Offline computation			Derive the lookup table that contains the optimal voltage and frequencies, given the activity factor range of on-chip core.
Online	OS	Rebalance routine	Invoke two routines at the beginning of each workload migration time interval (e.g. 20ms) to conduct inter-layer or intra-layer migrations.
		Scheduler routine	Monitor the activity factors of run-time processes using performance counters and determine the global power-thermal budgeting using run-time lookup table.
	Hardware	Local DVFS	Proactive distributed DVFS based on global guidance and local variation.
		Local clock	Reactive distributed clock throttling to guarantee thermal safety.

migration through the communication between different agents to deduce the task entering thermal emergency and the most appropriate physical location to have this task migrated to. In addition, this technique applies dynamic virtualized inter-task communication to minimize the impact of migrating a task to the overall throughput and performance.

4.3 Reactive Versus Proactive Management Schemes

This class of classification distinguishes between the various thermal management schemes by giving the answer to the following question; *when the manager applies the control action*. The first class of management schemes apply the control action based on the *current* measurement, hence called *reactive management*. The second class however, applies the control based on the *predicted* measurement trajectory, taking a *proactive management* decision. The main difference in operation between these two management schemes is illustrated in Fig. 4.4.

The reactive thermal management schemes are dependent on the current, and possibly the history, of the measured variables, which are mainly the 3D MPSoC temperatures. This is mainly useful when the measured variable cannot be easily predicted or when the prediction

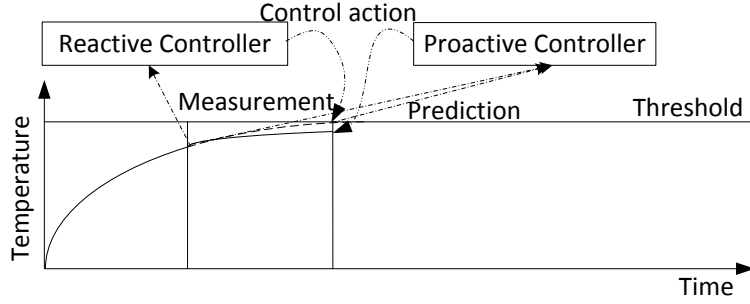


Figure 4.4: Schematic diagram showing the difference between reactive and proactive controllers.

would add instability to the control policy and controlled variable. Reactive management provides stability and robustness to the controlled process, but it may fail in handling extreme violation scenarios effectively.

On the other hand, proactive thermal management schemes mainly rely on the predicted trajectory of the measured variable. The prediction horizon can be of short or long time windows, and it can be a part of the policy derivation such as model predictive control [111], or the prediction is decoupled from the control action. Proactive management schemes prevent reaching any thermal runaway or hot spot threshold violation, but it has to be applied with very well-formulated problem to avoid any unstable outcome or oscillatory behavior.

In the following subsections we show the recent state-of-the-art management schemes that can be classified either *reactive* or *proactive* schemes.

4.3.1 Reactive Management

A recent paper proposes a temperature-aware scheduling method specifically designed for air-cooled 3D MPSoCs [12]. This scheduling methodology monitors the temperature history of each processing unit when it executes a workload of certain requirements. From this history,

the proposed policy places a probability factor that determines the likelihood of each core to receive a workload, which is also combined with the non-uniform heat dissipation nature of the processing units in 3D MPSoC. This probability ($P(k)$) is calculated for each time step (k) and updated as follows:

$$P(k) = P(k-1) + W \quad (4.1)$$

$$W = \begin{cases} \beta_{inc} \cdot W_{diff} \cdot \frac{1}{\alpha_i} & : W_{diff} \geq 0 \\ \beta_{dec} \cdot W_{diff} \cdot \alpha_i & : W_{diff} < 0 \end{cases} \quad (4.2)$$

$$W_{diff} = T_{pref} - T_{avg} \quad (4.3)$$

where W is the weight factor, T_{pref} is the preferred operating temperature, T_{avg} is the average temperature observed in the history window, α_i ($0 < \alpha_i < 1$) is the thermal index of core i that indicates the ability to dissipate heat from this core to the heat sink and β is an empirically determined constant to decide the rate of change in the probability values. If a certain core's temperature exceeds a predefined thermal threshold, the corresponding probability is set to zero to avoid any overhead in the upcoming time window. This policy is also combined with per-core DVFS which is applied based on the core's location in the 3D MPSoC and on the thermal state. Experimental results (cf. Fig. 4.6) show that this policy reduces thermal hot spots by up to 32% with better spatial and temporal thermal balancing compared to 2D MPSoC-based thermal balancing policies, when it is applied to the 2-tier and 4-tier 3D MPSoCs shown in Fig. 4.5. These policies include clock gating (CGate), temperature-triggered DVFS (DVFS_TT), utilization-based DVFS (DVFS_util), floorplan-induced DVFS (DVFS_FLP), task migration (Migr), and adaptive task allocation mechanisms (AdaptRand).

Another thermal management proposal that performs task scheduling on a reactive manner for 3D MPSoCs has been applied [112]. This proposal mainly targets thermal balancing by thread scheduling and migration from a core that has its temperature exceeding a certain threshold to a cold core. In addition this management scheme adds a constraint for accessing memory units from the processing units, and takes this constraint into consideration in the scheduling policy.

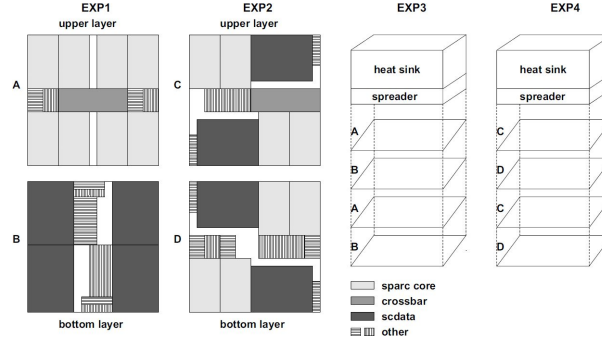


Figure 4.5: Schematic diagram of the floorplans used in experimental evaluation [12].

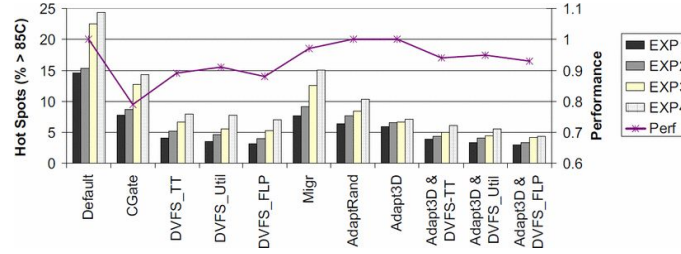


Figure 4.6: Thermal hot spots and performance of the proposed technique in [12], compared to other thermal management mechanisms.

4.3.2 Proactive Management

Proactive thermal management has been used for job allocation and scheduling in previous work [113]. This work tries to allocate tasks based on their relative thermal resistivity path to the heat sink, as well as the predicted thermal temperature of a certain allocation map. This work sorts the jobs based on the current thermal profile, either in ascending or descending orders. Then, for each job in the sorted list, it is tested by allocating it to a certain core. If the predicted temperature is below a certain threshold then allocation is a valid one. Otherwise, the job is examined for allocation to another core.

A more recent proactive management scheme that relies on model

predictive controller (MPC) has been proposed [114]. In this work, a thermal management scheme is developed that controls task scheduling, DVFS, and variable liquid flow rate for interlayer liquid cooled 3D MPSoC. At each time interval, a new set of workloads arrives, and the management scheme allocates these tasks to various cores and sets the corresponding flow rate such that the predicted peak temperature is reduced while minimizing the 3D MPSoC power consumption (cooling and computation power). Then for each processing element it applies MPC to the assigned workload such that the local predicted temperature is reduced while using the minimum computing energy possible via DVFS.

To evaluate the effectiveness of this thermal control, we compare it against different state-of-the-art thermal management techniques, which are as follows:

- **Liquid cooling with LB (LC_LB)** [115]: It applies the maximum cooling flow rate, while the jobs are scheduled with load balancing policy (LB). LB balances the workload by moving threads from a core's queue to another if the difference in queue lengths is over a threshold.
- **LUT-based flow rate control with LB (LC_VAR)** [99]: It dynamically changes the flow rate based on the predicted maximum temperature, while the jobs are scheduled with LB.
- **Fuzzy-logic control (LC_FUZZY)** [38]: This mechanism utilizes fuzzy logic in deriving thermal management mechanism that controls the variable liquid flow rate and DVFS.

In addition we refer to this management scheme as *LC_PROACTIVE* in the following paragraphs. In this evaluation of different thermal management policies, *LC_PROACTIVE* is compared with respect to the other management techniques mentioned above based on the:

- maximum and average temperatures; and
- computational and cooling power consumption.

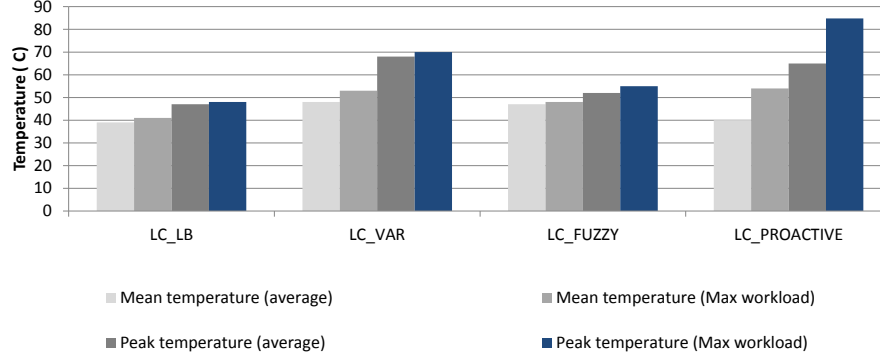


Figure 4.7: Peak and average temperatures observed using all the policies, both for the average case across all workloads and maximum workload on 4-tier 3D MP-SoC [114].

Thermal impact of all the policies on a 4-tier 3D MPSoC is shown in Fig. 4.7. This figure shows that LC_LB reduces the peak temperature to 47°C , whereas LC_FUZZY and LC_VAR push the system into a higher peak of 52°C and 67°C , respectively, but still avoid any hot-spots. This is the similar case in LC_PROACTIVE, where the peak temperature reaches 84°C . The alteration between the peak temperature comes from the fact that main target is to reduce the peak temperature to any value below 85°C . However, since each technique has a different management policy, with different control elements, the peak and average temperatures are affected.

Fig. 4.8 shows the total consumed power when running the various policies on the 4-tier MPSoC with the average workload [114]. Energy consumption values are normalized with respect to the load balancing policy on the 3D-MPSoC with LC_LB. In this figure, LC_PROACTIVE manages to reduce the cooling power and the overall system power by 60% and 23%, respectively, with respect to LC_LB. Moreover, LC_PROACTIVE even reduces the cooling energy more than LC_VAR and LC_FUZZY by 40% and 22%, respectively.

4.4 Heuristic Versus Algorithmic Management Schemes

The third classification of 3D MPSoC thermal management divides the schemes based on the question *how the management action is deduced*.

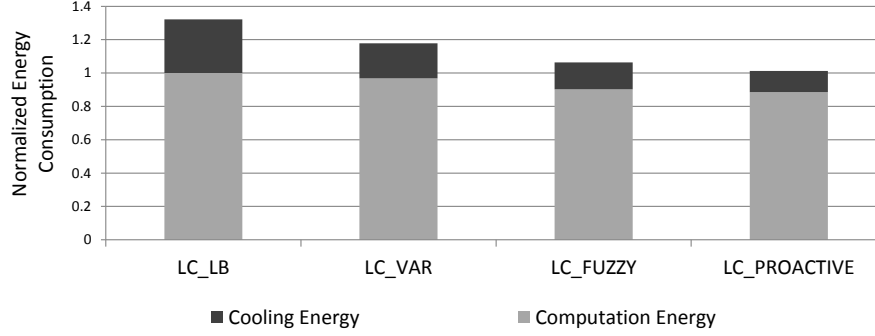


Figure 4.8: The normalized energy consumption in the whole system (chip and cooling network) [114].

The action can be taken either by using a set of heuristics, which can be derived from experimental observations, empirical analysis, rules of thumb, or even expert's knowledge. Heuristic-based management schemes are beneficial, hence utilized, in the following situations:

- The targeted problem is highly complex that limits developing an algorithmic policy, while a crisp set of rules of thumb can easily deduce a near-optimal management policy.
- The targeted problem is simply-formulated such that a heuristic approach would perform optimally with a significantly-reduced overhead and complexity compared to an algorithmic approach.

Another way of taking the management action is through algorithmic development. This approach utilizes the different algorithms for optimizing the management actions such as linear, convex, nonlinear, dynamic and adaptive dynamic programming, as well as optimal control theory approaches. Using algorithms is highly beneficial when the target problem demands a multivariate optimization, with certain constraints and guaranteed performance outcome. Algorithmic management schemes usually demand the target problem to be defined as an optimization one using state-space representation. In the following subsections we highlight the state-of-the-art management schemes that follow either approach.

4.4.1 Heuristic Management

Heuristic-based thermal management scheme developed for interlayer liquid-cooled 3D MPSoCs has been recently proposed [99]. This approach targets load balancing while controlling the flow rate of the injected fluid. The load balancing policy is based on assigning a temperature-induced weight to each processing unit task queue length, such that the processing capability of each core matches its corresponding thermal conductivity. These weights are computed off-line based on empirical analysis and used as a heuristic to balance the loads to different processing units. Then, the liquid flow rate is deduced empirically and stored in look-up tables such that the corresponding flow rate is used when the peak temperature is predicted to be a certain value.

Our previous work on thermal management of liquid-cooled 3D MPSoC has been primarily developed on a thorough experimental analysis of a 3D test-bed with interlayer liquid cooling [38]. We analyse the thermal impact of the various control knobs, namely task scheduling, DVFS, and liquid flow rate on 3D MPSoC designs with respect to the worst case and typical operating conditions. In the analysis of DVFS and varying flow rate, we use an infinite thread input with full utilization. The threads are executed on a variable number of active cores in the 3D test-bed, which is shown in Fig. 4.9(c). This 4-tier 3D test chip prototype (presented in [36, 41]) is shown in Fig. 4.9(a) and Fig. 4.9(b). The 3D MPSoC used in the analysis consists of two tiers, where each tier contains four main hot spot sources modeling high performance processors. The remaining area contains background heaters playing the role of caches, interconnects, and other blocks.

After performing the design-time thermal analysis, a set of rules are derived that are used in this run-time thermal management. This scheme uses fuzzy logic to derive the run-time control actions. Although the same approach can be adapted to use different cyber-physical controllers, we opt to use rule-based fuzzy controller instead of other linear multi-input multi-output controllers (MIMO) [116] as with this technique, we are able to achieve effective control with a straightforward, low-complexity, and flexible implementation. Various low-complexity techniques can be used for deriving the fuzzy rules, such as off-line

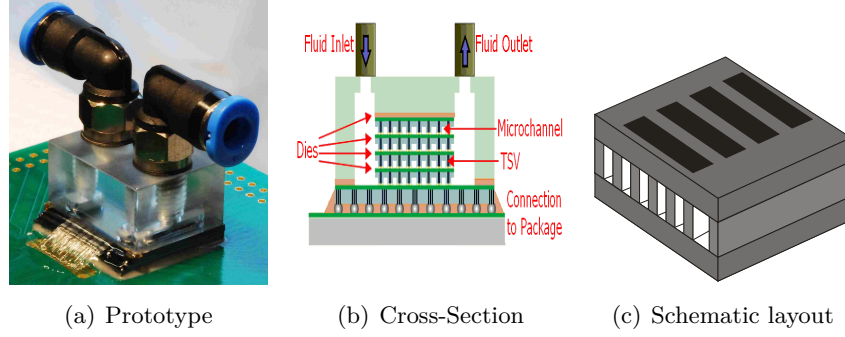


Figure 4.9: The manufactured prototype, cross-section, and layout of the test-bed used in this exploration. The 3D test-bed (c) has four hot spots sources (black), liquid microchannels (white), and background heaters (gray) [38].

analytical analysis and on-line learning mechanisms [117], and fuzzy control can be implemented at the software-level with low overhead. Moreover, fuzzy control operates efficiently at run-time with inputs that have a degree of uncertainty in describing the system state [118], which is the case in 3D MPSoCs where various inputs can be affected by a number of conditions (e.g., ambient temperature changes, unexpected workloads, temperature sensors inaccuracy, stack degradation, etc.).

When applied to various 3D MPSoC architectures, the fuzzy control proposal achieves up to 67% cooling energy savings and up to 30% computational energy savings, compared to baseline liquid-cooled 3D MPSoC with no thermal management mechanism. While the results of this proposal seems effective, there is no proof that this technique guarantees the optimal operating condition. Indeed, this is the downfall of the heuristic mechanisms that favors developing an algorithmic approach, given that it is feasible to develop an algorithmic scheme.

4.4.2 Algorithmic Management

Recent work proposes convex optimization-based thermal management policy for interlayer liquid-cooled 3D MPSoC [119]. This proposal uses the reduced order system of the target 3D MPSoC thermal model. Then, based on thermal sensors measurement, a thermal estimation using Kalman filter is performed. This thermal estimation is used in

the formulated convex optimization problem that manages the DVFS and liquid flow rate of the target 3D MPSoC. The problem is formulated to minimize the energy consumption of the system subjected to temperature and performance constraints. The formulation of this problem is stated as follows:

$$J = \sum_{\tau=1}^h \left(\|\mathbf{R}\mathbf{p}_\tau\| + \|\mathbf{T}\mathbf{u}_\tau\| \right) \quad (4.4)$$

$$\min J \quad (4.5)$$

$$\text{subject to : } f_{\min} \preceq \mathbf{f}_\tau \preceq \mathbf{f}_{\max} \quad \forall \tau \quad (4.6)$$

$$\mathbf{x}_{\tau+1} = \mathbf{A}\mathbf{x}_\tau + \mathbf{B}\mathbf{p}_\tau \quad \forall \tau \quad (4.7)$$

$$\tilde{\mathbf{C}}\mathbf{x}_{\tau+1} \preceq \mathbf{t}_{\max} \quad \forall \tau \quad (4.8)$$

$$\mathbf{u}_\tau \succeq \mathbf{0} \quad \forall \tau \quad (4.9)$$

$$\mathbf{u}_\tau = \mathbf{w}_\tau - \mathbf{f}_\tau \quad \forall \tau \quad (4.10)$$

$$\mathbf{l}_\tau \succeq \mu \mathbf{f}_\tau^2 \quad \forall \tau \quad (4.11)$$

$$-\mathbf{w} \preceq \mathbf{m}_{\tau+1} - \mathbf{m}_\tau \preceq \mathbf{w} \quad \forall \tau \quad (4.12)$$

$$0 \preceq \mathbf{m}_\tau \preceq \mathbf{1} \quad \forall \tau \quad (4.13)$$

$$\mathbf{p}_\tau = [\mathbf{l}_\tau; \mathbf{m}_\tau] \quad \forall \tau \quad (4.14)$$

where matrices \mathbf{A} , \mathbf{B} are related to the overall 3D MPSoC system description. The horizon of this predictive policy is defined as h [120]. Then, the objective function J is expressed by a sum over the horizon.

In the cost function (Eq (4.4)), the first term $\|\mathbf{R}\mathbf{p}_\tau\|$ is the norm of the power input vector p weighted by matrix \mathbf{R} . Matrix \mathbf{R} contains the maximum value of the power consumption of the tiers and the cooling system. The second term $\|\mathbf{T}\mathbf{u}_\tau\|$ is the norm of the required workload, but not yet executed. To this end, the weight matrix \mathbf{T} quantifies the importance that executing the required workload from the scheduler has in the optimization process. Then, Inequality 4.6 defines a range of working frequencies to be used.

Equation 4.7 defines the evolution of the 3D MPSoC according to the present state and inputs. Equation 4.8 states that temperature constraints should be respected at all times and in all specified locations. Since the system cannot execute jobs that have not arrived, every entry

of \mathbf{u}_τ has to be greater than or equal to 0 as stated by Equation 4.9. The undone work at time τ , u_τ is defined by Equation 4.10. Equation 4.11 defines the relation between the power vector \mathbf{l} and the working frequencies. μ is a technology-dependent constant.

Then, Equations 4.12-4.13 define constraints on the liquid cooling management. The normalized pumping power value (\mathbf{m}) scales, at any time instance τ , from 0 (no liquid injection) to 1 (power at the maximum pressure difference allowable), as shown in Equation 4.13. Moreover, the maximum increment/decrement change in the pumping power value from time (τ) to ($\tau + 1$) is limited by a another normalized value \mathbf{w} , as shown in Equation 4.12, which models the mechanical dynamics of the pump.

Finally, the control problem is formulated over an interval of h time steps, which starts at current time τ . Indeed the result of the optimization is an optimal sequence of future control moves (i.e., amount of workload to be executed in average for each tier of the 3D MPSoC which is stored in vector \mathbf{f}). Then, only the first samples of such a sequence are applied to the target 3D MPSoC, while the remaining moves are discarded. Thus, at each next time step, a new optimal control problem based on new temperature measurements and required frequencies is solved over a shifted prediction horizon (e.g., the "receding-horizon" [120] mechanism), which represents a way of transforming an open-loop design methodology into a feedback one, as at every time step the input applied to the process depends on the most recent measurements. This work manages to reduce the 3D MPSoC energy consumption and temperature, compared to a heuristic-based management scheme [38] by up to 16%. However, this algorithmic approach would suffer from an increased solving time, hence degrading the 3D MPSoC performance by a considerable value.

Another proposal that uses algorithmic-based management is proposed for air-based 3D MPSoC [121]. This work tries to maximize the 3D MPSoC throughput, while meeting the power and thermal constraints. This is achieved via controlling the operating voltage and frequency level of each processing element. To achieve the desired metrics, the problem is formulated as a polynomial programming problem.

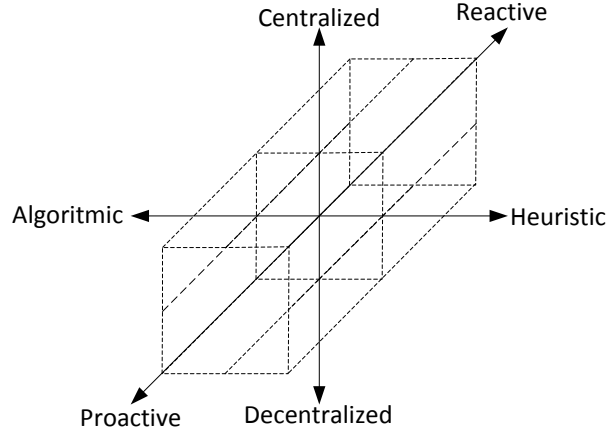


Figure 4.10: 3-dimensional space of the thermal management mechanisms based on the classifications in sections 4.2, 4.3, and 4.4.

4.5 The Global Perspective on Thermal Management

We have shown in the previous sections the various classifications of thermal management mechanisms. However, these classifications are not decoupled from each other. In fact, these classification are orthogonal to each other, and if combined with the various control knobs classes, the thermal management research hyperspace can be formed (cf. Fig. 4.10).

It is expected that each thermal management scheme can be classified by either of these classifications. Thus, we report for each of the elaborated work its classification in all the classes in Table 4.2. This table shows that the development trend for 3D MPSoC thermal management schemes is towards centralized, proactive and heuristic-based schemes. Moreover, the state-of-the-art favors handling workloads than controlling the thermal package. Most of the work on control the thermal package of 3D MPSoC deals with interlayer liquid-cooled designs. These classification choices have been widely used due to the symmetry, the complexity of heat transfer, and the prevention of severe thermal runaway situations in 3D MPSoC.

Table 4.2: Classification of selected state-of-the-art in all the classes

Management Scheme	Classification 1		Classification 2		Classification 3		Control Knobs	
	Centralized	Decentralized	Reactive	Proactive	Heuristic	Algorithmic	Workload	Circuit
Zhou et al. [109]	✓		✓		✓		✓	
Qian et al. [91]	✓			✓	✓			
Zhu et al. [21]		✓		✓	✓		✓	✓
Ebi et al. [110]		✓		✓	✓		✓	
Coskun et al. [12]	✓		✓		✓		✓	
Wang et al. [112]	✓		✓		✓		✓	
Liu et al. [113]	✓			✓	✓		✓	
Zanini, Sabry et al. [114]		✓		✓		✓	✓	✓
Coskun et al. [99]	✓			✓	✓		✓	✓
Sabry et al. [38]	✓			✓	✓		✓	✓
Zanini et al. [119]	✓			✓		✓	✓	✓
Zhao et al. [121]	✓			✓		✓	✓	✓

However, as our understanding of the heat propagation on 3D MP-SoCs evolves, other design parameters are selected in developing more robust, effective, and efficient thermal management schemes.

4.6 Summary

We have explored in this section the various run-time thermal management mechanisms for 3D MPSoCs. Run-time thermal management strategies are crucial to adapt the thermal state of the target system based on the dynamic changes of operating conditions. In the context of MPSoCs (2D and 3D), the main dynamic factor is the workload variation. The adaptation of the thermal state mainly targets thermal balancing by efficiently-using the existing temperature-affecting resources. We have provided a classification framework that can identify the developed management schemes in literature based on several aspect, namely, *where*, *when*, and *how* the thermal management action is taken. We also show that these classifications are orthogonal to each other and hence each management strategy can be classified by all of these aspects.

Conclusion

Vertically-integrated 3D multiprocessors systems-on-chip (3D MP-SoCs) are a key enabling architectural paradigm for enduring the integration of several computational modules within a unit area to provide higher performance computing units. However, 3D MPSoCs bring several challenges, where temperature-based ones are crucial due to the augmented power density. Temperature challenges have opened several research directions to undermine the corresponding impact of high temperatures. In this review, we have explored the state-of-the-art to tackle temperature attenuation in a generalized framework. We have shown the several advanced cooling technologies specifically designed for 3D MPSoCs and the state-of-the-art thermally-aware design-time optimization and run-time management schemes. In particular, we have included in this survey the following:

- We have explored the existing thermal modeling frameworks for 3D MPSoCs, with the emphasis on liquid-cooled 3D MPSoCs modeling, in Section 2. We start in this section by explaining the fundamentals of heat transfer in solids and fluids. Then, we explore the various modeling approaches, namely finite-element and compact finite-difference modeling frameworks, which exist in literature.
- We have investigated the temperature-aware design-time opti-

mization mechanisms in Section 3. In this section, we expand the design-time research direction to include techniques that optimize the heat generation and dissipation within the 3D MPSoC, which is referred to as *On-Chip* optimizations. In addition, we have also explored the research direction that optimizes the heat dissipation paths from 3D MPSoCs to the heat sink layers, which we refer to as *Off-Chip* optimizations.

- We have provided a detailed overview of the run-time thermal management schemes found in literature in Section 4. We have mapped the management directions to a hyper-space by answering the questions *where*, *when*, and *how* the management action is taken. We have shown several works mapped to this hyper-space, highlighting the general trend in the state-of-the-art thermal management schemes.

While this review is primarily written to cover the existing methodologies and techniques, the survey actually provides insights on possible future directions related to this research discipline (temperature-aware design and management). We briefly mention these directions as follows:

- Thermal modeling of advanced cooling technologies. While several works on single-phase liquid cooling modeling of 3D MPSoCs exist, the amount of work on two-phase cooling modeling is not significant for a mature modeling framework outcome. This is in fact related to the increased modeling complexity of this cooling technology, which is superior than single-phase cooling. However, there are several on going works [53, 88, 122] that would eventually lead to providing a valid model of this advanced cooling technology.
- Complementary to modeling advanced cooling technologies, corresponding design-time optimizations and run-time management schemes would be required for these technologies. The introduction of these technologies would demand higher complexity of the developed schemes to harness the potential of these new technologies.

- Cross-layer and formal methods development of design-time optimization and run-time management schemes. While the previous works use several knobs simultaneously in achieving highly efficient optimization schemes, at design-time and run-time, exploiting the inter-communication between these knobs (e.g., application characteristics, application mapping, DVFS, active cooling control,...) is yet to be examined. This is indeed a great opportunity for prospective researchers, as in the case of the emerging cross-layer reliability and error resilience [123, 124].

Acknowledgements

The authors would like to thank Dr. Arvind Sridhar from EPFL, Prof. Ayse K. Coskun from Boston University, and Dr. Bruno Michel and Dr. Thomas Brunschweiler from IBM Zurich, for their support with the discussions on the liquid cooling modeling and experiments of the different thermal management strategies of 3D MPSoC stacks with liquid cooling support. This work was supported in part by the Swiss Confederation under the TRANSCEND Nano-Tera Strategic Action and the CMOSAIC RTD project, as well as the EC in the 7th Framework Program under the GreenDataNet and PRO3D STREP Projects.

References

- [1] M. Pedram. Energy-efficient datacenters. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 31(10):1465–1484, 2012.
- [2] G. E. Moore. Cramming more components onto integrated circuits. *Electronics Magazine*, 38(8):114–117, 1965.
- [3] W. Knight. Two heads are better than one [dual-core processors]. *IEE Review*, 51(9):32–35, 2005.
- [4] P. P. Gelsinger. Microprocessors for the new millennium: Challenges, opportunities, and new frontiers. In *ISSCC*, pages 22–25, 2001.
- [5] L. Benini and G. De Micheli. Networks on chips: a new soc paradigm. *Computer*, 35(1), 2002.
- [6] T. Hey et al. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research, 2009.
- [7] A. Burns et al. Shimmer a wireless sensor platform for noninvasive biomedical research. *IEEE Sensors Journal*, 10(9), 2010.
- [8] X. Xie et al. A low-power digital ic design inside the wireless endoscopic capsule. *IEEE Journal of Solid-State Circuits*, 41(11), 2006.
- [9] K. Nomura et al. Performance analysis of 3d-ic for multi-core processors in sub-65nm cmos technologies. In *ISACS*, pages 2876–2879, 2010.
- [10] W. R. Davis et al. Demystifying 3d ics: the pros and cons of going vertical. *IEEE Design and Test of Computers*, 22(6):498–510, 2005.

- [11] H. S. Lee and K. Chakrabarty. Test challenges for 3d integrated circuits. *IEEE Design and Test of Computers*, 26(5):26–35, 2009.
- [12] A. K. Coskun et al. Dynamic thermal management in 3D multicore architectures. In *DATE*, pages 1410–1415, 2009.
- [13] J. Meng and A. K. Coskun. Analysis and runtime management of 3d systems with stacked dram for boosting energy efficiency. In *DATE*, 2012.
- [14] A. Leon et al. A power-efficient high-throughput 32-thread SPARC processor. *ISSCC*, 42(1):7 – 16, 2007.
- [15] Yole Development. 3d ic and tsv report. yole.fr/pagesAn/products/Report_sample/3DIC.pdf, 2007.
- [16] M. Horowitz et al. Scaling, power, and the future of cmos. In *IEDM*, pages 7–15, 2005.
- [17] R. G. Dreslinski et al. Near-threshold computing: Reclaiming moore’s law through energy efficient integrated circuits. *Proc. of the IEEE*, 98(2), 2010.
- [18] E. Pop. Energy dissipation and transport in nanoscale devices. *Nano Research*, 3(3):147–169, 2010.
- [19] J. Donald and M. Martonosi. Techniques for multicore thermal management: Classification and new exploration. In *ISCA*, pages 78–88, 2006.
- [20] A. K. Coskun, T. Simunic Rosing, and K. Whisnant. Temperature aware task scheduling in MPSoCs. In *DATE*, pages 1659–1664, 2007.
- [21] C. Zhu et al. Three-dimensional chip-multiprocessor run-time thermal management. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 27(8):1479–1492, August 2008.
- [22] P. G. Del Valle and D. Atienza. Emulation-based transient thermal modeling of 2d/3d systems-on-chip with active cooling. *Microelectronics Journal*, 42:564–571, 2011.
- [23] D. Atienza, G. De Micheli, L. Benini, J. L. Ayala, P. G. Del Valle, M. DeBole, and V. Narayanan. Reliability-aware design for nanometer-scale devices. In *ASPDAC*, pages 549–554, 2008.
- [24] S. M. Lim et al. Effects of BTI during AHTOL on SRAM VMIN. In *IRPS*, 2011.
- [25] T. Grassler et al. Simultaneous extraction of recoverable and permanent components contributing to bias-temperature instability. In *IEDM*, pages 801–804, 2007.

- [26] E. Takeda and N. Suzuki. An empirical model for device degradation due to hot-carrier injection. *IEEE Electron Device Letters*, 4(4):111–113, 1983.
- [27] P. Heremans et al. Temperature dependence of the channel hot-carrier degradation of n-channel mosfet’s. *IEEE Transactions on Electron Devices*, 37(4):980–993, 1990.
- [28] R. Moazzami et al. Temperature acceleration of time-dependent dielectric breakdown. *IEEE Transactions on Electron Devices*, 36(11):2462–2465, 1989.
- [29] D. Qian and D. J. Dumin. The electric field, oxide thickness, time and fluence dependences of trap generation in silicon oxides and their support of the e-model of oxide breakdown. In *IPFA*, 1999.
- [30] J. R. Black. Electromigration—a brief survey and some recent results. *IEEE Transactions on Electron Devices*, 16(4), 1969.
- [31] M. Pedram and S. Nazarian. Thermal modeling, analysis, and management in vlsi circuits: principles and methods. *Proceedings of the IEEE*, 94(8), 2006.
- [32] Y. Taur et al. Cmos scaling into nanometer regime. *Proceedings of the IEEE*, 85(4):486–504, 1997.
- [33] T. Sato et al. On-chip thermal gradient analysis and temperature flattening for soc design. In *ASP-DAC*, pages 1074–1077, 2005.
- [34] Jayanth Srinivasan et al. The case for lifetime reliability-aware microprocessors. In *ISCA’04*, pages 276–287, 2004.
- [35] J. Cong and Y. Zhang. Thermal via planning for 3-d ics. In *ICCAD*, 2005.
- [36] T. Brunschwiler et al. Interlayer cooling potential in vertically integrated packages. *Microsyst. Technol.*, 15(1):57 – 74, 2009.
- [37] T. Brunschwiler et al. Hotspot-optimized interlayer cooling in vertically integrated packages. In *MRS Fall Meeting*, 2009.
- [38] M. M. Sabry et al. Energy-Efficient Multi-Objective Thermal Control for Liquid-Cooled 3D Stacked Architectures. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 30(12):1883–1896, 2011.
- [39] M. M. Sabry et al. Towards thermally-aware design of 3d mpsoes with inter-tier cooling. In *DATE*, pages 1466–1471, 2011.
- [40] M. M. Sabry, A. Sridhar, and D. Atienza. Thermal balancing of liquid-cooled 3d-mpsoes using channel modulation. In *DATE*, 2012.

- [41] A. Sridhar et al. 3D-ICE: Fast compact transient thermal modeling for 3D-ICs with inter-tier liquid cooling. In *ICCAD*, pages 463–470, 2010. <http://esl.epfl.ch/3d-ice.html>.
- [42] Y. Zhan et al. Thermally aware design. *Foundatin and Trends in Electronic Design Automation*, 2(3):255–370, 2008.
- [43] J. Kong et al. Recent thermal management techniques for microprocessors. *ACM Computing Surveys*, 44(3):13:1–13:42, 2012.
- [44] G. Sun et al. Three-dimensional integrated circuits: Design, eda, and architecture. 5(1-2):1–151.
- [45] John Lienhard-IV and John Lienhard-V. *A heat transfer textbook*. Phlogiston Press, Cambridge, Massachusetts, 2006.
- [46] Wei Huang, S. Ghosh, S. Velusamy, K. Sankaranarayanan, K. Skadron, and M.R. Stan. Hotspot: a compact thermal modeling methodology for early-stage vlsi design. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, 14(5):501–513, may 2006.
- [47] F. Incropera, D. Dewitt, T. Bergman, and A. Lavine. *Fundamentals of heat and mass transfer*. John Wiley and Sons, 2007.
- [48] Z. Chen, X. Luo, and S. Liu. Thermal analysis of 3D packaging with a simplified thermal resistance network model and finite element simulation. In *ICEPT-HDP*, pages 737–741, aug. 2010.
- [49] T Wang and C Chen. 3-D thermal-ADI: a linear-time chip level transient thermal simulator. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 21(12):1434–1445, December 2002.
- [50] www.ansys.com/products/fluid-dynamics/cfx.
- [51] J.-M. Koo, S. Im, L. Jiang, and K. E. Goodson. Integrated microchannel cooling for three-dimensional electronic circuit architectures. *Journal of Heat Transfer*, 127(1):49–58, 2005.
- [52] H. Mizunuma et al. Thermal modeling and analysis for 3D-ICs with integrated microchannel cooling. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 30(9):1293–1306, 2011.
- [53] Y. J. Kim et al. Thermal characterization of interlayer microfluidic cooling of three-dimensional integrated circuits with nonuniform heat flux. *Journal of Heat Transfer*, 132(4):1–9, 2010.
- [54] A. Sridhar et al. Compact transient thermal model for 3D-ICs with liquid cooling via enhanced heat transfer cavity geometries. In *THER-MINIC*, pages 1–6, 2010.

- [55] A. Sridhar et al. 3D-ICE: A compact thermal model for early-stage design of liquid-cooled ics. *(to appear in) IEEE Transactions on Computers*, 2013.
- [56] H. Qian et al. Thermal simulator of 3d-ic with modeling of anisotropic tsv conductance and microchannel entrance effects. In *ASPDAC*, 2013.
- [57] A. H. Ajami et al. Analysis of non-uniform temperature-dependent interconnect performance in high performance ics. In *DAC*, 2001.
- [58] A. H. Ajami, K. Banerjee, and M. Pedram. Modeling and analysis of nonuniform substrate temperature effects on global ULSI interconnects. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 24(6):849–861, June 2005.
- [59] H. Su et al. Full-chip leakage estimation considering power supply and temperature variations. In *ISLPED*, pages 78–83, 2003.
- [60] M. Ni et al. An analytical study on the role of thermal tsvs in a 3dic chip stack. In *DATE*, 2010.
- [61] M. Healy et al. Multiobjective microarchitectural floorplanning for 2-D and 3-D ICs. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 26(1), Jan 2007.
- [62] M. S. Bakir et al. 3d heterogeneous integrated systems: Liquid cooling, power delivery, and implementation. In *CICC*, pages 663 – 670, 2008.
- [63] K. Kota, P. Hidalgo, Y. Joshi, and A. Glezer. Thermal management of a 3d chip stack using a liquid interface to a synthetic jet cooled spreader. In *THERMINIC*, 2009.
- [64] A. Furmanczyk et al. Multiphysics modeling of integrated microfluidic-thermoelectric cooling for stacked 3d ics. In *SEMI-THERM*, pages 35–41, 2003.
- [65] H. N. Phan and D. Agonafer. Experimental analysis model of an active cooling method for 3d-ics utilizing a multidimensional configured thermoelectric. In *SEMI-THERM*, pages 55–58, 2010.
- [66] P. Greenalgh. Big.little processing with arm cortex-a15 and cortex-a7. www.arm.com/files/downloads/big.LITTLE_Final.pdf.
- [67] N. Xu et al. Thermal-aware post layout voltage-island generation for 3d ics. *Journal of Computer Science and Technology*, 28(4):671–681, 2013.
- [68] K. Puttaswamy and G. H. Loh. Thermal herding: Microarchitecture techniques for controlling hotspots in high-performance 3d-integrated processors. In *HPCA*, pages 193–204, 2007.

- [69] M. Bao et al. Temperature-aware idle time distribution for energy optimization with dynamic voltage scaling. In *DATE*, 2010.
- [70] B. C. Schafer et al. Temperature-aware compilation for vliwprocessors. In *RTCSA*, 2007.
- [71] Y. Han et al. Temperature aware floorplanning. In *Workshop on Temperature Aware Computing Systems*, 2005.
- [72] K. Sankaranarayanan, S. Velusamy, M. Stan, and K. Skadron. A case for thermal-aware floorplanning at the microarchitectural level. *Journal of Instruction-Level Parallelism*, 8:1–16, 2005.
- [73] W.-L. Hung et al. Thermal-aware floorplanning using genetic algorithms. In *ISQED*, 2005.
- [74] J. Cong, J. Wei, and Y. Zhang. A thermal-driven floorplanning algorithm for 3D-ICs. In *ICCAD*, pages 306–313, 2004.
- [75] W.-L. Hung et al. Interconnect and thermal-aware floorplanning for 3D microprocessors. In *ISQED*, pages 98–104, 2006.
- [76] M. Ekpanyapong et al. Thermal-aware 3d microarchitectural floorplanning. Georgia Institute of Technology, 2004.
- [77] X. Zhou et al. Temperature-aware register reallocation for register file power-density minimization. *ACM TODAES*, 14(2), 2009.
- [78] M. M. Sabry et al. Thermal-aware compilation for system-on-chip processing architectures. In *GLSVLSI*, 2010.
- [79] A. Jorio et al. *Carbon nanotubes: advanced topics in the synthesis, structure, properties and applications*. Springer, 2008.
- [80] M. Freitag et al. Energy dissipation in graphene field-effect transistors. *Nano Letters*, 9(5):1883–1888, 2009.
- [81] A. A. Balandin et al. Superior thermal conductivity of single-layer graphene. *Nano Letters*, 8(3):902–907, 2008.
- [82] B. Goplen and S. Sapatnekar. Thermal via placement in 3d ics. In *ISPD*, pages 167–174, 2005.
- [83] X. Li et al. Lp based white space redistribution for thermal via planning and performance optimization in 3d ics. In *ASPDAC*, pages 209–212, 2008.
- [84] T. Zhang et al. Temperature-aware routing in 3d ics. In *DAC*, 2006.
- [85] Hai Wei et al. Cooling three-dimensional integrated circuits using power delivery networks. In *IEDM*, pages 14.2.1–14.2.4, 2012.

- [86] M. Saini and R. L. Webb. Heat rejection limits of air cooled plane fin heat sinks for computer cooling. *IEEE Transactions on Components and Packaging Technologies*, 26(1):71–79, 2003.
- [87] S. Zimmermann et al. Aquasar: A hot water cooled data center with direct energy reuse. *Energy*, 43(1):237–245, 2012.
- [88] S. Szczukiewicz et al. Two-phase flow boiling in a single layer of future high-performance 3d stacked computer chips. In *ITHERM*, 2012.
- [89] T. Brunschweiler et al. Angle-of-attack investigation of pin-fin arrays in nonuniform heat-removal cavities for interlayer cooled chip stacks. In *SEMI-THERM*, pages 116–124, 2011.
- [90] F. Alfieri et al. 3d integrated water cooling of a composite multilayer stack of chips. *Journal of Heat Transfer*, 132(12), 2010.
- [91] H. Qian et al. Cyber-physical thermal management of 3D multi-core cache-processor system with microfluidic cooling. *ASP Journal of Low Power Electronics*, 7(1):1–12, 2011.
- [92] H. Qian, Ch. Chang, and H. Yu. An efficient channel clustering and flow rate allocation algorithm for non-uniform microfluidic cooling of 3d integrated circuits. *Integration, the VLSI Journal*, 46(1):57–68, 2013.
- [93] B. Shi et al. Non-uniform micro-channel design for stacked 3D-ICs. In *DAC*, pages 658–663, 2011.
- [94] B. Shi and A. Srivastava. Cooling of 3d-ic using non-uniform micro-channels and sensor based dynamic thermal management. In *Annual Allerton Conf. Communication, Control, and Computing*, 2011.
- [95] H. Qian and C. H. Chang. Microchannel splitting and scaling for thermal balancing of liquid-cooled 3dic. In *ISCAS*, 2013.
- [96] M. M. Sabry et al. Greencool: An energy-efficient liquid cooling design technique for 3d mpsoes via channel width modulation. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 32(4):524–537, 2013.
- [97] R. Shah and A. London. *Laminar flow forced convection in ducts*. New York: Academic Press, 1978.
- [98] Y. Tan et al. Modeling and simulation of the lag effect in a deep reactive ion etching process. *Journal of Micromechanics and Microengineering*, 16, 2006.
- [99] A. K. Coskun et al. Energy-efficient variable-flow liquid cooling in 3D stacked architectures. In *DATE*, pages 111–116, 2010.

- [100] A. K. Coskun, T. Simunic Rosing, K. A. Whisnant, and K. C. Gross. Static and dynamic temperature-aware scheduling for multiprocessor socs. *IEEE Transactions on VLSI*, 16(9):1127–1140, Sept. 2008.
- [101] A. Bartolini et al. A distributed and self-calibrating model-predictive controller for energy and thermal management of high-performance multicores. In *DATE*, 2011.
- [102] J. Kim et al. Correlation-aware virtual machine allocation for energy-efficient datacenters. In *DATE*, 2013.
- [103] Y. U. Ogras, R. Marculescu, D. Marculescu, and E. G. Jung. Design and management of voltage-frequency island partitioned networks-on-chip. *IEEE Transactions on VLSI*, 17(3):330–341, 2009.
- [104] P. Bogdan, S. Jian, R. Tornero, and R. Marculescu. An optimal control approach to power management for multi-voltage and frequency islands multiprocessor platforms under highly variable workloads. In *ISNoC*, pages 35–42, 2012.
- [105] J. Choi et al. Thermal-aware task scheduling at the system software level. In *ISLPED*, 2007.
- [106] A. K. Coskun et al. Temperature management in multiprocessor socs using online learning. In *DAC*, pages 890–893, 2008.
- [107] Festo electric automation technology. <http://www.festo-didactic.com/ov3/media/customers/1100/00966360001075223683.pdf>.
- [108] T. Emi et al. Tape: Thermal-aware agent-based power economy for multi/many-core architectures. In *ICCAD*, pages 302–309, 2009.
- [109] X. Zhou et al. Thermal management for 3D processors via task scheduling. In *ICPP*, pages 115–122, 2008.
- [110] T. Ebi et al. Agent-based thermal management using real-time i/o communication relocation for 3d many-cores. In *PATMOS*, pages 112–121, 2011.
- [111] F. Allgöwer and A. Zheng. *Nonlinear Model Predictive Control*. Birkhäuser, 2000.
- [112] H. Wang et al. Thermal management via task scheduling for 3d noc based multi-processor. In *ISOC*, pages 440–444, 2010.
- [113] S. Liu et al. Thermal-aware job allocation and scheduling for three dimensional chip multiprocessor. In *ISQED*, pages 390–398, 2010.
- [114] F. Zanini, M. M. Sabry, D. Atienza, and G. De Micheli. Hierarchical thermal management policy for high-performance 3d systems with liquid cooling. *IEEE JETCAS*, 1(2):88–101, 2011.

- [115] F. Mulas et al. Thermal balancing policy for multiprocessor stream computing platforms. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 28(12):1870–1882, 2009.
- [116] Y. Diao et al. Using mimo linear control for load balancing in computing systems. In *ACC*, pages 2045–2050, 2004.
- [117] M. Su and H. Chang. Application of neural networks incorporated with real-valued genetic algorithms in knowledge acquisition. *Fuzzy Sets and Systems*, 112(1):85 – 97, 2000.
- [118] H. T. Nguyen and N. R. Prasad. *Fuzzy modeling and control, selected works of M. Sugeno*. CRC press, 1999.
- [119] F. Zanini, D. Atienza, and G. De Micheli. A combined sensor placement and convex optimization approach for thermal management in 3d-mpsoc with liquid cooling. *Integration, the VLSI journal*, 46:33–43, 2013.
- [120] A. Bemporad et al. The explicit linear quadratic regulator for constrained systems. *Automatica*, 38(1):3 –20, 2002.
- [121] G. Zhao et al. Processor frequency assignment in three-dimensional mp-socs under thermal constraints by polynomial programming. In *APC-CAS*, pages 1668–1671, 2008.
- [122] A. Sridhar et al. Steam: a fast compact thermal model for two-phase cooling of integrated circuits. In *ICCAD*, 2013.
- [123] P. Gupta et al. Underdesigned and opportunistic computing in presence of hardware variability. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 32(1), 2013.
- [124] J. Henkel et al. Design and architectures for dependable embedded systems. In *CODES+ISSS*, 2011.