

Modeling and Optimization of Chip Cooling with Two-Phase Vapor Chambers

Zihao Yuan¹, Geoffrey Vaartstra², Prachi Shukla¹, Sherief Reda³, Evelyn Wang², and Ayse K. Coskun¹

¹Boston University, Boston, MA - {yuan1z, prachis, acoskun}@bu.edu

²Massachusetts Institute of Technology, Cambridge, MA - {gvaartst, enwang}@mit.edu

³Brown University, Providence, RI - {sherief_reda}@brown.edu

Abstract—Ultra-high power densities that are expected in future processors cannot be efficiently mitigated by conventional cooling solutions. Using two-phase vapor chambers (VCs) with micropillar wick evaporators is an emerging cooling technique that can effectively remove high heat fluxes through the evaporation process of a coolant. Two-phase VCs with micropillar wicks offer high cooling efficiency by leveraging a capillary-driven flow, where the coolant is passively driven by the wicking structure that eliminates the need for an external pump. Thermal models for such emerging cooling technologies are essential to evaluate their impact on future processors. Existing thermal models for two-phase VCs use computational fluid dynamics (CFD) modules, which incur long design and simulation times. This paper presents a fast and accurate compact thermal model for two-phase VCs with micropillar wicks. Our model achieves a maximum error of 1.25°C with a speedup of 214x in comparison to a CFD model. Using our proposed thermal model, we build an optimization flow that selects the best cooling solution and its cooling parameters to minimize the cooling power under a temperature constraint for a given processor and power profile. We then demonstrate our optimization flow on different chip sizes and hot spot distributions to choose the optimal cooling technique among VCs, microchannel-based two-phase cooling, liquid cooling via microchannels, and a hybrid cooling technique with thermoelectric coolers and liquid cooling with microchannels.

I. INTRODUCTION

Localized power densities on chips are expected to reach and surpass $1\text{-}2\text{ KW}/\text{cm}^2$ in future processors [1] and result in thermal hot spots that can significantly degrade system performance and reliability. Current cooling techniques such as forced air cooling via fans or traditional heat sinks are not sufficient to mitigate these high power densities efficiently. Therefore, cooling is becoming more critical in future processors.

Researchers have been exploring new cooling techniques such as liquid cooling via microchannels [2], thermoelectric coolers (TEC) [3], two-phase cooling [4], [5], and hybrid cooling [6]. Two-phase cooling with vapor chambers (VCs) is particularly attractive as it offers many advantages over the other techniques: (i) it reduces thermal gradients, (ii) the evaporator in VCs removes heat passively and saves pumping power (in contrast to liquid cooling via microchannels), and (iii) it has a higher cooling efficiency [5], [7]. In this technique, phase change from liquid to vapor takes place inside an enclosure called VC. The bottom surface of the VC has a porous wick that sustains thin-film evaporation supplied by passive, capillary-driven flow [5]. In this work, we focus on an evaporator wick composed of a micropillar array that has periodic and precisely defined geometry (e.g. [8], [9]).

Having fast and accurate thermal models is essential to enable power-efficient cooling optimization for processors. Researchers have developed fast models for various cooling methods including liquid cooling via microchannels and hybrid cooling (of liquid cooling and TEC) [10], [11]. To select and optimize a cooling solution for a given chip and power profile, there is a need for fast thermal modeling approach for two-phase VCs with micropillar wick evaporator. Simulations for two-phase VCs are typically carried out using computational fluid dynamics (CFD) modules in COMSOL and ANSYS (e.g., [5]). However, these tools are computationally expensive and experience long solution times along with large memory requirements [12]. These limitations make CFD tools unsuitable for modeling the

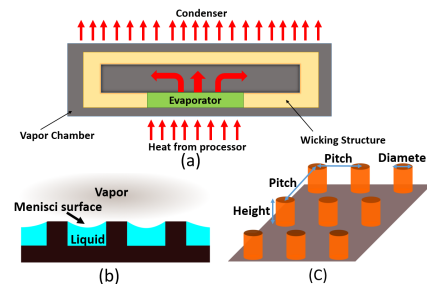


Fig. 1: (a) Vapor chamber structure view, (b) micropillar wick side view, (c) micropillar wick side view overall view.

cooling technique together with realistic processor architectures and applications.

In this work, we present a modular compact thermal model (CTM) (i.e., a lumped RC network model) of two-phase VCs with micropillar wick evaporators to enable speedy and accurate steady-state analysis of this cooling technique on realistic chip designs. In addition, we use the proposed CTM to design an optimization flow that maximizes the cooling power savings by selecting the best cooling solution as well as optimizing its cooling parameters. The main contributions of this paper are as follows:

1. We propose a steady-state CTM for two-phase VCs with micropillar wick evaporators and use the CTM to compare the cooling performance (i.e., hot spot temperature reductions and thermal gradients) against liquid cooling via microchannels and microchannel-based two-phase cooling (Sec. III and V-B).
2. We design a cooling optimization flow that selects the most power-efficient cooling solution and its cooling parameters for a given target chip and power profile, such that the cooling power (cooling power refers to the power used to reject the heat out of the cooling system to the chiller/condenser) is minimized and the chip temperature stays below a safe threshold (Sec. IV-C).
3. We demonstrate the practicality of the optimization flow on two realistic chips. We show that our optimization flow can select the best cooling technique (among VCs, liquid cooling via microchannels, hybrid cooling, and microchannel-based two-phase cooling) along with its optimal parameters to efficiently cool these chips (Sec. V).

II. RELATED WORK

High chip temperatures have been major concerns for several decades. As a result, researchers are actively working on finding new solutions to maintain safe chip temperatures. Some works have proposed design-time thermal management techniques (e.g., [13], [14]), while others have focused on runtime policies that use control knobs such as dynamic voltage and frequency scaling, task scheduling, and thread migration (e.g., [15]). Although these policies help reduce thermal violations, they also incur performance losses.

Another body of work focuses on emerging cooling techniques to help reduce chip temperatures and avoid performance degradation. In liquid cooling via microchannels, a coolant is pumped through the microchannels that are etched on the back of the silicon. Heat generated by a chip is absorbed by the coolant as it flows through

the microchannels. This technique is an attractive cooling solution for 3D-stacked architectures as it enables inter-tier cooling. Several works have introduced accurate CTMs of liquid cooling [10], [16].

TECs operate on Peltier effect such that when a biased current passes through this device, heat is absorbed on one side and rejected from the other. A recent work presents a CTM for TECs to investigate its capability in targeting high density hot spots [11]. Recently, a hybrid cooling design with TECs and liquid cooling via microchannels has also been explored to target high density hot spots [11].

Phase-change cooling is another area where latent heat is used to dissipate large amounts of heat from the chip. One type of phase-change is from solid to liquid, which is utilized in PCMs. Both coarse-grained and fine-grained PCM CTMs have been developed to demonstrate its use in computational sprinting [17], [18]. Another type of phase-change is from liquid to vapor. Sridhar et al. introduce a CTM simulator that models the liquid-vapor phase change inside microchannels by conserving energy and mass in two-phase [19].

The key innovation in our work is that we create a CTM to analyze the cooling performance of two-phase VCs with micropillar wick evaporators, which have some significant advantages over the emerging cooling methods described above. We also design an optimization flow to select the best cooling solution and its cooling parameters to optimize the cooling performance under a user-defined thermal constraint for a given chip and power profile. Our CTM enables a speedy evaluation of the cooling efficiency of two-phase VCs for realistic chips and comparisons to other cooling techniques.

III. PROPOSED TWO-PHASE COOLING MODEL FOR VAPOR CHAMBERS WITH MICROPILLAR WICK EVAPORATORS

Background on Vapor Chambers: The schematic of a VC is shown in Fig. 1 (a). On the bottom side of VC, there is an evaporator that consists of a thin porous wick. The evaporator is placed directly on top of the heat source (i.e., the processor). As the saturated coolant flows within the porous wick, the coolant absorbs the heat generated by the chip and evaporates. Above the VC, there is a condenser (e.g., a heat sink) that condenses the saturated vapor back to liquid phase. The condensed liquid is recirculated in the VC by the additional wicking structures along its side walls [5]. Fabrication and implementation details of VCs can be found in previous works [5], [20]. VCs have been shown to achieve significant cooling performance on electronic devices [5] and are already being used as cooling solutions for CPUs and GPUs [5], [21]. There are two major metrics that determine the performance in VCs: (i) HTC, and (ii) dry-out heat flux. *HTC* is the rate of heat transfer per unit temperature difference between the evaporator and the environment. *Dry-out heat flux* refers to the thermal limit of a two-phase device, beyond which the coolant ceases to exist in two-phases and instead, is found in only vapor phase. Micropillar wick evaporators have been shown to improve cooling efficiency and enhance dry-out heat flux owing to their high capillary pumping budget and extended menisci evaporation area [8], [9].

Proposed Two-Phase VCs CTM: We adopt a CTM approach to implement our proposed model. The entire system including the VC is divided into small grid cells. The default grid cell is shown in Fig. 2 (a). This figure shows that the virtual temperature node, which represents the temperature of the grid cell, is placed on the bottom surface. The micropillar wick evaporator is modeled as a separate layer and is placed directly above the processing layer. In Fig. 2 (b), we demonstrate how the grid cells of the two layers are connected in the model. We assume that inside the VC there is only saturated vapor and all the liquid is contained in the wicking structures and evaporator as shown in Fig. 1 (b). To model the saturated vapor conditions, we place an additional virtual temperature node on top of each micropillar wick layer grid cell as shown in Fig. 2 (b). The temperature of this node is set to the saturated temperature of the coolant, T_{Sat} , and therefore, depends on coolant properties and pressure inside the VC. The micropillar wick layer along with T_{Sat}

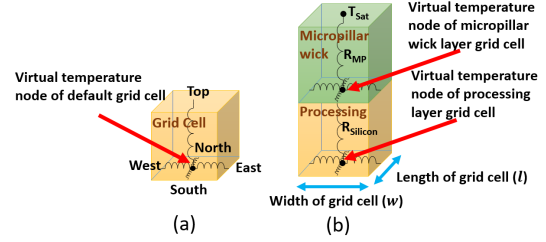


Fig. 2: (a) Default grid cell, (b) proposed grid cells for modeling two-phase VCs with micropillar wick evaporators

nodes represent the VC. Since we assume that T_{Sat} is maintained at a constant temperature, we do not need to model a condenser.

We assign silicon properties to the processing layer grid cells and represent the lateral and vertical thermal resistance, respectively, of each cell as $R_{Silicon}$ as shown in Fig. 2 (b). In the micropillar wick layer, determining the vertical thermal resistance of each grid cell is complicated because the coolant exists in two phases. We use a previously established relationship between the HTC and thermal resistance of a grid cell to represent the vertical thermal resistance of a micropillar grid cell, R_{MP} [10], [12]. Micropillar wick HTC is highly dependent on the coolant, VC pressure, and micropillar wick geometry (micropillar height, diameter, and pitch as shown in Fig. 1 (c)) [5]. We use a COMSOL model to extract the HTC of a wide range of micropillar geometries, coolants, and VC pressures. This COMSOL model is detailed in a prior work [22]. In this COMSOL model, the authors separate the CFD simulation into the fluid domain and heat transfer domain. By coupling these two domains, they iteratively obtain the temperature distribution. They use the Young-Laplace equation to relate the curvature of the liquid-vapor interface to local pressure. In addition, they also use Darcy's law and a volumetric loss function to model fluid flow in a uniform porous medium and the evaporative flux, respectively. Furthermore, they parametrically derive the permeability and HTC for each micropillar wick geometry. The extracted HTCs are stored in HTC lookup tables for various coolants and geometries. For simplicity, we assume a flat evaporation surface in our model instead of a menisci evaporation surface as shown in Fig. 1 (b). The flat evaporation surface is a conservative assumption that is widely used to define and simplify the liquid-vapor interface [9]. Such an assumption enables us to employ a uniform HTC across the micropillar wick layer.

Compared to other state-of-the-art compact modeling methodologies [10], [11], [19], the distinctions of our CTM are as follows: i) our proposed CTM models two-phase cooling in VCs, a passive cooling technique that requires no cooling power on the evaporator side; ii) we place the virtual temperature node at the bottom of the VC grid cells to model heat transfer happened on the evaporator and since there is no pumping power at the evaporator side, there is no need for voltage-controlled current sources; and iii) since the HTC is stored in a lookup table, our modeling methodology can be generalized to model different two-phase cooling devices.

IV. OPTIMIZATION FLOW FOR POWER-EFFICIENT COOLING

In this section, we first show an analytical model of dry-out heat flux for the VCs. This dry-out heat flux model is used to determine the optimal micropillar wick geometry. We then perform parametric studies to understand the relationships between HTC, dry-out heat flux, micropillar geometries, and chip sizes. Finally, we present our cooling optimization flow that selects the most power-efficient cooling solution and its parameters for a given chip and power profile. This optimization flow incorporates the micropillar wick geometry, liquid flow velocity, TEC current, and mass flow velocity as tuning parameters to optimize cooling solutions based on the cooling power and temperature constraint.

A. Dry-out Heat Flux Analytical Model

One major concern while designing VCs is to prevent dry-out. Dry-out heat flux for square chips is defined in Eqs. (1) and (2) [8]. The coolant and micropillar parameters are listed in Table I.

$$q''_{dry-out} = (40/3)\psi M \cos\theta_{rec} \quad (1)$$

$$M = \frac{\sigma_{lv}\rho_l h_{lv}}{\mu_l} \quad (2)$$

ψ is a dimensionless function of micropillar geometry that lumps the effect of the geometry on heat transfer capacity [8]. M is a figure of merit for the coolant. We use the above equations to calculate the dry-out heat flux, $q''_{dry-out}$. Since the dry-out heat flux is sensitive to micropillar geometry and chip dimensions [5], our optimization flow selects a geometry that provide the highest HTC while also preventing dry-out for a given chip and its power profile.

B. Parametric Study

Recall that high HTC and high dry-out heat flux deliver better cooling performance of two-phase cooling in VCs. In order to understand the relationships among the HTC, dry-out heat flux, micropillar geometries, and chip sizes, we perform parametric studies using a COMSOL model [22]. Across all studies, we use a coolant, R134a, at T_{Sat} of $50^\circ C$ under 13.2 bar pressure and vary the chip size from 4 mm^2 to 100 mm^2 . In the first study, we vary the micropillar height (h) from $20 \mu\text{m}$ to $70 \mu\text{m}$, with diameter (d) and pitch (i) set to $10 \mu\text{m}$ and $20 \mu\text{m}$, respectively. Fig. 3 (a) shows the inverse relationship between HTC and dry-out heat flux as micropillar height changes. The dry-out heat flux increases with increasing micropillar height because the wick becomes more permeable, which lowers the viscous resistance to capillary flow. On the other hand, HTC decreases with increasing micropillar height since the liquid film becomes thicker, thus increasing conduction resistance. In the second study, we vary i from $60 \mu\text{m}$ to $150 \mu\text{m}$, with h and d set to $55 \mu\text{m}$ and $10 \mu\text{m}$, respectively. The results of this simulation are shown in Fig. 3 (b), in which we can see that as the pitch increases, both dry-out heat flux and HTC decrease. In this particular regime, the loss of capillary pumping budget due to increasing the pitch is more significant than the increase in permeability. Therefore, increasing the pitch leads to earlier dry-out. The HTC decreases with increasing pitch because the solid fraction ($c = \frac{\pi i}{4} (\frac{d}{i})^2$) diminishes, forcing more heat to conduct through the liquid film. In the third study, we fix h to $50 \mu\text{m}$ and i to $30 \mu\text{m}$, and vary d from $5 \mu\text{m}$ to $17 \mu\text{m}$. We observe that both the dry-out heat flux and HTC increase with increasing micropillar diameter. This is because a larger diameter leads to a higher capillary pumping budget that increases the dry-out heat flux, while the increase in solid fraction is favorable for conduction, thus enhancing the HTC. Based on the above studies, we see that both the HTC and dry-out heat flux have nontrivial relationships with different

TABLE I: Coolant and micropillar parameters

h	Height of micropillar
d	Diameter of micropillar
i	Pitch of micropillar wick
P_{bg}	Background power density
P_{hs}	Hot spot power density
ρ_l	Liquid density
h_{lv}	latent heat of vaporization
μ_l	Dynamic viscosity
q''	Heat flux
$q''_{dry-out}$	Dry-out heat flux
θ_{rec}	Receding contact angle for fluid-solid pair
p	Pressure
κ	Permeability of the wick
T_{Sat}	Saturated temperature of the coolant
σ_{lv}	Surface tension
ΔT	Thermal gradients across the chip
MG_{opt}	Optimal micropillar geometry
$P_{cooling}$	Cooling power
T_{hs}	Hot spot temperature
T_{limit}	User-defined temperature limit
u	Liquid flow velocity
I	TEC current
G	Mass flow velocity

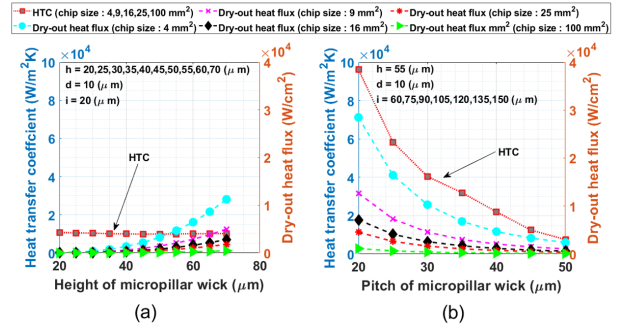


Fig. 3: Parametric study for different micropillar wick geometries. Dry-out heat flux is shown on the right axes.

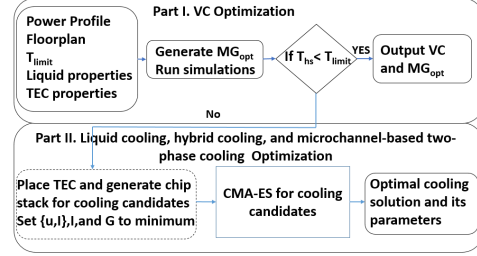


Fig. 4: Proposed optimization flow.

geometry parameters. As a result, we need an optimal micropillar geometry to enhance the cooling performance of two-phase VCs.

C. Proposed Optimization Flow

The goal of our optimization flow is to select the best cooling method along with its cooling parameters, for a target chip and its power profile, that minimizes the cooling power under a temperature constraint. We incorporate liquid cooling via microchannels, hybrid cooling (of liquid cooling via microchannels and TEC), two-phase VCs with micropillar wick evaporator, and microchannel-based two-phase cooling as cooling solution candidates. These cooling methods have been shown to achieve higher cooling performance than the traditional heat sink and forced air cooling via fan [5]–[7], [10]. In addition, they are also compatible for processor cooling. We adopt the CTMs for liquid cooling via microchannels, hybrid cooling, and microchannel-based two-phase cooling from recent works [11], [12], [19].

$$\min \alpha P_{cooling, norm} + \beta (\max(T_{hs} - T_{limit}, 0), norm) \quad (3)$$

The objective function of our optimization is to minimize the cooling power under a temperature constraint (Eq. (3)). The cooling power for liquid cooling via microchannels and microchannel-based two-phase cooling (i.e., the pumping power) are calculated based on the pressure drop along the channel and the liquid volumetric flow rate [16], while TEC cooling power is the difference between heat absorbed and rejected on the cold and hot sides. The maximum attainable liquid flow velocity and TEC current are set to 2.6 m/s and 7 A , respectively, owing to system constraints [3], [10]. For microchannel-based two-phase cooling, the maximum allowed mass flow velocity is set to $560 \text{ kg/m}^2\text{s}$ [4]. In Eq. (3), the cooling cost is normalized with respect to the maximum cooling power (i.e., $u = 2.6 \text{ m/s}$ and $I = 7 \text{ A}$). The difference between T_{hs} and T_{limit} is normalized with the user-defined maximum on-chip temperature, i.e., T_{limit} . α is the user-specific weight factor with no unit and β is the penalty weight that we set to a large value to prevent violation of the temperature constraint. We set $\alpha = 0.05$ and $\beta = 0.95$ according to our system. For two-phase VCs with micropillar wick evaporator, we need to add the dry-out constraint (Eq. (1)) that prevents dry-out by ensuring that the dry-out heat flux of the selected micropillar wick geometry is greater than or equal to the hot spot power density.

In addition to that, we incorporate micropillar geometry constraints ($h/i \geq 0.2$, $0.06 < d/i < 0.6$) to get high capillary pressure [23].

Our proposed optimization flow is shown in Fig. 4. We divide the optimization flow into two parts. In the first part, we determine the micropillar geometry with the highest HTC under the dry-out constraint. We then simulate the chip using two-phase VCs with micropillar wick evaporators as the cooling solution. If a wick geometry satisfies the temperature constraint, we select two-phase VC with the micropillar wick geometry as the optimal technique. The reason is that VC is a passive cooling device that requires no additional power on the evaporator side. Otherwise, we use a covariance matrix adaptation evolution strategy (CMA-ES) to find the optimal $\{u, I\}$ pair for hybrid cooling, optimal I for liquid cooling via microchannels and optimal G for microchannel-based two-phase cooling (see Table I). CMA-ES is a stochastic, derivative-free sampling method which does not require a numerical objective function to converge to an optimal solution. The pseudo code of our CMA-ES implementation for hybrid cooling is shown in Algorithm 1.

The optimization flow then enters the second part when the first part fails to find a wick geometry that satisfies the temperature constraint. In each iteration, the algorithm first samples the $\{u, I\}$ pairs based on a multivariate normal distribution and then runs thermal simulations for hybrid cooling with those sample points (lines 2-4). Next, all the $\{u, I\}$ pairs are sorted in increasing cost (line 5). In lines 6-7, the algorithm assigns a weight vector w to the $\{u, I\}$ pairs to update the mean of the sampling distribution [24]. This step ensures that the sampling distribution for the next iteration moves closer to the $\{u, I\}$ pair that gives the minimum cost in the current iteration [24]. Next, the algorithm updates the evolution paths, p_c and p_σ , which conceptually stand for the search paths in the CMA-ES algorithm (line 8-9). It then updates the covariance matrix and step-size based on the p_c and p_σ . All the update functions ($update_{p_\sigma}$, $update_{p_c}$, $update_C$, and $update_\sigma$) used in this algorithm are adopted from a recent work [24]. After max_{iter} number of iterations, the algorithm generates the optimal $\{u_{opt_hybrid}, I_{opt_hybrid}\}$ pair that results in the minimum cost (line 13) along with its corresponding on-chip temperature, $T_{hs_opt_hybrid}$. Similarly, we apply CMA-ES algorithm to liquid cooling via microchannels and microchannel-based two-phase cooling to get the optimal I_{liquid} , $T_{hs_opt_liquid}$, G and $T_{hs_opt_mic}$. We compare $T_{hs_opt_hybrid}$, $T_{hs_opt_liquid}$, and $T_{hs_opt_mic}$ to T_{limit} , respectively, and select the cooling solutions that satisfy the temperature constraint. Finally, we compare the cooling power of the selected solutions to output the one with the minimum cooling power as the optimal choice.

V. EXPERIMENTAL RESULTS

In this section, we first validate the proposed two-phase CTM against a COMSOL model to show the accuracy and speedup obtained by our model. We then compare the cooling performance of two-phase VCs with micropillar wick evaporators with that of liquid cooling via microchannels by using our proposed CTM. Next, we apply our proposed optimization flow to various chip floorplans and power profiles to evaluate the cooling efficiency of two-phase VCs, liquid cooling via microchannels and hybrid cooling (liquid cooling via microchannels and TEC). Finally, we use our proposed optimization flow on two realistic chips to select the optimal cooling techniques for each of these chips to show the efficiency of our optimization flow. We implement and integrate our proposed CTM in the HotSpot-6.0 thermal simulator [25].

A. Validation of the Proposed Model

We model a $2 \times 2 \text{ mm}^2$ chip with a thickness of $100 \mu\text{m}$ in COMSOL and HotSpot to validate the accuracy of our proposed CTM. We run two sets of simulations in COMSOL: (i) processing layer with a uniform power density, and (ii) processing layer with a non-uniform power density with a $500 \times 500 \mu\text{m}^2$ hot spot placed

Algorithm 1: CMA-ES

Input : Number of samples per iteration, λ
Input : Objective function (Eq. (3)), $cost$
Input : Maximum number of iterations, max_{iter}
Initialize: Sampling distribution mean, μ
Initialize: Step-size, σ
Initialize: Covariance matrix, $C = \text{Identity matrix}$
Initialize: Cumulation for σ and C , $p_\sigma = 0$, $p_c = 0$
Initialize: $k = 0$

- 1 **while** $k < max_{iter}$ **do**
- 2 **for** i in $1 \dots \lambda$ **do**
- 3 $\{u, I\}_i = \mathcal{N}(\mu, \sigma^2 C)$
- 4 $cost_i = \text{Thermal Simulation}(\{u, I\}_i)$
- 5 Sort $\{u, I\}$ based on increasing $cost$ (objective function)
- 6 $\mu' = \mu$
- 7 $\mu = \sum_{i=1}^{\lambda/2} (w_i \{u, I\}_i)$
- 8 $p_\sigma = update_{p_\sigma}(p_\sigma, \sigma^{-1} C^{-1/2} (\mu = \mu'))$
- 9 $p_c = update_{p_c}(p_c, \sigma^{-1} (\mu = \mu', ||p_\sigma||))$
- 10 $C = update_C(C, p_c, (\{u, I\}_1 - \mu)/\sigma, \dots, (\{u, I\}_\lambda - \mu)/\sigma)$
- 11 $\sigma = update_\sigma(\sigma, ||p_\sigma||)$
- 12 $k++$

13 Generate $\{u_{opt_hybrid}, I_{opt_hybrid}\}, P_{cooling_opt_hybrid}$, and $T_{hs_opt_hybrid}$ based on the last iteration simulation results

TABLE II: Comparison between proposed CTM and COMSOL.

Simulations	Coolant	P_{bg}	P_{hs}	{h,d,i}	Avg error ($^{\circ}\text{C}$)	Max error ($^{\circ}\text{C}$)
Uniform Power	Water	100	100	{30,12,36}	0.20	0.19
		100	100	{40,16,48}	0.22	0.21
		100	100	{50,20,60}	0.25	0.25
	R134a	20	20	{30,12,36}	0.65	0.71
		20	20	{40,16,48}	0.75	0.77
		20	20	{50,20,60}	0.80	0.82
Non-uniform Power	Water	50	100	{30,12,36}	0.23	0.23
		50	200	{40,16,48}	0.31	0.26
		50	300	{50,20,60}	0.49	0.45
	R134a	20	25	{30,12,36}	0.99	1.11
		20	50	{40,16,48}	1.02	1.12
		20	75	{50,20,60}	1.04	1.24

at the center. Each simulation set has three different micropillar wick geometries and two coolants: water and R134a. In these validation experiments, since water has a better HTC compared to R134a, to ensure the maximum temperatures are less than 85°C , we select higher power densities for water and lower power densities for R134a. In the experiment with uniform power density, we set the power density equal to 100 W/cm^2 for water and 20 W/cm^2 for R134a. In the non-uniform power density simulations, we set the background power density to 50 W/cm^2 for water and 20 W/cm^2 for R134a. The hot spot power density is set to 100, 200, and 300 W/cm^2 for water and 25, 50, and 75 W/cm^2 for R134a. To prevent extremely high chip temperatures, we reduce the T_{Sat} of water to 50°C , by setting the pressure inside VC to 0.124 bar. As for R134a, we set its T_{Sat} to 50°C under 13.2 bar pressure. For each COMSOL simulation, we use 592 nodes to simulate the fluid domain and 1106 nodes to simulate the heat transfer domain. Among all the simulation cases, it takes a minimum of 4 iterations to converge and finish the simulation. We model the same chip in HotSpot using the grid model, and use 64×64 grids to compute the steady-state temperatures. The simulation time of the COMSOL CFD model is 45s, while it only takes 0.21s to simulate our proposed CTM. Table II compares the temperatures obtained in the above simulations. For both uniform and non-uniform simulations, the proposed model achieves high accuracy with both the maximum and average errors less than 0.5°C for water and 1.25°C for R134a while achieving a speed up of $214\times$ when compared to COMSOL CFD simulations. Typical accuracies for CTMs of various cooling technologies range from 89.9% to 97.3% [10], [11], [19]. Our proposed CTM provides a 98.5% accuracy which is similar to the approaches mentioned above. These simulations show that our model can significantly reduce simulation time with only a small tradeoff in accuracy.

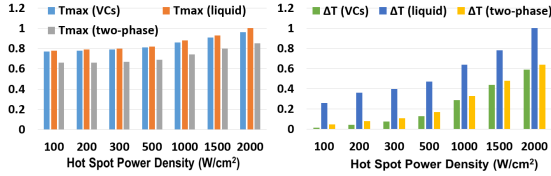


Fig. 5: Comparison of cooling performance of two-phase VCs with micropillar evaporators (VC), liquid cooling via microchannels (liquid), and microchannel-based two-phase cooling (two-phase) when flow velocity = 2.6 m/s and mass flow velocity = $300 \text{ kg/m}^2\text{s}$. Results are normalized to liquid cooling when $P_{hs} = 2000 \text{ W/cm}^2$.

B. Cooling Performance Evaluation

To evaluate the cooling performance of two-phase VCs with micropillar wick evaporators, we run thermal simulations to compare its cooling performance and cooling power with that of liquid cooling via microchannels and microchannel-based two-phase cooling. The simulated chip is $20 \times 20 \text{ mm}^2$ large with a $500 \times 500 \mu\text{m}^2$ hot spot placed at the center. The background power density is set to 50 W/cm^2 and the hot spot power density varies from 100 to 2000 W/cm^2 . For a fair comparison, we select water as a coolant with a saturated temperature of 50°C (pressure = 0.124 bar). We vary the liquid flow velocity from 0.5 to 2.6 m/s [10] and mass flow velocity from 100 to $560 \text{ kg/m}^2\text{s}$ [4]. We sandwich the liquid microchannel layer between the processing layer and the packaging layer. Structural properties of liquid cooling via microchannels and microchannel-based two-phase cooling are shown in Table III. Table IV shows the optimal micropillar geometries selected by the optimization flow along with the estimated pumping power for liquid-cooling via microchannels and microchannel-based two-phase cooling. Fig. 5 shows the simulation results when liquid flow velocity is set to 2.6 m/s and mass flow velocity is set to $560 \text{ kg/m}^2\text{s}$. Two-phase VCs achieve lower hot spot temperature when compared to liquid cooling via microchannels. Since microchannel-based two-phase cooling is an active microfluidic two-phase cooling method, it provides higher hot spot temperature reductions than two-phase VCs. Most importantly, two-phase VCs provide higher reductions on thermal gradients by up to 11.78°C in comparison to liquid cooling via microchannels, and 1.5°C in comparison to microchannel-based two-phase cooling, without additional pumping power on the evaporator side.

C. Optimization Flow Results

Table V shows the results of our optimization flow for different chip floorplans, P_{hs} s, and distribution of hot spots. The background power density is set to 50 W/cm^2 . We select water as a coolant with a saturated temperature of 50°C . We select square and symmetric chips as our floorplans as shown in Fig. 6. Our optimization flow can also be used for non-square and asymmetric chips (Sec. V-D). For hybrid cooling chip stack, we add an additional TEC layer between liquid microchannel layer and processing layer. In the TEC layer, a $3.5 \times 3.5 \text{ mm}^2$ TEC unit is placed directly above each hot spot in the processing layer. The rest of the TEC layer comprises silicon. For the

TABLE III: Structural properties and simulation parameters.

Processing layer thickness	$750 \mu\text{m}$
Microchannel height	$200 \mu\text{m}$
Microchannel width (same as wall width)	$50 \mu\text{m}$
TEC layer thickness	$100 \mu\text{m}$
Packaging layer (bulk silicon)	$40 \mu\text{m}$
higher power density chip size	$20 \times 20 \text{ mm}^2$
Intel SCC core size	1.129 mm^2
lower power density chip size	$18 \times 14.1 \text{ mm}^2$

TABLE IV: Optimal geometries (h, d, i) of two-phase VCs and estimated pumping power of liquid cooling via microchannels and microchannel-based two-phase cooling ($G = 560 \text{ kg/m}^2\text{s}$).

$P_{hs} \text{ (W/cm}^2\text{)}$	100	200	300	500	1000	1500	2000	
$MG_{opt} \text{ (}\mu\text{m)}$	20,10.5	25,10.5	30,10.5	30,10.5	35,10.5	45,10.5	45,10.5	
$u \text{ (m/s)}$	0.5	1	1.5	2.6	$G \text{ (kg/m}^2\text{s)}$	100	300	560
$P_{pump} \text{ (W)}$	0.17	0.66	1.5	4.5	$P_{pump} \text{ (W)}$	0.2	0.41	1.14

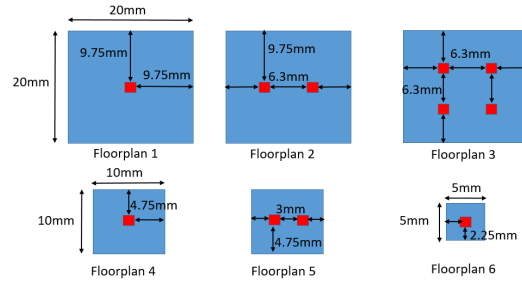


Fig. 6: Tested floorplans (Red square indicate hot spots.).

optimization flow, we set max_{iter} and λ to 20 and 40, respectively. We also initialize the sampling distribution mean and step-size to 0.5 and 0.2, respectively.

As shown in Table V, our optimization flow selects microchannel-based two-phase cooling for smaller chips with relatively higher power densities. In these cases, both microchannel-based two-phase cooling and hybrid cooling can reduce the hot spot temperatures below 85°C . Since hybrid cooling incurs an additional TEC power, microchannel-based two-phase cooling is more power-efficient than hybrid cooling. For larger chips with multiple high density hot spots, hybrid cooling is selected owing to its hot spot mitigation ability. However, for smaller chips with lower power densities, VCs are selected as the optimal cooling solution. VC is a passive cooling device that can efficiently spread the generated heat laterally. However, its performance starts to degrade in smaller chips and it is not able to target hot spots. This is because the total amount of heat transfer through lateral convection diminishes with decreasing chip size. Note that, liquid cooling via microchannels are not shown in any of the test cases. Generally, heat absorption during evaporation process can remove more heat than sensible heat absorption [7], which means that microchannel-based two-phase cooling consumes lesser pumping power to remove same amount of heat than liquid cooling via microchannels.

D. Evaluation with Realistic Chips

In this section, we apply the proposed optimization flow to realistic chips in order to demonstrate the efficiency of our optimization flow as a tool for optimal cooling solution selection and cooling parameters optimization. We simulate a 256-core processor inspired by the Intel SCC scaled to 22 nm (higher power density chip) [13], and a 45 nm 12-core AMD Magny-Cours processor (lower power density chip) [26]. In order to apply the proposed model to both realistic chips, we use a 64×64 grid size to simulate both the chips and set water as a coolant. Detailed specifications of both chips are given in Table III.

For the higher power density chip, the core architecture is based on IA-32 core [27]. We obtain power profiles of a simulated SCC-based chip from a recent work [13]. Their work uses Sniper [28] and McPAT [29] to obtain power profiles by running various multi-threaded benchmarks under different processor frequencies, followed by steady-state thermal simulations with HotSpot to obtain the

TABLE V: Optimization results when $T_{limit} = 85^\circ\text{C}$ (N/A stands for none of cooling solution can reduce the maximum chip temperature under 85°C). VC{h, d, i} represents two-phase VCs, hy{u, I} stands for hybrid cooling, and mic{G} stands for microchannel-based two-phase cooling.

Floorplan	Hot Spot Power Density (W/cm ²)					
	100	300	1000	1300	1700	2000
1	VC{20, 5, 10} 0W	VC{25, 5, 10} 0W	VC{35, 5, 10} 0W	VC{40, 5, 10} 0W	hy{1.56, 4.2} 3.75W	hy{2.49, 1.49} 4.30W
2	VC{20, 5, 10} 0W	VC{25, 5, 10} 0W	VC{35, 5, 10} 0W	VC{40, 5, 10} 0W	hy{2.49, 6.93} 9.85W	hy{2.57, 5.95} 13.54W
3	VC{20, 5, 10} 0W	VC{25, 5, 10} 0W	VC{35, 5, 10} 0W	VC{40, 5, 10} 0W	N/A	N/A
4	VC{20, 5, 10} 0W	VC{20, 5, 10} 0W	VC{20, 5, 10} 0W	mic{183} 0.033W	mic{222} 0.0581W	mic{261} 0.073W
5	VC{20, 5, 10} 0W	VC{20, 5, 10} 0W	VC{20, 5, 10} 0W	mic{204} 0.042W	mic{246} 0.0651W	mic{303} 0.091W
6	VC{20, 5, 10} 0W	VC{20, 5, 10} 0W	mic{111} <0.01W	mic{132} <0.01W	mic{168} <0.01W	mic{180} 0.012W

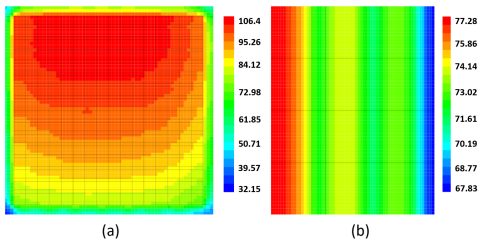


Fig. 7: Thermal maps for (a) higher power density chip, and (b) lower power density chip.

thermal profile of the chip. For our simulations, we select the power profile that results in the highest thermal gradient and chip temperature of the SCC-based chip, to extract the most interesting thermal profile of the chip. The selected power profile has a hot spot power density of 216.6 W/cm^2 , which results in a maximum thermal gradient of 16.1°C and maximum chip temperature of 125.78°C when using the default heat sink in HotSpot. If we set the temperature limit to 107°C , then the optimization flow outputs hybrid cooling with $\{2.6 \text{ m/s}, 6.5 \text{ A}\}$ as the optimal $\{u, I\}$ pair with a cooling power of 48.2 W . Thermal map for this case is shown in Fig. 7 (a). As for lower power density chip, its hot spot has a power density of 69.8 W/cm^2 [18]. If we set the temperature constraint to 85°C , then the optimization flow selects two-phase VCs as the most power-efficient cooling solution with the optimal micropillar geometry of $h, d,$ and i as $20 \mu\text{m}, 5 \mu\text{m},$ and $10 \mu\text{m}$, respectively. Thermal map for lower power density chip is shown in Fig. 7 (b). The maximum temperatures showed on these two thermal maps are $< 1^\circ\text{C}$ lower than the temperature constraints we set for these two realistic chips, which shows the effectiveness of our optimization flow.

For each realistic chip, it takes less than 16 minutes for the optimization flow to converge on the optimal cooling result. For the lower power density chip, our optimization flow selects two-phase VCs with micropillar wick evaporator as the optimal cooling technique. However, for higher power density chip, since two-phase VCs, microchannel-based two-phase cooling, and liquid cooling via microchannels are not able to reduce the temperature below 107°C , the optimization flow selects hybrid cooling as the optimal cooling solution.

VI. CONCLUSION

In this paper, we present a CTM for two-phase VCs with micropillar wick evaporators. We validate our proposed model against a COMSOL model and integrate the CTM into the HotSpot 6.0 thermal simulator. Our proposed model achieves a maximum error of less than 1.25°C and speedup of up to 214x when compared to COMSOL. We also discuss the tradeoffs between HTC and dry-out heat flux for various micropillar wick geometries. We then build an optimization flow that selects the most power-efficient cooling solution and its parameters to achieve minimum cooling power under a temperature limit for a given power profile and chip. We further use our proposed optimization flow on two realistic chips. For both chips, the optimization flow is able to find the most power-efficient cooling solution under a user-defined temperature constraint.

ACKNOWLEDGEMENT

This project has been partially funded by the NSF CRI (CI-NEW) grant #1730316/1730003/1730389.

REFERENCES

- [1] M. Schultz *et al.*, "Embedded two-phase cooling of large three-dimensional compatible chips with radial channels," *Journal of Electronic Packaging*, vol. 138, no. 2, p. 021005, 2016.
- [2] B. Dang *et al.*, "Integrated microfluidic cooling and interconnects for 2D and 3D chips," *IEEE Transactions on Advanced Packaging*, vol. 33, no. 1, pp. 79–87, 2010.
- [3] I. Chowdhury *et al.*, "On-chip cooling by superlattice-based thin-film thermoelectrics," *Nature nanotechnology*, vol. 4, no. 4, p. 235, 2009.

- [4] S. G. Kandlikar, "Fundamental issues related to flow boiling in minichannels and microchannels," *Experimental Thermal and Fluid Science*, vol. 26, no. 2-4, pp. 389–407, 2002.
- [5] M. Bulut, S. G. Kandlikar, and N. Sozbir, "A review of vapor chambers," *Heat Transfer Engineering*, pp. 1–23, 2018.
- [6] K. Yazawa *et al.*, "Cooling power optimization for hybrid solid-state and liquid cooling in integrated circuit chips with hotspots," in *IEEE Proc. of 13th Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm)*, 2012, pp. 99–106.
- [7] J. Thome, "Heat transfer engineering data book III," 2010.
- [8] S. Adera, D. Antao, R. Raj, and E. N. Wang, "Design of micropillar wicks for thin-film evaporation," *International Journal of Heat and Mass Transfer*, vol. 101, pp. 280–294, 2016.
- [9] M. Wei *et al.*, "Optimization and thermal characterization of uniform silicon micropillar based evaporators," *International Journal of Heat and Mass Transfer*, vol. 127, pp. 51–60, 2018.
- [10] A. Sridhar, A. Vincenzi, D. Aienza, and T. Brunschwiler, "3D-ICE: A compact thermal model for early-stage design of liquid-cooled ICs," *IEEE Transactions on Computers*, vol. 63, no. 10, pp. 2576–2589, 2014.
- [11] F. Kaplan, S. Reda, and A. K. Coskun, "Fast thermal modeling of liquid, thermoelectric, and hybrid cooling," in *Proc. of Intersociety Conf. on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm)*, 2017, pp. 726–735.
- [12] Z. Yuan *et al.*, "Two-phase vapor chambers with micropillar evaporators: a new approach to remove heat from future high-performance chips," in *Proc. of Intersociety Conf. on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm)*, 2019.
- [13] F. Eris *et al.*, "Leveraging thermally-aware chiplet organization in 2.5D systems to reclaim dark silicon," in *IEEE Proc. of Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2018, pp. 1441–1446.
- [14] K. Dev, I. Paul, W. Huang, Y. Eckert, W. Burleson, and S. Reda, "Implications of integrated cpu-gpu processors on thermal and power management techniques," *arXiv preprint arXiv:1808.09651*, 2018.
- [15] H. F. Sheikh, I. Ahmad, Z. Wang, and S. Ranka, "An overview and classification of thermal-aware scheduling techniques for multi-core processing systems," *Sustainable Computing: Informatics and Systems*, vol. 2, no. 3, pp. 151–169, 2012.
- [16] A. K. Coskun, J. L. Ayala, D. Aienza, and T. S. Rosing, "Modeling and dynamic management of 3D multicore systems with liquid cooling," in *17th IFIP International Conference on Very Large Scale Integration (VLSI-SoC)*, IEEE, 2009, pp. 35–40.
- [17] A. Raghavan *et al.*, "Computational sprinting," in *IEEE Proc. of 18th International Symposium on High Performance Computer Architecture (HPCA)*, 2012, pp. 1–12.
- [18] F. Kaplan *et al.*, "Modeling and analysis of phase change materials for efficient thermal management," in *IEEE Proc. of 32nd International Conference on Computer Design (ICCD)*, 2014, pp. 256–263.
- [19] A. Sridhar *et al.*, "STEAM: A fast compact thermal model for two-phase cooling of integrated circuits," in *IEEE Proc. of International Conference on Computer-Aided Design (ICCAD)*, 2013, pp. 256–263.
- [20] J. Hsieh, H. Huang, and S. Shen, "Experimental study of microrectangular groove structure covered with multi mesh layers on performance of flat plate heat pipe for led lighting module," *Microelectronics Reliability*, vol. 52, no. 6, pp. 1071–1079, 2012.
- [21] "Graphics Reinvented: NVIDIA GeForce RTX 2080 Ti Graphics Card," <https://www.nvidia.com/en-us/geforce/graphics-cards/rtx-2080-ti/>.
- [22] G. Vaartstra, Z. Lu, and E. N. Wang, "Simultaneous prediction of dryout heat flux and local temperature for thin film evaporation in micropillar wicks," *International Journal of Heat and Mass Transfer*, vol. 136, pp. 170–177, 2019.
- [23] C. Byon and S. J. Kim, "The effect of meniscus on the permeability of micro-post arrays," *Journal of Micromechanics and Microengineering*, vol. 21, no. 11, p. 115011, 2011.
- [24] A. Auger and N. Hansen, "Tutorial cma-es: evolution strategies and covariance matrix adaptation," in *GECCO (Companion)*, 2012.
- [25] K. Skadron *et al.*, "Temperature-aware microarchitecture," in *IEEE Proc. of 30th Annual International Symposium on Computer Architecture (ISCA)*, 2003, pp. 2–13.
- [26] P. Conway *et al.*, "Blade computing with the AMD Opteron processor ("magny-cours")," in *Hot Chips 21 Symposium (HCS)*, 2009, pp. 1–19.
- [27] J. Howard *et al.*, "A 48-core IA-32 processor in 45 nm CMOS using on-die message-passing and DVFS for performance and power scaling," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 1, pp. 173–183, 2011.
- [28] T. E. Carlson, W. Heirman, and L. Eeckhout, "Sniper: Exploring the level of abstraction for scalable and accurate parallel multi-core simulation," in *ACM Proc. of International Conference for High Performance Computing, Networking, Storage and Analysis*, 2011, p. 52.
- [29] S. Li *et al.*, "McPAT: An integrated power, area, and timing modeling framework for multicore and manycore architectures," in *IEEE/ACM Proc. of 42nd Annual International Symposium on Microarchitecture (MICRO)*, 2009, pp. 469–480.