LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-CONF-474345

Fighting Fire with Fire: Modeling the Datacenter-Scale Effects of Targeted Superlattice Thermal Management

S. Biswas, M. Tiwari, L. Theogarajan, T. P. Sherwood, F. T. Chong

March 18, 2011

International Symposium on Computer Architecture
San Jose, CA, United States
June 4, 2011 through June 8, 2011

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

Fighting Fire with Fire: Modeling the Datacenter-Scale Effects of Targeted Superlattice Thermal Management

Susmit Biswas*, Mohit Tiwari†, Timothy Sherwood†
Luke Theogarajan‡, Frederic T. Chong†

Lawrence Livermore National Laboratory, Livermore, CA - 94550, USA*
Department of Computer Science, UC Santa Barbara, USA†
Department of Electrical and Computer Engineering, UC Santa Barbara, USA‡

biswas3@llnl.gov*, {tiwari, sherwood, chong}@cs.ucsb.edu†, ltheogar@ece.ucsb.edu‡

ABSTRACT

Local thermal hot-spots in microprocessors lead to worst-case provisioning of global cooling resources, especially in large-scale systems. However, efficiency of cooling solutions degrade non-linearly with supply temperature, resulting in high power consumption and cost in cooling – 50∼100% of IT power. Recent advances in active cooling techniques have shown on-chip thermoelectric coolers (TECs) to be very efficient at selectively eliminating small hot-spots, where applying current to a superlattice film deposited between silicon and the heat spreader results in a Peltier effect that spreads the heat and lowers the temperature of the hot-spot significantly to improve *chip reliability*. In this paper, we propose that hot-spot mitigation using thermoelectric coolers can be used as a *power management mechanism* to allow global coolers to be provisioned for a better worst case temperature leading to substantial savings in cooling power.

In order to quantify the potential power savings from using TECs in data center servers, we present a detailed power model that integrates on-chip dynamic and leakage power sources, heat diffusion through the entire chip, TEC and global cooler efficiencies, and all their mutual interactions. Our multi-scale analysis shows that, for a typical data center, TECs allow global coolers to operate at higher temperatures without degrading chip lifetime, and thus save ∼27% cooling power on average while providing the same processor reliability as a data center running at $288K$.

Categories and Subject Descriptors

B.8.1 [**Performance and Reliability**]: Reliability, Testing, and Fault-Tolerance; I.6.5 [**SIMULATION AND MODELING**]: Model Development

General Terms

Design, Management, Reliability, Measurement

Keywords

Data center, cooling power, active cooling, TEC

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. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
ISCA'11, June 4–8, 2011, San Jose, California, USA.
Copyright 2011 ACM 978-1-4503-0472-6/11/06 ...$10.00.

1. Introduction

The running costs of data centers are dominated by the need to dissipate heat generated by thousands of server machines. Higher temperatures are undesirable as they lead to premature silicon wear-out; in fact, mean time to failure has been shown to decrease exponentially with temperature (Black's law [7]). Although other server resources also generate heat, microprocessors still dominate in most server configurations [4] and are also the most vulnerable to wear-out as the feature sizes shrink. Even as processor complexity and technology scaling has increased the *average* energy density inside a processor to maximally tolerable levels, modern microprocessors make extensive use of hardware structures such as the load-store queue and other CAM-based units, and the *peak* temperatures on chip can be much worse than even the average temperature of the chip. In recent studies, it has been shown that *hot-spots* inside a processor can generate $\sim 800W/cm^2$ heat flux whereas the average heat flux is only $10-50W/cm^2$ [36], and due to this disparity in heat generation, the temperature in hot spots may be up to 30°C more than average chip temperature.

The key problem processor hot-spots create is that in order to prevent some critical hardware structures from wearing out faster, the air conditioners in a data center have to be provisioned for worst case requirements. Worse yet, air conditioner efficiencies decrease non-linearly as the desired ambient temperature decreases relative to the air outside. As a result, the global cooling costs in data centers are directly correlated with the maximum hot-spot temperatures of processors, and there is a distinct requirement for a cooling technique to mitigate hot-spots selectively so that the global coolers can operate at higher temperatures while providing the same chip reliability.

We observe that localized cooling via superlattice microrefrigeration presents exactly this opportunity whereby hot-spots can be cooled selectively, allowing global coolers to operate at a higher temperature with higher efficiency. Recent advances in processor cooling technologies have demonstrated that thermoelectric coolers, which use the Peltier effect to form heat pumps, can be used to reduce the temperature of hot spots, and thereby increase the reliability of processors. By placing a thermoelectric cooler layer between the heat spreader and the processor die, and applying current selectively to the coolers over the hot spots, heat from the hot-spots can be spread much more effectively. The ability to implement such thermoelectric coolers on a real silicon device has been demonstrated recently [11], albeit for small prototype chips. In this paper, we propose that superlattice coolers can be used for active cooling power management by mitigating hot-spots and

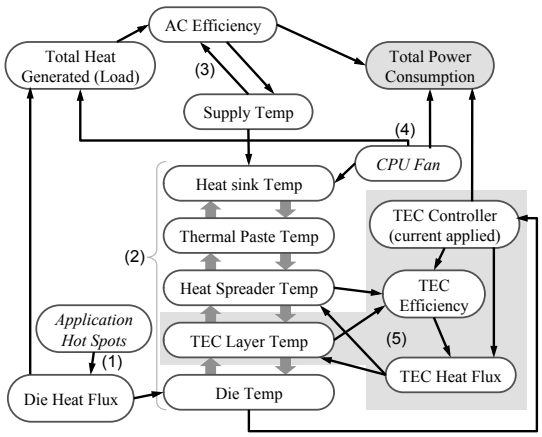


Figure 1: Heat Ecosystem

running the data centers at a higher temperature while attaining the same level of chip reliability. Before such thermoelectric coolers can be integrated in commodity server processors, we must ask the key question: "What is the potential for superlattice microrefrigeration to reduce global cooling costs in data centers?". In order to answer this question, our research makes the following specific contributions:

- We propose to use superlattice thermo-electric coolers (TEC) as an active *cooling power management* device instead of as a reliability enhancing device. By mitigating hot-spots in servers, TEC devices enable global coolers to maintain power-efficient ambient temperatures, while the TEC keeps hot-spots closer to temperatures usually attained through more aggressive air conditioning.
- We present a comprehensive analysis of the impact of thermoelectric coolers on global cooling costs. In Figure 1, we show an overview of the modeled system. Our analysis covers all aspects of cooling a server in a data center, and integrates on-chip dynamic and leakage power sources with a detailed heat diffusion model of a processor (that models the silicon to the thermoelectric cooler to the heat spreader and the heat sink) and finally the data-center cooling efficiency.
- We find that TEC devices are very effective in spreading the heat away from hot-spots in a processor, but careful design choices are required to use them as cooling power management mechanism. Evaluating over 43 application phases from SPEC CPU 2000 benchmark suite, we report that by adding an energy efficient TEC layer in chip package and increasing the supply air temperature in a data center from a typical temperature of $294K$ (70°F), on average 12% cooling power could be saved. In conservative data centers with $288K$ (60°F) supply air temperature, (*e.g.* HPC data centers [1]), we observe ∼27% cooling power savings by running the data center at $297.5K$(75.2°F) temperature without affecting life span of processors. We find that selectively activating TECs can be effective for low-efficiency TEC devices. When using energy-efficient TECs [11], however, more cooling power can be saved by increasing data center temperature even higher while switching on all or majority of the TEC blocks.

The remainder of the paper is organized as follows. We motivate this work in Section 2, present an overview of the models in Section 3 and further details in Section 4. We describe the experimental methodology in Section 5, present the results in Section 6, and finally conclude in Section 8.

2. Motivation

Data centers house thousands of servers for running Internet and high-performance applications. When an application executes in a processor, often the components in the processor do not get used uniformly, leading to hot-spots in it. The heat fluxes from different segments of a processor differ significantly, often by an order of magnitude as reported by Chrysler [12]. For example, an average heat flux inside a processor is $10-50W/cm^2$ [36], where as the heat flux in a specific $400\mu m$ x $400\mu m$ area can be as high as $800W/cm^2$. Due to this difference in heat flux density, the temperature inside a chip can vary 5°C ∼ 30°C.

In a data center, the cooling subsystem is designed considering the hottest point in the processors. However, a cooling unit operates at low efficiency if the target supply air temperature is low, and requires large amount of power to cool a data center. By increasing the temperature of the supply air, the cooling power could be reduced significantly. In a data center, this results in saving millions of dollars, but running the data center at a higher temperature can induce higher component failure, which increases exponentially with the operating temperature following Black's equation [7]. An increase in 10°C in a component can reduce its lifetime to $1/2$ or even less. Therefore, cooling units in many state of the art data centers (primarily HPC data centers) supply air at 15°C (∼ 59°F) or lower temperature to the servers [1, 14, 20], though newer commercial data centers supply relatively warm air (∼ 70°F) to reduce cooling power at the cost of reduced reliability.

Recent advances in active cooling technology have demonstrated the feasibility of using Peltier-effect-based Thermoelectric Cooling (TEC) [11, 3, 36] devices for eliminating the hot-spots from a processor. A typical TEC device is $\sim 70\mu m$ in thickness [11] and resides between heat spreader and the silicon die. Metal contacts are deposited on n and p substrates, which are soldered to electroplated Cu contacts. The other ends of the junctions are connected to power supply to form a transistor-level Peltier micro-cooling device, as shown in Figure 2. TEC blocks are most efficient when their sizes are small ($< 100\mu m$ a side)[36]. In order to cover a larger hot-spot, a collection of tiles of these coolers can be built as a superlattice on top of the die as shown in Figure 2. TEC devices have been demonstrated to have $\sim 5\mu s$ response time [3], which makes them suitable for rapid cooling of hotspots. While pumping the heat out of the die, the flow of current through the superlattice produces a heating effect (the Joule effect) that adds to the heat flux, but overall the hot spot temperature and hence the temperature profile of the die becomes nearly uniform. These coolers are able to sustain heat flux up to $1250W/cm^2$ [11]. By absorbing heat from the areas close to hot-spots, TECs are able to reduce the temperature of the hot-spots on die, improving the mean time to failure (MTTF) of the die.

In this paper, we propose that TEC devices can be used in data centers as a cooling power management mechanism instead of a way to enhance reliability. By cooling hot-spots selectively, a data center can be run at a higher ambient temperature and still attain similar hot-spot temperature achievable by aggressive air-conditioning, leading to a reduction in cooling power without affecting chip life span. We build and integrate thermal models of different components of a whole

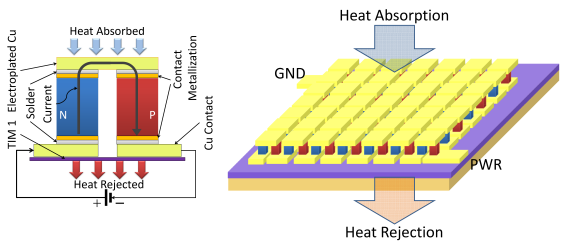


Figure 2: Individual TEC cell to superlattice

data center – from the silicon in a chip, embedded TECs, to the air conditioners – and quantify the expected savings in cooling power in the whole data center.

3. Model of Datacenter Cooling: From Die to Air Conditioner

In a typical data-center the cooling units in individual floors or rooms are provided with cold air or water to extract heat from them. In a chilled water cooling system, a chiller plant generates cold water centrally using cooling towers, and pipes water to air handling units (computer room air conditioning or CRAC units) that supply cold air to servers.

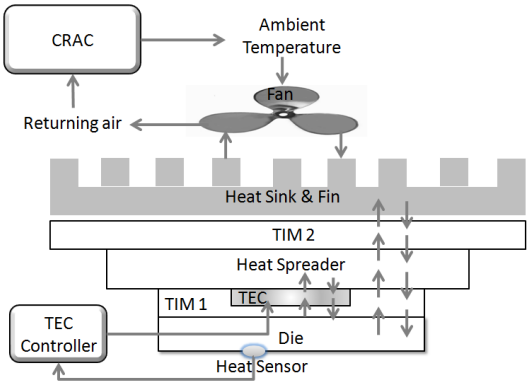


Figure 3: System Interaction

Our goal is to quantify the impact of using micro-level thermoelectric coolers on the power for cooling an entire data center, which is typically $50-100\%$ of the IT power [26]. Figure 3 shows details of all the components in the modeled system and how they interact among themselves. At the bottom most layer, the die is housed in a ceramic case. A thermal interface material (TIM) layer is applied on top of it, which is used to connect the heat spreader to die. Similarly another layer of TIM is applied between the heat spreader and the heat sink. We model a slab-fin heat sink in this paper, but our model could be easily used for a pin-fin heat sink as well. To incorporate active cooling, the TEC tiles are housed on the bottom side of heat spreader. In Figure 5, we show the placement of individual TEC tiles, which consume $9mm^2$ area ($< 10\%$ of chip area). A controller could use sensor readings from the hot spots to trigger TECs. In our experiments we assume that thermal sensors are placed in the middle of TEC blocks to activate them. In older Intel®processors, a thermal sensor is placed near hotspots to measure die temperature, which could be used to trigger TECs when needed. In Intel®Core^{TM}2 Duo and newer processors a digital thermometer is fabricated that measures temperatures from multiple points distributed across

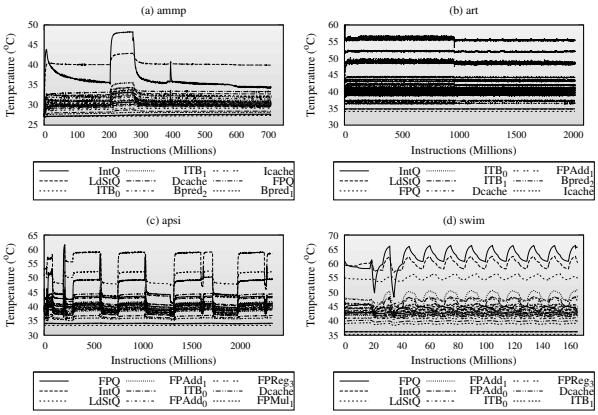


Figure 6: Temperature distribution of sample SPEC CPU2000 benchmarks using Minnespec input sets. The temperature trace demonstrates that there are phases in applications. Different unit might be the hottest across applications or phases in them. For example, integer-queue unit is the hottest in ammp and art, whereas floating-point queue is the hottest unit in apsi and swim. In these figures we only show the nine hottest units sorted in descending temperatures. Other units have significantly lower temperatures.

the die [27]. Inputs from the sensors could be used to activate individual TEC devices more selectively. In the chip package, the heat generated from computations on the chip, spreads up to the heat sink, and is ultimately absorbed by the air outside (that then needs to be cooled using the CRAC). CRAC supplies cold air to the server room, which is driven by a fan to the heat sink. The supply air temperature of the air affects the efficiency of the CRAC. Therefore, we create a model that integrates detailed chip-level heat diffusion with power efficiencies of server and data center-level components. Adding a layer of TECs requires an investment of energy in order to improve the heat transfer from the die to the heat spreader, but decreases the peak temperature on the die. Our task then is to determine how much heat has to be absorbed by the CRAC before and after introducing thermo-electric cooling within the chip stack.

We approach the problem of modeling this multi-scale system in a bottom-up fashion. Servers generate heat from dynamic switching as well as static leakage. The consumed power, both from the dynamic and static components, dissipates as heat to the chip, increasing its temperature. Due to the positive feedback loop between the temperature and the leakage power of the chip, we need numerical simulations for estimating the thermal profile of the chip. We model leakage power as a function of temperature by using ITRS [2] data and HSPICE simulation on logic and SRAM cells. To estimate dynamic power, we obtain power traces for each architectural component using the Wattch tool [8] on the SPEC CPU2000 benchmark suite. Figure 6 illustrates that applications show phase-behavior during its executions. As numeric simulation restricts us from analyzing entire execution, we choose representative points from the execution according to the phases. We use the power consumption of individual units at those execution points to create sources of appropriate power densities at each architectural component. The sources are used to simulate heat diffusion beginning from the die and all the way to the heat sink

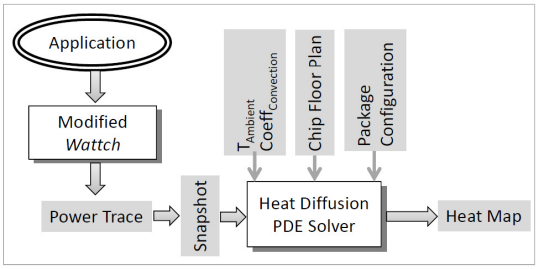


Figure 4: Toolchain to generate thermal map of a chip package: dynamic power sources on-chip are determined using Wattch [8], leakage power is modeled based on ITRS data. A thermal map , generated by solving heat diffusion equations, is used to compute reliability and power.

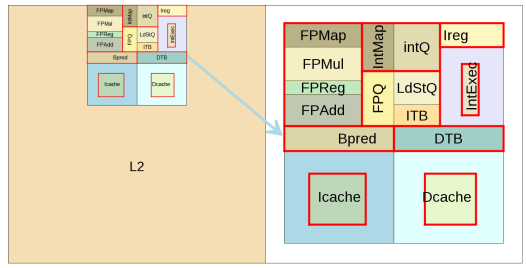


Figure 5: Locations and sizes of TEC tiles are shown by thick red lines in the Figure. The heuristics used in this process is based on the usage of units and their temperature profiles obtained using HotSpot tool.

in order to determine the temperature-map and leakage power consumption of the entire chip as shown in Figure 4. This total power generated by the chip, along with power generated by other server components, has to be removed by the CRAC, whose efficiency depends on the temperature of supplied cold air in the data center. The net result is that a data center can be modeled as a closed system where power densities on the die and power provided by the TEC form the heat sources, the power provided by the CRAC forms the heat sink, and a heat diffusion model through the chip yields the hot-spot temperature. Our goal here is to find out whether investing some power in the TEC to reduce the peak temperature on the die is worth increasing the leakage power of the chip (as the CRAC has to absorb it). We obtain the heat-map for the active layer of the processor for each application for two cases – (1) without the TEC layer in chip package, running the data center at a low ambient temperature and (2) with the TEC layer packaged inside a processor running in a warmer data center. Then, we compute the benefit of using TECs to reduce global cooling power and the effect on chip-reliability.

4. Detailed Models

We present details on modeling the sources of heat (*i.e.* computation on the chip and other sources on a server), heat flows through the processor chip, power spent by the TEC layer, cooling power consumption by the CRAC and finally the processor reliability model in the following.

4.1 Modeling Sources of Power

Dynamic Power: We use Wattch [8] to estimate power consumed by different architectural components for an Alpha EV6 processor running SPEC-CPU 2000 programs with MinneSPEC large inputs. Though there are data-center servers that use embedded processors to reduce power consumptions, a major segment of data-centers still use high power processors for performance [24]. Wattch, which is validated against 3 processors ($10-13\%$ errors in modeling reported [8]), provides a power trace and we use it to obtain the power density over time for each component. We account for this error by performing a sensitivity study on the heat flux rate. We choose representative points in the execution by analyzing temperature traces collected using HotSpot tool. We define a phase change if temperature of any unit changes by 8% on average in a window of 50 million instructions. We found this heuristic to detect the phases correctly, which we validated with visual inspection. The computed power density is provided as an input power source to our chip heat diffusion model. Thus we evaluate the system for representative points during execution.

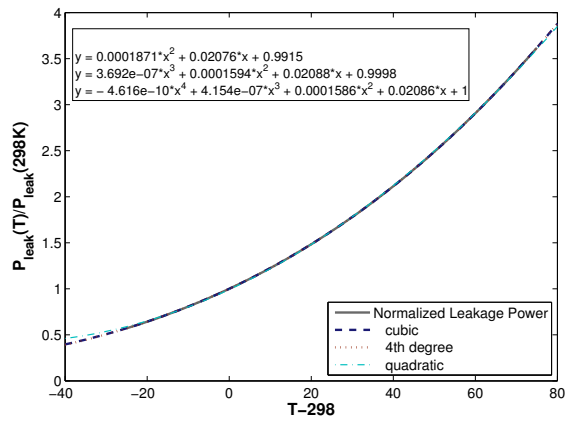


Figure 7: Leakage dependence on temperature obtained by simulating a SRAM cell at 32nm technology node using HSPICE for various temperatures, and then performing curve-fitting with Matlab

Leakage Power Estimation: In order to build a model for how leakage power varies with temperature, we simulated an SRAM cell at 32nm technology node using HSPICE for various temperatures and obtained the leakage power relative to 298K. In order to determine leakage power *density*, we first estimate the leakage power density at $298K$, and then scale this number for different temperatures using relative leakage power values from the HSPICE results. According to ITRS data [2], the leakage current at 298K for 32nm technology is $60nA/\mu m$ (at 1.1V). We assume that transistors cover 70% of the overall area [2] and that $1/2$ of the CMOS transistors are off at any given time. Since the length of a smallest sized transistor for a 32nm technology node is $0.096\mu m$, the leakage power density can be estimated as $0.7*0.5*1.1*(0.06/0.096)$ $\mu W/\mu m^2$ *i.e.* $24.05W/cm^2$ at nominal temperature $298K$. Post curve-fitting, we find that a 3rd-order model represents the relationship of leakage power with temperature keeping the error rate below 0.25%.

4.2 Modeling Heat Diffusion

We model the diffusion of heat flux through the silicon and other layers through solving heat equation of a system using iterative methods. The temperature inside the package at point (x,y,z) and at time $t+1$ is modeled with equation (9). The

boundary condition is modeled using convection process to air with equation (10). The heat sink is modeled by treating it as a part of the chip stack [10]. In our model we derive the behavior from thermal properties and the geometry of the heat sink. We treat the points outside the surface as independent of each other though in a practical situation, their behavior will be affected by each other. Fans are modeled by using a higher value of the convection heat transfer coefficient ($80W/m^2\,^\circ$C) [10]. We derive these equations in Appendix A, which can be referred to for further details.

We model the TEC layer as part of the chip stack, treating the hot side as a heat source and the cold side as a heat sink. We transform the heat-pumping-capability equation of TECs to the discrete domain and incorporate in our numerical PDE solver. Please refer to Appendix A for further details.

4.3 TEC Power Modeling

Tne efficiency of TEC, termed as coefficient of performance (COP), determines the cost of cooling using TEC. The maximum value of COP [3] can be calculated as:

$$COP_{max} = \frac{T_{cold}}{T_{hot} - T_{cold}} \frac{(1 + ZT_{avg}^{0.5}) - T_{hot}/T_{cold}}{(1 + ZT_{avg}^{0.5}) + 1} \quad (1)$$

where $T_{avg} = (T_{cold} + T_{hot})/2$

Chowdhury *et al.* [11] have demonstrated efficient TECs with ZT (figure of merit) as high as 2.5 along with reduction of 15°C temperature. Conservatively with an assumption of 10°C temperature difference with the cold side at $350K$, we find COP_{max} of ≈ 10.25.

4.4 CRAC Power Modeling

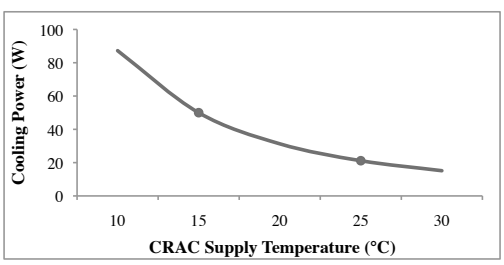


Figure 8: Efficiency model of a water chilled CRAC unit [22]: two points on the graph show the benefit of increasing the supply air temperature from 288K to 298K to extract 100W of heat

In order to maintain the temperature of the server room, the CRAC unit needs to remove at least equal amount of heat as generated by the server. We adapt the efficiency model for a *water chilled* CRAC unit as presented by Moore *et al.* [22] in this work, which is shown in Figure 8. In a typical HPC data center, due to reliability requirements, cold air of $\sim 288K$ temperature is provided to the servers, as recommended by ASHARE [14] and used in many industry standard data centers [20]. In our work we experiment with different supply air temperatures as newer data centers promote supplying relatively warmer air.

4.5 Modeling Other Server Components

Server Fans: Power consumed in server fans can vary anywhere between 8W and 40W as CPU load changes. As fan speed has a cubic relationship with its power consumption, running the fan at higher speed consumes more amount of power. In this analysis, we assume that the fan speed is fixed at $10W$ [4].

Memory: Changes in temperature have very little effect on the power consumption of the memory system. Subthreshold leakage is not the primary source of leakage as opposed to logic components. Leakage in DRAM is found to be slower and having lower dependence on temperature, and therefore, we exclude it from our model and assume a power consumption of $5W$ [4].

Misc. Server Components: The effect of temperature on power consumption of other components is very small compared to the processor and fans. Some devices may even consume less power e.g. hard drives spinning consumes less power as the viscosity of the grease inside it decreases with increase in temperature. We assume that rest of the server consumes $40W$ [4].

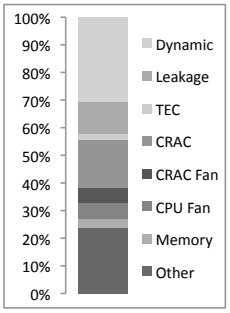


Figure 9: Contribution of components at $294K$

Chiller plan, CRACs and Fans: The efficiency of the cooling subsystem, primarily the chiller plant and CRACs, gets affected heavily with changes in supply air temperature. We consider that the CRAC fans consume a fixed amount of power: 10KW per CRAC [22] *i.e.* $8.93W$ per server. In Figure 9, we show the contribution of each component to the power consumption of the system.

4.6 Modeling Processor Reliability

We follow the RAMP [32] reliability modeling methodology in this work. In a processor, hard failures arise from several phenomena – Electromigration (EM), Stress migration, Time-dependent dielectric breakdown (TDDB) and Thermal cycling. Of these, EM and TDDB are reported to be most critical for small feature sizes and have strong correlations with temperature. Therefore, we select these two factors of failure for evaluating life span (MTTF) of a processor. Following the RAMP methodology, we exclude L2 caches while computing MTTFs as they often have ECC and can be bypassed to preserve correctness. For details, please refer to Appendix B.

5. Experimental Methodology

We collect power traces by running SPEC CPU2000 benchmarks on Wattch [8]. The configuration of the modeled Alpha-EV6-like processor is shown in Table 1. We select representative simulation points for each benchmark following the method described in Section 4.1, and then compute power density for each block of the floorplan for each such point, which we feed to our numerical PDE solver along with the configuration of the environment (ambient temperature and convective heat transfer coefficient of air flow) and package configuration. In this paper, we evaluate the benefits for 43 application phases across SPEC CPU2000 benchmarks. We simulate a chip package where we make realistic assumptions in its configuration and physical parameters (listed in Table 2) and estimate the temperature distribution inside the package.

With different ambient temperatures, we obtain the heat map of the die layer for the case where TEC is present and the case

I-Fetch Q	8	**Issue/Commit Width**	4/4
RUU Size	128	**iALU/iMult/fAlu/fMult**	2/2/1/1
LSQ Size	64	**L1 I,D-Cache Ports**	4
Branch Predictor	2-level, 1024 Entry History Length 10	**BTB size**	2048
		RAS entries	16
L1 I-Cache	32KB + 32 KB, Direct Mapped, 64 byte lines	**L1 Latency**	1 Cycle
L1 D-Cache		**Branch Penalty**	3 Cycles
L2 Cache	4MB, 8 way, 64B lines	**L2 Latency**	6 Cycles

Table 1: Configuration of simulated processor in Wattch

Layer	**Area** (mm^2)	**Height** (mm)	**Specific Heat** (J/kgK)	**Density** (kg/m^3)	**Thermal Conductivity** (W/mK)
Die	10x10	0.5	712	2330	120
TIM 1	10x10	0.2	230	7310	30
IHS	30x30	1.8	385	8930	390
TIM 2	30x30	0.2	2890	900	6.4
Heat Sink	60x60	6.4	385	8930	360
Slab-fin	0.1x0.1	20			
$TEC(Bi_2Te_3/Sb_2Te_3)$ [34], $\alpha = 301\mu V/K$, $\rho_e = 108 \times 10^{-5}\Omega m, capacity = 400W/cm^2$					
TEC	9	0.09	162.5	7100	17.0

Table 2: Properties of chip package layers

with passive cooling only, *i.e.* without TEC. Using the CRAC model of efficiency, we report the estimated power saving in the cooling infrastructure. In this step we also consider the power consumed for powering the TEC. With our reliability model, we estimate the life-span of a processor without using TECs and compare the its power consumption with the case of a TEC packaged processor running at a air temperature higher supply without affecting life-span (MTTF due to EM and TDDB).

6. Results

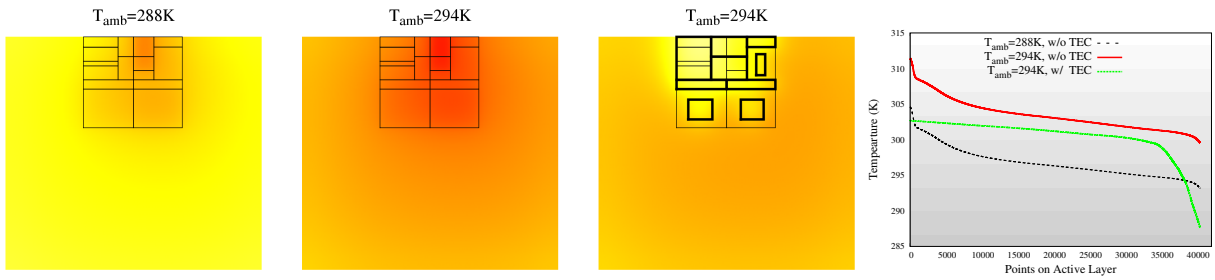
In this section, we present the results from our experiments to quantify the effect of micro-level coolers on global cooling and also on the lifetime of the processors. In our experiments, we use lower ambient temperature than prior work [33] in simulating thermal behavior, as the ambient temperature in a data center is significantly lower (15 °C vs. 45 °C) when compared to ambient temperature for a desktop server.

In Figure 11, we show the heat flux in individual components across all 43 application phases, illustrating the diversity in our evaluation scenarios. For each of these simulation points we estimate the cooling power consumption following the methodology described in Section 4 with and without TEC devices. Due to the positive feedback nature of leakage on temperature of the chip, overall temperature of the chip increases when the supply air temperature is changed from 288K to 294K as shown in Figure 10(a). Using TECs aligned to the hot regions of the processor die, the peak temperature of the chip is reduced as shown also in Figure 10(a). Looking into the temperature distribution across points in the active layer provides us with the insight behind the effectiveness of TECs as shown in Figure 10(b). By increasing the supply air temperature from $288K$ to $294K$, the temperature of the points in the active layer increase in the same pattern. By using the TECs with a higher ambient temperature cuts off the peaks, reducing the peak temperature, though the temperature of L2 cache increases as shown in Figure 10(a).

Recent endeavors in reducing data center cooling power have explored the possibility of increasing supply air temperature, which increases the COP of cooling units, and thereby, reduces cooling power significantly. However, as shown on the secondary vertical axis in Figure 12, chip reliability worsens considerably when the supply air temperature is increased from $288K$ (60 °F); sacrificing 14% $MTTF_{EM}$ and 7% $MTTF_{TDDB}$ to save 12% cooling power at supply air temperature of $290K$ (63 °F), and 37% $MTTF_{EM}$ and 20% $MTTF_{TDDB}$ to save 27.5% cooling power with $294K$ (70 °F) air temperature on average across 43 application phases. By switching on TECs over the hot regions of the die, the reliability of the processor is not sacrificed, while retaining significant cooling power savings, as shown in Figure 12(b). In an aggressive use of TECs, where all TEC blocks are activated, 19% cooling power could be saved on average by raising supply air temperature to $294K$, while increasing reliability by 23.2% from EM failures and 11.4% from TDDB failures. On a more selective activation scenario of TEC blocks, where only TECs over relatively hot areas are switched on, a savings of 24% in cooling power could be achieved at the cost of 4.5% $MTTF_{EM}$ and 2.5% $MTTF_{TDDB}$, which is a small loss considering the average life span of processors [32].

Since Wattch and HotSpot tools introduce some errors in evaluation, up to $10-13\%$ from Wattch [8] and $0.2K$ from HotSpot [35], we conduct a sensitivity experiment. We estimate the variation in cooling power reduction for a subset of application phases by setting heat flux values by $\pm 10\%$, as shown in Figure 13. In this experiment we activate all TEC blocks to ensure high reliability. Selective activation of TECs provides higher benefits ($\sim$24%), as shown in Figure 12(b). We find that TECs work better for higher power density, as the scope for savings becomes larger, though the variation is marginal. We also examine the sensitivity of savings in power with the efficiency of TEC devices. With COP of 1, *i.e.* spending $1W$ power to extract $1W$ of heat, even selective activation TECs blocks may consume $\sim 17W$ power on average, and thereby increasing cooling cost by $\sim 21\%$. However, with a low COP of 2 we start observing benefits as shown in Figure 14. With high efficiency TECs ($COP = 10$), $\sim$24% cooling power could be saved without sacrificing chip reliability. With TECs of moderate efficiency ($COP = 5$), we find that switching all TEC blocks lead to only 8% cooling power savings, whereas selective activation of TEC blocks lead to 18.5% savings with negligible loss of 6% EM and 3% TDDB MTTF.

Finally, we perform a study on one application phase over a range of supply air temperature to estimate the savings. We select $apsi_1$ for this study, which is representative of average behavior, as shown in Figure 12(b). In order to save cooling power, the supply air temperature could be raised, which leads to reduction in MTTF. TEC devices improve the reliability, thereby enabling operation of CRAC to supply air at higher temperature. However, due to power consumption in the TECs and increase in leakage power, the benefit of increasing target supply air temperature decreases, as shown in Figure 15. Due to this overhead in cooling, TEC device packaged chips require more cooling power at the same supply air temperature, but the chip reliability is enhanced significantly. Running the data center with higher ambient temperature and still achieving same reliability as low supply air temperature yields significant cooling power savings. It can be observed that TEC devices always provide better pareto optimal choice over the case where ambient temperature is increased while not using TECs. By comparing two points with equal y-coordinate, one lying on the trend line for TEC packaged chip and the other on the trend line corresponding to the case without TECs, we measure the savings in power.



(a) Heat Map (b) Temperature Distribution

Figure 10: Temperature profile and effect of TECs. (a) The heatmap of the active layer with an air temperature of 288K and without TECs is shown on the left. The chip becomes hotter by increasing the ambient temperature to 294K, which is shown in the middle heatmap. By using TECs, temperature peaks are cut off leading to reduction in peak temperature as shown in the rightmost heatmap and the linear profile in (b).

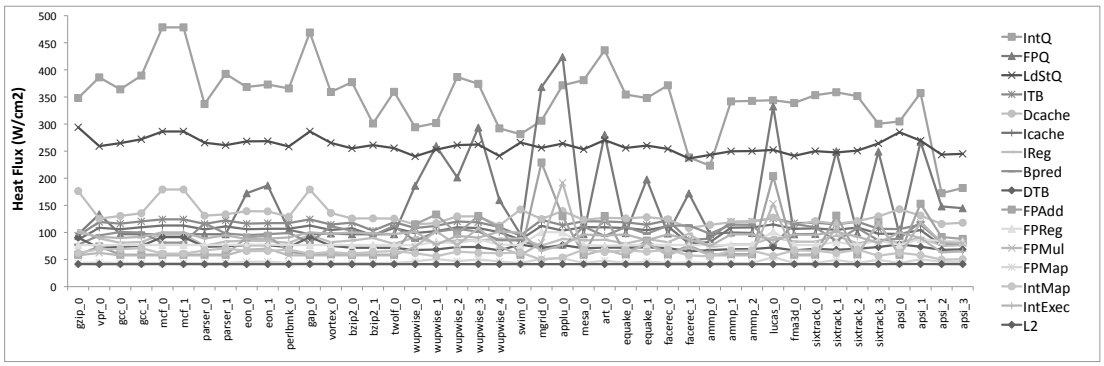


Figure 11: Heat flux in different architectural components estimated across benchmarks using Wattch

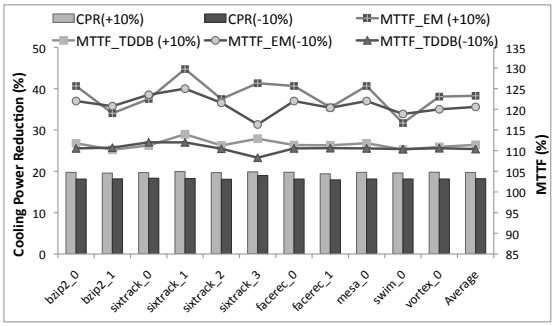


Figure 13: Sensitivity to 10% variation in heat flux with all TECs active

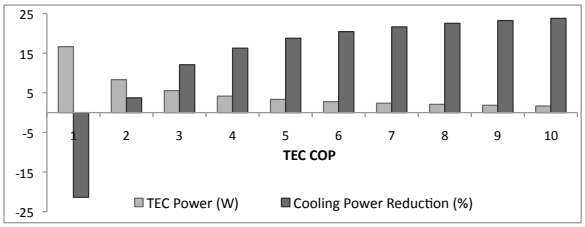
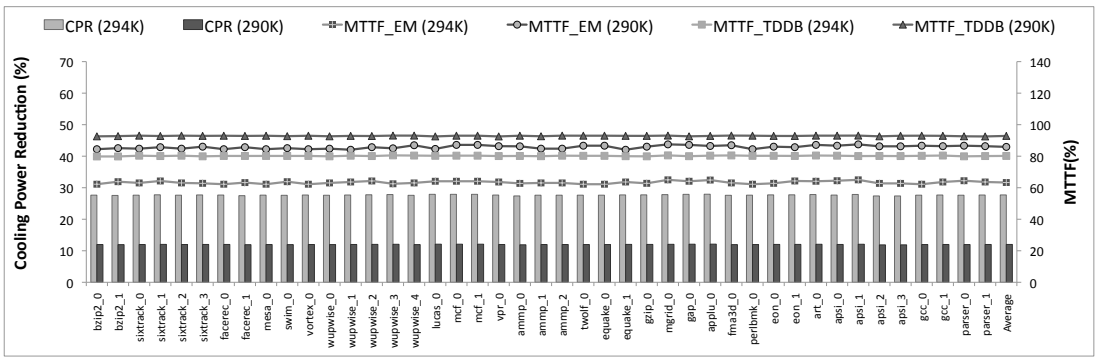


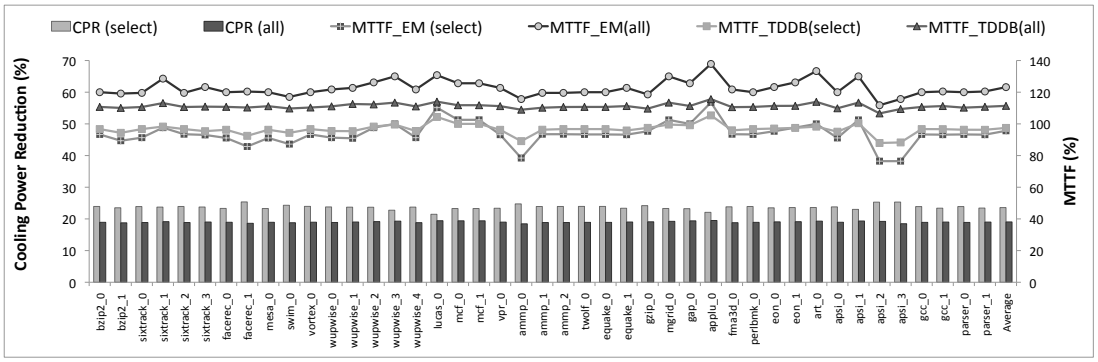
Figure 14: For TECs with COP of 10, over 23.5% cooling power can be saved by supplying air at 294K instead of 288K, whereas > 12% savings is achievable by activating all TECs having low efficiencies (COP=3) as well.

If the CRAC unit is set to supply cold air at 292K to save 24% cooling power with a reduction in chip life time by 25% due to EM failures, deploying TECS with COP of 5 does not result in cooling power savings by raising data center temperature even up to 302K, as shown in Figure 15(a). In a data center running at 294K (70°F), cooling power increases if TEC of COP 5 or lower is used. From our simulations we find that if one is ready to sacrifice chip reliability by supplying air at 300K (82.4°F), there may not be any benefit of using TECs if the COP is < 10, though the chip life time may be reduced to 42% of a chip cooled with air at 288K (60°F). For a data center running at 294K ambient temperature, we observe 12% reduction in cooling power with a TEC COP of 10, while providing the same reliability. With TECs of even higher COP = 15, 16% cooling power could be saved in such a scenario. In more conservative data centers, for example HPC data centers, where CRAC supplies air at 288K ($\sim$ 60°F) [1], TEC devices of COP = 10 and COP = 15 could lead to 27.5% and 30% cooling power savings respectively without affecting chip reliability.

Static vs. dynamic activation: We compare the two ways of using TECs; (1) using minimal number of TECs to achieve cooling power reduction without affecting chip reliability, and (2) activating all TECs and at the same time supplying even warmer air to achieve the same level of reliability. The savings in cooling power from CRAC units by supplying warmer air, and the increase in leakage power and the TEC power consumption decide the suitable method. We show the benefits of the first method in Figure 15 and the second method in Figure 14. We find that with TECs of low or moderate COP (*e.g.* 5) provide higher benefit with selective activation (19% vs.

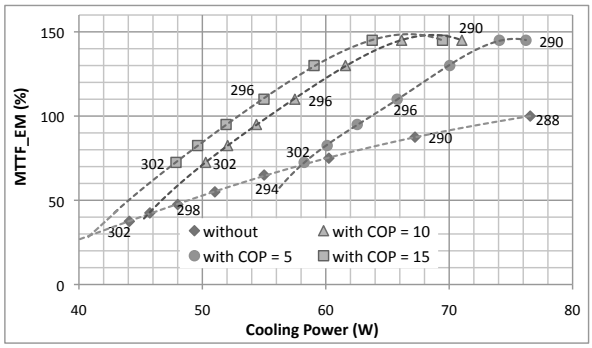


(a) Without TEC

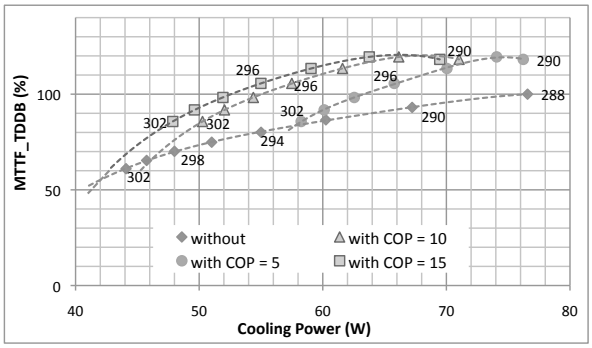


(b) With TEC

Figure 12: (a) Without using TECs 12% *and* 27.5% *cooling power could be saved by raising supply air temperature from* 288*K to* 290*K and* 294*K respectively, but at the cost of* 7% *and* 14% $MTTF_{TDDB}$ *and* 20% *and* 37% $MTTF_{EM}$. *(b) TECs may lead to* 19% *cooling power reduction while improving* $MTTF_{EM}$ *by* 22.5% *and* $MTTF_{TDDB}$ *by* 11% *if* all *TEC blocks are powered up. A* selective *activation of TECs could lead to* 24% *cooling power reduction at the cost of* 6% $MTTF_{EM}$ *and* 3% $MTTF_{TDDB}$.



(a) Cooling Power and $MTTF_{EM}$



(b) Cooling Power and $MTTF_{TDDB}$

Figure 15: By increasing the supply air temperature cooling power may be reduced significantly even without using TEC, but at the cost of reduced reliability. As shown in (a) and (b), MTTF corresponding to EM and TDDB effects decrease with increase in supply air temperature (indicated with labels to the points in these graphs). By spreading heat from hotspots, TEC devices achieve higher reliability while reducing cooling power consumption. In an aggressive data center running at 294*K (*70°F*), cooling power even increases if TEC of COP 5 or lower is used. With a COP of 10, we observe* 12% *reduction while not reducing chip life span. With a TEC of even higher COP of* 15, 16% *cooling power could be saved. In a more conservative data center, for example HPC data centers running at* 288*K (*60°F*), TEC devices of COP = 10 lead to* 27.5% *cooling power savings without loss of MTTF.*

14% savings in cooling power) with minor loss of chip reliability (6% from EM failures). However, TECs with high efficiency (COP = 10 [11]) are more effective (27% vs. 24%) if more of them are switched on and the supply air temperature is even more increased, as shown in Figure 15. In architectures, where hot-spots switch dynamically [13] (*e.g.* between ALU and FPU), selective activation of TECs may lead to higher savings in cooling power.

Summary: We find that TEC devices present us with an opportunity to reduce global cooling power in data centers with-

out affecting chip life time. However, energy efficient superlattice coolers are required to leverage this savings. TECs with low or moderate efficiency (e.g. COP = 5) may not provide any reduction in cooling power due to the increase in leakage power when supply air temperature is increased. Even with power efficient TECs (e.g. COP = 10), the reduction in cooling power will be negligible if reliability is already largely sacrificed in the base case to save cooling power by raising data center temperature to $\sim$85°F. If chip reliability is important, *e.g.* in HPC data centers, efficient TEC devices (with COP = 10, demonstrated by Chowdhury *et al.* [11]), could lead to savings in cooling power by $\sim$27% while providing the same chip life span as running the data center at $288K$ (60°F).

7. Related Work

Our work touches on different areas of research including thermal modeling of processors, hot spot management and techniques for reducing the power usage in data centers. In this section, we describe prior work in these domains and compare them to our work.

Thermal Modeling of Processors: The increase of power density inside processors has led to higher device failure rates. At the architecture level, Brooks *et al.* [8] introduced Wattch, which estimates power usage of a processor on the basis of its component activity. The components are modeled at a block level and coarse-grain power usage statistics are derived. Stan *et al.* [33] proposed the HotSpot model, which uses a lumped modeling technique to model the thermal paths inside a processor and computes the temperature at a block granularity using power density statistics. Li *et al.* [19] use geometric multigrid methods to efficiently solve a large number of heat PDEs efficiently. Monchiero *et al.* [21] evaluate thermal/power/performance design choices in multicore architectures by modeling heat diffusion in them. Our work builds upon the full chip thermal model using fundamental heat equations and solving it for the steady state. While the previous work models processors in isolation, our model integrates TEC and CRAC efficiencies to determine power savings at the data center level.

Hot-Spot Removal: There has been a significant body of work in finding solutions to mitigate hot spots. Processors (e.g. Transmeta Crusoe) often use dynamic voltage scaling (DVS) or deactivation techniques to control the temperature of processors. Skadron *et al.* [31] introduced thermal-behavior-aware microarchitecture and floorplanning techniques to reduce hot-spots. Puttaswamy *et al.* [25] propose *Thermal Herding* technique that reduces 3D power density by placing highly switching 16-bits closer to the heat sink, and thereby, reducing the occurrences of hot-spots as well. Since hot spots arise from high transistor switching rates, Heo *et al.* [15] propose that migrating the computations frequently across multiple locations can reduce the hot-spots. Task migration approach is orthogonal to active cooling technique that we analyze in this paper. In fact, both of these techniques could work well with each other, where TEC could be used to prevent wear out at high loads. Task migration could be used in presence of redundant structures. Adding such structures, e.g. extra ld/st queue) would be intrusive to processor design. Huang *et al.* [16] propose a framework to control the performance of a chip for keeping the temperature of the processor below a target. We find that since TECs target hot-spots directly, it can be used in conjunction with the above more indirect means of mitigating hot-spots.

Modeling Microprocessor Reliability: We build our reliability models following the RAMP [32] modeling methodology, which considers four phenomena behind processor wear-outs. In this work, we consider failures due to EM and TDDB as other mechanisms have lower effect on modern microprocessors [32]. We extend this model by computing the joint probability distribution of failure by incorporating the MTTF function for EM and TDDB, which are found to show lognormal [18] and Weibull distributions [17] respectively.

Reducing the Power Usage in Data Centers: On a data center level, where thousands of processors are cooled constantly using CRAC units, conserving power has both ecological and financial benefits. Patel *et al.* [23] and Schmidt *et al.* [28] modeled CFD simulations of a data center air flow and report that careful design and air cooling provisioning is important for sustained operation of a data center. Chaparro *et al.* [9] present a quantitative model of data center power efficiency and report that reducing supply air temperature provides larger energy savings than running the blowers at high speed. However, due to non-uniformity in resource usage, hot spots at lower granularities cannot be removed. Sharma *et al.* [29] propose thermal monitoring and thermal load balancing techniques by scheduling jobs to balance the temperature across a data center through prediction of utilization of server resources. Moore *et al.* [22] propose techniques (zone-based discretization and minimized heat recirculation) to create prioritized lists of servers for proper scheduling of jobs, taking into account the efficiency of CRAC units for supplying cold air. Bash *et al.* [5, 6] propose an architecture and control mechanism for reducing cooling cost using dynamic thermal management based on a cooling cost model. Sharma *et al.* [30] report that power consumption for cooling a data center can be reduced significantly by designing the air flow path to prevent mixing of hot and cold air and present non-dimensional parameter based models of the air flow inside aisles. These techniques improve the efficiency of the global coolers and should work well in conjunction with direct hot-spot mitigation using TECs.

8. Conclusions

In a data center with thousands of servers, the cooling system plays a major role in reliability of the servers, and also in the cost of running a data center, which is typically in the order of millions of dollars per year. In a typical data center, the cooling power is nearly equal to the total power of running the IT equipment. Cooling units show higher efficiency with the increase in the target supply air temperature. However, the cooling units are provisioned for the worst case, *i.e.* the hot-spots inside the processors. As reliability is exponentially related to the operating temperature, increasing the supply air temperature reduces the lifetime of the hottest processor component, and in turn, of the entire processor.

Localized cooling via superlattice thermoelectric coolers provides the capability to reduce the temperature of the hot-spots, and thereby, increase the lifetime of the critical components. Recent advances suggest the feasibility of integrating high efficiency thermoelectric coolers [11] with the chip package to create high efficiency heat pumps for cooling hot-spots. In this paper, we propose that TEC devices could be used as a *cooling power management mechanism* to reduce global cooling cost. We integrate a multi-scale model of the system, considering different thermal components of a data center, from

chip package to the cooling units. Then, we perform numerical analysis to obtain the chip thermal behavior and quantify the benefit of incorporating a TEC inside the chip package. We show that ∼27% of global cooling power could be saved, compared supply air temperature of $288K$, by running data centers at a higher temperature while cooling the hot-spots selectively with superlattice TECs.

Acknowledgements

Part of this work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344 (LLNL-CONF-474345), and Grant No. FA9550-07-1-0532 (AFOSR MURI), a Google Directed Research Grant for Energy-Proportional Computing and NSF 0627749 to Frederic T. Chong, Grant No. CCF-0448654, CNS-0524771 and CCF-0702798 to Timothy Sherwood.

9. References

[1] Comp turns up the heat on energy conservation. `https://newsline.llnl.gov/_rev02/articles/2009/oct/10.02.09-energy.php`.

[2] The International Technology Roadmap for Semiconductors. http://www.itrs.net/.

[3] A. Bar-Cohen and P. Wang. On-Chip Thermal Management and Hot-Spot Remediation. In *Nano-Bio-Electronic, Photonic and MEMS Packaging*, pages 349–429. Springer Science+Business Media, LLC, 2010.

[4] L. A. Barroso and U. Hölzle. The Case for Energy-Proportional Computing. *IEEE Computer*, 40(12):33 –37, Dec. 2007.

[5] C. Bash and G. Forman. Cool Job Allocation: Measuring the Power Savings of Placing Jobs at Cooling-efficient Locations in the Data Center. In *ATC'07: 2007 USENIX Annual Technical Conference*.

[6] C. Bash, C. Patel, and R. Sharma. Dynamic Thermal Management of Air Cooled Data Centers. In *ITHERM '06*, pages 445–452.

[7] J. Black. Electromigration - A Brief Survey and Some Recent Results. *IEEE Transactions on Electron Devices*, 16(4):338–347, 1969.

[8] D. Brooks, V. Tiwari, and M. Martonosi. Wattch: A Framework for Architectural-level Power Analysis and Optimizations. *ACM SIGARCH Computer Architecture News*, 28(2):83–94, 2000.

[9] P. Chaparro, J. Gonzalez, and A. Gonzalez. Thermal-aware Clustered Microarchitectures. In *ICCD 2004*, pages 48–53.

[10] C. Chen, C. Wu, and C. Hwang. Optimal Design and Control of CPU Heat Sink Processes. *IEEE Transactions on Components and Packaging Technologies*, 31(1):184, 2008.

[11] I. Chowdhury, R. Prasher, K. Lofgreen, G. Chrysler, S. Narasimhan, R. Mahajan, D. Koester, R. Alley, and R. Venkatasubramanian. On-chip Cooling by Superlattice-based Thin-film Thermoelectrics. *Nature Nanotechnology*, 2009.

[12] G. Chrysler. Building Blocks for Thermal Management of Electronics. *Next-Generation Thermal Management Materials and Systems*, 2002.

[13] S. Chung and K. Skadron. Using On-Chip Event Counters For High-Resolution, Real-Time Temperature Measurement. In *ITHERM'06*, pages 114–120.

[14] D. Garday and D. Costello. Air-Cooled High-Performance Data Centers: Case Studies and Best Methods. Technology@Intel Magazine, `http://www.intel.com/it/pdf/air-cooled-data-centers.pdf`.

[15] S. Heo, K. Barr, and K. Asanović. Reducing Power Density Through Activity Migration. In *ISPLED 2003*, pages 217–222.

[16] M. Huang, J. Renau, S. Yoo, and J. Torrellas. The Design of DEETM: a Framework for Dynamic Energy Efficiency and Temperature Management. *Journal of Instruction-Level Parallelism*, 3:1–31, 2002.

[17] M. Kimura. Field and Temperature Acceleration Model for Time-Dependent Dielectric Breakdown. *IEEE Transactions on Electron Devices*, 46(1):220–229, 2002.

[18] M. Lane, E. Liniger, and J. Lloyd. Relationship Between Interfacial Adhesion and Electromigration in Cu Metallization. *Journal of Applied Physics*, 93:1417, 2003.

[19] P. Li, L. Pileggi, M. Asheghi, and R. Chandra. Efficient Full-chip Thermal Modeling and Analysis. In *ICCAD 2004*, pages 326–331.

[20] P. Mathew, S. Greenberg, S. Ganguly, D. Sartor, and W. Tschudi. How Does Your Data Center Measure Up? Energy Efficiency Metrics and Benchmarks for Data Center Infrastructure Systems. Technical Report LBNL-1960E, April 2009.

[21] M. Monchiero, R. Canal, and A. Gonzalez. Power/Performance/Thermal Design-Space Exploration for Multicore Architectures. *IEEE Transactions on Parallel and Distributed Systems*, 19(5):666 –681, 2008.

[22] J. Moore, J. Chase, P. Ranganathan, and R. Sharma. Making Scheduling "Cool": Temperature-Aware Workload Placement in Data Centers. In *ATEC'05: 2005 USENIX Annual Technical Conference*.

[23] C. Patel, R. Sharma, C. Bash, and A. Beitelmal. Thermal Considerations in Cooling Large Scale High Compute Density Data Centers. In *ITHERM 2002*, pages 767–776.

[24] J. Pflueger and A. Esser. The Energy Smart Data Center. in Dell Power Solutions, `http://www.dell.com/downloads/global/power/ps1q08-20080179-CoverStory.pdf`, February 2008.

[25] K. Puttaswamy and G. Loh. Thermal Herding: Microarchitecture Techniques for Controlling Hotspots in High-performance 3d-Integrated Processors. In *HPCA*, pages 193–204, 2007.

[26] N. Rasmussen. Electrical Efficiency Modeling of Data Centers. APC. White paper, 2006.

[27] E. Rotem, J. Hermerding, A. Cohen, and H. Cain. Temperature measurement in the Intel (R) CoreTM Duo Processor. *Arxiv preprint arXiv:0709.1861*, 2007.

[28] R. R. Schmidt, E. E. Cruz, and M. K. Iyengar. Challenges of Data Center Thermal Management. *IBM Journal of Research and Development*, 49(4/5):709–723, 2005.

[29] R. K. Sharma, C. E. Bash, C. D. Patel, R. J. Friedrich, and J. S. Chase. Balance of Power: Dynamic Thermal Management for Internet Data Centers. *IEEE Internet Computing*, 9(1):42–49, 2005.

[30] R. K. Sharma, C. E. Bash, and R. D. Patel. Dimensionless Parameters For Evaluation Of Thermal Design And Performance Of Large-Scale Data Centers. In *8th ASME/AIAA Joint Thermophysics and Heat Transfer Conference*, 2002.

[31] K. Skadron, M. Stan, K. Sankaranarayanan, W. Huang, S. Velusamy, and D. Tarjan. Temperature-aware Microarchitecture: Modeling and Implementation. *ACM TACO*, 1(1):94–125, 2004.

[32] J. Srinivasan, S. V. Adve, P. Bose, and J. A. Rivers. The Case for Lifetime Reliability-Aware Microprocessors. In *ISCA '04*, pages 276–287.

[33] M. R. Stan, K. Skadron, M. Barcella, W. Huang, K. Sankaranarayanan, and S. Velusamy. HotSpot: a Dynamic Compact Thermal Model at the Processor-Architecture Level. *Microelectronics Journal*, 34:1153–1165, 2003.

[34] F. Takahashi, Y. Hamada, T. Mori, and I. Hatta. Thermal Characteristics at Interface of Bi_2Te_3/Sb_2Te_3 Superlattices. *Japanese Journal of Applied Physics*, 43(12):8325–8330, 2004.

[35] S. Velusamy, W. Huang, J. Lach, M. Stan, and K. Skadron. Monitoring temperature in FPGA based SoCs. In *ICCD 2005*, pages 634 – 637.

[36] Y. Zhang, J. Christofferson, A. Shakouri, G. Zeng, J. Bowers, and E. Croke. On-Chip High Speed Localized Cooling Using Superlattice Micro-Refrigerators. *IEEE Transactions on Components and Packaging Technologies*, 29(2):395, 2006.

APPENDIX

A. Modeling Heat Diffusion

From the second law of thermodynamics, we use equation (2), where A is the surface area and n is the direction normal to it. T is the temperature distribution and k is the thermal conductivity of the material. We assume k to be a constant across layer. It has been shown that the temperature dependence of thermal conductivity of silicon is low, and even a change is temperature by 20°C results in only 6.67% change in k, which does not affect the results significantly. Similarly, from Newton's law of cooling we find equation (3) for convection heat loss from surface, where h is the convection-heat-transfer-coefficient, $T_{surface}$ is the temperature at the surface and T_{amb} is the temperature a distance of infinity from that surface. Equation (2) and (3) can be rewritten in a continuous form as a partial differential equation as equations (4) and (5). In these two equations, C_p is the heat capacity (J/gmK), k is the thermal conductivity (W/mK) and ρ is the density (gm/m^3) of the material. $g(x,y,z,t)$ is the heat generation rate (J/m^3s) at point (x,y,z), which includes the leakage and dynamic power. For simplicity, we assume k to be constant across a layer and g to be a step function inside the layer. At the hot spot, $g(x,y,z)$ is larger to other points by an order of magnitude. Transforming the equations to a discrete form we obtain equations (6) and (7), where M_x, M_y, M_z are linear operator in x, y and z directions as described in equation (8).

We assume that there is no heat flow towards the bottom of the die as the rate of heat flow in that direction negligible compared to other directions. We model it by Neumann boundary condition where M_z is treated as $2 \times (T_{1,y,z} - T_{0,y,z})$. Simplifying the equations further by assuming the unit length in all directions to be same, we obtain equations 9 and 10. Thermoelectric cooler's heat pumping capability is described by equation (12), where ρ_e is the electrical resistivity of the TEC material. In order to model the heat flow in TEC layer, following the law of conservation of energy, we rewrite equation (12) as equation (11). Rewriting the equation in a discrete form, we obtain equations (13) and (14). We approximate the solution by ignoring the x and y direction heat propagation through Peltier effect in thermoelectric cooling layer as the geometry dictates flow in vertical direction primarily. The model has been simplified by assuming that the thickness of the TEC layer is the length of the cubes in the 3D mesh. The simple explanation of the TEC layer is that the cold side absorbs the heat at the cold side and transmits to the hot side (which is larger in area and thus spreads the heat, reducing the peak temperature of the hot spot). The temperature at the cold side continues decreasing until the Peltier heat flux from cold to hot side equals the conduction heat flux from hot side to the cold side. We solve this problem iteratively to find the stable temperature.

$$Rate\ of\ conduction\ (W) = -kA\frac{\partial T}{\partial n} \quad (2)$$

$$Rate\ of\ heat\ convection\ (W) = hA(T_{surface} - T_{amb}) \quad (3)$$

$$\rho C_p \frac{\partial}{\partial t} T(x,y,z,t) = \nabla.[k(x,y,z,t)\nabla T(x,y,z,t) + g(x,y,z,t)] \quad (4)$$

$$k(x,y,z,t)\frac{\partial}{\partial n} T(x,y,z,t) = h[T(x,y,z,t) - T_{amb}] \quad (5)$$

$$T^{t+1} = T^t + \Delta t\left(\frac{k}{\rho C_p}\right)\left[\left(\frac{M_x}{\Delta x^2} + \frac{M_y}{\Delta y^2} + \frac{M_z}{\Delta z^2}\right)T^t\right] + \Delta t\frac{g}{\rho C_p} \quad (6)$$

$$T_n = T_{amb} + (T_{surface} - T_{amb}).e^{-\frac{hn}{k}} \quad (7)$$

$$M_x T_{x,y,z} = T_{x-1,y,z} + T_{x+1,y,z} - 2.T_{x+1,y,z} \quad (8)$$

$$T^{t+1} = T^t + \Delta t\left(\frac{k}{\rho C_p}\right)\left[\left(\frac{1}{\Delta x^2}\right)(M_x + M_y + M_z)T^t\right] + \Delta t\frac{g}{\rho C_p} \quad (9)$$

$$T_{surface+1} = T_{surface} - \Delta x.\frac{h}{k}.(T_{surface} - T_{amb}) \quad (10)$$

$$\rho C_p \frac{\partial T}{\partial t} = \nabla.\frac{1}{l}\left(0.5\frac{\alpha^2}{\rho_e}T^2_{cold-side} - \nabla^2.kT)\right) \quad (11)$$

$$q = \frac{1}{l}\left(0.5\frac{\alpha^2}{\rho_e}T^2_{cold-side} - k(T_{hot-side} - T_{cold-side})\right) \quad (12)$$

$$T^{t+1}_{cold-side} = T^t_{cold-side} + \frac{k\Delta t}{\rho C_p \Delta x^2}\left[(M_x + M_y + M_z)T^t\right] - \frac{\Delta t}{\rho C_p \Delta x^2}\frac{\alpha^2}{2\rho_e}T^2_{cold} \quad (13)$$

$$T^{t+1}_{hot-side} = T^t_{hot-side} + \frac{k\Delta t}{\rho C_p \Delta x^2}\left[(M_x + M_y + M_z)T^t\right] + \frac{\Delta t}{\rho C_p \Delta x^2}\frac{\alpha^2}{2\rho_e}T^2_{cold} \quad (14)$$

In this work we use the explicit method and solve it by using a threshold for terminating simulation. We define steady state when the average temperature of the structure does not change by more than $1e^{-6}$°C.

B. Failure Model

In this section we present the reliability model based on RAMP [32]. Two major factors for processor failure, which have strong correlation with operating temperature, are Electromigration and Time-dependent dielectric breakdown (TDDB). Stress migration, which arises due to mechanical stress and Thermal cycling, which is caused by processor state change (e.g. power state change, shut-down, power-up etc.), have lower effect than Electromigration and TDDB for processors with small feature size [32]. Electromigration occurs due to mass transport of metal atoms in copper interconnects, resulting in depletion of metal in one region and pile up in other, which might lead to resistance variation or open circuits. According to Black's equation, mean time to failure at temperature $\hat{T}$ with respect to temperature T can be modeled as equation (15), where E_a is the activation energy and k is a constant depending on the interconnect metal. TDDB or gate-oxide breakdown is due to the breakdown of gate dielectric layer leading to a conductive path. MTTF due to TDDB, as shown in RAMP methodology, could be modeled as equation (16), where V is the V_{dd}, and a, b, X, Y and Z are fitting parameters with values of $78, -0081, 0.759ev, -66.8evK$ and $-8.37e^{-4}ev/K$ respectively (adapted from RAMP model [32]).

$$\frac{MTTF_{EM}(\hat{T})}{MTTF_{EM}(T)} = e^{\frac{E_a}{k\hat{T}} - \frac{E_a}{kT}} \quad (15)$$

$$\frac{MTTF_{TDDB}(\hat{T})}{MTTF_{TDDB}(T)} = (\frac{1}{V})^{(b\hat{T} - bT)}e^{\frac{(X+Y/\hat{T}+Z\hat{T})}{k\hat{T}} - \frac{(X+Y/T+ZT)}{kT}} \quad (16)$$

In a system with independent components, the failure of the chip will be related to the fastest failing point in it. Therefore the probability of a working chip at time t can be formulated as $P_{working}(t) = \prod^{c \in C}(1 - CDF_{failure}(c,t))$, where C represents all components in a chip and $CDF_{failure}(c,t)$ is the cumulative distribution function for failure rate of component c at time t. MTTF due to Electromigration follow lognormal distribution [18], for which we use $\sigma = 0.25$, and failures due to TDDB follow Weibull distribution [17], for which we use $k = 9$. We estimate MTTF of the processor when the probability of its working is > 0.95. We have performed a sensitivity study on these parameters and found similar correlation with processor reliability.