Attaining Single-Chip, High-Performance Computing through 3D Systems with Active Cooling

THIS ARTICLE EXPLORES THE BENEFITS AND THE CHALLENGES OF 3D DESIGN AND DISCUSSES NOVEL TECHNIQUES TO INTEGRATE PREDICTIVE COOLING CONTROL WITH CHIP-LEVEL THERMAL-MANAGEMENT METHODS SUCH AS JOB SCHEDULING AND VOLTAGE FREQUENCY SCALING. USING 3D LIQUID-COOLED SYSTEMS WITH INTELLIGENT RUNTIME MANAGEMENT PROVIDES AN ENERGY-EFFICIENT SOLUTION FOR DESIGNING SINGLE-CHIP MANY-CORE ARCHITECTURES.

Ayse K. Coskun
Jie Meng
Boston University
David Atienza
Mohamed M. Sabry
École Polytechnique Fédérale de Lausanne

Performance demands are increasing in data centers and high-performance computing clusters, which today run various applications from document and media processing to scientific computing and complex modeling. In tandem, the industry has moved into building many-core systems, where a single chip has dozens of cores. Intel's Single-Chip Cloud Computer[1] and Tilera's 64-core processors are recent examples of such systems. Although many-core systems could provide immense computational capacity, achieving high performance in such systems is highly challenging. Communication latency and memory bandwidth limit many-core performance. Many-core systems have other challenges as well. Because of the larger die sizes, the manufacturing yield is lower, high reliability is more difficult to achieve, process variations are more severe, and the production cost is higher compared to smaller chips. These challenges accelerate with smaller technology nodes.

An emerging design technique called 3D stacking addresses these challenges. First, because a 3D stacked system has a smaller footprint (that is, per-chip area), the manufacturing yield is higher, reducing the overall design cost.[2,3] The on-chip interconnects are shorter in 3D systems, reducing the wire delay and capacitance. Because the through-silicon vias (TSVs) connecting the layers do not adhere to the chip's pin-out restrictions, we can build interconnect architectures with higher bandwidth between cores and memory blocks. However, 3D design accelerates the thermal challenges because of the higher thermal resistivity resulting from vertical stacking. In fact, high temperatures and cooling challenges are among the major gating factors in building high-performance 3D systems.[4]

Prior research has addressed the thermal challenges in 3D systems through design techniques such as temperature-aware floorplanning[5] and dynamic management techniques such as temperature-aware job

0272-1732/11/$26.00 © 2011 IEEE Published by the IEEE Computer Society

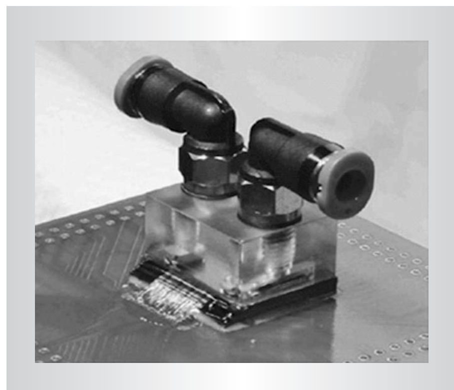

Figure 1. Liquid-cooled 3D chip prototype, built by IBM Zürich and École Polytechnique Fédérale de Lausanne (EPFL).[8] Liquid cooling can more effectively remove heat compared to conventional heat sinks and fans.

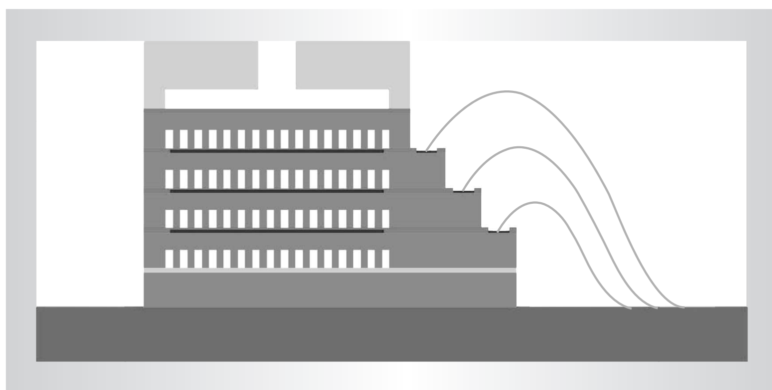

Figure 2. Side view of the 3D liquid-cooled system prototype showing the built-in microchannels. The liquid arrives at the chip through a pipe, gets distributed among the microchannels, flows through the chip, and is collected into another pipe at the outlet.

scheduling and dynamic voltage and frequency scaling (DVFS).[6] Although these approaches provide substantial benefits in reducing the temperatures below critical levels, they are not always sufficient or energy efficient for managing 3D many-core systems. Temperature can increase dramatically in 3D systems (for example, above 100°C), making well-known thermal-management techniques inadequate for controlling temperature without considerably hurting performance.

Active cooling—in which liquid (for example, water) flowing through built-in microchannels (or cold plates) cools the chip—has emerged as a viable cooling alternative for high-performance 3D systems in the past decade.[7] Recently, a prototype 3D system with built-in microchannels was manufactured[8] (see Figures 1 and 2). Liquid cooling has a higher efficiency of removing heat compared to conventional heat sinks and fans, and therefore can address the pressing thermal challenges in 3D systems. Liquid-cooled 3D systems, however, bring novel challenges in cooling control and in efficient integration with chip-level thermal-management techniques.

Many-core 3D design with active cooling is highly complex, with a number of constraints such as cost, peak power, energy, reliability, and yield. Solving these challenges through design-time optimization adds to the area overhead, increases time-to-market and cost, and often results in suboptimal operation owing to dynamic variations in workload. We propose integrating active-cooling control with thermally aware job scheduling and DVFS to optimize energy efficiency of high-performance 3D systems while maintaining low and stable temperature profiles. Temperature and energy benefits as compared to conventional cooling systems are remarkable: we reduce the peak temperature to 50°C from critically high levels, while achieving over 2× cooling-energy savings on a 64-core 3D system.

A manufacturing perspective on 3D design

Recent chip sizes for many-core systems reach 500 to 600 mm^2. An important advantage of 3D stacking comes from silicon economics: individual chip yield, which is inversely correlated with area, increases when a larger number of chips with smaller areas are manufactured.[2] We can estimate the cost of manufacturing a 3D system, C, using the wafer cost C_{wafer}; wafer utilization U_{3D} (that is, how many 3D systems can be cut from the wafer, considering the chip, TSV, and scribe area); and the yield Y_{system} (see Equation 1).[3]

$$C = \frac{C_{\text{wafer}}}{U_{3\text{D}} \cdot Y_{\text{system}}} \quad (1)$$

We calculate a 3D system's yield by extending the negative binomial-distribution model.

Equation 2 shows the yield computation for 3D systems with known-good-die (KGD) bonding, where the dies are tested prior to bonding. In the equation, D is the defect density, with a typical range between 0.001 per mm^2 and 0.005 per mm^2 ($D = 0.001/mm^2$ in this work). A is the total chip area to be split into n layers, α is the defect clustering ratio (we set α to 4),[3] A_{TSV} is the area overhead of the TSVs, P_{stack} is the probability of having a successful stacking operation for KGDs, and n is the number of stacks. In wafer-to-wafer bonding, the yield loss is higher because individual dies are not tested prior to bonding.

$$Y_{\text{system}} = \left[1 + \frac{D}{\alpha}\left(\frac{A}{n} + A_{\text{TSV}}\right)\right]^{-\alpha} P^{n}_{\text{stack}} \quad (2)$$

Figure 3 shows the manufacturing yield and cost per 3D system for building a 64-core chip using 45-nm technology. We assume a standard 300-mm wafer with an estimated cost of $3,000, and we compare 3D designs with two and four layers to a single-layered many-core system with a total area of 321 mm^2. Table 1 provides the system properties, which are based on the sizes and architectures of the cores and caches in the Intel 48-core Single-Chip Cloud Computer (SCC).[1] This example illustrates the benefits of 3D stacking for designing big chips with respect to cost and yield: building the same many-core chip as a four-tier 3D system instead of a 2D system decreases the per-system manufacturing cost by 24 percent. This analysis assumes a mature, reliable bonding process with a probability of success value (P_{stack}) of 0.99.[3,9] If the stacking success is lower, system yield drops with the number of layers, increasing the cost of 3D design (see Equation 2). Therefore, a key aspect for making 3D design cost-effective is the deployment of a mature, reliable bonding process, which is being pursued by manufacturers in both high-performance computing and embedded markets.

Building liquid-cooled systems requires an additional etching phase for the microchannels. This process is similar to the existing process for the TSV design, which includes etching and filling up the etched space with copper. Based on our discussions with industry, the additional etching phase for the microchannels incurs negligible additional cost with respect to a 3D system without liquid cooling. The bonding phase for liquid-cooled 3D systems introduces additional complexity, because the microchannels remove part of the surface touching the chip, and the gluing of the layers must be performed more carefully. This complexity translates to around 20 percent additional

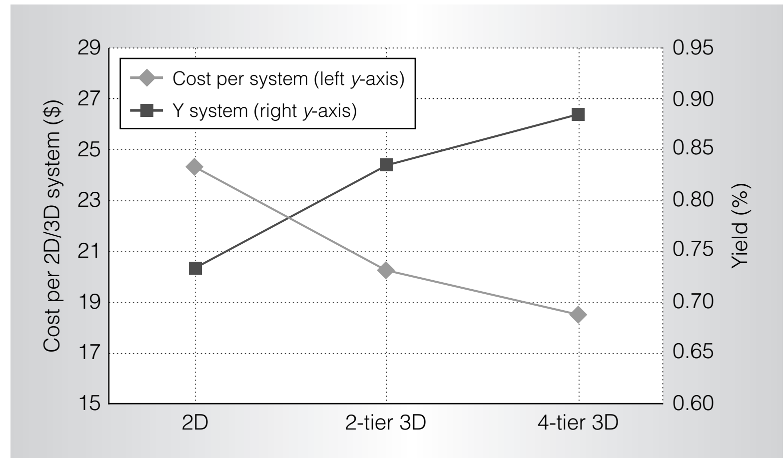

Figure 3. Comparing the cost and yield for building a 64-core chip using 2D and 3D design, assuming a mature stacking process (see Table 1 for chip properties). Building the 64-core chip as a four-tier 3D system decreases the manufacturing cost by 24 percent compared to the 2D chip. This decrease is mainly due to the increase in yield. For the four-tier system, wafer utilization increases slightly as well (by 9 percent), because smaller chips use the wafer area more efficiently.

Table 1. Properties of the 64-core big chip.

Parameter	Value
Process technology	45 nm
No. of cores	64
No. of private Level 2 (L2) caches	64
L2 cache size and area	256 Kbytes, 0.74 mm^2
Core architecture	Similar to Pentium-class cores in Intel Single-Chip Cloud Computer (SCC)[1]
Core area	4.3 mm^2
Through-silicon via (TSV) dimensions	50 μm, × 50μm, 100-μm pitch
On-chip bus width	128 bits

manufacturing costs compared to 3D systems with TSVs only (without microchannels). Recently, researchers have implemented and tested a practical implementation of silicon-based microchannel coolers for high-power chips.[10]

Not only does 3D design improve the manufacturing yield for big chips, but it also helps integrate a larger number of transistors within the same footprint without scaling the technology node. In addition, 3D design allows for putting tested KGDs together to build a large chip. Moving to process nodes beyond 45 nm will be prohibitively expensive for some manufacturers. Therefore, 3D design is both a promising method to pack more functionality on a single chip and a key enabler for the design of big chips.

Thermal challenges of 3D design

3D systems improve manufacturing yield, reduce design cost, and enable high-density integration. Stacking, however, makes it difficult to cool systems effectively through conventional heat sinks and fans, because the thermal resistivity is high for the layers away from the heat sink. According to *International Technology Roadmap for Semiconductors* (*ITRS*) predictions, junction-to-air thermal resistance must drop dramatically below 0.18°C/W to cool future high-performance 3D chips. In practice, such cooling efficiency is nearly impossible to achieve at acceptable packaging and cooling costs and dimensions. Even though conventional heat-sink-based cooling is inadequate for high-performance 3D systems, it is viable for lower-power 3D designs. Therefore, 3D systems developed for embedded-computing environments, where a liquid-cooling infrastructure is not feasible, rely on thermal-management methods to find reliable, energy-efficient operating points.

Thermally aware floorplanning plays a crucial role in controlling peak temperature in 3D systems. Fundamental layout guidelines for temperature-aware 3D design include avoiding placing power-hungry components close together and making effective use of low-power components (such as memory units) to help the heat spread. Thermal modeling is a key component of thermally aware design and runtime management. Today, well-known tools such as HotSpot include 3D-modeling features.[11] In recent work, we extended the HotSpot simulator to model the thermal impact of TSVs.[12] Because copper has higher thermal conductivity than silicon or the interlayer glue material, chip areas with a high TSV density observe a reduction in temperature. This temperature decrease, although usually limited to a few degrees, can be effective in reducing the peak temperature in hot zones.

Although temperature-aware design has considerable benefits in reducing peak and average temperatures, it is often insufficient for eliminating the thermal hot spots in 3D systems, especially when there are more than two layers in the stack. Thus, dynamic thermal-management policies, such as job scheduling and DVFS, are essential in 3D systems, especially at the absence of active cooling.

In Figure 4, we show the effects of dynamic thermal-management policies on 3D systems with conventional cooling infrastructures (that is, only heat sinks and fans, no liquid cooling). Table 2 summarizes the package and floorplan parameters used in the thermal simulations. Table 3 includes the liquid-cooling-related design assumptions (we discuss results with active cooling in the next section). We modeled a heat sink with a convection resistance of 0.1 K/W, representing the heat sinks in modern CPUs. In these experiments, we leveraged a 64-core processor manufactured at 45 nm as our target system. Each core has two-way issue out-of-order execution, two integer units and one floating-point unit, a 16-Kbyte private Level 1 (L1) instruction cache, a 16-Kbyte private L1 data cache, and a 256-Kbyte private L2 cache. The core architecture is based on the cores used in Intel SCC[1] (see Table 1).

We run eight benchmarks from PARSEC[13] (`blackscholes`, `bodytrack`, `canneal`, and `fluidanimate`) and NAS[14] (`cg`, `dc`, `mg`, and `ua`) parallel benchmark suites on the M5 performance simulator[15] to collect performance traces for a range of workloads running on the 64-core system. We use McPAT to derive the power values corresponding to the performance traces.[16]

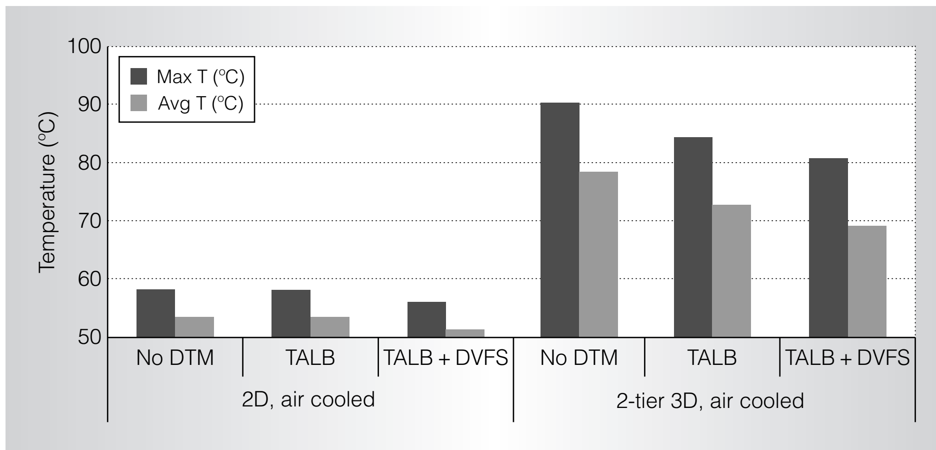

Figure 4. Comparing the peak and average temperatures for the 64-core 2D and 2-tier 3D systems without active cooling for no thermal management (No DTM), temperature-aware load balancing (TALB), and TALB combined with dynamic voltage and frequency scaling (TALB + DVFS). TALB reduces the peak temperature below the critical value of 85°C. TALB + DVFS reduce the temperatures further. Peak and average temperatures exceed 100°C and 90°C, respectively, for the four-tier 3D system even with TALB + DVFS, making the design infeasible.

In McPAT, we set the supply voltage to 1.14 V and the frequency to 1 GHz.[1] We calibrate the power results from McPAT to match the reported power values for the cores in the Intel SCC. We assume the cores have the following additional voltage-frequency pairs available: (0.75 V, 366 MHz), (0.9 V, 590 MHz), and (1.05 V, 866 MHz). The average core power at the highest speed (dynamic and leakage) is 1.8 W, whereas the leakage component is 0.7 W at 75°C. Each L2 cache consumes 0.3 W dynamic power and 0.17 W leakage power at 75°C. We compute the temperature dependence of leakage power using a common exponential formulation.[17]

Table 2. Floorplan and package parameters used in the thermal simulations.

Parameter	Value
Silicon conductivity	130 W/(m·K)
Silicon capacitance	1,635,660 J/(m^3·K)
Wiring-layer conductivity	2.25 W/(m·K)
Wiring-layer capacitance	2,174,502 J/(m^3·K)
Heat sink conductivity	10 W/K
Heat sink capacitance	140 J/K
Total area of each tier in two-tier, 64-core 3D stack	176 mm^2
Total area of each tier in four-tier, 64-core 3D stack	88 mm^2

In the 2D floorplan, each core is next to its private L2 cache on an 8 × 8 grid. In the two-tier and four-tier 3D systems, each layer has 32 and 16 core cache pairs, respectively. In this 64-core system, the core and cache areas differ greatly, so we do not place the cores and caches onto separate layers to avoid wasting substantial area. When core and cache sizes are similar, however, it is beneficial to separate the core and memory layers to exploit the heat dissipation from hotter cores to cooler caches. In fact, we expect temperature increases to be within manageable ranges for low-power cache layer stacking options.

The temperature-aware load balancing (TALB) in Figure 4 balances the *weighted utilization* among the cores,[12] which is the actual utilization of the cores multiplied by a thermal factor representing the average thermal stress on the core. In this way, cores that are at locations more prone to hot spots (such as cores on the layers away from a heat sink) receive a lower workload compared to cores at easier-to-cool locations

Table 3. Liquid-cooling design assumptions used in the thermal simulations.

Parameter	Value
Water conductivity	0.6 W/(m·K)
Water capacitance	4,183 J/(kg·K)
Channel width	50 μm
Channel thickness	100 μm
Channel pitch	150 μm
Channel length	11 mm
Number of channels per cavity in the two-tier, 64-core 3D stack	106
Number of channels per cavity in the four-tier, 64-core 3D stack	53

on the chip. The DVFS policy we implement adjusts the voltage frequency level according to the estimated utilization of the cores. TALB has a negligible performance cost and can reduce the temperature below 85°C for all workloads. Combining TALB with DVFS reduces the temperatures further. Such dynamic-management techniques are essential for providing low-cost thermal management of air-cooled 3D systems, and they also offer opportunities to improve cooling efficiency in active-cooling environments.

High-performance 3D systems with active cooling

Researchers have studied the use of convection in microchannels to cool high-power density chips since the initial work by Tuckerman and Pease.[18] Their liquid-cooling system can remove 1,000 W/cm^2; however, the volumetric flow rate and pressure drop are large, making the approach unsuitable for chip-level implementation. Recent work shows that backside liquid cold plates, such as staggered microchannel and distributed return-jet plates, can handle up to 400 W/cm^2 in single-chip applications.[19] Using pin fin structures at a chip size of 1 cm^2, the heat-removal performance is more than 200 W/cm^2 for TSV pitches larger than 50 μm.[7] Although the cooling capacity of active cooling is remarkable, several key challenges must be addressed to enable reliable and efficient operation.

One such challenge is dynamic flow-rate control. The advanced cooling capacity comes with a pumping-energy cost to push the fluid into the microchannels. For example, for a cluster with 60 computing stacks, as proposed in the *Aquasar* data center design, the cooling infrastructure consumes up to 70 W (similar to the power consumption of a many-core chip). Workload changes significantly over time in real-life systems, and there are many idle cycles—especially for data centers, which must operate at medium utilization levels to handle unexpected rises in service requests. Thus, adjusting the liquid flow dynamically to meet the cooling demand saves cooling energy compared to setting a sufficiently high flow rate for handling the worst-case temperatures. Adjusting the flow rate of pumps, however, typically has an overhead in the range of several hundred milliseconds. This overhead requires implementing a predictive control strategy to adjust the cooling before thermal emergencies occur. We also must minimize the flow rate changes for ensuring stable thermal profiles and reliable pump operation.

Another challenge is integrating flow-rate control with chip-level thermal-management techniques. Chip-level techniques such as DVFS and job scheduling successfully reduce and balance the temperature,[6] improving cooling efficiency. Combining various control knobs in a single low-overhead controller, however, is challenging because the control parameters differ in their time constants, performance and energy overheads, and benefits. For example, changing the DVFS setting typically takes tens to hundreds of microseconds, whereas flow-rate changes take hundreds of milliseconds. The overhead for workload scheduling is typically low. However, when the system is highly utilized, job scheduling is not sufficient to control the temperature, requiring more aggressive techniques such as DVFS or flow-rate control. Simultaneously using distinct control knobs at runtime requires a controller capable of making intelligent decisions quickly.

To address these challenges, we propose a dynamic thermal-management strategy to enable high-performance, energy-efficient, and reliable operation of liquid-cooled 3D systems. The dynamic approach uses a

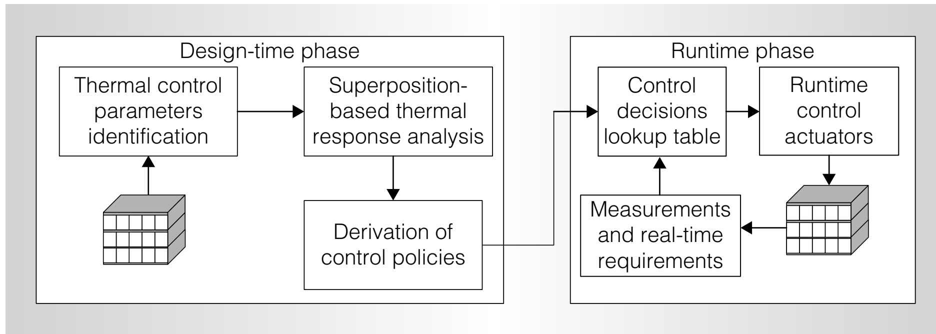

Figure 5. Our proposed design-time and runtime thermal-management approach using a fuzzy-logic controller for DVFS and flow-rate control. Our goal is to minimize energy consumption while maintaining the desired performance and keeping temperature within safe bounds.

design-time thermal analysis of the liquid-cooled 3D system with respect to flow rate, DVFS, and task-scheduling decisions, exploring the thermal impact of tuning each thermal-control knob on the temperature profile. Then, the integrated controller monitors input variables (such as temperature and core utilization), allowing a degree of uncertainty, and tunes the control knobs through a set of predefined rules.

Figure 5 demonstrates the schematic diagram of the combined design-time and runtime thermal-management approach. This approach starts with identifying the input parameters to the controller, the control knobs, and each knob's reaction time. Using this initial analysis, we derive a set of management rules for each knob and combine the rules in a superposition phase to create a global lookup table of control decisions. The thermal controller uses this set of rules at runtime to dynamically set each core's DVFS setting and the chip's coolant flow rate. The objective is to minimize the energy consumption of the cooling infrastructure and the performance degradation while keeping the 3D stack's temperature within the safe bounds.

To enable the 3D system's design-time thermal analysis, we use an automated thermal-modeling tool, 3D-ICE,[20] which can model the microchannels and stacked layers. The modeling principles are similar to those in compact automated thermal models such as HotSpot,[11] except that the interlayer material in our model is a heterogeneous infrastructure to address the differences in the thermal-resistivity values of the microchannels carrying the liquid and the glue material. Additionally, depending on the flow rate provided to push water in the microchannels, the junction temperature of the cells adjacent to the channels changes according to the heat absorption along the channel. We compute the total temperature rise on the junction $\Delta T\mathrm{j}$ as

$$\Delta T_{\mathrm{j}} = \Delta T_{\mathrm{conduction}} + \Delta T_{\mathrm{convection}} + \Delta T_{\mathrm{sensibleheat}}$$

The thermal gradient due to heat conduction through the back-end-of-line (BEOL) layer is computed using the BEOL resistivity. The convection rate depends on the heat transfer coefficients of the materials and is independent of the flow rate when the system reaches boundary conditions. The temperature change resulting from the absorption of sensible heat along the microchannel depends on the volumetric flow rate and is computed iteratively along the channel. In fact, the heat absorption along the channels typically causes a noticeable thermal gradient across the die, as shown in Figure 6. Our prior work provides a detailed understanding of the thermal model for liquid-cooled 3D systems.[12,20]

We leverage a multiple-input, multiple-output fuzzy controller to integrate flow-rate control with DVFS for enabling energy-efficient temperature management on liquid-cooled 3D systems. The controller is

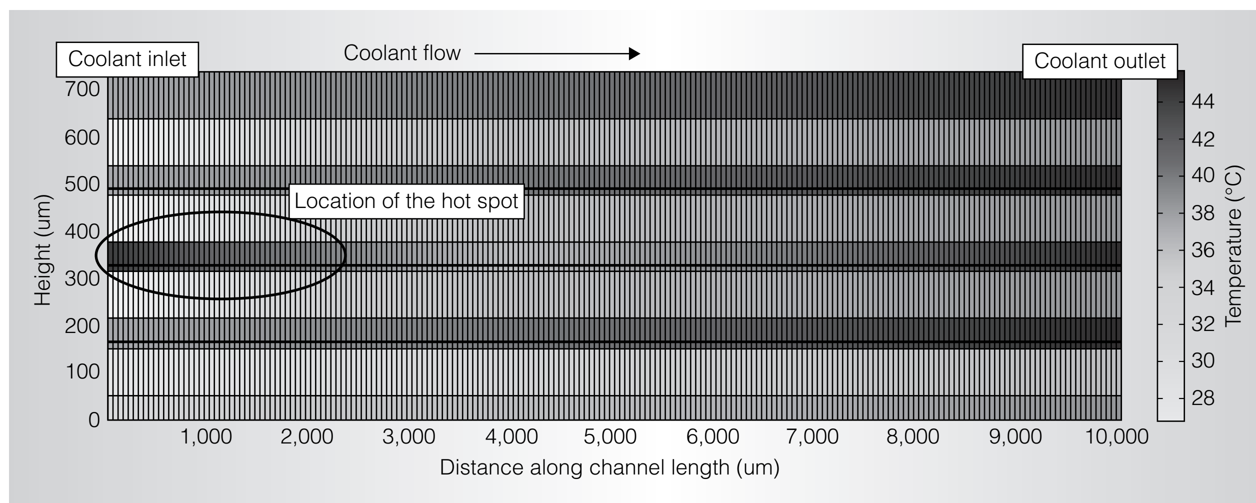

Figure 6. Temperature change across the microchannels on the five-tier liquid-cooled prototype developed by IBM Zürich and EPFL.[7] The temperature change across the channel reaches 15°C for this system. 3D-ICE explicitly models the gradients across the channels.

based on the Takagi-Sugeno Fuzzy model,[21] and its inputs include the distance from the inlet port, predicted maximum chip temperature, and utilization for each core. Thermal prediction is performed through regression-based methods.[12] The outputs are the flow rate of the liquid going into the stack and DVFS settings for each core. The controller makes decisions using the rule base formed during the offline-analysis stage. For example, if the distance from the port is long (that is, cores are in the hotter zone), core utilization is high, and temperature is high, we apply a high flow rate and a medium DVFS level. If a core is close to the inlet port, we can maintain the temperature at a low level using a high DVFS level (incurring no performance overhead) and a low flow-rate setting (please refer to Sabry et al. for the complete set of rules[21]). The global controller can explore energy-efficient solutions and trade-offs that would not be possible using independent control mechanisms.

Figure 7 compares the thermal profiles in the liquid-cooled 64-core chip with dynamic control (Fuzzy + TALB) against liquid-cooled systems that have a static flow-rate setting to handle worst-case temperatures (No DTM and TALB). Table 3 provides the characteristics of the microchannels we used in these experiments. In addition, because each 3D system's footprint determines the total number of microchannels, we include the number of available microchannels per cavity (a cavity is the cooling layer, including a set of microchannels between two adjacent active layers) in the table. The number of microchannels per cavity is lower in the four-tier system, whereas the number of cavities is larger compared to the two-tier system.

We use a maximum flow-rate setting of 46.9 ml/minute and 23.5 ml/minute per cavity for the two-tier and four-tier 3D systems, respectively. We selected these flow rates on the basis of the pumps suitable for liquid-cooled 3D systems (such as 12 V miniature gear pumps), with a pressure drop of 1 bar. The per-cavity flow rate is lower in the four-tier system because we keep the overall flow rate coming into the stack the same in both cases. We use straight microchannels, each with a cross section of 50 μm × 100 μm and a 150-μm pitch to account for TSV placement in the channel walls. Notice from Figure 7 that all the experiments with liquid cooling reduce the peak and average temperatures to below 50°C. The fuzzy controller combines dynamic flow-rate adjustment with DVFS and saves cooling energy by letting the temperature rise within safe margins. We could reduce the temperature to a similar level as Fuzzy + TALB by aggressively applying DVFS or other

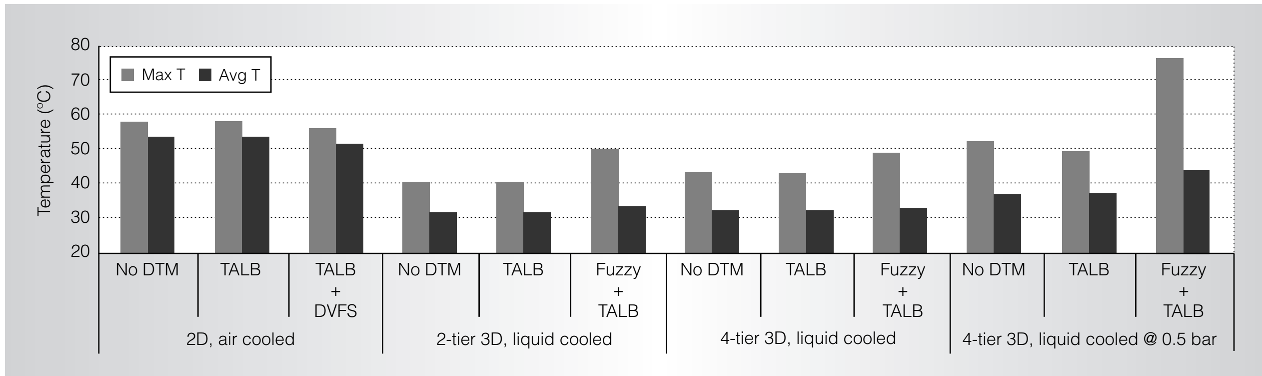

Figure 7. The plot compares maximum and average temperatures between liquid-cooled 3D systems and the 2D air-cooled baseline. Liquid cooling dramatically reduces temperatures across all the workloads. No DTM and TALB policies on the liquid-cooled systems use a static flow rate to cool the chip even under worst-case temperatures. Our fuzzy controller combined with TALB trades off temperature reduction with cooling energy; we allow temperature to increase within safe margins by reducing the flow rate when the cooling demand is lower. The 0.5-bar setting reduces the liquid flow to relax temperature constraints further for larger energy savings.

thermal-management techniques, at the cost of severe performance degradation. Fuzzy + TALB applies liquid-cooling management with performance-aware scheduling and DVFS, limiting the performance cost to less than 1 percent of the original (No DTM) case.

Fuzzy + TALB prevents overcooling and reduces the associated cooling energy by adjusting the flow rate to match the system's cooling needs. For example, for the four-tier 3D liquid-cooled system, cooling energy is reduced by 61 percent compared to the cooling energy spent in No DTM and TALB. Because fuzzy control includes DVFS, the chip-level energy drops as well—for example, 15 percent on average for the four-tier system. The overall energy savings (chip and cooling) on the two-tier and four-tier systems are 16 and 21 percent, respectively. For reducing cooling energy further, we also experiment with a lower flow rate (0.5-bar pressure drop with a flow rate of 16.45 ml/minute) to let the temperature rise without exceeding 80°C. In comparison to the 1-bar setting, the low flow-rate setting reduces cooling energy by 23 percent. However, we see a substantial increase in thermal gradients: while the original setting maintains gradients below 15°C, the 0.5-bar setting causes spatial gradients on the chip to grow significantly higher. Large spatial gradients decrease the cooling efficiency, and are not desirable because they can lead to design complexity or timing problems on the die.

3D systems with active cooling, when intelligently managed by adaptive techniques aware of the physical properties and constraints, offer a cost-efficient solution for building high-performance many-core big chips. Many interesting research problems remain in this area. Our results indicate the need for developing novel temperature-aware floorplanning methods for 3D chips with liquid cooling, addressing the thermal impact of both TSVs and the thermal heterogeneity caused by the liquid flow. We have experimented with straight microchannels in our work, following the recent developments and prototypes in the industry. Two-phase cooling and different microchannel geometries bring several new areas to explore, along with new challenges in manufacturing and control. In addition, studying the reliability of 3D systems is crucial for developing future 3D stacks. For liquid-cooled systems, we expect interesting reliability challenges due to the presence of water. Furthermore, DRAM stacking is a promising feature for enabling higher performance. To thoroughly understand the performance and temperature impact of memory stacking, we need various levels of design space exploration and potentially new modeling techniques. Finally, for

future data centers containing many liquid-cooled stacks, we need approaches to enable reuse of the wasted cooling energy in the hot water coming out of chip stacks. Recent research contributions in this direction include the IBM Aquasar liquid-cooled server rack (including 2D nodes), which recovers a significant part of the cooling energy by reusing the hot water for building heating. MICRO

Acknowledgments

This research has been partially funded by the Nano-Tera RTD project CMOSAIC (ref. 123618), financed by the Swiss Confederation and scientifically evaluated by SNSF, and the PRO3D European Union FP7-ICT-248776 project. We thank Thomas Brunschwiler and Bruno Michel from the Advanced Packaging Group of IBM Zürich, as well as Yusuf Leblebici from the Microelectronic Systems Laboratory of EPFL for their contributions in experimentally validating the thermal model for 3D ICs with intertier liquid cooling.

References

1. J. Howard et al., ''A 48-Core IA-32 Message-Passing Processor with DVFS in 45 nm CMOS,'' *Proc. IEEE Int'l Solid-State Circuits Conf.,* 2010, IEEE Press, pp. 108-109.
2. S. Reda, G. Smith, and L. Smith, ''Maximizing the Functional Yield of Wafer-to-Wafer 3D Integration,'' *IEEE Trans. Very Large Scale Integration Systems,* vol. 17, no. 9, 2009, pp. 1357-1362.
3. A.K. Coskun, A.B. Kahng, and T. Rosing, ''Temperature- and Cost-Aware Design of 3D Multiprocessor Architectures,'' *Proc. 12th Euromicro Conf. Digital System Design, Architectures, Methods, and Tools,* IEEE CS Press, 2009, pp. 183-190.
4. G.H. Loh and Y. Xie, ''3D Stacked Microprocessor: Are We There Yet?'' *IEEE Micro,* vol. 30, no. 3, 2010, pp. 60-64.
5. M. Healy et al., ''Multiobjective Microarchitectural Floorplanning for 2D and 3D ICs,'' *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems,* vol. 26, no. 1, 2007, pp. 38-52.
6. C. Zhu et al., ''Three-Dimensional Chip-Multiprocessor Runtime Thermal Management,'' *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems,* vol. 27, no. 8, 2008, pp. 1479-1492.
7. T. Brunschwiler et al., ''Interlayer Cooling Potential in Vertically Integrated Packages,'' *Microsystem Technologies,* vol. 15, no. 1, 2008, pp. 57-74.
8. T. Brunschwiler et al., ''Validation of the Porous-Medium Approach to Model Interlayer-Cooled 3D-Chip Stacks,'' *Proc. IEEE Int'l Conf. 3D System Integration,* IEEE CS Press, 2009, pp. 1-10.
9. X. Dong and Y. Xie, ''System-Level Cost Analysis and Design Exploration for Three-Dimensional Integrated Circuits (3D ICs),'' *Proc. Asia and South Pacific Design Automation Conference,* IEEE Press, 2009, pp. 234-241.
10. E.G. Colgan et al., ''A Practical Implementation of Silicon Microchannel Coolers for High Power Chips,'' *IEEE Trans. Components and Packaging Technologies,* vol. 30, no. 2, 2007, pp. 218-225.
11. K. Skadron et al., ''Temperature-Aware Microarchitecture,'' *Proc. 30th Ann. Int'l Symp. Computer Architecture,* ACM Press, 2003, pp. 2-13.
12. A.K. Coskun et al., ''Energy-Efficient Variable-Flow Liquid Cooling in 3D Stacked Architectures,'' *Proc. Design Automation and Test in Europe,* IEEE Press, 2010, pp. 111-116.
13. C. Bienia, ''Benchmarking Modern Multiprocessors,'' doctoral dissertation, Computer Science Dept., Princeton University, 2011.
14. D. Bailey et al., *The NAS Parallel Benchmarks,* tech. report RNR-94-007, NASA Ames Research Center, 1994.
15. N.L. Binkert et al., ''The M5 Simulator: Modeling Networked Systems,'' *IEEE Micro,* July/Aug. 2006, pp. 52-60.
16. S. Li et al., ''McPAT: An Integrated Power, Area, and Timing Modeling Framework for Multicore and Many-Core Architectures,'' *Proc. 42nd Ann. IEEE/ACM Int'l Symp. Microarchitecture,* ACM Press, 2009, pp. 469-480.
17. S. Heo, K. Barr, and K. Asanovic, ''Reducing Power Density through Activity Migration,'' *Proc. Int'l Symp. Low-Power Electronics and Design,* ACM Press, 2003, pp. 217-222.
18. D.B. Tuckerman and R.F.W. Pease, ''High-Performance Heat Sinking for VLSI,'' *IEEE*

Electron Device Letters, vol. 2, no. 5, 1981, pp. 126-129.
19. T. Brunschwiler et al., ''Direct Liquid-Jet Impingement Cooling with Micron-Sized Nozzle Array and Distributed Return Architecture,'' *Proc. 10th Intersociety Conf. Thermal and Thermomechanical Phenomena in Electronics Systems,* IEEE Press, 2006, pp. 196-203.
20. A. Sridhar et al., ''3D-ICE: Fast Compact Transient Thermal Modeling for 3D-ICs with Intertier Liquid Cooling,'' *Proc. IEEE/ACM Int'l Conf. Computer-Aided Design,* IEEE Press, 2010, pp. 463-470.
21. M. Sabry, A.K. Coskun, and D. Atienza, ''Fuzzy Control for Enforcing Energy Efficiency in High-Performance 3D Systems,'' *Proc. IEEE/ACM Int'l Conf. Computer-Aided Design,* IEEE Press, 2010, pp. 642-648.

Ayse K. Coskun is an assistant professor in the Electrical and Computer Engineering Department at Boston University. Her research interests include energy-efficient computing, multicore architectures, 3D stacked architectures, embedded systems, and software. Coskun has a PhD in computer science and engineering from the University of California, San Diego. She's a member of IEEE and the ACM.

Jie Meng is a PhD student in electrical and computer engineering at Boston University. Her research interests include many-core and 3D stacked architectures, focusing on energy awareness and performance improvement for future systems. Meng has an MS in electrical engineering from McMaster University in Canada.

David Atienza is a professor of electrical engineering and the director of the Embedded Systems Laboratory at the École Polytechnique Fédérale de Lausanne, and an adjunct professor in the Computer Architecture Department at the Complutense University of Madrid (UCM). His research interests include system-level design methodologies for high-performance multiprocessor systems on chips (MPSoCs) and embedded systems, including 2D and 3D thermally aware design, wireless sensor networks, hardware and software reconfigurable systems, dynamic-memory optimizations, and network-on-chip design. Atienza has a PhD in computer science and engineering from the Inter-University Microelectronics Center, Belgium, and the UCM. He's an associate editor of *IEEE Transactions on CAD*, *IEEE Embedded Systems Letters*, and *Integration: The VLSI Journal*.

Mohamed M. Sabry is a PhD student in the Electrical Engineering Department and a member of the Embedded Systems Laboratory at the École Polytechnique Fédérale de Lausanne. His research interests include system design and resource management methodologies in embedded systems, and MPSoCs, especially temperature and reliability management of 2D and 3D MPSoCs. Sabry has an MS in electrical engineering from AinShams University, Egypt.

Direct questions or comments about this article to Ayse K. Coskun, Electrical and Computer Engineering Department, Boston University, 8 Saint Mary's Street, Boston, MA 02215; acoskun@bu.edu.

Selected CS articles and columns are also available for free at http://ComputingNow.computer.org.