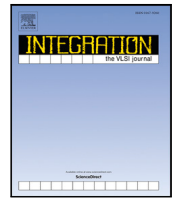




Contents lists available at ScienceDirect

Integration, the VLSI Journal

journal homepage: www.elsevier.com/locate/vlsi

Hot-spot aware thermoelectric array based cooling for multicore processors[☆]

Jinwei Zhang^{a,*}, Sherif Sadiqbatcha^a, Liang Chen^a, Cuong Thi^a, Sachin Sachdeva^a, Hussam Amrouch^b, Sheldon X.-D. Tan^a

^a Department of Electrical and Computer Engineering, University of California, Riverside, CA 92521, USA

^b Department of Electrical and Computer Engineering, University of Stuttgart, Stuttgart, Germany

ARTICLE INFO

Keywords:

Thermal modeling
Thermoelectric cooler array
Power modeling
Thermal simulation
Low power design

ABSTRACT

In this paper, we propose a hot-spot aware Thermoelectric Cooler (TEC)-based active cooling technique, called *TEC-Array*, which can perform targeted cooling of the spatially and temporally changing on-chip hot spots in any multi-core processor. It is in contrast to many existing works, where TECs were used for spatially homogeneous cooling, which leads to high energy costs. The proposed cooling system involves a 2D array of TEC modules where each TEC module is individually controllable. This enables the ability to spatially vary the cooling for hot spots across the chip's surface while the processor is under load. This way, the areas of the chip consuming more power, and consequently generating more heat, can be cooled more intensely than the areas that are consuming very little power. To dynamically control the TEC-Array, the proposed approach applies machine learning predicted full-chip power-density maps to generate discrete voltage maps which are in turn used to dynamically configure the voltage settings of the TEC-Array. In addition, we also propose a numerical simulation framework, which employs an accurate 3D coupled multi-physics model for TEC devices to consider Peltier, heat transfer, Joule heating and complex electro-thermal coupling effects by solving the coupled heat conduction and current continuity equations. The numerical results on an Intel quad-core chip shows that the proposed TEC-Array cooling can substantially reduce the peak temperatures compared to the traditional passive heat sink cooling method. Furthermore, compared to the existing single TEC module based cooling method, the proposed method can reduce both TEC power and temperature gradients across chips under the same maximum temperature constraints, which can further reduce spatial temperature induced stress such as thermal cycling, thermo-migration and unbalanced aging etc. As a result, the new TEC-Array cooling can enable more aggressive chip performance with increased thermal design power (TDP) while maintaining the chip design lifetime.

1. Introduction

Rapid integration and feature size scaling continue to increase power densities in modern high-performance processors. As a consequence, thermal management has become a challenging issue. Managing the temperature of the processor not only affects the performance of the chip, but also its reliability. It has been shown that long-term reliability effects, such as negative bias temperature instability and electromigration, worsen exponentially with temperature [1]. Therefore, effective and efficient cooling of high-performance processors is an important area of research.

Traditional cooling systems such as passive (natural convection) and active (forced convection) heatsinks have been used for many years as they offer a relatively simple, and cost-effective solution to

the problem at hand. However, with the aforementioned increase in power-densities, it becomes increasingly difficult to effectively manage the temperature of high-performance processors using heatsink cooling; this is especially true for server-grade chips with high core counts. Increasing the size of the heatsink is one solution to this problem, however, it is not a realistic solution due to space constraints. In addition, heatsink cooling can, at best, offer cooling up to the ambient temperature. Hence, there is a need for more advanced, area and power efficient, cooling systems for modern high performance processors both at the consumer level and at the server/super-computer level.

Thermoelectric coolers (TEC) are a promising technology that offer high cooling density with low area overheads. TECs are two-sided solid-state devices that use the Peltier effect to generate a heat flux with an

[☆] This work is supported in part by the National Science Foundation (NSF) grant under No. CCF-2007135 and in part by NSF grants under No. CCF-2113928 and No. OISE-1854276.

* Corresponding author.

E-mail address: jzhan319@ucr.edu (J. Zhang).

<https://doi.org/10.1016/j.vlsi.2022.11.006>

Received 29 March 2022; Received in revised form 7 October 2022; Accepted 6 November 2022

Available online 11 November 2022

0167-9260/© 2022 Elsevier B.V. All rights reserved.

applied potential difference between the two terminals of the device. TECs effectively transfer heat from one side (cold-side) to the other (hot-side) in proportion to the applied voltage. Hence, the rate of heat flow can be precisely controlled, allowing for the design of dynamically controlled cooling systems. Dynamically modulating the TEC voltage for on-demand cooling is crucial for making TECs a viable cooling method for microprocessors since TECs incur high power overheads.

Many studies have explored the use of TECs for the application of cooling high-performance processors. Dousti et al. proposed a hybrid cooling system that uses both TEC devices and traditional fans with the use of an optimization algorithm to dynamically vary the TEC voltage and fan speed in accordance with the processor's operating temperature [2]. Jayakumar et al. proposed a dynamic thermal management scheme (DTM) that enables the joint control of the processor's frequency and voltage, TEC voltage and the fan speed with the aim of maximizing the performance of the chip under temperature and power constraints [3]. While these methods were effective in cooling the processor, they are not energy-efficient implementations of TEC cooling as the entire die is cooled at the same rate. Meaning, energy is wasted in cooling the areas of the chip that are consuming very little power (generating very little heat). In addition, these methods do not address the concern of high thermal gradients across the chip's surface, which over-time results in some areas of the chip aging at a faster rate than in other areas.

Alternatively, Long et al. proposed a method of embedding Thin-Film TEC devices within the chip's package where TECs are placed only over hot spots [4]. While this method offers spatially varying cooling, it does not offer temporally varying cooling, which is an important factor in reducing the power consumption of the TECs. The reason is that hot spots are not active (producing heat) at all times as they are workload-dependent. Hence, Dousti et al. [5] proposed the use of TEC clusters located over the functional units that regularly produce hot spots on the chip. The TEC clusters are established using a clustering algorithm that aims to minimize power waste from unnecessary use of TECs. A current bypass switch technique is used to turn the clusters on and off as needed while the processor is under load. This method allows both spatially and temporally varying TEC cooling. However, the drawback of this method is that this system needs to be specially designed for each processor model and therefore is not a generic cooling system that can be widely adopted. Moreover, this method requires integrating the TEC clusters within the processor's package, hence requiring additional efforts in package design. There is still a lack of an externally mounted TEC-based on-demand cooling system that is generic enough to be used for a variety of microprocessors. Such a system can be quickly and easily adapted by industry with minimal overheads.

In this article, we aim to mitigate the aforementioned problems by proposing a hot-spot-aware TEC-based active cooling technique, called *TEC-Array*, which can perform targeted cooling of active hot spots that vary spatially and temporally in multi-core processors. The proposed system involves an array of TEC modules, controlled by a power-density based control scheme. Our new contributions are summarized below:

1. First, we propose to use the current state-of-the-art power modeling techniques to estimate the full-chip power-density distribution across the processor die. The estimated power-density maps are then converted to discrete voltage-maps, which are in turn used to dynamically configure the voltage settings of the TEC-Array. The proposed approach allows dynamically varying targeted cooling where hot spots are cooled more intensely than non-hot-spots. The proposed TEC-Array is especially beneficial for the running conditions in multi-core processors when tasks are mapped to a few cores while the others remain idle. In this scenario, only the cores that are under load can be cooled with minimal cooling applied to the idle cores which in turn leads to significant energy savings over time and a reduction of the aforementioned thermal gradients.
2. Second, to validate the proposed the TEC-Array based active cooling and power management method, we further propose a numerical simulation framework, which employs a more accurate 3D coupled multiphysics model for TEC devices to consider Peltier, heat transfer, Joule heating, and complex electro-thermal coupling effects by solving the coupled heat conduction and current continuity equations.
3. Lastly, with numerical results on an Intel quad-core chip, we show that the proposed TEC-Array cooling can substantially reduce the peak temperatures compared to the traditional passive heat sink cooling method. Furthermore, compared to the existing single TEC module based cooling method, the proposed method can reduce both TEC power and temperature gradients across chips under the same maximum temperature constraints, which can further reduce spatial temperature induced stress such as thermal cycling, thermo-migration, unbalanced aging etc. As a result, the new TEC-Array cooling can enable more aggressive chip performance with increased thermal design power (TDP) while maintaining the chip's design lifetime.

We remark that the primary goal of this work is to allow the chip to run at higher performance modes with increased TDP without suffering the thermal gradient induced aging impacts compared to the existing heat sink and single TEC device-based coolers. It should be acknowledged that the TEC devices do indeed incur additional power overheads compared to traditional heatsink based coolers. On the other hand, TEC devices do provide significantly higher cooling potential compared to traditional heat-sinks, hence optimizations can be carried out with the objective of maximizing the chip's performance with respect to the power consumption of the cooling system. If the total power consumption (chip power plus cooling power) is a concern, we want to mention that the proposed technique is orthogonal to the other online chip power reduction/management techniques such as dynamic voltage and frequency scaling (DVFS), and task scheduling and migration which can be leveraged to optimize the overall power consumption of the system (chip plus cooling power) as demonstrated by many existing works. For example, in [6] Amrouch et al. propose a thermal management scheme for Neural processing units that involves combined control of a TEC-based cooler as well as modulating the chip's frequency, and inference precision in order to find the optimal setting to maximize the chip's throughput, while minimizing the overall power consumption of the system. Similarly, in [7] Lundquist et al. proposed using a TEC cooler and heatsink with a fan to dynamically control the processor cooling. Here, the TEC voltage and fan speeds are controlled adaptively to sustain the chip temperature under a predetermined threshold in order to avoid performance degradation while minimizing the power consumption of the system. Such combined control can be used in conjunction with the proposed TEC-Array in order to realize further optimization to reduce the power consumption of the entire system while maximizing performance relative to the consumed power.

The article is organized as follows. Section 2 reviews the existing work in real-time full-chip temperature and power estimation, as well as the existing simplified 1D energy equilibrium model for TEC. Section 3 summarizes one of the existing thermal and power-map estimation techniques which can be used to estimate the full-chip power density maps in real-time. This information is in turn used to dynamically modulate the voltage levels of the TEC-Array. Section 4 introduces the architecture and the power-density based control flow of the proposed TEC-Array. Section 4.3 describes the details of the modeling and simulation setup of the proposed TEC-Array on a commercial multiphysics simulator. Section 5 presents the simulation results for the TEC-Array as well as comparisons with the traditional TEC cooling and passive heatsink cooling. Section 6 concludes this article.

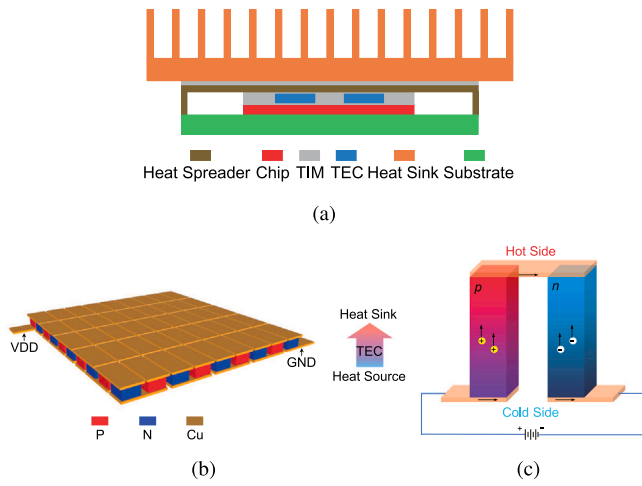


Fig. 1. (a) The side view of the chip package. (b) 3D view of thin-film TEC devices. (c) Peltier effect for an N-P pair in the TEC devices.

2. Related works

2.1. Thermoelectric effects in a nutshell

Thermoelectric cooling uses the Peltier effect to create a heat flux at the junction of two different types of materials. A Peltier cooler, heater, or thermoelectric heat pump is a solid-state active heat pump which transfers heat from one side of the device to the other, with consumption of electrical energy, depending on the direction of the current.

Thermoelectric coolers (TEC) are based on the principles of the thermoelectric (TE) effect where heat flux is generated at the intersection of two materials with different electron densities, as shown in Fig. 1. TECs have two sides, namely the cold side and the hot side, where heat is transferred from the cold side to the hot side along the direction of DC current flow. In cooling applications, the hot side of the TEC is typically attached to a heatsink in order to keep it at ambient temperature allowing the cold side to go below ambient temperature. While the basic principles of the TE effect is intuitive, the physics governing this phenomenon is complex and therefore challenging to model and simulate.

The TE effect is an energy conversion phenomenon that consists of the Seebeck, Peltier, Thomson and Joule heating effects. We assume that the two ends of the TE material are excited by an ideal voltage source, which is capable of supplying and maintaining the same voltage. Under these conditions, the Seebeck effect cannot impact the voltage and current of TE materials. Therefore, we do not consider the Seebeck effect. Thomson effect is typically ignored for the constant Seebeck coefficient. Hence, the dominant factors for modeling the TE effect are as follows.

First, under the Peltier effect, the current density \mathbf{J} (A/m²) driven by the voltage across two ends produces heat flux \mathbf{q}_p moving from cold side to hot side, which is expressed as

$$\mathbf{q}_p = P\mathbf{J} \quad (1)$$

where $P = S \times T$ is the Peltier coefficient, S is the Seebeck coefficient (V/K), T is the temperature, and heat flux \mathbf{q}_p is a flow of heat energy per unit area per unit time (W/m²).

Second, Fourier's law describes that the heat flux flows from high temperature to low temperature, which is expressed as

$$\mathbf{q}_{\text{Fourier}} = -\kappa \nabla T \quad (2)$$

where κ is the thermal conductivity, and ∇ is the gradient vector operator.

Last, the power per unit volume (W/m³) due to Joule heating effect is calculated by

$$g_{\text{Jh}} = \mathbf{J} \cdot \mathbf{E} = \frac{\|\mathbf{J}\|^2}{\sigma} \quad (3)$$

where σ is electrical conductivity, \mathbf{E} is the electric field, “ \cdot ” is the dot product operator, $\|\mathbf{J}\|^2$ is the dot product of \mathbf{J} and itself, and \mathbf{J} is equal to the product of σ and \mathbf{E} . To model the TE effect, two models were primarily investigated in existing work [8–10]. One is a simplified 1D energy equilibrium model and the other is a more complicated 3D coupled multiphysics model.

Simplified one-dimensional energy equilibrium model was first proposed to characterize the cooling heat flux of the TEC device based on energy balance on the cold side. The model accounts for the Peltier, Joule heating and heat transfer effects. Here cooling heat flux, which is a key parameter that represents the cooling capacity of TEC, is given by [8,10,11]

$$q_c = q_p - q_{\text{Jh}} - q_{\text{Fourier}} = ST_c J - \frac{1}{2} \frac{J^2 L}{\sigma} - \frac{\kappa}{L} (T_h - T_c) \quad (4)$$

where T_c and T_h are the temperatures at the cold and hot side, respectively. L is thickness of TEC leg. q_{Fourier} is essentially the heat flux from high temperature to low temperature, which is Fourier's law expressed by

$$q_{\text{Fourier}} = -\kappa \nabla T \approx -\kappa \frac{\Delta T}{\Delta x} = \kappa \frac{T_h - T_c}{L} \quad (5)$$

The effective heat flux caused by Joule heating at cold side is calculated by

$$q_{\text{Jh}} \approx \frac{1}{2} \frac{Q_{\text{Jh}}}{A} = \frac{1}{2} \frac{g_{\text{Jh}} \mathcal{V}}{A} = \frac{1}{2} \frac{J^2 L}{\sigma} \quad (6)$$

where A and \mathcal{V} are cross-sectional area and volume of the TEC leg, respectively, and g_{Jh} is defined in Eq. (3); and Q_{Jh} is the heat produced by Joule heating.

However, this model is overly simplified with approximated expressions to describe the Peltier, Joule heating and Fourier heat transfer effects. The results produced by the model are inaccurate when the thermal gradient is large because the formula is too simple to describe these complicated phenomena. An accurate TEC model needs to be developed to consider spatial temperature as we will discuss in Section 4.3.

2.2. Runtime thermal-map estimation

As we will discuss in the next section, intelligently controlling the proposed TEC-Array will require real-time knowledge of the spatial power distribution (power-density map) across the chip. The power-density map will be used to determine the amount of cooling needed in different areas of the chip. For example, the areas of the chip consuming more power will need more cooling than the areas of the chip consuming less power.

Real-time estimation of the power-density map, however, is a challenging task. Alternatively, if the real-time spatial thermal-map (or heat-map) across the chip's surface area can be determined, then the power-density map can be easily estimated using a thermal-to-power transformation method (to be discussed in the next subsection). Real-time full-chip thermal-map estimation has been well studied in the past. In general, three strategies have been explored. One strategy is based on the placement of on-chip temperature sensors and reconstruction of the full-chip thermal-maps from the sensor data [12–14]. The second strategy is to estimate the full-chip thermal-maps using a thermal model. These thermal models can be built using the so-called “bottom-up” approaches such as HotSpot [15] based simplified finite difference methods, finite element methods [16], equivalent thermal RC networks [17], and recently proposed behavioral thermal models based on the matrix pencil method [18] and the subspace identification method [19,20]. The third, which is our own recent work,

leverages the recent advancements in machine learning to propose a completely post-silicon approach to real-time thermal-map estimation of existing commercial off-the-shelf processors [21]. Here, we exploit the correlation between the processor's temperature and system-level variables such as frequency, voltage, and other performance metrics that have been shown to be effective in the past. Here, the correlation between the transient behavior of high-level performance metrics and the processor's thermal profile is automatically learned via training a long-short-term-memory-based recurrent-neural-network.

Ultimately, adapting one of the aforementioned strategies to derive a thermal model that can be deployed in the processor enables real-time estimation of the full chip thermal-maps while the processor is in use.

2.3. Runtime power-map estimation

Real-time estimation of the spatial power-maps (or power-density maps) is essential for controlling the proposed TEC-Array. And real-time estimated thermal-maps are highly valuable since they can be used to estimate the spatial power-density maps using a thermal-to-power transformation method [22,23]. In [22], the authors developed a general blind power identification, called *BPI* method for power-maps from thermal measurements. Whereas [23] proposed a first-principle based efficient power-map estimation utilizing thermal measurements. In essence, the latter work simply calculates spatial power density map using the steady state thermal diffusion equation shown in Eq. (7).

$$-\kappa \nabla^2 T = g_T \quad (7)$$

Here T is the spatial thermal-map (in K), κ is the thermal conductivity, g is the volumetric power-density (W m^{-3}), which can be transformed to spatial power-density map by multiplying the silicon thickness factor, and ∇^2 is the Laplace operator. Hence, if T can be estimated using one of the methods mentioned in Section 2.2, then g_T can be calculated using Eq. (7). Note, κ is the effective thermal conductivity of the processor die. This can either be calculated by the manufacturer, with the knowledge of the exact material composition of the die, or experimentally estimated by a third-party using infrared measurements and using Eq. (7) as demonstrated in [23]. For example, Fig. 2 shows the spatial thermal-map (T) of an Intel i7-8650U (Fig. 2(a)), in one thermal steady state, and the corresponding power-map (Fig. 2(b)) calculated using Eq. (7). Here, the effective thermal conductivity was calculated to be $\kappa = 174 \text{ W m}^{-1} \text{ K}^{-1}$ in [23].

3. Power-map estimation overview

3.1. Power-map estimation summary

Accurately estimating the spatial power-maps of the chip is important for effectively controlling the proposed TEC Array. In this article, we present simulation results from a commercial physics simulator, where the power density maps are known. However, in reality, when the processor is operational, determining the power density maps in real-time is not a trivial task. Alternatively, one can estimate the full-chip thermal-maps and use the thermal-map to power-map conversion methodology to estimate the power-maps as discussed in Section 2.3.

Several methods exist for estimating the full chip thermal-maps of a microprocessor while it is under load (Section 2.2). Here, we summarize our own machine learning based framework published in [21]. The approach in [21] is based on training a recurrent-neural-network model with thermal data recorded directly from a processor. This approach requires two critical pieces of data that have to be collected in synchronous while the processor is under load. First is the time-stamped thermal-maps collected from the chip at a steady sampling rate (e.g. 60 Hz), and the second is a suite of high-level performance metrics captured at the same time instances as the temperature data. This requires an advanced IR thermography setup that records lucid thermal-maps of the processor under test. This setup will be discussed

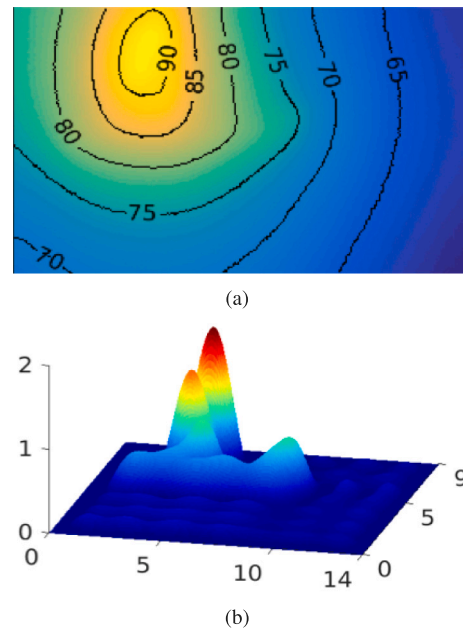


Fig. 2. Thermal-map to power-map conversion: (a) Experiment measured thermal-map (T); (b) Estimated power-density map (g_T) in 3D view [23].

in detail in the next subsection. At the same time, a suite of high-level performance metrics must be recorded in synchronous with the capture rate of the IR camera.

Once sufficient data is acquired, a specialized Recurrent-Neural-Network (RNN) architecture called Long-Short-Term-Memory (LSTM) network can be employed to train the online thermal model. Fig. 3(a) illustrates the data preparation and training phase of the approach in [21]. Once trained, the thermal model will use the processor's real-time performance metrics M_t as inputs to estimate the thermal-maps T_t . The model estimates the first 36 most-dominant discrete cosine transformation (DCT) coefficients f_i of the thermal-maps and then recover the thermal-maps via inverse DCT transformation. Deployment flow of full-chip thermal-maps in real-time is shown in Fig. 3(b). The thermal-maps can then be converted to power-density maps using Eq. (7) as shown in Section 2.3. RNN-based model structure is illustrated as Fig. 3(c). The network structure we deployed is a two-layer network with k ($k = 86$) nodes and 60 time-steps of feedback in the LSTM layer, and 36 nodes in the linear output layer. The inference time is only 0.41 ms per inference and memory overhead for storing the network weights is only 557 Kb. Training process is not done online, but done on another computing system hence the training effort does not impact the online control performance. For the technical details on this approach, please refer to [21].

We remark that from a practical application perspective, the machine learning based full-chip power density map estimation approach discussed above can be trained using measured data from a thermal IR imaging system (shown in Fig. 4) or other numerical simulation methods, during the design time, considering practical cooling, package and thermal conditions. Such modeling only needs to be done once during the design time or post-silicon phase and the resulting models can be applied to different cooling conditions (with heat sink, without heat sink, etc.) as demonstrated recently in [23]. Different cooling conditions will change the temperature of the chip, but they do not change the locations and shapes of hot spots, and thus (relative) power density values of those hot spots. As a result, the power density map (relative distribution of power across the die area) will be similar under all cooling conditions as it fundamentally represents the spatial power consumption of silicon with respect to the chip's utilization. For TEC-based control, such relatively valued power density map will be

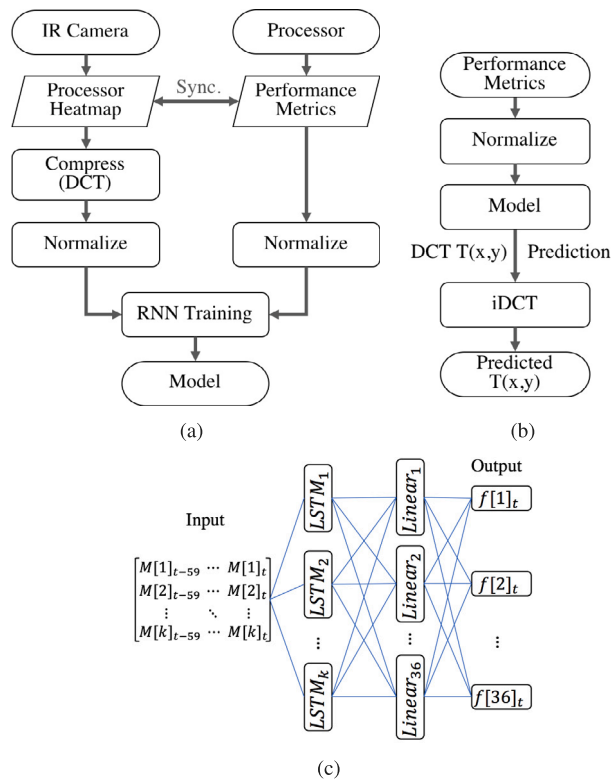


Fig. 3. Full chip thermal model: (a) Training; (b) Deployment; (c) RNN model structure [21].

sufficient as it provides the **key differential values** for different TEC devices.

3.2. IR thermography setup

As mentioned, the aforementioned heat-map estimation approach [21] relies on accurate time-series data of the full-chip heat-maps measured directly from the processor under test. Externally measuring this information is challenging, especially when the processor is under load as it requires the processor to be operated without the traditional front-mounted cooling systems (i.e. heat-sink). To address this issue, we have built an advanced infrared (IR) thermography setup shown in Fig. 4. This setup is based on the thermal imaging setup proposed in [24]. The setup features a thermo-electric device mounted on the PCB directly beneath the processor allowing it to be cooled from underneath: leaving the front side fully exposed to the IR camera without any interference layer in-between. Unlike the traditional oil-based front-cooling methods, no de-embedding [25] is required in this setup. Instead, a programmable power supply is used to control the heat flow through the thermo-electric device so that the operating conditions can be matched to the baseline cooling unit using a heat-sink [24].

A detailed description of the IR thermography setup is as follows. The thermal camera used in this setup is a FLIR A325sc (16-bit 320 × 240px, 60 Hz). The camera is rated for the temperature range of 0 °C to 328 °C, and a spectral range of 7.5 μm to 13 μm. In order to enhance the spatial resolution of the camera, a microscope lens is used providing a spatial resolution of 50 μm/px. The IR camera has an internal waveform generator that outputs a square waveform in synchronous with the capture frequency of the camera. An I/O device is then employed to interface the waveform generator to the processor under test so that the performance metrics (recorded on the processor) can be synchronized with the thermal data recorded by the IR camera. The processor under test is an Intel i7-8650U, which has 4 cores

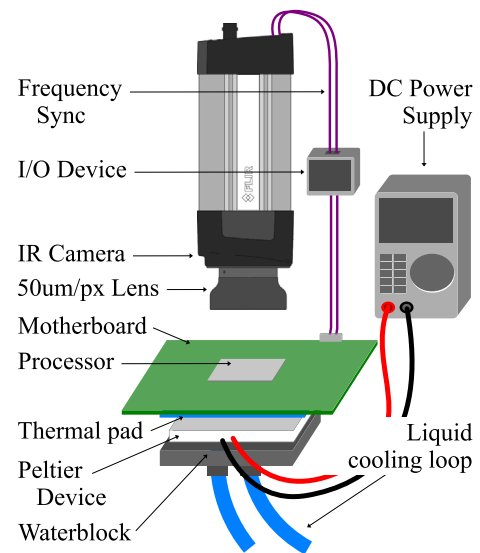


Fig. 4. Illustration of our IR Thermography Setup.

with 2 threads per core. Mounted on the PCB directly underneath the processor is the thermo-electric-based cooling system which includes a Peltier device powered by a programmable power source. A liquid cooling loop is used to cool the hot side of the Peltier device and ensure proper operation.

4. Proposed TEC-Array control framework

4.1. TEC-Array architecture

The proposed TEC-Array consists of a number of small TEC modules arranged to form a 2D array. In our case, we will consider a 24 unit array (4 × 6 TEC modules) as shown in Fig. 5(a). The cold side of the TEC-Array is to be affixed over a microprocessor (as shown in Fig. 5(b)), with a traditional heatsink (or liquid cooling unit) affixed to the hot-side (as shown in Fig. 5(c)). The TEC-Array will “pump” the heat from the surface of the chip to its hot side, which is in turn extracted by the heatsink and ultimately transferred to ambient air through natural or forced convection.

The proposed TEC-Array is designed such that the rate of heat transfer through each TEC module in the array can be controlled by varying the amplitude of the voltage applied to each module. Such granularity in control can be possible because the proposed TEC-Array is affixed over the processor’s heat-spreader (outside the chip packaging). For methods where the TEC modules are integrated into the packaging, this level of granular control is not possible due to design constraints [5].

The proposed setup enables the ability to precisely control the cooling across the surface area of the processor die in a non-homogeneous manner. Meaning, the areas of the chip consuming more power (hence generating more heat) can be cooled more intensely than the areas of the chip that are consuming less power (hence generating less heat). This level of control significantly aids in reducing the thermal gradients across the chip’s surface (difference in temperature in one area of the chip vs. another), consequently normalizing the rate of aging effects across the chip’s surface as previously mentioned. This is in contrast with the traditional TEC-based coolers where a single large TEC module is used, consequently limiting spatial control. We note that the dimension of the TEC-Array is not particularly chosen. This dimension is used for the demonstration of the proposed cooling technique, other configurations should work for this work as well. Ideally, a smaller size of the individual TEC module, i.e. the larger dimension of the TEC-Array would lead to finer thermal control in general.

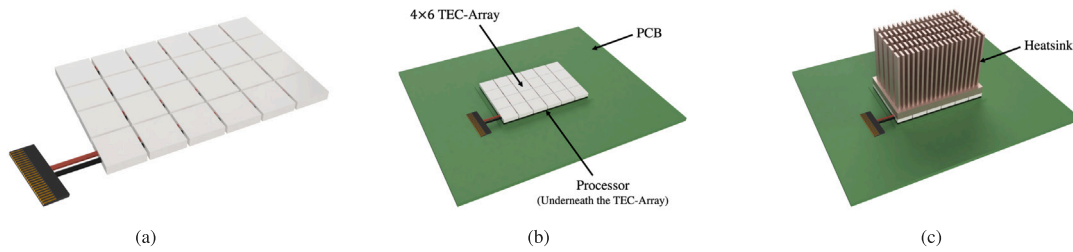


Fig. 5. (a) TEC-Array. (b) TEC-Array affixed over a processor. (c) Heatsink affixed over the TEC-Array.

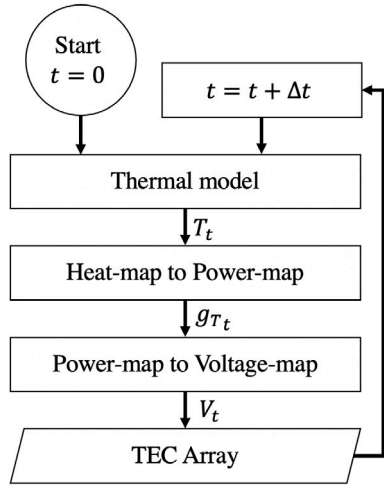


Fig. 6. Proposed TEC-Array control flow.

4.2. TEC-Array control flow

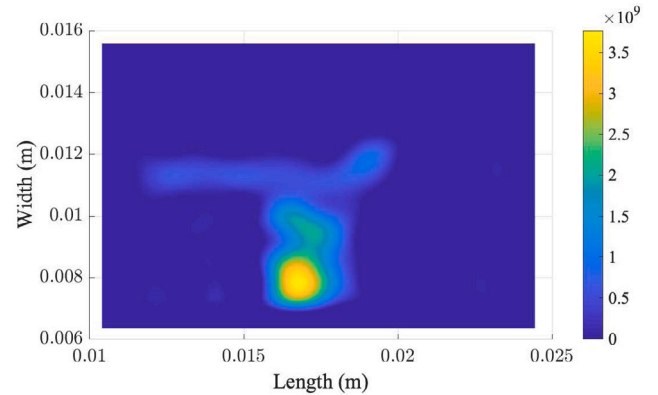
Controlling the TEC-Array can be done by generating a discrete voltage map (V) with the same dimensionality as the TEC-Array (4×6 in our case), where each index of V denotes the voltage that is to be applied to the corresponding TEC-module. For effective thermal control, at each time-step t , V_t should be proportional to the power-density map (gT_t).

To this end, we propose the TEC-Array control flow illustrated in Fig. 6. Here, at each time-step (t), the thermal-map of the chip (T_t) is estimated using a thermal model (Section 2.2), which is then used to calculate the power-density map gT_t . Once gT_t is determined, V_t can be calculated as per Eq. (8).

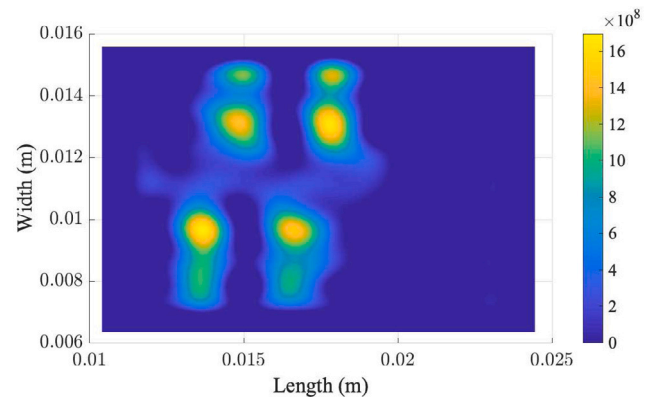
$$V = \text{resize}\left(\frac{gT_t - g_{min}}{g_{max} - g_{min}} \times V_{max}\right) \quad (8)$$

Here, g_{min} and g_{max} denote the minimum and maximum power density observed across all time-steps, and V_{max} is the maximum usable voltage rating of the TEC modules. The resize function is used to reduce the dimensionality of V to match the TEC-Array (4×6 elements).

Eq. (8) generates V such that the TEC modules located in the areas of the chip with higher power density are assigned higher voltages than the TEC modules located in the areas of the chip with lower power density. For example, Fig. 7 shows two power-density maps (gT_t) of an Intel i7-8650U quad-core processor. Fig. 7(a) shows gT_t when the processor is executing a single-threaded workload (only one core active), whereas Fig. 7(b) shows gT_t when the processor is executing a multi-threaded workload (all four cores active). Here we see a single hot spot (location of high power density) in Fig. 7(a) and four distinct hot spots in Fig. 7(b), with significant portions of the chip under lower power-density in both cases. With the proposed TEC-Array and the aforementioned control scheme, the TEC modules located over the



(a)



(b)

Fig. 7. Two power-maps (W/m^2) of an Intel i7-8650U: (a) Single threaded workload; (b) Multi threaded workload.

hot spots will be assigned higher voltages than the ones located over non-hot-spot as shown in Fig. 8 (voltage-maps V).

The size of hot spots or the type of the system (homogeneous/heterogeneous) do not affect the control algorithm since the voltage map of TEC-Array is calculated from the package level full-chip power-density map (Eq. (8)). Power peaks shown on the power-density map are defined as hot spots in our case, which essentially generate intense heat and cause high temperatures, usually the same as temperature hot spots. Hence, TEC modules within the array are turned on and controlled as long as the hot spot locations are mapped to the TEC modules according to the power-density map. In contrast to cooling the entire chip at the same rate, this method allows targeted cooling of hot spots which in turn results in significant savings in the energy consumption of the cooling system as we will demonstrate in Section 5.

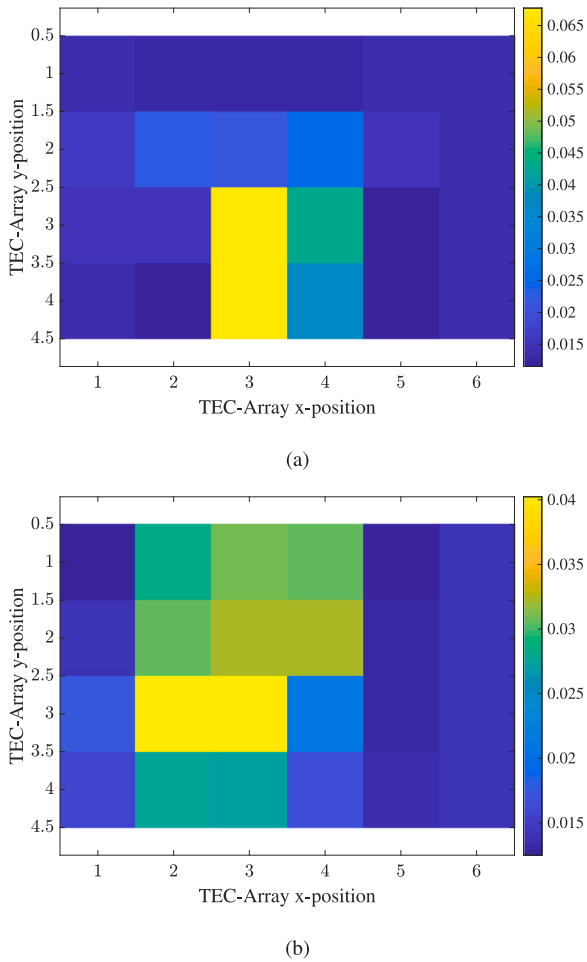


Fig. 8. Voltage maps ($V/position$) generated using Eq. (8): (a) From g_T of Fig. 7(a); (b) From g_T of Fig. 7(b).

4.3. 3D multiphysics model for TEC devices

COMSOL Multiphysics, a commercial finite-element-based physics simulation software, was used to simulate the proposed cooling system. The setup illustrated in Fig. 5 was configured in COMSOL using the built-in 3D modeling graphical user interface. In terms of the material properties, COMSOL's default materials for the printed circuit board (PCB), silicon (for the processor) and copper (for the heatsink) were used. The default conductive boundary was assumed between the PCB, processor, TEC-Array and heatsink, with natural convection between the heatsink and static ambient air. The ambient temperature was set at 20 °C. The processor is set as the heat source with spatial power-density set as per g_T for each time-step t . The 3D geometry of the processor's silicon die is $10 \times 15 \times 0.5$ mm. The 3D geometry of the 4×6 TEC-Array was designed with the TEC leg height of 0.5 mm and the side length of 2.5 mm for each individual module. We remark that the routing and design of control wires inside the TEC-Array is not the focus of this work and hence not illustrated. The physics of the TEC modules are configured as described below.

A 3D coupled multiphysics model is employed to characterize the behavior of TEC devices accurately [9,10,26–28]. Considering thermal and electrical fields, the thermoelectric analysis is an electric-thermal co-simulation, meaning the two fields are solved simultaneously. Therefore, this model is called a 3D coupled multiphysics model, where Peltier effects build the relationship between thermal and electrical fields. Specifically, for electric field, we have

$$\mathbf{E} = -\nabla\phi \quad (9)$$

where ϕ is the potential in the TEC. Then, the current continuity equation is formulated as

$$\nabla \cdot \mathbf{J} = \nabla \cdot (-\sigma \nabla \phi) = 0 \quad (10)$$

One side is set to ground boundary condition $\phi(0) = 0$ and the other side is excited by an ideal voltage source $\phi(L) = V$.

For thermal field, we integrate the Peltier effect into the heat flux

$$\mathbf{q} = \mathbf{q}_{\text{Fourier}} + \mathbf{q}_P = -\kappa \nabla T + P\mathbf{J} \quad (11)$$

Then, considering the Joule heating, the transient heat conduction equation is rewritten as

$$\nabla \cdot \mathbf{q} = \nabla \cdot (-\kappa \nabla T + ST\mathbf{J}) = \frac{\|\mathbf{J}\|^2}{\sigma} \quad (12)$$

The heat sink is placed at one side to dissipate the heat, and we set this side to a constant temperature boundary condition

$$T(0) = T_0 \quad (13)$$

where T_0 is the reference temperature (hot side for TEC device). Another side is used to remove the heat from the chip (the cold side from TEC device), and boundary condition is set to the heat flux boundary condition

$$-\mathbf{n} \cdot \mathbf{q} = -\mathbf{n} \cdot (-\kappa \nabla T + ST\mathbf{J}) = q_c \quad (14)$$

where \mathbf{n} is the unit outward normal vector of the boundary surface, q_c is the cooling heat flux.

Based on a partial differential equation, the 3D coupled multiphysics model considers thermal gradients and can simulate the heat conduction more accurately compared with a simplified 1D energy equilibrium model. In this work, the Mathematics Module in COMSOL Multiphysics is used to numerically solve the 3D coupled multiphysics model based on Eq. (10) and (12). In COMSOL Multiphysics, we apply those equations as the 3D coupled governing equations to the electric field and heat transfer mechanism of each individual TEC module within the array, as shown in Fig. 5.

5. Results and discussions

A total of 190 power-density maps calculated using thermal-maps of an Intel i7-8650U test chip [23] are pipelined as the dynamically varying heat sources for the thermal-electric co-simulation for this work. At each time-step, one power-density map is applied to the chip's top surface as the heat source. For comparison purposes, three scenarios were considered. First is the processor cooled using the proposed TEC-Array and the control scheme from Section 4. For the second scenario, we consider the traditional TEC cooling approach where a single large TEC module is used as opposed to the proposed array of TEC modules. In both cases, the same heatsink (Fig. 5(c)) was affixed over the TEC modules. For the third scenario, we remove the TEC layer and affix the heatsink directly over the processor to mimic standard passive heatsink cooling. All other parameters are kept exactly the same for all three scenarios.

All three scenarios were simulated for a total of 190 time-steps, using the aforementioned 190 power-density maps of the Intel i7-8650U. For the results, shown in Fig. 9, we present three metrics (T_{max} , ΔT , and $P_{Array}/P_{Trad.}$) to characterize the benefits of the new cooling technique.

Fig. 9(a) shows the max temperature (T_{max}), observed across the area of the chip. The max temperature is crucial because it directly impacts the performance of the processor. On-board dynamic voltage and frequency scaling (DVFS) controllers dynamically throttle the processor's voltage and frequency according to max temperature. On Intel chips, the DVFS controller allows the frequency to exceed the chip's rated frequency (i.e. 4 GHz) when the max temperature is below a threshold (called Intel thermal velocity boost [29]); the DVFS controller will also scale the voltage and frequency down significantly if T_{max} is

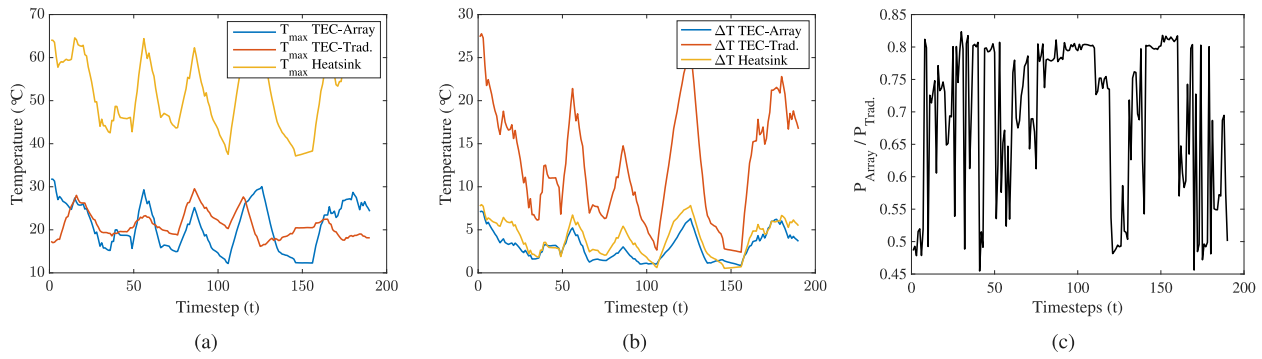


Fig. 9. (a) Max temperature. (b) Spatial temperature range. (c) Relative power consumption.

high in order to prevent the temperature from exceeding the maximum junction temperature of the given technology node [29]. Hence, T_{max} is an important factor that affects the processor's performance. From Fig. 9(a), we see that the processor cooled using the TEC-Array can maintain T_{max} similar to the traditional TEC system (T_{max} TEC-Trad.) where a single large TEC module is used to cool the entire surface of the chip at the same rate (also voltage modulated based on peak power). Note, both the TEC-Array and TEC-Trad outperform the heatsink-only cooling as expected. In our experiments, the max temperature over the hot spot controlled with TEC-Array and traditional single TEC module (TEC-Trad) are both around 30°C as shown in Fig. 9(a). To achieve finer control, the end user can apply a smaller module size, i.e. larger array dimension, and tune the voltage modulation coefficients to control the temperature over the hot spots based on their needs.

The second metric presented in Fig. 9(b) shows ΔT . It is the difference between the maximum (T_{max}) and minimum temperature (T_{min}) across the die area. ΔT represents the spatial thermal gradients, which have huge impacts on the chip reliability due to effects such as thermal cycling [30], thermo-migration [31,32] and timing violation due to uneven stressing or aging of the chip. For better visualization, we present Fig. 10 that shows the simulated thermal-map (spatial temperature distribution across the die area) at time-step $t = 53$. Fig. 10(a) illustrates the thermal-map under TEC-Array cooling, while Fig. 10(b) shows the thermal-map under TEC-Trad cooling. Here we can see that, because the proposed TEC-Array enables spatially variable cooling, where the hot spots are cooled more intensely than the non-hot-spots, ΔT is relatively low. This is in contrast with TEC-Trad which can have much larger temperature gradients. As a result, the proposed TEC-Array cooling actually can allow the chip to run in higher performance modes via DVFS thanks to the increased thermal design power (TDP) as shown in Fig. 9(a) and at the same time maintaining or even increasing chip design lifetime due to the reduced thermal gradients across the chip. This can significantly reduce the thermal gradient induced stress such as thermal cycling for devices, thermo-migration for interconnects, and unbalanced thermal induced aging [1].

In addition, we observe that TEC-Trad lead to less efficient power consumption for TEC devices as the non-hot-spots are cooled as intensely as the hot spots, resulting in a high ΔT and consequently incurring a significant wastage of power used by the cooling system. This can be seen in Fig. 9(c) where the power consumption of the TEC-Array is shown in proportion to the power consumption of the TEC-Trad (P_{Array}/P_{Trad}). The data shows that the cumulative energy consumption of the TEC-Array is 66.23% of TEC-Trad across the 190 time-steps. Hence, over the long term, implementing the proposed TEC-Array will result in significant energy savings, compared to TEC-Trad, while yielding the same benefits (i.e. T_{max}).

Furthermore, In our experiment, the computations shown in Figs. 6 and 3(b) are done on an Intel i7-8650U test processor, which has 4 cores with 2 threads per core. The overall computation time for

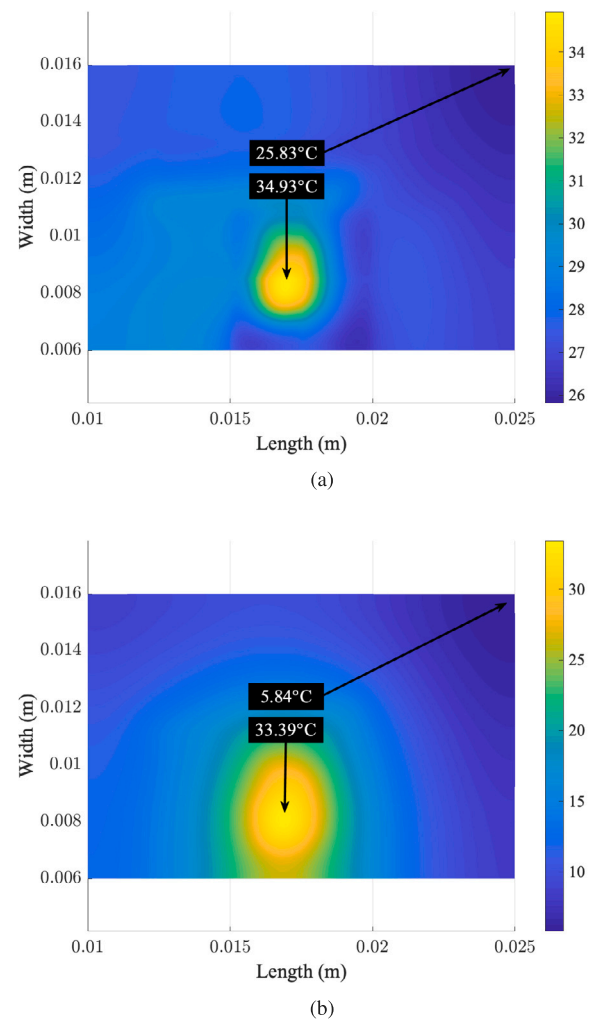


Fig. 10. 2D Thermal-map (temperature in °C) at time-step 53 under (a) TEC-Array and (b) TEC-Trad.

obtaining the TEC-Array's control voltage map regarding one time-step is no more than 8 ms, which is considerably shorter than the thermal control cycle (time-step), i.e. 16 ms (60 Hz) in our case. Hence, the computation overhead is reasonably small for practical use.

6. Conclusion

In this paper, we presented a general hot-spot-aware thermo-electric-based active cooling solution for multicore processors. The

proposed cooling system involves a 2D array of TEC modules, called TEC-Array, that enables the ability to spatially vary the rate of cooling across the surface of the processor die. The TEC-Array is controlled using a power-density based control scheme that allows the hot-spots to be cooled more intensely than non-hot-spots. Moreover, the general nature of the proposed system allows it to be easily adopted for any commercially available processor. The numerical results on an Intel quad-core chip show that the proposed TEC-Array cooling can substantially reduce the peak temperatures compared to the traditional passive heat sink cooling method. Furthermore, compared to the existing single TEC module based cooling method, the proposed method can reduce both TEC power and temperature gradients across the chip under the same maximum temperature constraints. As a result, the new TEC-Array cooling can enable more aggressive chip performance with increased thermal design power (TDP) while maintaining the chip design lifetime. The proposed TEC power management is also orthogonal to other power management techniques, such as DVFS, which can be used to reduce the total power of a chip.

CRedit authorship contribution statement

Jinwei Zhang: Writing – original draft, Data curation, Methodology, Investigation. **Sheriff Sadiqbacha:** Conceptualization, Methodology, Investigation, Writing – original draft, Visualization. **Liang Chen:** Formal analysis. **Cuong Thi:** Software. **Sachin Sachdeva:** Software. **Hussam Amrouch:** Supervision. **Sheldon X.-D. Tan:** Supervision, Writing – review & editing.

Declaration of competing interest

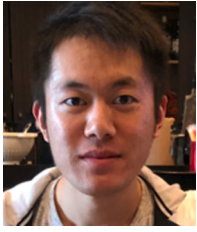
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

- [1] S.X.-D. Tan, M. Tahoori, T. Kim, S. Wang, Z. Sun, S. Kiamehr, *VLSI Systems Long-Term Reliability – Modeling, Simulation and Optimization*, Springer Publishing, 2019.
- [2] M.J. Dousti, M. Pedram, Power-aware deployment and control of forced-convection and thermoelectric coolers, in: 2014 51st ACM/EDAC/IEEE Design Automation Conference (DAC), 2014, pp. 1–6, <http://dx.doi.org/10.1145/2593069.2593186>.
- [3] S. Jayakumar, S. Reda, Making sense of thermoelectrics for processor thermal management and energy harvesting, in: 2015 IEEE/ACM International Symposium on Low Power Electronics and Design (ISLPED), 2015, pp. 31–36, <http://dx.doi.org/10.1109/ISLPED.2015.7273486>.
- [4] J. Long, S. Ogrenic Memik, M. Grayson, Optimization of an on-chip active cooling system based on thin-film thermoelectric coolers, in: 2010 Design, Automation Test in Europe Conference Exhibition (DATE 2010), 2010, pp. 117–122, <http://dx.doi.org/10.1109/DATE.2010.5457225>.
- [5] M.J. Dousti, M. Pedram, Power-efficient control of thermoelectric coolers considering distributed hot spots, in: 2015 Design, Automation Test in Europe Conference Exhibition (DATE), 2015, pp. 966–971, <http://dx.doi.org/10.7873/DATE.2015.0517>.
- [6] H. Amrouch, G. Zervakis, S. Salamin, H. Kattan, I. Anagnostopoulos, J. Henkel, Npu thermal management, *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.* 39 (11) (2020) 3842–3855, <http://dx.doi.org/10.1109/TCAD.2020.3012753>.
- [7] C. Lundquist, V. Carey, Microprocessor-based adaptive thermal control for an air-cooled computer cpu module, in: Seventeenth Annual IEEE Semiconductor Thermal Measurement and Management Symposium (Cat. No. 01CH37189), 2001, pp. 168–173, <http://dx.doi.org/10.1109/STHERM.2001.915174>.
- [8] D.M. Rowe, *Thermoelectrics Handbook: Macro To Nano*, CRC Press, Boca Raton, FL, USA, 2005.
- [9] E.E. Antonova, D.C. Looman, Finite elements for thermoelectric device analysis in ansys, in: Proc. 24th Int. Conf. Thermoelectrics, ICT, 2005, pp. 215–218.
- [10] R.A. Kishore, A. Nozariasbmarz, B. Poudel, M. Sanghadasa, S. Priya, Ultra-high performance wearable thermoelectric coolers with less materials, *Nature Commun.* 10 (1) (2019) 1–13.
- [11] I. Chowdhury, R. Prasher, K. Lofgreen, G. Chrysler, S. Narasimhan, R. Mahajan, D. Koester, R. Alley, R. Venkatasubramanian, On-chip cooling by superlattice-based thin-film thermoelectrics, *Nat. Nanotechnol.* 4 (4) (2009) 235–238.
- [12] R. Cochran, S. Reda, Spectral techniques for high-resolution thermal characterization with limited sensor data, in: Proceedings of the 46th Annual Design Automation Conference, DAC '09, ACM, New York, NY, USA, 2009, pp. 478–483, <http://dx.doi.org/10.1145/1629911.1630037>.
- [13] J. Ranieri, A. Vincenzi, A. Chebira, D. Atienza, M. Vetterli, Eigenmaps: Algorithms for optimal thermal maps extraction and sensor placement on multicore processors, in: Proceedings of the 49th Annual Design Automation Conference, DAC '12, ACM, New York, NY, USA, 2012, pp. 636–641, <http://dx.doi.org/10.1145/2228360.2228475>.
- [14] X. Li, X. Li, W. Jiang, W. Zhou, Optimising thermal sensor placement and thermal maps reconstruction for microprocessors using simulated annealing algorithm based on pca IET circuits, *Dev. Syst.* 10 (6) (2016) 463–472, <http://dx.doi.org/10.1049/iet-cds.2016.0201>.
- [15] W. Huang, S. Ghosh, S. Velusamy, K. Sankaranarayanan, K. Skadron, M.R. Stan, HotSpot: A compact thermal modeling methodology for early-stage VLSI design, *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* 14 (5) (2006) 501–513.
- [16] S.P. Gurrum, Y.K. Joshi, W.P. King, K. Ramakrishna, M. Gall, A compact approach to on-chip interconnect heat conduction modeling using the finite element method, *J. Electron. Packag.* 130 (2008) 031001.1–031001.8.
- [17] Y.C. Gerstenmaier, G. Wachutka, Rigorous model and network for transient thermal problems, *Microelectron. J.* 33 (2002) 719–725.
- [18] D. Li, S.X.-D. Tan, E.H. Pacheco, M. Tirumala, Parameterized architecture-level dynamic thermal models for multicore microprocessors, *ACM Trans. Des. Autom. Electron. Syst.* 15 (2) (2010) 1–22, <http://doi.acm.org/10.1145/1698759.1698766>.
- [19] T. Eguia, S.X.-D. Tan, R. Shen, D. Li, E.H. Pacheco, M. Tirumala, L. Wang, General parameterized thermal modeling for high-performance microprocessor design, *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* (2011).
- [20] Z. Liu, S.X.-D. Tan, H. Wang, Y. Hua, A. Gupta, Compact thermal modeling for packaged microprocessor design with practical power maps, integration, *VLSI J.* 47 (1) (2014) in press, online access: <http://www.sciencedirect.com/science/article/pii/S0167926013000412>.
- [21] S. Sadiqbacha, J. Zhang, H. Amrouch, S.X.-D. Tan, Real-time full-chip thermal tracking: A post-silicon, machine learning perspective, *IEEE Trans. Comput.* (2021) <http://dx.doi.org/10.1109/TC.2021.3086112>.
- [22] S. Reda, K. Dev, A. Belouchrani, Blind identification of thermal models and power sources from thermal measurements, *IEEE Sens. J.* 18 (2) (2018) 680–691.
- [23] J. Zhang, S. Sadiqbacha, M. O'Dea, H. Amrouch, S.X.-D. Tan, Full-chip power density and thermal map characterization for commercial microprocessors under heat sink cooling, *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.* (2021) 1, <http://dx.doi.org/10.1109/TCAD.2021.3088081>.
- [24] H. Amrouch, J. Henkel, Lucid infrared thermography of thermally-constrained processors, in: 2015 IEEE/ACM International Symposium on Low Power Electronics and Design (ISLPED), 2015, pp. 347–352.
- [25] K. Dev, A.N. Nowroz, S. Reda, Power mapping and modeling of multi-core processors, in: International Symposium on Low Power Electronics and Design, ISLPED, 2013, pp. 39–44.
- [26] M. Jaegle, Multiphysics simulation of thermoelectric system—modeling of peltier cooling and thermoelectric generator, in: Proc. COMSOL Conf., 2008, pp. 4–6.
- [27] C. Goupil, *Continuum Theory and Modeling of Thermoelectric Elements*, Wiley-VCH, Weinheim, Germany, 2015.
- [28] Y. Shi, Y. Wang, D. Mei, Z. Chen, Numerical modeling of the performance of thermoelectric module with polydimethylsiloxane encapsulation, *Int. J. Energy Res.* 42 (2018) 1287–1297.
- [29] Intel, Technical resources: Intel core processors, <https://www.intel.com/content/www/us/en/products/docs/processors/core/core-technical-resources.html>.
- [30] Failure mechanisms and models for semiconductor devices, in: JEDEC Publication JEP122-a, JEDEC Solid State Technology Association, 2002.
- [31] A. Abbasinasab, M. Marek-Sadowska, RAIN: A tool for reliability assessment of interconnect networks—physics to software, in: Proc. Design Automation Conf., DAC, ACM, New York, NY, USA, 2018, pp. 133:1–133:6.
- [32] M. Kavousi, L. Chen, S.X.-D. Tan, Electromigration immortality check considering joule heating effect for multisegment wires, in: Proc. Int. Conf. on Computer Aided Design, ICCAD, 2020, pp. 1–8.



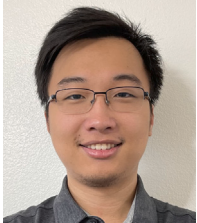
Jinwei Zhang is a current Ph.D. candidate who joined VSCLAB in the Department of Electrical and Computer Engineering at the University of California Riverside in Fall 2018. He received his B.S. degree in Electrical and Control Engineering from Beijing Institute of Technology, China in 2014 and M.S. degree in Electrical Engineering from Washington University in Saint Louis, United States in 2016. His research interests are in Electronic Design Automation (EDA), system level thermal and power modeling for VLSI and silicon validation with modern machine learning techniques.



Sheriff Sadiqbacha received his B.S. degree (most outstanding graduate of the year) in Computer Engineering from California State University, Bakersfield in 2016. He received M.S. degree (magna cum laude) and Ph.D. degree in Electrical Engineering from the University of California, Riverside in 2019 and 2022. He is currently a Ph.D. student and Associate Instructor in the Department of Electrical and Computer Engineering at the University of California, Riverside. His research interests include pre-silicon reliability (Electromigration), and applied machine-learning in the area of Electronic Design Automation (EDA), post-silicon thermal, power, and reliability modeling and control.



Liang Chen was born in 1992. He received the B.E. degree in electromagnetic field and wireless technology from Northwestern Polytechnical University, Xi'an, China, in 2015 and his Ph.D. degree in electronic science and technology with Shanghai Jiao Tong University, Shanghai, China. He is currently a Post-Doc with the VLSI System and Computation Laboratory, Department of Electrical and Computer Engineering, University of California at Riverside, Riverside, CA, USA. His current research interests include signal integrity of high-speed interconnects, electrothermal co-simulation of 3-D integrated packages, and electromigration reliability.



Cuong Thi was born in 1996 in Viet Nam. He migrated to the United States in 2010 with his family. He attended University of California, Riverside in 2019. His goal is to achieve a Bachelor's degree in Computer Engineering. His areas of interest include Computer Hardware, Control Algorithms, Thermal Simulation and 3-D Modeling.



Sachin Sachdeva was born in 1995. He received his B. Tech degree in Electronics and Communication Engineering in 2016 from National Institute of Technology, Hamirpur, India. He is currently a Ph.D. student in Electrical Engineering at University of California Riverside, joining in 2021 Fall. His research interests are in electronic design automation (EDA), pre-silicon reliability and thermal modelling, and machine learning applications in VLSI.



Hussam Amrouch is a Junior Professor for the Semiconductor Test and Reliability (STAR) chair within the Computer Science, Electrical Engineering Faculty at the University of Stuttgart as well as a Research Group Leader at the Karlsruhe Institute of Technology (KIT), Germany. He received his Ph.D. degree with distinction (Summa cum laude) from KIT in 2015. His main research interests are design for reliability and testing from device physics to systems, machine learning, security, approximate computing, and emerging technologies with a special focus on ferroelectric devices.

He holds seven HiPEAC Paper Awards and three best paper nominations at top EDA conferences: DAC'16, DAC'17 and DATE'17 for his work on reliability. He currently serves as Associate Editor at Integration, the VLSI Journal. He has served in the technical program committees of many major EDA conferences such as DAC, ASP-DAC, ICCAD, etc. and as a reviewer in many top journals like T-ED, TCAS-I, TVLSI, TCAD, TC, etc. He has 115 publications (including 43 journals) in multidisciplinary research areas across the entire computing stack, starting from semiconductor physics to circuit design all the way up to computer-aided design and computer architecture. He is a member of the IEEE. ORCID 0000-0002-5649-3102.



Sheldon X.-D. Tan received his B.S. and M.S. degrees in electrical engineering from Fudan University, Shanghai, China in 1992 and 1995, respectively and the Ph.D. degree in electrical and computer engineering from the University of Iowa, Iowa City, in 1999. He is a Professor in the Department of Electrical Engineering, University of California, Riverside, CA. He also is a cooperative faculty member in the Department of Computer Science and Engineering at UCR. His research interests include Machine and deep learning for VLSI reliability modeling and optimization at circuit and system levels, machine learning for circuit and thermal simulation, thermal modeling, optimization and dynamic thermal management for many-core processors, parallel computing and adiabatic and Ising computing based on GPU and multicore systems. He has published more than 320 technical papers and has co-authored 6 books on those areas.

Dr. Tan received NSF CAREER Award in 2004. He received Best Paper Awards from ICSICT'18, ASICON'17, ICCD'07, DAC'09. He also received the Honorable Mention Best Paper Award from SMACD'18. He was a Visiting Professor of Kyoto University as a JSPS Fellow from Dec. 2017 to Jan. 2018. He is serving as the TPC Chair for ASPDAC 2021, and the TPC Vice Chair for ASPDAC 2020. He is serving or served as Editor in Chief for Elsevier's Integration, the VLSI Journal, the Associate Editor for three journals: IEEE Transaction on VLSI Systems (TVLSI), ACM Transaction on Design Automation of Electronic Systems (TODAES), Elsevier's Microelectronics Reliability and MDPI Electronics, Microelectronics and Optoelectronics Section.