The Case for Water-Immersion Computer Boards

Michihiro Koibuchi
National Institute of Informatics,
The Graduate University for Advanced Studies, 2-1-2, Hitotsubashi,
Chiyoda-ku, Tokyo, JAPAN 101-8430
koibuchi@nii.ac.jp

Ikki Fujiwara
National Institute of Informatics,
2-1-2, Hitotsubashi, Chiyoda-ku,
Tokyo, JAPAN 101-8430
ikki@nii.ac.jp

Naoya Niwa
Keio University, 3-14-1 Hiyoshi,
Kohoku-ku, Yokohama, Kanagawa,
JAPAN 223-8522
naoya@am.ics.keio.ac.jp

Tomohiro Totoki
Keio University, 3-14-1 Hiyoshi,
Kohoku-ku, Yokohama, Kanagawa,
JAPAN 223-8522
totoki@am.ics.keio.ac.jp

Shoichi Hirasawa
National Institute of Informatics,
2-1-2, Hitotsubashi, Chiyoda-ku,
Tokyo, JAPAN 101-8430
hirasawa@nii.ac.jp

ABSTRACT

A key concern for a high-power processor is heat dissipation, which limits the power, and thus the operating frequencies, of chips so as not to exceed some temperature threshold. In particular, 3-D chip integration will further increase power density, thus requiring more efficient cooling technology. While air, fluorinert and mineral oil have been traditionally used as coolants, in this study, we propose to directly use tap or natural water due to its superior thermal conductivity. We have developed the "in-water computer" prototypes that rely on a parylene film insulation coating. Our prototypes can support direct water-immersion cooling by taking and draining natural water, while existing cooling requires the secondary coolant (e.g. outside air in cold climates) for cooling the primary coolants that contact chips. Our prototypes successfully reduce by 20 degrees the chip temperature of commodity processor chips. Our analysis results show that the in-water cooling increases the acceptable amount of power density of chips, thus achieving higher operating frequencies of chips. Through a full-system simulation, our results show that the water-immersion chip multiprocessors outperform the counterpart water-pipe cooled and oil-immersion chips by up to 14% and 4.5%, respectively, in terms of execution times of NAS Parallel Benchmarks.

CCS CONCEPTS

• **Hardware** → *Temperature optimization*; *3D integrated circuits*.

KEYWORDS

In-water computers, liquid immersion cooling, thermal-aware chip multiprocessors (CMPs)

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
ICPP 2019, August 5–8, 2019, Kyoto, Japan
© 2019 Association for Computing Machinery.
ACM ISBN 978-1-4503-6295-5/19/08…$15.00
https://doi.org/10.1145/3337821.3337830

ACM Reference Format:
Michihiro Koibuchi, Ikki Fujiwara, Naoya Niwa, Tomohiro Totoki, and Shoichi Hirasawa. 2019. The Case for Water-Immersion Computer Boards. In *48th International Conference on Parallel Processing (ICPP 2019), August 5–8, 2019, Kyoto, Japan.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3337821.3337830

1 INTRODUCTION

As the power density increases in a chip, a key concern for designing a high-power chip is heat dissipation, which limits power consumption due to enforced temperature thresholds. For example, recent accelerator chips reach over two-hundred Watts, e.g. 245 Watts in Intel Knight Landing 7290, and 500 Watts in an ASIC Bitcoin-mining processor chip [2]. Shortly the 3-D chip integration of a conventional chip multiprocessor (CMP) will be mature that further increases the power density, e.g. 425 Watts in a conventional CMP in 2033 taken from IRDS roadmap[14]. There is a strong need for more efficient cooling on a chip.

Computers have historically been air-cooled. However, liquid coolants have been used successfully in some High-Performance Computing (HPC) and datacenter platforms (e.g., Fluorinert for the Cray-2 series, Mineral oil for the Tsubame-KFC supercomputer [9], Fluorinert for the kukai cluster at Yahoo Japan). These coolants, due to thermal conductivity much higher than that of air, make it possible to increase chip power consumption without violating operating thermal constraints. Furthermore, liquid coolants benefit from natural convection effects, cooling chips more efficiently than air cooling even with high fan speeds. Power usage effectiveness (PUE) as low as 1.03 been reported for liquid-immersion HPC systems [12].

In this study, we present yet another liquid cooling, namely a water-immersion method. Water-immersion cooling is attractive due to (i) high thermal conductivity (when compared to air, mineral oil, and fluorinert); (ii) possibly direct immersion cooling; (iii) lowered safety concerns (when compared to mineral oil and fluorinert); and/or (iv) lower cost of coolants (when compared to mineral oil and fluorinert).

The first advantage of our approach is the high thermal conductivity for better cooling. Figure 1 shows maximum chip operating frequency vs. number of stack chips. The temperature constraint is 78°C taken from its specifications. These results are obtained in simulation, based on a model of the Xeon E5-2667v4 (1.2-3.6

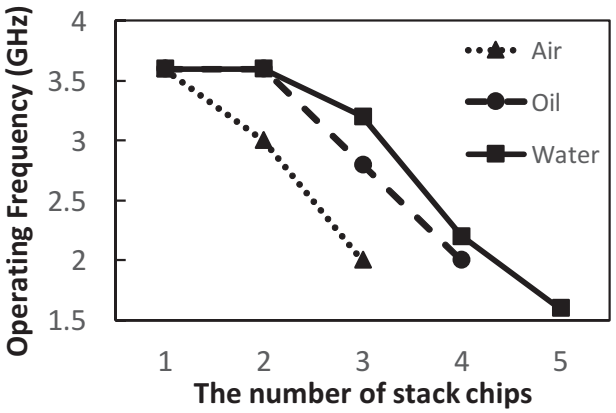


Figure 1: Maximum frequency vs. number of stacked Xeon E5 chips for air, mineral oil and water cooling.

GHz) chip. The maximum power consumption of each voltage-and-frequency scaling (VFS) step is measured with the Intel RAPL (Running Average Power Limit) tool, when executing one `stress` command (which calculates π) on each core. Using the obtained power profile and its physical layout, which we compute based on a high-resolution photo of the die, we obtain the maximum temperature in the CMP layout using HotSpot v6.0 [24] (see upcoming sections for full details on our experimental methodology). The main finding is that air cooling, due to poor heat dissipation properties, limits operating frequencies of 3 stacked chips to 2.0 GHz, and does not enable a 4-chip layout. This limitation is one of the reasons why air-cooled 2.5-D implementations are currently considered for high-power chip integration. Mineral oil dissipates heat better, and thus allows operating 3 chips at 2.8 GHz or 4 chips at 2.0 GHz. Water cooling would improve these frequencies to 3.2 GHz and 2.2 GHz, respectively.

The second advantage of our approach is to enable "direct" cooling. The existing liquid cooling requires the secondary cooling system for cooling the primary coolant, naturally degrading its total efficiency. Our in-water computer prototypes can face natural water. Thus, we can take and drain the waters of river for the primary coolant, or they can be placed under natural river. Our direct cooling does improve its total efficiency.

Existing liquid coolants have the advantage of providing electric insulation because water except pure water provides no electric insulation. We explore in-water computer prototypes that rely on a μm-film insulation coating for electric isolation.

Besides real cooling prototypes for proof of in-water concept, we evaluate the efficiency of our water cooling on 3-D chip multiprocessors (CMPs) using power analysis (McPAT v1.3), thermal analysis (HotSpot v6.0) and full-system simulator (gem5). The water cooling attempts to increase total power consumption: given a number of chips, a temperature threshold, it maximizes the application performance.

This work makes the following contributions:

- We demonstrate the feasibility of in-water commodity computers by using a 120-150μm parylene film to cover motherboards fully. Measured temperature reduction on Xeon E3-1270v5 processors, when compared to air cooling, is about 20°C when executing CPU-intensive workloads. Another finding is that the lifetime of the coated computer is expected to be a couple of years when some parts of motherboard, i.e. area of memory slots, is not coated (is masked when coating). In this case, we should set up it as follows: the hottest area around processors is underwater, while the uncoated area is placed above the surface of water. (Section 2)
- We demonstrate the impact of the water cooling on the application performance of thermal-aware CMPs. The water-immersion CMP improves the chip operating frequency given temperature constraints on HotSpot 6.0 temperature modeling analysis [24][1], leading to up to 14% and 4.5% execution-time reduction for the NAS Parallel Benchmark (NPB), when compared to water-pipe and mineral oil cooling, respectively. We find that the thermal constraint strongly affects the application performance, especially when the number of chips increases on the 3-D integration. (Section 3)
- We explore the impact of coolant heat transfer coefficient on the maximum chip operating frequency and the influence of thermal-aware floorplan optimization on the maximum chip operating frequency. Also, we qualitatively compare direct cooling enabled by our approach to existing liquid cooling at the macro-system point of view. (Section 4)

The remainder of the paper is organized as follows. Section 5 discusses related works. Section 6 concludes with a summary of our findings and perspectives on future work.

2 IN-WATER μ-FILM APPROACH

The main challenge is that water does not provide electric insulation. In some existing computer systems, water cooling has only been used in *heat pipes*, i.e., water pipes placed near the processor, or in radiators for secondary cooling [21, 23, 26]. By contrast, we propose the full immersion of processors in non-pure (e.g., tap) water. Historically, heat pipe approach using water that goes through the surface of all processor chips has been used, however, their efficiency is, of course, not better than that of in-water cooling (see the results in Figures 7 and 8).

2.1 Design

We propose to allow high power-density chips underwater by the use of a μm-film coating technology that is originally designed for electric insulation purposes. After attempting several combinations of film-coating materials, including epoxy resin, we eventually realized working prototypes by covering commodity computers with 120μm and 150μm parylene films. Especially, we use the diX C Plus film provided by KISCO Ltd. [17]. We experimented with thinner parylene films, e.g., 50μm, but the underwater computers failed after only a few hours in operation and could never be booted again after this initial failure.

[1] Our extension to support (1) chip rotation on 3-D integration and (2) floorplans of Intel Xeon E5-2667v4 and Xeon Phi7290 processors is available from [30].

Given a motherboard that includes processors, memory, and I/O cables, our approach consists in fully covering the motherboard with the parylene film, including the I/O connectors and the first few cms of the I/O cables. The coating is applied using a vacuum deposition process known as chemical vapor deposition (CVD). Parylene films applied by CVD provide high electric insulation, high moisture-proofing, and high chemical resistance. This technology therefore widely applied to various commodity components in fields such as aerospace, automobile, and healthcare. Importantly for our purpose, CVD operates at room temperature and thus does not damage ICs. Furthermore, although a motherboard has an uneven non-convex surface, the gaseous coating material penetrates the entire volume and results in an almost uniformly thick film over the entire surface. The parylene material itself is cheap, and the total coating cost would become nominal if a commodity CVD production line were developed for the purpose of producing in-water computers.

The drawbacks of our proposed in-water approach are: (i) a faulty component on the motherboard cannot be replaced since the film coats the entire motherboard; (ii) the coating may raise chip temperature; and (iii) the parylene film has unknown lifetime when used for in-water computers. The first drawback is inherent to our approach. To mitigate the second drawback we simply break the parylene film on each chip's heat-spreader surface, and replace it by TIM (Thermal Interface Material) and a heatsink that is tightly applied to the heat-spreader surface. Our experiments have shown no water leakage due to this technique. To address the third drawback we have developed several prototypes, described in the next section.

2.2 Test Boards

We designed a test board, shown in Figure 2, to investigate the lifetime of each component on a computer. The reason for developing this test board is that it is difficult to identify faulty components in commercial motherboards. The test board uses five distinct voltage supply units and the following seven components: USB, Ethernet (RJ45), mPCIe, PCIex4, CR2032 (micro cell), PGA (pin grid array) and mega-AVR microcontrollers. We picked these components because they have complex physical shapes, and thus may have short film coating lifetimes. In case a component causes a short / leak of electricity due to being immersed in water, the test board identifies the component and measures the amount of leakage.

We have been running 5 boards covered by the 120 and 150 μm parylene film for over 2 years, and counting. To date, the only component faults have been for all five PCIex4 components, one RJ45, and one mPCIe for which small amounts of leakage have been measured. All five CR2032s seem to be electrically discharged. We thus recommend to put PCIex4, RJ45 and mPCIe components above the surface of the water and to remove microcell components from the motherboard.

2.3 Servers

Besides the test-board experiments, we have used the parylene film to cover various-sized motherboards: Intel NUC6i7KYK, ASRock Q1900M (Intel Quad-Core Celeron Processor J1900), AS-1341G (Intel Atom E660T) and FUJITSU Server PRIMERGY TX1320 M2 (Xeon E3-1270v5) (Figure 3). These computers have worked underwater for up to a half year (and counting). However, in the case

Figure 2: In-water Test Board.

of the FUJITSU server, on the 7th day, the server could no longer be booted successfully. This particular FUJITSU server includes an integrated Remote Management Controller (iRMC), through which we were able to determine that the problem was due a memory module ("Memory module failed (disabled) (CRITICAL)"). Interestingly, the iRMC has been working for over 18 months and reporting that other components of the server are still functional. The same phenomenon appeared in a FUJITSU server worked only in air. Even in the case of onboard memory, i.e. AS-1341G, the memory module failure appeared both in-water and air cooling after it works for five months.

Our anecdotic evidence indicates that our proposed approach works better when memory slots are not covered by the parylene film; then we will obtain a longer lifetime than a couple of years. It is possible to implement it by masking them when coating a motherboard by the parylene film.

In our experiments, we have encountered problems with PCIex4, mPCIe, RJ45 connectors and with a memory slot for 2 years in total. They can be resolved by uncoating them (, and placing them above the surface of the water).

2.4 Temperature Evaluation

To show the advantage of full in-water immersion, we have measured chip temperature for the film-coated PRIMER-GY TX1320 M2 server (Xeon E3-1270v5(3.6 GHz), 4 GB DDR4 2133 MHz Unbuffered DIMM×2) for three options: (i) air cooling (the motherboard is placed right next to a high-speed fan); (ii) immersing only the heatsink in water; and (iii) full immersion in water using our approach. We ran the `stress` command, operating chips at their maximum frequencies, and obtained temperature information directly from the OS (Ubuntu Linux). We measured temperature for air cooling at 76 °C. Immersing only the heatsink in water leads to a 5 °C reduction down to 71 °C. Immersing the whole board in water lowers the temperature by 15 °C down to 56 °C. These results show the significant potential thermal benefits of full in-water immersion.

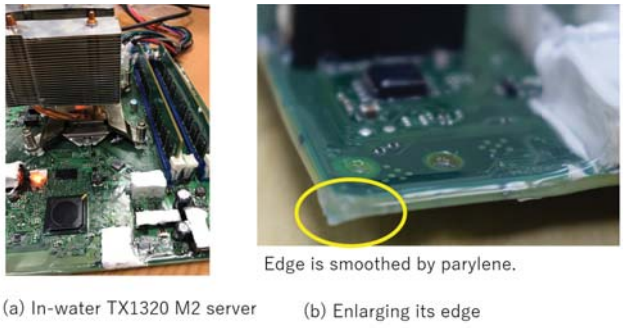


Figure 3: PRIMERGY TX1320 M2 server covered with 150μm parylene film (film is too thin and transparent to be seen on photo). We use ANNEX Thermal Grizzly Kryonaut as thermal interface material (TIM). The white bond is RTV KE-45 for padding the unused I/O (later, we find out that the bond is not needed for electric insulation. Only parylene film is enough for this purpose.).

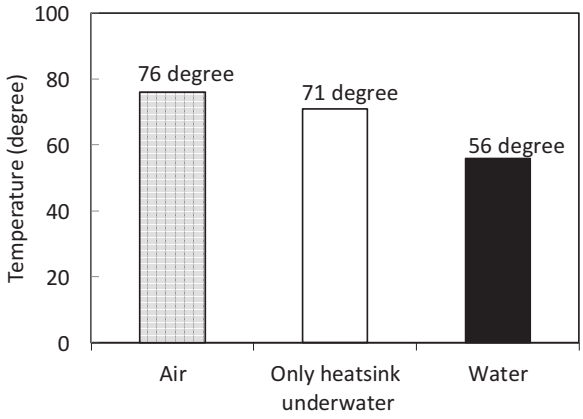


Figure 4: Chip temperature for the film-coated PRIMERGY TX1320 M2 server vs. cooling option.

3 SIMULATION EVALUATION

The previous section illustrates the impact of water cooling on the chip temperature using our prototypes and lifetime. To evaluate the benefits of in-water computing over other cooling alternatives for high-power chips, in this section, we quantify gains in operating frequencies and in parallel (OpenMP) application performance on temperature-constraint 3-D chip multi-processors (CMPs).

3.1 Baseline Chip Multi-Processor

We use standard simulation tools to evaluate our designs quantitatively: McPAT v1.3 [20] to compute power consumption; HotSpot v6.0 [24] to compute maximum temperature (based on the power trace generated by McPAT simulations); and gem5 [4] to estimate application execution time.

For simplicity and generality, we consider a baseline 16-tile chip that we obtain by modifying the Xeon processor description file provided by McPAT, the details of which are given in Table 1 and Figure 5. The bottom part of the table shows network-on-chip (NoC) specifications.

Table 1: Specification of the baseline 2-D CMP.

Processor family	x86-64
Number of cores	4
L1 I/D cache size	32/128 KiB (line:64B)
L1 cache latency	1 cycle
L2 cache bank size	12 MiB (assoc:8)
L2 cache latency	6 cycles
Memory size	4 GiB
Memory latency	160 cycles
Area	169 mm^2
Maximum Power (low-power)	47.2 Watts @ 2.0 GHz
Maximum Power (high-frequency)	56.8 Watts @ 3.6 GHz
Router pipeline	[RC][VSA][ST/LT]
Buffer size	5 flits per VC
Protocol	MOESI directory
# of VCs	3 (one VC for each message class)
On-chip topology	4 *times*4 mesh
Control / data packet size	1 flits / 5 flits

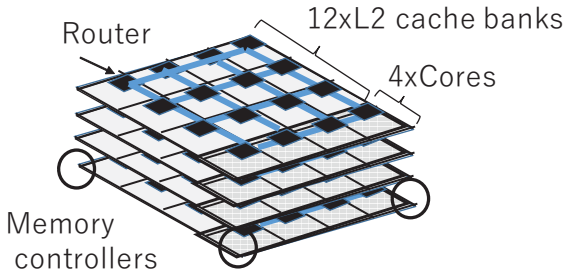


Figure 5: A 3-D CMP that consists of four chips, each of which has four cores and twelve L2 cache banks.

We obtain area and maximum (static and dynamic) power distribution of a chip using McPAT for 22nm technology with physical gate lengths configured for high-performance applications. Note that the McPAT simulation does not include the power consumption of vertical interconnects, such as through silicon via (TSV) and ThruChip interface (TCI). However, this power consumption is less than 0.3 W for a 256 Gbps (128 bit$\times$2.0 GHz) vertical link and so we neglect its impact (both in terms of power consumption and heat dissipation) [16].

We assume two designs of voltage-and-frequency scaling (VFS) on McPAT simulation for 11 steps from 1.0 GHz to 2.0 GHz, in 0.1 GHz increments, and 13 steps of VFS from 1.2 GHz to 3.6 GHz in 0.2 GHz increments, respectively. In the remain of this paper, the former is so named “low-power CMP” and the latter is “high-frequency CMP”.

Each pair of voltage and frequency is approximated as

$$T_{delay} \propto \frac{CV}{(V - V_{th})^{\alpha}} ,$$

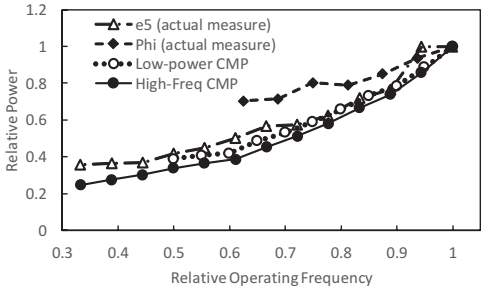


Figure 6: Power vs. Operating frequency, relative to maximum power-and-frequency behavior, for different CMPs.

where T_{delay} is the gate delay, C is the capacitance being switched, V is the supply voltage, V_{th} is the threshold voltage, and α is the velocity saturation index in a short channel MOSFET. The values for V and V_{th} are taken from the McPAT technology file, and we set α to 1.3. We have verified that the above model leads to frequency/power values that are consistent with actual measurements on the recent Xeon E5-2667-v4 processor and Xeon Phi7250 processors, as show in Figure 6.

High-end processors typically specify a recommended maximum operating temperature threshold above 70 °C (e.g., 78 °C for Xeon E5-2667-v4). In this study, we conservatively assume a temperature threshold of 80 °C.

3.2 Chip Frequencies

We consider 3-D CMPs. Given an 80°C temperature threshold, we determine the maximum chip operating frequency (assuming all chips operate at the same frequency) for 1 to 15 chips, based on HotSpot v6.0 simulation results. Although transient behaviors of computational applications are interesting and have been studied [1, 3, 10, 11], in this study, we only consider the worst-case study, i.e. the steady-state behavior in which each module fully works. Simulation parameters are listed in Table 2. We set heat transfer coefficients of air, mineral oil, fluorinert, and water to 14, 160, 180, 800 $W/(m^2K)$, respectively. Our results also include the water pipes cooling option, assuming that the heatsink is replaced by a typical closed-loop liquid CPU cooler.

Table 2: HotSpot v6.0 Simulation Parameters.

Heatsink	12×12×3 cm, 400 W/mK, 0.3024 m^2
Heat spreader	6×6×0.1 cm, 400 W/mK
Parylene film	120 μm, 0.14 W/mK
TIM / Glue	20 μm, 0.25 W/mK
Outside temp.	25°C

Figures 7 and 8 plot operating frequency vs. number of chips for each cooling option. In low-power CMPs, the air cooling and the water-pipe cooling can work at up to 4 and 7 chips, respectively, thus their plots for 5 and 8 chips cannot be drawn in the figure.

These results show that air or water-pipe cooling does not make it possible to operate more than 2 or 3 chips, respectively, and only at low frequency. Water cooling via immersion leads to the best results,

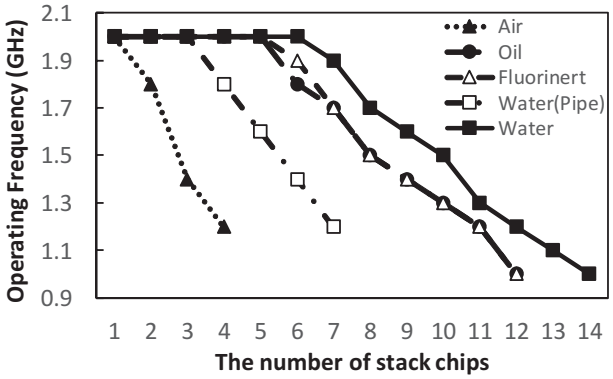


Figure 7: Maximum chip operating frequency vs. number of chips in a stack low-power CMP for different cooling options.

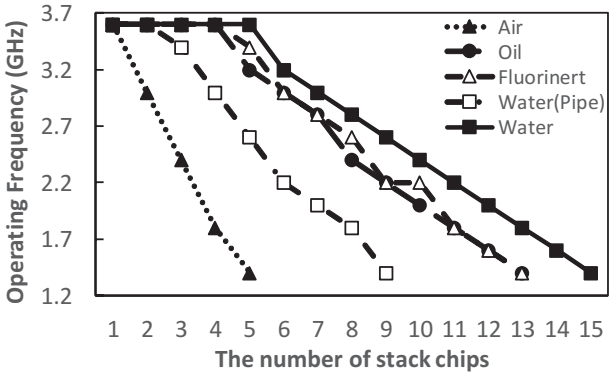
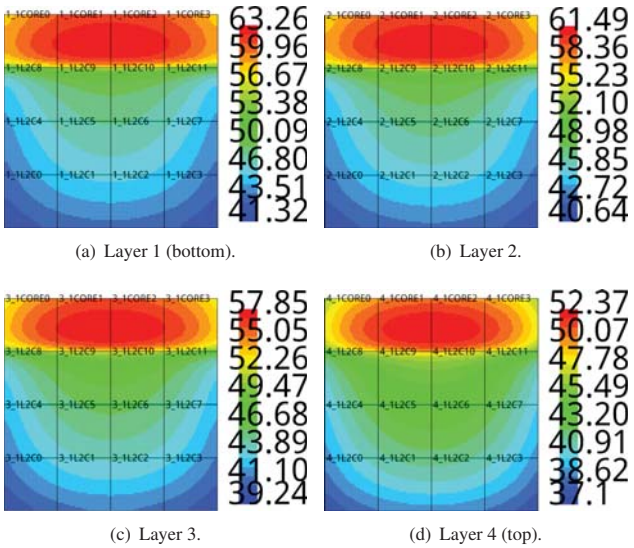


Figure 8: Maximum chip operating frequency vs. number of chips in a stack high-frequency CMP for different cooling options.

making it possible to run chips at higher frequencies than its closest competitors, fluorinert and mineral oil, when 6 or 5 chips or more are used in low-power or high-power CMPs, respectively. Although the high-frequency CMP has the higher power consumption than that in the low-power CMP, the high-frequency CMP can support the more significant number of chips than the low-power CMP if allowing lower operating frequency. This is because the high-frequency chip supports the broader range of VFS, thus supporting the lower power mode of the chip.

Figure 9 illustrates the thermal map of the high-frequency CMPs at 3.6GHz for water cooling. Since the processor cores (CORE1 - CORE4) have higher power density than L2 cache, the non-uniform thermal distribution is obtained on a chip. As expected, the upper tier has a lower temperature at the same position.

Through the HotSpot simulation results, we conclude that in-water CMPs should provide significant application performance advantages, which we further corroborate in the next section.



(a) Layer 1 (bottom). (b) Layer 2.

(c) Layer 3. (d) Layer 4 (top).

Figure 9: Thermal map of the 4-chip high-frequency CMP (3.6GHz) for water cooling. Notice that the color scales are not the same.

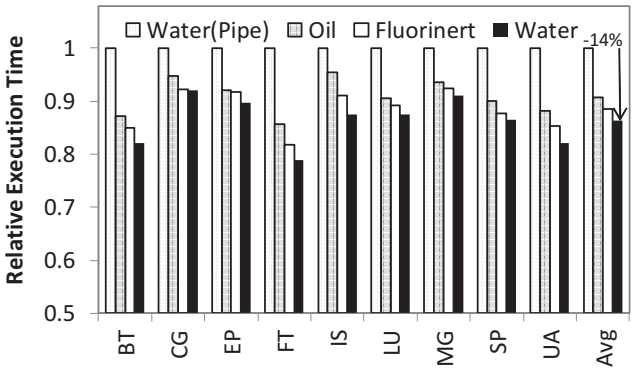


Figure 10: Benchmark execution times, relative to water pipes cooling for a 6-chip low-power CMP.

3.3 Parallel Application Performance

We perform a full-system simulation of 6 and 8-chip 3-D integrated CMPs using gem5 [4]. We simulate the execution of nine parallel programs from the OpenMP implementation of the NAS Parallel Benchmarks. These programs were compiled with GCC 4.4.7 and executed (in simulation) with a Linux kernel 2.6.22.9. We assume the executions to 24 or 32 threads in 6 or 8-chip CMPs, respectively.

We consider four cooling options: water pipe, mineral oil, fluorinert, and water. We omit air cooling as it cannot support 6 and 8 chips (see the previous section). Simulated application execution times for each benchmark are shown in Figures 10 to 13. The execution times are relative to execution times obtained with the water pipe for Figures 10, 12 and 13, and those with mineral oil cooling option for Figure 11 (lower values mean better performance). Since

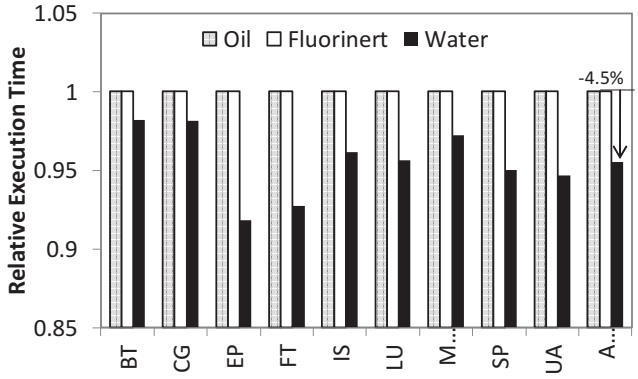


Figure 11: Benchmark execution times, relative to mineral oil cooling for an 8-chip low-power CMP.

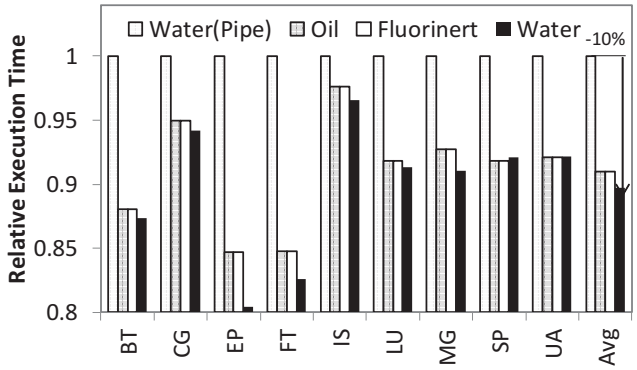


Figure 12: Benchmark execution times, relative to water pipes cooling for a 6-chip high-frequency CMP.

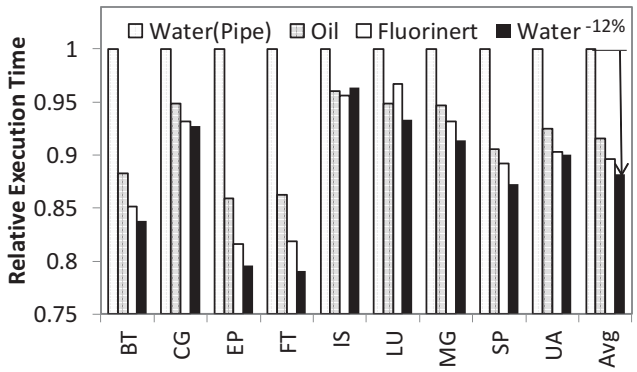


Figure 13: Benchmark execution times, relative to water pipes cooling for an 8-chip high-frequency CMP.

the water pipe cooling cannot support eight chip stack using the high-frequency CMPs, we can not plot ts result in Figure 11. In line with frequency results in the previous section, water cooling leads to the fastest operating frequency, thus obtaining fastest application execution times by up to 14% on average. In both cases for 6 and 8

chips, it is oblivious that the parallel-application performance tends to be proportional to the maximum operating frequency. We highly recommend using water cooling.

4 FURTHER CONSIDERATIONS

4.1 Heat Transfer Coefficient

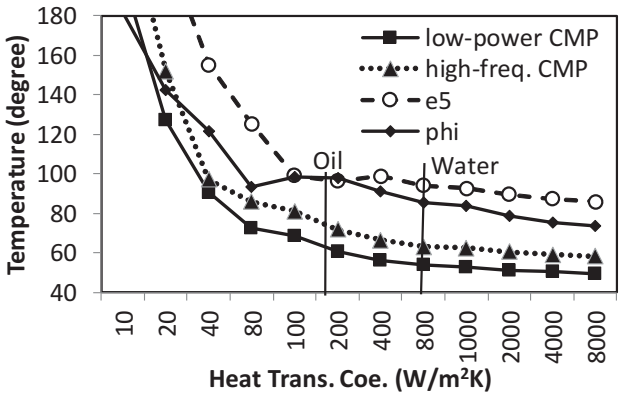


Figure 14: Temperature vs. heat transfer coefficient.

We use HotSpot to estimate CMP (maximum) temperature for a range of heat transfer coefficients for the coolant. Figure 14 shows temperature vs. heat transfer coefficient for the low-power and high-freq CMPs, Intel Xeon E5-2667-V4 ("e5"), and Phi7250 ("phi"), each operating at their maximum frequency. The number of stack chips is set to four. As expected, temperature decreases as the heat transfer coefficient increases. Interestingly, especially for a high-power chip like the Xeon e5, we find that non-negligible temperature reduction can be achieved for heat transfer coefficients higher than that of water. Therefore, even in the case of in-water computers, it could be worthwhile in practice to increase coolant flow speed (e.g., via turbines).

4.2 Influence of Thermal-aware Floorplan Optimization on Coolant Performance

The power, and thus heat, distribution of a chip is not uniform across its surface. For instance, a processor core typically has high power density, while the last level cache (LLC) has low power density, as shown in Figure 9.

There is room to improve thermal-aware chip layouts for the 3-D stack. In terms of thermal-aware chip design, 3-D floorplan algorithms have been proposed [7, 24], as well as algorithms that compute microchannel layouts. To simply understand its impact on the temperature of the conventional chips considered in Figure 5, it may thus be worthwhile rotating some chips, e.g., to make a low-power-density area of a chip overlap a high-power-density area of another chip [13].

In the high-frequency CMP, four processor cores are all located in the bottom tile row. To investigate the effectiveness of chip rotations, we consider the 3-D stack layout, for which we quantify the effect of rotating all chips in even layers by 180°. Although the 90° rotation seems better, chips that are rectangle cannot be stacked. Figure 15

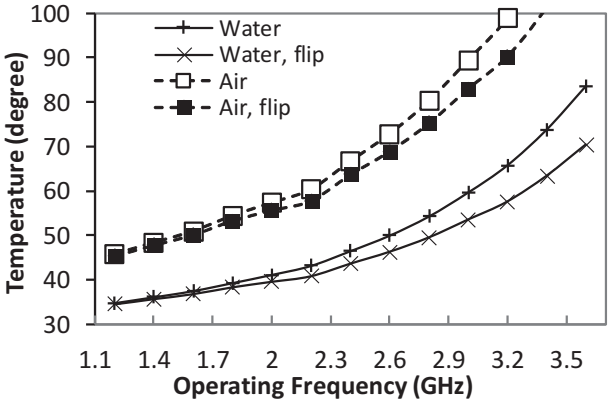
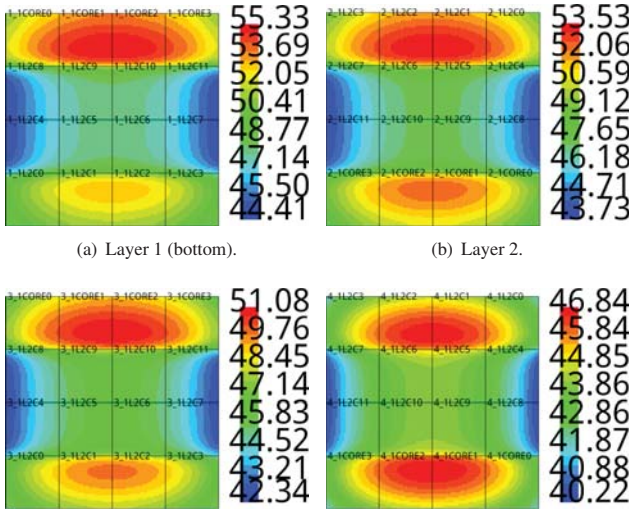


Figure 15: Operating frequency vs. temperature with and without chip rotation on the high-frequency CMP.



(a) Layer 1 (bottom). (b) Layer 2.

(c) Layer 3. (d) Layer 4 (top).

Figure 16: Thermal map of the 4-chip high-frequency CMP (flip, 3.6GHz) for water cooling. Notice that the color scales are not the same.

shows layout temperature, as determined via HotSpot simulations, vs. chip frequency, for the 4-chip stack with and without chip rotations (denoted by "flip"). Results are shown for air cooling and water cooling.

These results indicate that chip rotation has some effect on the temperature of air cooling and water cooling. For instance, considering a temperature threshold of 80°C, chip rotations make it possible to operate chips at 3.0 GHz instead of 2.8 GHz for air cooling. The water cooling does 3.6 GHz. The temperature reduction due to chip rotations when operating both chips at 3.6 GHz is 13°C.

Figure 16 illustrates the thermal map of the high-frequency CMP at 3.6GHz for water cooling with chip rotation, taken from the plot in Figure 15. When compared to Figure 9, we found that the

chip rotation distributes power through the entire chip surface more uniformly.

Our main finding is that the performance tendency in which water cooling outperforms the air cooling is maintained under both the 3-D stack layout in Figure 5 and the 3-D thermal-aware layout. Even using the flip, or optimizing chip layouts, we expect that there is a similar temperature gap among air, mineral oil and water cooling to those in the previous section.

4.3 Generalization

First, we focus on simulation accuracy. We ultimately evaluated and compared power analysis (McPAT), temperature analysis (HotSpot) and application performance (gem5). McPAT framework reported 22.61 % of power and 16.7% of area gap between its results and real Xeon Tulsa chips [20]. In the temperature evaluation, the accurate transient state analysis for a given application program is an exciting topic of modeling and computation acceleration [10, 11]. By contrast, in this study, we consider a worst-case design. Our results should be considered in early-stage design for parameter survey.

Second, we focus on the graph curves between frequency and the number of chips in different processor models. We measured their power consumption by RAPL when executing the stress command per core on Intel Xeon E5-2667-V4 and Phi7250. Using RAPL we limit the maximum operating frequency and measure the power. Using the obtained power profile and its physical layout of a high-resolution die photo, we obtained the maximum temperature in the CMP layout using HotSpot v6.0 [24]. Although their maximum power consumption depends on its application, the curves between operating frequency and the power consumption are similar. (We confirm this fact when executing NAS Parallel Benchmarks 3.3.1 OpenMP on Intel Xeon E5-2667-V4 and Phi7250.) We used the power trace of the stress command because it takes the average curves between operating frequency and the power consumption among the programs executed in this study. We compare the low-power and high-power CMP chips to Intel Xeon E5-2667-V4 in Figure 1 and Phi7250 in Figure 17. The water-pipe cooling and mineral oil cooling can work at up to two and three chips, respectively, thus their plots for three and four chips cannot be drawn in Figure 17. Figure 18 is its thermal map at 1.2 GHz. It interestingly illustrates that the thermal distribution is better than that in low-power and high-freq CMPs in Figures 9 and 16. This comes from that the number of processor cores is large and it is distributed through the entire chip, thus achieving more uniform thermal distribution. Since the both Intel Xeon E5 and Phi chips have higher power than the low-power and high-frequency chips, both Intel chips are accepted with the maximum operating frequency (i.e. 3.6 GHz for Intel Xeon E5 and 1.6 GHz for Intel Phi). We can draw the conclusion that the water immersion provides the same or higher operating frequencies for a given number of chips.

4.4 Application to Truly Direct Cooling Framework under Natural Water

What if a computer system works in natural water, e.g. in a river or in the ocean? then (1) heat would directly dissipate into the water and (2) a power usage effectiveness (PUE) of approximately 1.00 would

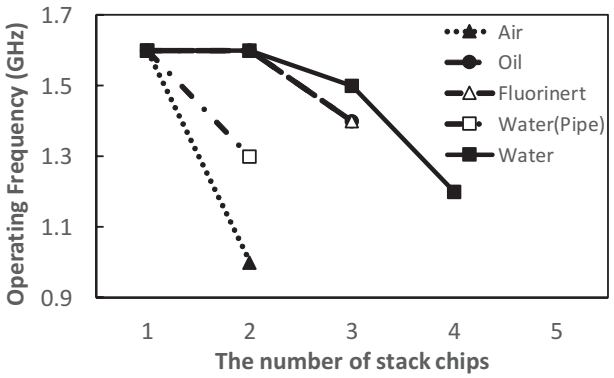
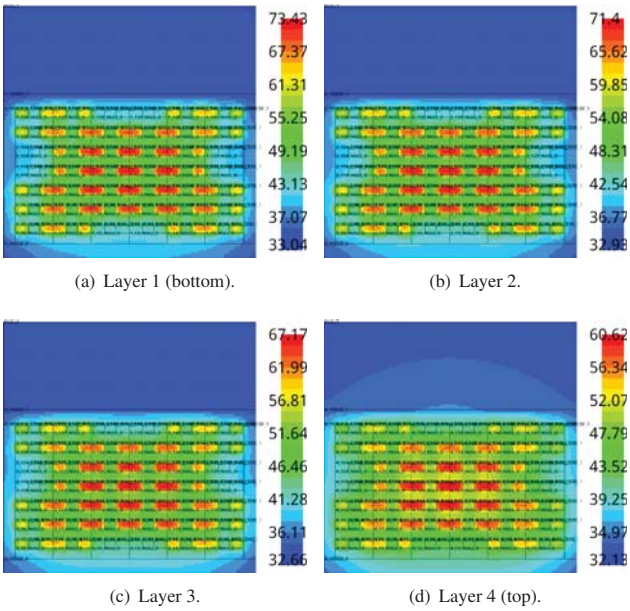


Figure 17: Maximum frequency vs. number of stacked Xeon Phi 7290 chips for air, mineral oil and water cooling.



(a) Layer 1 (bottom). (b) Layer 2.

(c) Layer 3. (d) Layer 4 (top).

Figure 18: Thermal map of the 4-chip Xeon Phi7290-model CMP (1.2 GHz) for water cooling. Notice that the color scales are not the same.

be achieved. We qualitatively consider the possible application of in-water cooling under natural water, and its efficiency.

4.4.1 No Secondary Coolants Enabled by Water-Immersion Cooling. The chip temperature can be computed based on an equivalent circuit of thermal resistances and capacitances. In this context, when the in-water cooling is placed under natural water, e.g. river, we can take the advantage of the removal of the secondary coolant (for cooling the primary coolants, i.e. air or oil, that contact chips) from the thermal circuit. Conventional air-cooled and oil-immersion computer system has the secondary coolant that has a certain amount of thermal resistances and capacitances. The secondary coolant is, of course, a lowest temperature, and the primary coolant is higher

temperature than that of the secondary coolant. Their cooling efficiency is not better than that of the in-water cooling under natural water that becomes the "primary" coolant.

4.4.2 Reducing PUE . The entire cooling facilities of datacenters should be carefully considered. Existing cooling systems for parallel computers have typically used both the primary and the secondary coolants. The secondary coolants should be cost-effective, and they can be not only air but also natural water, e.g. cold lake at upland or sea [22]. The Swiss National Supercomputing Centre (CSCS) has used natural water for cooling on a lake, and it has pumped over a distance of 2.8 km [29]. Also, AQUASER [26] and ABCI (AI Bridging Cloud Infrastructure) have used water-pipe cooling that faces processor chips, and they allow to use warm water, e.g. 60°C, that can reduce the cost of the secondary-cooling facilities, e.g. capacity of chillers and size of pumps. ABCI achieves the cooling capacity of 70kW per rack by using hot water primary cooling and air secondary cooling [23]. They usually rely on water pipe with pumps to contact hot chips, and the water in the pipe should be clean enough not to be stuck. By contrast, in-water computers can remove the above facilities of the secondary coolant and water pipes when they are directly deployed under natural water with approximately ideal PUE.

4.4.3 Experiment. We experiment upon our two in-water PCs with the μm parylene film, ASRock Q1900M on Tokyo Bay, and one with the parylene film worked for 53 days (that record is shorter than the case under-tapped water) under Tokyo Bay, as illustrated in Figure 19. In this experiment, the film coating is the same as the case under-tapped water. Although there were a large number of shellfishes on the yellow box that stores PCs. There is some seaweed growing in the box.

Existing cooling system does not use sea as the primary coolant, (The experiment [22] used sea as the secondary coolant), while our in-water approach can do. Although further development undersea is needed for stabilizing the long lifetime undersea, our proof-of-concept experiment may enable to widen the design option of the entire cooling facilities and their location under natural water.

Sea is an extreme example as the primary coolant. Our in-water cooling has long-term potential for designing the facilities that support natural water as the primary coolant. (This exploration is out of the scope of this paper.)

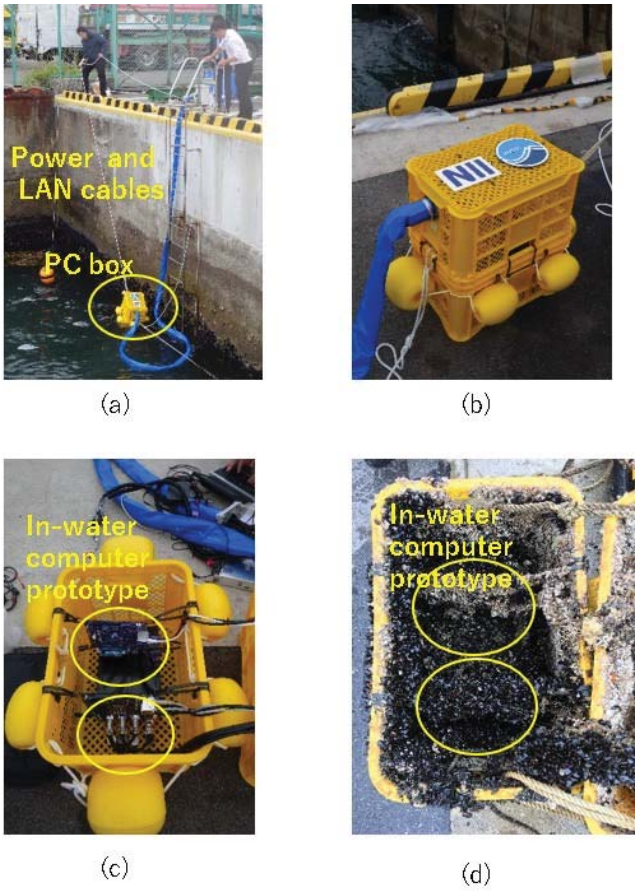


Figure 19: Proof of concept prototype under Tokyo Bay (a) the entire system, (b) the two motherboards are stored within a yellow box, (c) the status in the 1st day (d) the status after 55 days.

5 RELATED WORKS

5.1 Emerging Cooling Technologies

In this study, we propose using an unconventional cooling medium, namely, tap or natural water. Other unconventional cooling liquids, such as liquid nitrogen or dry ice, have been used to enable high chip frequencies (as seen in Master Overclocking Arena competitions). Also, phase-change cooling, in which a gas (or mix of gases) is compressed into a liquid, is more efficient than using only gases or only liquids. Electronics completely immersed in a dielectric (non-conductive) fluid, such as Novec, chip temperatures reach fluid boiling temperature [2].

As described in the previous section, there are a large number of the secondary indirect cooling techniques that make coolant that faces chips cool via the radiator, such as the use of sea in Microsoft Project Natick or the use of air outside in cold climates. The secondary cooling techniques are indirect and they are quite different from our in-water cooling; the chip temperature depends on the direct coolant that faces chips.

Cooling has also been considered at the chip design level. The application of microchannel (water) cooling to 2-D [18, 31] and 3-D ICs [6, 19, 25] has been shown to be efficient. This is because a large number of microchannels can be laid out around areas of high heat density, e.g., processor cores. Due to chip thickness considerations, it is unclear whether microchannel cooling can be applied to CMPs that rely on inductive coupling.

5.2 Thermal-aware Techniques

Many techniques have been proposed to manage heat dissipation at chip design time and at runtime. Some of them are inspired by the power-aware computing design. Dynamic Thermal Management (DTM) at runtime has also been investigated [5, 27]. It avoids exceeding thermal constraints by lowering power. For example, power gating for reducing static power and DVFS for reducing dynamic power can be used to implement DTM strategies at the hardware

level. DTM can also be achieved in software via job scheduling [8]. Although we assume the worst-case design for thermal analysis in this study, the thermal evaluation of the DTM requires the transient temperature distribution. The evaluation models that support the transient temperature distribution are helpful [15, 28].

In the system level, some recent supercomputers have employed the power gap *a.k.a. power knob*. If all the computer components work at the maximum performance, the aggregate power consumption becomes higher than that allowed by the supercomputer. However, each parallel program has typically computation, memory or rarely network intensive. Some hardware resources are not fully utilized in such a program, thus, its power consumption can be smaller than that allowed by the supercomputer thanks to DTM. Our work is orthogonal to DTM techniques but widens the overall design space for thermal-aware systems.

6 CONCLUSIONS

In this study, we have proposed the in-water cooling since water has lower cost, lower safety concerns, and higher heat transfer coefficient than currently used coolants. We have successfully developed proof-of-concept prototypes of in-water computers by entirely covering test boards as well as off-the-shelf motherboards with a 120-150μm parylene film. They successfully decrease by 20 °C chip temperature when compared to air cooling, when executing CPU-intensive workloads on Xeon E3-1270v5 processors. Another finding is that the lifetime of the coated PC would be a couple of years when some parts of the motherboard, e.g. around memory slots, are not coated. In the cooling point of view, processors should be underwater while the other parts can be above the surface of water.

For 3-D CMPs, we have shown that water cooling makes it possible to operate chips at higher clock rates than when using other coolants. These higher frequencies can translate to improvements up to 14% and 4.5% in the execution time of the NAS Parallel Benchmarks when compared to water-pipe cooling and mineral oil cooling, respectively. Overall, our results demonstrate the clear potential of our proposed approach. Our future work is (1) a more thorough exploration of the 3-D chip integration layout design that makes the best use of the water cooling capability, and (2) evaluation for the ability to densely pack compute nodes.

ACKNOWLEDGMENTS

The authors thank Prof. Henri Casanova, University of Hawai‘i at Manoa, for his valuable discussion about in-water computer prototypes. The authors give a special thanks to Dr. Shinichi Ishida, Cyber-Koubou LLC, Japan, for his fruitful implementation on the in-water test boards.

REFERENCES

[1] A Sridhar, A Vincenzi, M Ruggiero, T Brunschwiler, D Atienza. 2010. 3D-ICE: Fast compact transient thermal modeling for 3D-ICs with inter-tier liquid cooling. In *Proceedings of the International Conference on Computer-Aided Design*. 463–470.

[2] Alex Kampl. 2014. Bitcoin 2-Phase Immersion Cooling and the Implications for High Performance Computing. *Electronics Cooling Magazine* (March 2014), 24–29.

[3] ANSYS CFX. 2019. https://www.ansys.com/products/fluids/ansys-cfx.

[4] N. Binkert, B. Beckmann, G. Black, S. K. Reinhardt, A. Saidi, A. Basu, J. Hestness, D. R. Hower, T. Krishna, S. Sardashti, R. Sen, K. Sewell, M. Shoaib, N. Vaish, M. D. Hill, and D. A. Wood. 2011. The gem5 Simulator. *ACM SIGARCH Computer Architecture News* 39, 2 (May 2011), 1–7.

[5] D. Brooks and M. Martonosi. 2001. Dynamic Thermal Management for High-Performance Microprocessors. In *International Symposium on High-Performance Computer Architecture (HPCA)*. 171–182.

[6] G. Chen, J. Kuang, Z. Zeng, H. Zhang, E. F. Y. Young, and B. Yu. 2017. Minimizing Thermal Gradient and Pumping Power in 3D IC Liquid Cooling Network Design. In *Design Automation Conference (DAC)*. 70:1–70:6.

[7] J. Cong, J. Wei, and Y. Zhang. 2004. A Thermal-driven Floorplanning Algorithm for 3D ICs. In *IEEE/ACM International Conference on Computer-aided Design (ICCAD)*. 306–313.

[8] A. K. Coskun, T. S. Rosing, and K. Whisnant. 2007. Temperature aware task scheduling in MPSoCs. In *Design, Automation and Test in Europe Conference and Exposition (DATE)*. 1659–1664.

[9] T. Endo, A. Nukada, and S. Matsuoka. 2014. TSUBAME-KFC: A modern liquid submersion cooling prototype towards exascale becoming the greenest supercomputer in the world. In *20th IEEE International Conference on Parallel and Distributed Systems (ICPADS)*. 360–367.

[10] FloTHERM, Electronics thermal analysis software, Mentor Graphics. 2018. https://www.mentor.com/products/mechanical/flotherm/flotherm/.

[11] A. Fourmigue, G. Beltrame, and G. Nicolescu. 2014. Efficient transient thermal simulation of 3D ICs with liquid-cooling and through silicon vias. In *Design, Automation & Test in Europe Conference & Exhibition(DATE)*. 1–6.

[12] Green Revolution Cooling, white paper. 2015. Oil Submersion Cooling for Today's Data Centers. http://www.grcooling.com/wp-content/uploads/2015/06/GRC_WP-CLICK-Oil_Sub_DCc.pdf.

[13] T. Hanada, H. Sasaki, K. Inoue, and K. Murakami. 2011. Performance evaluation of 3D stacked multi-core processors with temperature consideration. In *IEEE International 3D Systems Integration Conference (3DIC)*. 1–5.

[14] International Roadmap for Devices and Systems (IRDS). 2019. https://irds.ieee.org/.

[15] D.-C. Juan, S. Garg, and D. Marculescu. 2011. Statistical thermal evaluation and mitigation techniques for 3D Chip-Multiprocessors in the presence of process variations. In *Design, Automation and Test in Europe, DATE*. 383–388.

[16] T. Kagami, H. Matsutani, M. Koibuchi, Y. Take, T. Kuroda, and H. Amano. 2016. Efficient 3-D Bus Architectures for Inductive-Coupling ThruChip Interfaces. *IEEE Trans. VLSI Syst.* 24, 2 (2016), 493–506.

[17] KISCO Ltd., diX Coating. 2014. http://www.kisco-net.com/business/dix-coating/.

[18] R.W. Knight, D.J. Hall, J.S. Goodling, and R.C. Jaeger. 1992. Heat sink optimization with application to microchannel. *IEEE Transactions on Components, Hybrids, and Manufacturing Technology* 15, 5 (1992), 832–842.

[19] J.M. Koo, S. Im, L. Jiang, and K.E. Goodson. 2005. Integrated Microchannel Cooling for Three-Dimensional Circuit Architectures. *ASME Journal of Heat Transfer* 127 (2005), 49–58.

[20] S. Li, J. H. Ahn, R. D. Strong, J. B. Brockman, D.M. Tullsen, and N. P. Jouppi. 2009. McPAT: An Integrated Power, Area, and Timing Modeling Framework for M ulticore and Manycore Architectures. In *the IEEE/ACM International Symposium on Microarchitecture*. 469–480.

[21] M. Matsuda and K.Mashiko and Y. Saito and T. Nguyen and T. Nguyen. 2015. Micro-channel Cold Plate Units for Cooling Super Computer. *Fujikura Technical Review* 44 (March 2015), 53–57.

[22] Microsoft, Project Natick. 2018. http://natick.research.microsoft.com/.

[23] National Institute of Advanced Industrial Science and Technology (AIST), AI Bridging Cloud Infrastructure (ABCI). 2019. https://abci.ai/.

[24] R. Zhang and M. R. Stan and K. Skadron. 2015. HotSpot 6.0: Validation,Acceleration and Extension. University o f Virginia, Tech. Report CS-2015-04.

[25] S. K. Roy, S. Mandal, C. Giri, and H. Rahaman. 2015. A thermal estimation model for 3D IC using liquid cooled microchannels and thermal TSVs. In *IFIP/IEEE International Conference on Very Large Scale Integration, VLSI-SoC*. 122–127.

[26] S. Zimmermann and I.Meijer and M. K.Tiwari and S.Paredes and B. Michel and D. Poulikakos. 2012. Aquasar: A hot water cooled data center with direct energy reuse. *ENERGY* 43 (July 2012), 237–245. Issue 1.

[27] K. Skadron, M. R. Stan, W. Huang, S. Velusamy, K. Sankaranarayanan, and D. Tarjan. 2003. Temperature-Aware Microarchitecture. In *International Symposium on Computer Architecture (ISCA)*. 2–13.

[28] A. Sridhar, M. M. Sabry, and D. Atienza. 2014. A Semi-Analytical Thermal Modeling Framework for Liquid-Cooled ICs. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 33, 8 (Aug 2014), 1145–1158.

[29] Swiss National Supercomputing Centre. 2019. https://www.cscs.ch/publications/fact-sheets/lake-water-to-cool-supercomputers/.

[30] T. Totoki, HotSpot-6.0 Extension. 2019. https://github.com/KoibuchiLab/HotSpot-6.0-3D-API.

[31] D.B. Tuckerman and R.F.W. Pease. 1981. High-performance heat sinking for VLSI. *IEEE Electron Device Letters* EDL-2, 5 (1981), 126–129.