A Novel Approach for Cooling Chiplets in Heterogeneously Integrated 2.5-D Packages Applying Microchannel Heatsink Embedded in the Interposer

György Bognár, Gábor Takács, and Péter G. Szabó

***Abstract*— In this article, an innovative approach to enhance heat transfer mechanisms in 2.5-D heterogeneous packages by integrating microchannels into the silicon interposer is presented. An analytical model is introduced to determine the main thermal properties of the cooling system. To demonstrate the cooling performance of the device, a 3-D model is created, and a total of 12 cases for four scenarios are investigated numerically. The results are compared with the classical cooling approach applying 10–100 cm^3/min water flow in the microchannels. It is demonstrated that in the case of a heterogeneously integrated system, incorporating a single-core chiplet as a central processing unit, the temperature dropped from 109 °C to 44 °C at the highest flow rate by applying the combination of the conventional heat sink and the integrated microchannel cooler. It is also shown that regarding a system with a four-core processing unit when only one core has a high load, the core temperature decreases by up to 72.7 °C. The contribution of the enhanced secondary heat flow path is obtained from computational fluid dynamics (CFD) simulation cases. The study shows that the importance of the secondary heat flow path has increased significantly, and the proposed novel cooling system can solve the thermal issues of the heterogeneously integrated 2.5-D devices.**

***Index Terms*— Embedded cooling, heterogeneous packaging, microchannels.**

I. Introduction

THE continuous breach of the "Red Brick Wall"—a term often used by the semiconductor industry [1]—with novel materials and compositions in integrated components together with 3-D architectures [2] results in new thermal design problems. Fin-FETs [3], gate-all-around MOSFETs [4], and modern trench capacitances [5] led us to the next phase of integration when the minimum feature size reached the nanometer scale. These new technologies are often incompatible with each other, and as a consequence, chips with different purposes (e.g., memory, arithmetic unit, and sensors) cannot be fabricated on the same substrate. To overcome this, researchers and electronics package designers are creating novel packaging assemblies that can integrate multiple functions into one package; otherwise, the assembly would require an entire printed circuit board (PCB). Thus, in modern 2.5-D or 3-D electronics packaging, more and more components are placed into a single package forming a complete system. In system-on-chip (SoC) realization, each functional block of a system is placed into a single semiconductor die [6], but in other solutions [e.g., in system-in-package (SiP) or system-on-package (SoP) implementations], these blocks can be placed onto different dies. These dies can usually be realized on different types of semiconductor materials and manufactured with varying semiconductor technology and minimal feature sizes (MFSs). Therefore, each integrated electronic circuit is fabricated with the proper technology and MFS on each die, which is the most outstanding achievement of heterogeneous integration.

With the advance of heterogeneous integration, these differently manufactured discrete chiplets can be placed in the same package on a passive or active silicon interposer [7], [8], [9] which allows high-speed connection between chiplets, unlike in SiP architectures with organic interposers [10], [11].

However, in the case of heterogeneous integration, we have the application as follows:

1) chiplets with diverse geometric parameters made by different semiconductor manufacturing technologies;
2) stacked-die structures;
3) different interconnection technologies (e.g., wire bonding, tape-automated bonding, and flip-chips) can be possible.

In addition, these result in increased dissipation per unit area, leading to cooling problems unlike before.

The traditional method [e.g., through the semiconductor dies itself and through the lid in a flip chip ball grid array (FCBGA) package] appears to pose challenges for heat transfer from the junctions to the ambient. A summary of the most current research on heat transfer can be found in [12]. To overcome these issues, other heat flow paths should be examined to determine whether there are possibilities to enhance their efficiency.

Generally, in the case of modern SoC devices, flip-chip packaging technology is commonly used. The main heat path points from the junction through the semiconductor die toward the backside of the chip and to the heatsink and

Manuscript received 12 May 2023; revised 7 June 2023 and 30 June 2023; accepted 17 July 2023. Date of publication 24 July 2023; date of current version 20 September 2023. This work was supported by the K_20 Grant of the National Research, Development, and Innovation Office (NKFIH) under Project 135224. Recommended for publication by Associate Editor L. Codecasa upon evaluation of reviewers' comments. *(Corresponding author: Péter G. Szabó.)*

The authors are with the Department of Electronics Devices, Budapest University of Technology and Economics, 1111 Budapest, Hungary (e-mail: bognar.gyorgy@vik.bme.hu; takacs.gabor@vik.bme.hu; szabo.peter@vik.bme.hu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCPMT.2023.3298378.

Digital Object Identifier 10.1109/TCPMT.2023.3298378

fan (HSF) assembly. In pure passive cooling solutions, high-thermal conductivity insulators, e.g., AIN or h-BN, can be applied instead of thermally resistive interlayer dielectrics in 3-D ICs [13]. However, this modification reduces the hotspot temperature by 20% only. In [14], a novel microgap dielectric coolant manifold has been designed and studied numerically for 2.5-D stacked integrated circuits with multiple high-power dies. Rajan et al. [15] successfully demonstrated a monolithic microfluidic cooling device for 2.5-D packages, where deionized water was used as a coolant in micropin fin heatsinks etched directly into the backside of the chiplets and covered by 3-D printed manifold.

To ensure the low junction-to-ambient thermal resistance, Hoang et al. [16] designed a 3-D-printed cold plate on the top of the device (a copper block with a top surface area of $1'' \times 1''$ to mimic a computer chip) and characterized it with water as the coolant. They managed to print a strong metal fin structure which can handle the high pressure of the fluid at high flow rates.

The application of phase-change materials (PCMs) as coolants has been extensively studied in the past decades [17], [18]. Bar-Cohen et al. [19] report on the fundamentals of evaporative cooling physics and a numerical modeling approach to enable the co-design of such solutions in emerging computing and communications systems. Wu et al. [20] propose the application of PCM inside the chip package instead of the conventional thermal interface material (TIM) layers. This solution can achieve a better heat transfer from the junction to the case and lower the thermal resistance within the internal structures.

In addition, CMOS compatibility of microchannel forming technology is still an emerging topic. Dang et al. [21] present the fabrication technology, assembly steps, and test results of a silicon chip with CMOS process compatible microchannel heatsink and thermofluidic chip input–output interconnect using wafer-level batch processing for 2-D and 3-D chips, as the microchannel heatsink is fabricated directly on the backside of each chip. One of the most interesting solutions can be the low-temperature electroplating process published in [22] and [23], where microchannels formed by electroplating were used for concentrated photovoltaic cell cooling.

Some researchers proposed a cooling method by applying thermal-sensitive materials (e.g., hydrogel) [24], which can self-adapt to the changes in the thermal conditions (hotspots), improving the thermal transfer inside the cooler.

It should be highlighted that a secondary heat flow path also exists that transfers a significant amount of the heat toward the package substrate and the mainboard, which can act as a heat-emitting structure (a passive cooling structure). Peng et al. [25] addressed the double-side cooling of power modules. They have proved the importance of the secondary heat flow path. Matsumoto et al. [26] presented how to realize the effective cooling from the bottom (substrate) side of chips, assuming a substrate that consists of organic dielectric materials and copper in a 3-D package for high-end server applications.

The heat transfer ratio between the primary (junction-to-case, toward the HSF) and secondary (junction-to-board, toward the PCB) heat flow paths is still to be determined in heterogeneously integrated structures, which is the main objective of some ongoing studies.

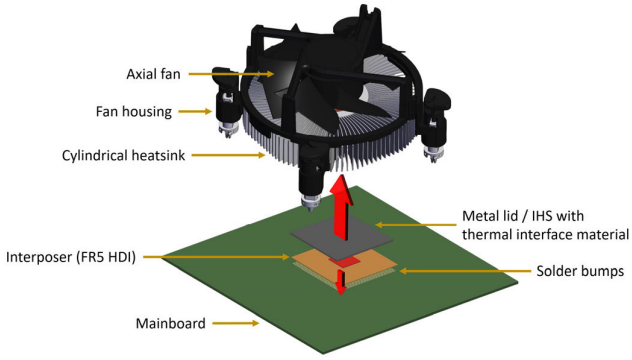

Fig. 1. Exploded view of an FCPGA structure with HSF, where the red arrows indicate the primary and secondary heat flow paths.

Based on these challenges imposed by insufficient cooling, our team proposes a novel concept to improve the heat transfer in modern—heterogeneous—packages through the secondary heat flow path. The objective of our research is to enhance junction-to-board heat transfer by creating a simple stack-up and integrating microchannel structures into the interposer layer. This article presents the investigation of the effect of these microchannels on circulating fluid in heterogeneously integrated systems by an industrial electronics cooling numerical simulator.

In [27] and [28], the fabrication and characterization of embedded microchannel structures were presented in detail. These microstructures were fabricated directly in the backside of the active silicon semiconductor dies by applying CMOS-compatible wet chemical etching process steps [27]. Hence, the heat conduction path between the junction and the wall of the microchannels, where the heat transfer happens between the die and the circulated fluid material, will be shortened by avoiding several TIM layers. The main aim was to avoid using the compounds of alkali metal materials, like potassium and sodium, due to their tendency to interstitially diffuse into the silicon, causing changes in the threshold voltage of MOS-FETs.

In the case of traditional FCBGA packaging structures (Fig. 1), the majority of the heat is conducted from the junction to the backside of the die itself, then through a couple of TIM layers and the metal lid—also called integrated heat spreader or IHS—to the heatsink. Finally, through the walls of the warmed-up heatsink, the heat is transferred to the circulated air. Since a common heatsink has a couple of square meters of surface area, the typical R_{thjc} value is less than 0.5 K/W.

It is important to highlight that not only one but also at least two significant heat flow paths exist in FCPGA structures. A secondary heat flow path can be identified between the active area and the ambient, where the heat flows through the microbumps, the interposer, and the bumps toward the mainboard, which in this case can be considered as a heatsink. However, the amount of dissipated heat flowing in this direction is not determined and still poses many questions.

Forming a microchannel heatsink structure in the backside of the flip-chip itself is applicable only in those cases when the traditional heatsink has to be replaced, e.g., in

space-constrained applications or in rack devices. The heat transfer coefficient will be significantly higher, but the heat transfer area will be decreased by at least one or two orders of magnitude. If the flip-chip technique cannot be applied (e.g., in a simple dual-in-line package), the microchannel-based cooling solutions could also be a good answer for thermal challenges. In these cases, the die is placed on a lead frame, and the electrical connections are created by applying different wirebond techniques. The heat can only be transferred to the ambient by conduction (through the lead frame toward the pins) and by natural convection. If microchannels are formed, R_{thja} can be decreased significantly by presenting a secondary heat flow path.

In the frame of our preceding research work, it was proven that embedded microscale cooling solutions could effectively improve the overall junction to ambient heat flow by increasing the heat transfer through the secondary heat flow path toward the microchannels. In [27], the manufacturing process for fabricating microchannel structures in silicon dies based on wet etching process steps was presented in detail.

In the case of 2.5-D or 3-D packaging, flip-chip and wirebond techniques are commonly and alternately applied on a common interposer. However, the traditional FCBGA package structure cannot be applied. The main problem is the varying height of the dies placed on the common interposer. In this case, the TIM layers with different thicknesses should be applied to realize thermal contact between the die and the common metal lid/IHS. The TIM layer has the worst heat conductance in the traditional heat flow path. Hence the thinnest possible layer would have to be chosen. However, another problem also arises. If microchannels are fabricated in flip-chip devices, the fluid supply is difficult to manage. Either a double-walled lid is applied, or inlets/outlets are opened through the metal lid traditional heatsink, and fan structures cannot be used anymore.

The only possible solution is to create a secondary heat flow path while maintaining the traditional FCBGA packaging technique. In the case of 2.5-D packaging and heterogenous integration, a silicon or organic interposer can be applied on which several dies with different thicknesses and with different bonding (e.g., flip-chip, wire bond) techniques are placed. One possible way is to enhance the role or create a secondary heat flow path by applying microchannel-based cooling. It is important to place these microchannels as near to the junction as possible to keep the overall thermal resistance between the junction and the ambient – in this case, between the junction and the fluid circulated in the channels – as low as possible.

The previously presented methodology for fabricating embedded microchannel structures in a silicon die is obviously applicable to silicon interposers. In the case of organic (e.g., FR5) interposers, additional process steps are necessary to fabricate channels in the epoxy-based material. In the first step, trenches are created by mechanical drilling using a benchtop milling machine in a pure epoxy glass fiber board (mainly FR4 material). Then, this board is stuck to the FR4/FR5 board or interposer by applying epoxy-based adhesives to form the channels.

However, several new design problems have to be dealt with for the application of microchannels. The dimensions of the side edges of the interposer are an order of magnitude larger than the side edge of the chips, typically 20–30 mm. That is why the lengths of the channels are also longer than in the case of in-chip realization, resulting in increased hydrodynamic resistance and higher pressure drop. Therefore, wider and deeper channels or more channels should be used.

Furthermore, it must be taken into account that in the case of parallel channels, the dimensions of each channel should be equal to the others: the length and the characteristic dimensions should be the same. It is essential to avoid that the mass flow rates in the channels are significantly different due to different hydrodynamic resistances of the channels. Different mass flow rates and velocities result in different heat transfer coefficients (h) and heat transfer. It could happen that different channel lengths result in only a single channel dominating the heat transfer. That is why hydrodynamic modeling is essential in the phase of designing the channels' layout and architecture.

These hydrodynamic design aspects should be considered at the first step of planning the microchannel structure to obtain an effective cooling solution. However, besides these aspects, it is also important to model and investigate the thermal behavior of the structure.

The aim of our work is to present a novel approach to enhance the heat transfer by applying integrated microscale channels embedded in the silicon interposer in 2.5-D packages, which will serve as fluid cooling vehicles. Section II presents an analytical model prior to the numerical simulation that will also be used to create equivalent circuit models of the heat transfer, besides cross-verifying the numerical results. Section III presents the approach to the new cooling concept together with the relevance of the secondary heat flow path. Section IV shows the full extent of the envisioned embedded cooling vehicle together with the 2.5-D heterogenous packaging and different cases of the numerical investigations and setup. The results and comparisons of our analytical and numerical investigations are presented in Section V, followed by the conclusion.

II. Analytical Model

Applying thermal modeling or simulation tools in the earliest phase of thermal-aware design flow is inevitable to obtain pressure drop, flow rate, and thermal resistance values as fast as possible, since it is a key factor for proper design and time-to-market. The applied process steps and technology determine the design rules (e.g., the minimum and maximum width, depth, and length), but the layout of the channels is based on preliminary calculation and modeling.

An analytical electrothermal–hydrodynamic modeling tool was developed to evaluate the thermal and hydrodynamic behavior of the structure with embedded channels. With these models, the dimensions and layout of the channels can be determined several times faster than by using any computational fluid dynamics (CFD) tools.

In order to derive the analytical model of heat transfer between the channel walls and the coolant, it was necessary to break down each channel into small segments and calculate the Nusselt number and heat transfer coefficient values for each segment. The main aim is to determine the thermal resistance between the walls of a single channel and the fluid.

There are differences in effective heat transfer values not only among the parallel channels but also within different

segments of a single channel. Near the inlets, the local Nusselt number and, hence, the heat transfer coefficient are significantly higher. h heat transfer coefficient can be calculated as follows:

$$h = \frac{k_f \cdot Nu}{D_H} \tag{1}$$

where k_f is the thermal conductivity of the fluid, and D_H is the hydraulic diameter [29]. The D_H hydraulic diameter can be determined for a square-based column as follows:

$$D_H = \frac{2 \cdot a \cdot b}{a + b}. \tag{2}$$

There are several analytical equations [30] to determine the average Nusselt number of a single channel, which depends on the cross-sectional geometry and the channel length; hence, the average heat transfer coefficient of a single channel can be determined. In our investigations, the laminar-type flow was typical in trapezoidal cross-shaped channel structures. All parallel channels have the same inlet and outlet points. Based on these assumptions, the Nusselt number can be calculated as follows:

$$\mathrm{Nu} = 5.14 + \frac{0.065 \cdot (D_H/L) \cdot \mathrm{Re} \cdot \mathrm{Pr}}{1 + 0.04 \cdot \left[(D_H/L) \cdot \mathrm{Re} \cdot \mathrm{Pr}\right]^{\frac{2}{3}}} \tag{3}$$

where L is the length of the channel, Re is the Reynolds number, and Pr is the Prandtl number, which can be calculated as follows:

$$\mathrm{Re} = \frac{D_H \cdot \frac{\mathrm{d}m}{\mathrm{d}t}/A}{\mu} = \frac{D_H \cdot \rho \cdot \frac{\mathrm{d}V}{\mathrm{d}t}/A}{\nu} = \frac{D_H \cdot \rho \cdot v}{\nu} \tag{4}$$

$$\mathrm{Pr} = \frac{c_p \cdot \nu}{k_f} \tag{5}$$

where ν is the dynamic, μ is the kinematic viscosity of the fluid, ρ is the density of the fluid, L is the length of the investigated channel, and v is the mean velocity of the fluid.

If we focus only on a dx length segment of the channel, the heat flows through the walls of the channel and warms up the fluid. Based on the conservation of energy, we can write

$$h \cdot \left(T_w - T_f\right) \cdot p \cdot \mathrm{d}x = \frac{\mathrm{d}m}{\mathrm{d}t} \cdot c_p \cdot \mathrm{d}T_f \tag{6}$$

where c_p is the specific heat of the fluid, dT_f is the temperature difference of the dm mass fluid when it enters and leaves the dx length segment of the investigated channel per unit time, dm/dt is the constant mass flow rate of the fluid, and T_w and T_f are the walls and fluid temperature, respectively.

After the rearrangement of (6), solving the resulting linear differential equation, the thermal resistance between the walls of a single channel and the fluid can be calculated as follows [32]:

$$R_{\mathrm{th\_uch}} = \frac{1}{\frac{\mathrm{d}m}{\mathrm{d}t} \cdot c_p \cdot \left(1 - e^{-\frac{h \cdot A}{\frac{\mathrm{d}m}{\mathrm{d}t} \cdot c_p}}\right)}. \tag{7}$$

It can be seen that $R_{\mathrm{th\_uch}}$ depends on the mass flow rate, A is the heat exchange area (the total sum of the walls' area), c_p is the specific heat of the fluid, and the heat transfer coefficient.

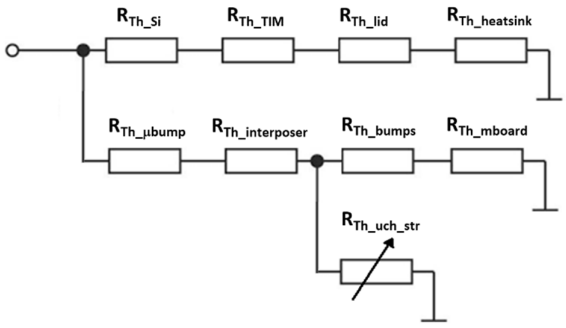

Fig. 2. Lumped element model of the heat map.

The overall $R_{\mathrm{th\_uch\_str}}$ value can be determined for parallel microchannel structures by adding the $G_{\mathrm{th\_uch}}$ conduction values of each channel and calculating the reciprocal value [32]

$$R_{\mathrm{th\_uch\_str}} = \frac{1}{\sum_{i=1}^{N} G_{\mathrm{th\_uch}\_i}}. \tag{8}$$

In [27], a novel methodology was presented to determine the $R_{\mathrm{th\_uch}}$ partial thermal resistance and cross-verify the analytically calculated and simulated values. This novel methodology is based on Joint Electron Device Engineering Council (JEDEC) JESD51 1-14 standard test method [31]. The successful validation showed that the maximum deviation between the analytical model and measured results is below 8%, but at high flow rates, the deviation decreases below 2%.

Based on these assumptions and calculations, an analytical model was built, and a hydrodynamic–thermal modeling tool was developed in ANSI C [32]. The hydrodynamic properties and the $R_{\mathrm{th\_uch}}$ thermal resistance of each channel can be determined within seconds. During the calculation of the pressure drops, minor losses are also taken into account. These minor losses are based on the local energy losses caused by the disruption of the flow due to possible inlets, outlets, bends, expansions, and contractions of the channel.

In the case of embedded microchannel heatsink structures, the heat conduction map can be generated (Fig. 2) by identifying each resistance. It can be clearly seen that there are two main heat flow paths. In the case of a standard FCBGA package without microchannel cooling structures, the primary heat flow path is toward the heatsink and the fan. Without a heatsink, the secondary heat flow path toward the motherboard becomes significant. In this latter case, heat can be transferred to the ambient by natural convection only, which means a low-heat transfer coefficient (h = 5–6 W/m^2·K), increasing the thermal resistance between the lid and the ambient, which at last results in elevated chiplet temperatures.

The embedded microchannel structures in the interposer can decrease the overall $R_{\mathrm{th\_ja}}$ by adding a low-thermal resistance parallel branch to the secondary heat flow path. Its value depends on the mass or volumetric flow rate; that is why the variable resistance symbol was used in Fig. 2.

III. Relevance and Enhancement of the Secondary Heat Flow Path

Removing the dissipated heat from the chiplets is one of the most challenging problems in modern 2.5-D and 3-D packaged systems. Currently, the application of stacked-die

structures and dies with different thicknesses makes the secondary heat flow path more relevant. Using TIMs with varying thicknesses between the chiplets and the lid [33] seems to be an easy solution. However, the limited thermal conductivity of TIMs leads to increasing partial thermal resistance and results in elevated junction temperatures, especially where higher distances should be bridged. A more effective technique to avoid elevated temperatures would be the application of integrated microscale cooler structures within the interposer or the package substrate itself. This way, the balance between the traditional main and secondary heat flow path will be shifted.

The secondary heat flow path thermal resistance depends mainly on the following:

1) the physical properties of materials and dimensions of the interposer and package substrate;
2) the type of materials and dimension of the joints between the chiplets and interposers;
3) number per unit area of through silicon via (TSV); number per unit area of the joints (bumps, c2c).

Furthermore, by the application of integrated microchannel structures, other parameters will also influence the cooling performance of the secondary heat flow path:

1) sizes and cross-sectional shape of the channels;
2) the number and length of the microscale channels;
3) physical properties of the applied fluid (e.g., thermal conductivity and specific heat, density, and dynamic viscosity);
4) the flow rate of the working fluid (volumetric or mass flow).

IV. Stack-Up and Simulation Models

During our investigations, the numerical model (3)-D and thermal model) of a 2.5-D integrated heterogeneous system with embedded microchannels was created and setup. A conventional FCBGA with multiple dies placed on an organic interposer formed the base of the system, which is placed onto a printed-wired board (PWB) that meets JEDEC-JESD-51-7 standards [34]. A chiplet/die arrangement was created based on Intel Kaby Lake-G architecture [35] inside the package. Three dies are placed with different designations: a CPU, a GPU, and a memory module. These chiplets were placed on a silicon interposer, in which microchannel structures were formed.

The simulation model was built up in Simcenter Flotherm, which is an industrial electronics cooling CFD designer tool by Siemens [36]. The test board is a 2s2p type JEDEC high-thermal conductivity card with two signal layers (top and bottom layers) and two internal power planes (VDD and GND) made of copper. These internal power layers can help distributing the heat from the package with an area of 30 × 30 mm. The package has 480 pieces of SnAg bumps (SAC305) represented as cuboids with 500 × 500 μm size and a thickness of 400 μm. In the center, a bank without bumps was formed, which means that the banks or fields of bumps were only created near the sides of the package. The main reason is that microchannels are formed in the interposer directly underneath the silicon dies; thus, no space is available for realizing dense copper connections.

The thickness of the interposer is 400 μm, including the 70-μm-deep and 170-mm-long parallel microchannels. At the inlet and outlet, there are two channels with a 1-mm width, forming a common rail. These are distributed into 61 pieces of 200-μm-wide channels in the middle, and two pieces of 300-μm-wide parallel microchannels at the sides over the size of 24.5× 24.5 mm^2 (see Fig. 3).

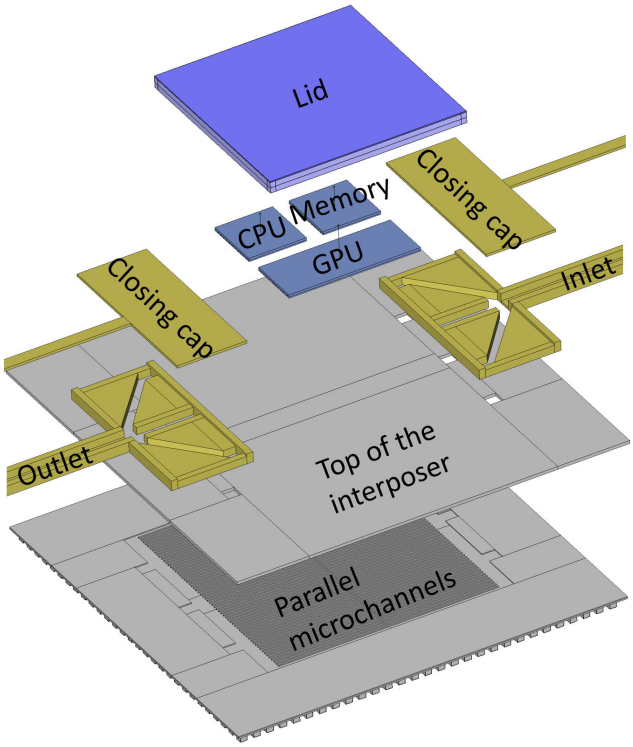

Fig. 3. Exploded view of the heterogeneous package.

The chiplets are connected to the interposer by SnAg microbumps in a 150 × 150 μm^2 area with 120-μm thickness. These microbumps are uniformly placed below the dies with SU-8 underfill. Over them, each die has a thickness of 300 μm with active area of 10 μm in thickness. The size of the GPU is 6.25 × 16 mm^2 with a total power of 30 W, while both the CPU and the memory module have a size of 7 × 7 mm^2; the CPU has a total dissipated power of 30 W, and the memory module operates at 5 W. To enable mounting the active fan-based cooling assemblies, the chiplets are covered by a 500-μm-thick stainless steel (Cr18/Ni8) lid with an 80-μm-thick TIM layer (Laird Tech-Tgrease 980) for good thermal contact (low thermal resistance). The thermal conductivity values of the materials mentioned can be seen from Table I.

During the mesh setup for the 3-D stack-up, special attention was paid in order to achieve appropriate resolution inside the channels and the high-aspect-ratio layers that can be found in the heat flow path. The minimal number of cells for each layer was at least 8 and 10 for the fluid domain. With a 4-mm maximum cell size, the whole domain can feature more than ten million cells.

Open boundary conditions around the simulation domain were set, meaning that the air can freely leave and enter the domain. There is a 38-mm air gap left over the lid up to the boundary to have space for the air to develop fully. The cooling

TABLE I
MATERIAL PROPERTIES

Material	Thermal conductivity [W/(m·K)]
FR4 (JEDEC test board dielectric material)	0.3
Copper (plane layers inside the JEDEC test board)	385
13.8% copper coverage (top layer inside the JEDEC test board)	55.2
20% copper coverage (bottom layer inside the JEDEC test board)	80
Silicon (interposer and dies)	150 at 27°C 99 at 127°C 62 at 327°C
SnAg (bumps, microbumps)	78.4
Underfill	0.2
Laird Tech - Tgrease 980 thermal interface material	3.8
Stainless steel (Cr18/Ni8) (lid)	16.3
Aluminium (heatsink)	201
ID-TG25 (thermal interface material)	10.5

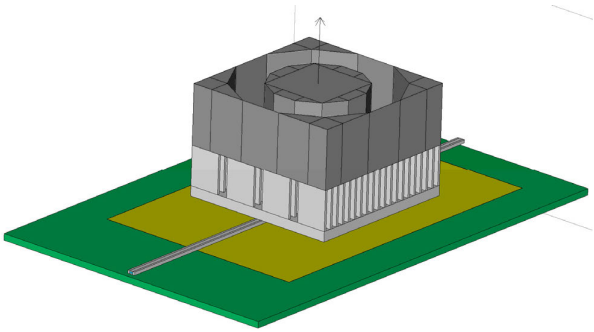

Fig. 4. HSF assembly over the package.

fluid is deionized water at 20 °C with volumetric flow rates of 0, 10, and 100 cm^3/min.

Four different scenarios were investigated and are presented in this article.

In the first scenario, only a common lid covered the chiplets, and no additional heatsink was applied. In this case, it was supposed that only chiplets with equal thickness were used, and only flip-chip interconnection technology was applied.

In the second scenario, an additional HSF cooler assembly was placed on top of the package to increase the heat transfer toward the lid of the package (Fig. 4). The heatsink is a 40 × 40 mm pin-fin aluminum structure with a base thickness of 2.5 mm, and 10 mm fin height in a 4 × 20 fin arrangement. A 0.1-mm-thick ID-TG25 TIM was defined between the heatsink and the package. On top of them, a Sanyo Denki 40 × 40 × 15 mm nonlinear axial fan was mounted.

In the third and fourth scenarios, the CPU was considered as a four-core processing unit, dividing its overall area into four equal dissipating parts. The architecture was implemented imposing a power limitation of 30 W on the whole CPU and assuming that only one core (designated as Core No. 1) was mostly active, dissipating 70% of the maximum overall power (21 W) while the other three cores (designated as Core Nos. 2–4), mostly inactive, only dissipate 10% each (3 W) of the maximum overall power. This power profile presents a common limitation when individual cores would be suitable to operate on higher performance (higher frequency), but the temperature limitations pose constraints due to the high dissipation.

TABLE II
JUNCTION TEMPERATURES OVER THE CHIPLETS AT THE GIVEN FLOW RATE FOR THE FIRST SCENARIO (OTL—OVER-THE-LIMIT: 300 °C+)

Volumetric flow rate [CCM]	CPU junction temperature min – max [°C]	GPU junction temperature min – max [°C]	Memory junction temperature min – max [°C]
0	OTL	OTL	OTL
10	107 – 150	56.6 – 123	42.5 – 64.2
100	49.2 – 60.1	33.5 – 46.2	24.6 – 30.6

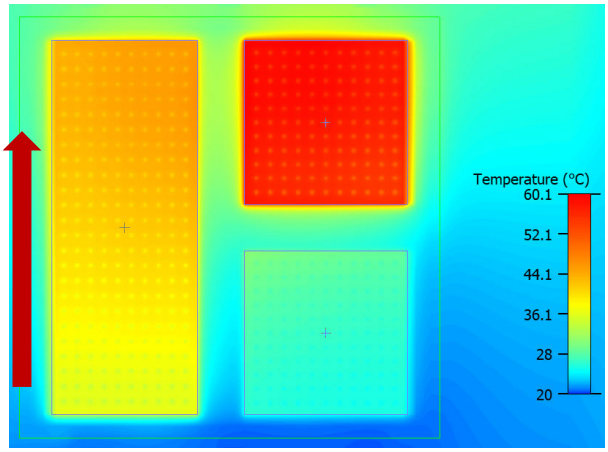

Fig. 5. Temperature distribution along the dies in the first scenario. (Red arrow indicates the direction of the fluid flow.)

V. RESULTS

In this work, three cases (fluid flow at different flow rates) of the four scenarios have been investigated (overall 12 cases for four scenarios). The simulation results show that a significant amount of heat can be transferred by the fluid flow through the microchannels. The results for the first scenario when no heatsink is applied on the top of the assembly can be found from Table II, while the temperature distribution can be observed from Fig. 5.

It can be seen that for the three cases examined (0–10–100 cm^3/min), the junction temperatures significantly decrease with increasing flow rate. Furthermore, since the heat transfer efficiency increases by increasing the volumetric flow rate, the temperature starts to increase significantly if the volumetric flow rate drops below a certain level. In Fig. 6, the pressure drop can be observed, where the difference between the inlet and the outlet is approximately 2 × 10^5 Pa. In Fig. 7, the fluid velocity distribution can be seen where, thanks to the uniform pressure distribution caused by the wide main channels at the inlet and outlet, the speed in each channel is in the same range. It should be noted that the cases for 0 cm^3/min in the second and fourth scenarios can be used to compare against traditional HSF-based cooling, since the fluid does not transfer any heat in these cases.

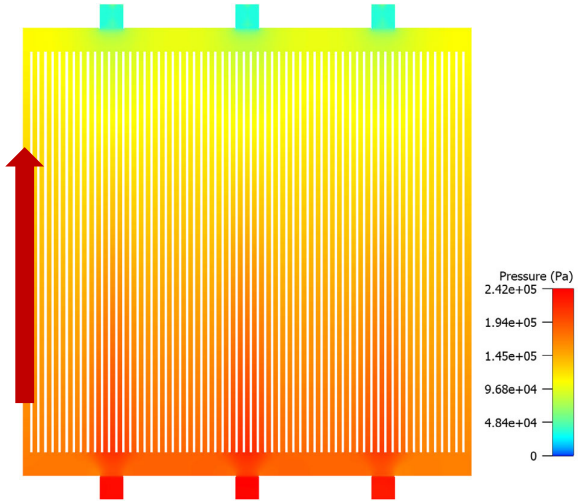

Fig. 6. Pressure drop along the microchannels at 100 cm^3/min volume flow rate in the first scenario.

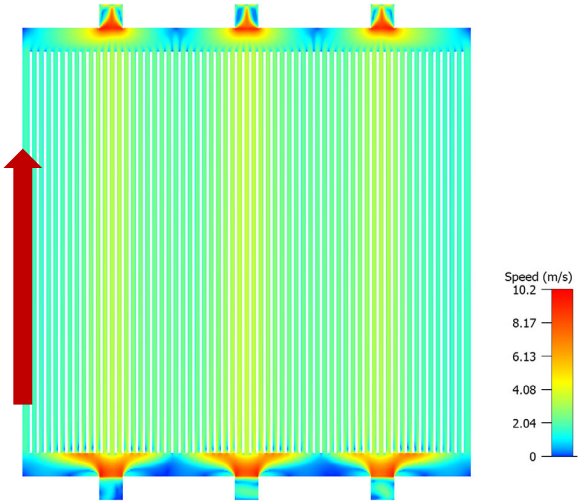

Fig. 7. Fluid velocity in the microchannels at 100 cm^3/min volume flow rate in the first scenario.

TABLE III
JUNCTION TEMPERATURES IN THE MIDDLE OF THE CHIPLETS WITH THE GIVEN FLOW RATE FOR THE FIRST SCENARIO (MICROCHANNELS ONLY) COMPARED WITH THE SECOND SCENARIO (MICROCHANNELS AND HSF ONLY)

	Microchannel			Microchannel + HSF		
Vol. flow rate [ccm]	CPU temp. [°C]	GPU temp. [°C]	Memory temp. [°C]	CPU temp. [°C]	GPU temp. [°C]	Memory temp. [°C]
0	OTL	OTL	OTL	109	93	79.5
10	134	97	50.5	81	65.5	48
100	54	42	26	44	37	27

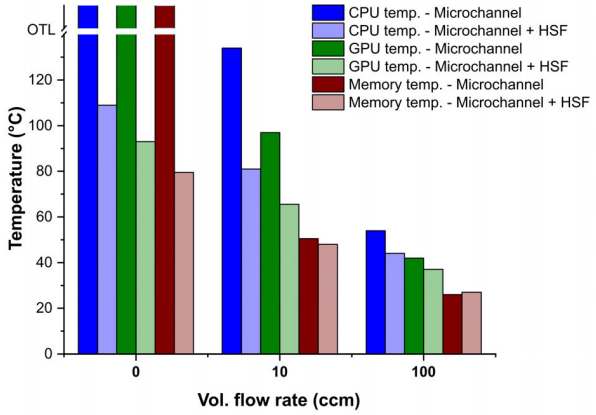

Fig. 8. Junction temperatures in the middle of the chiplets with the given flow rate for the first scenario compared with the second scenario.

The results obtained with the in-house hydrodynamic–thermal modeling tool are in good agreement with the results of the simulation steps. The average Reynolds number is 183; thus, laminar flow type was assumed. At 100-cm^3/min constant volumetric flow rate, the average fluid velocity is around 2.2 m/s in each channel, and the pressure drop along the channels is $\sim 1.04 \times 10^5$ Pa. The calculated cumulative partial thermal resistance of microchannel structure $R_{\text{th\_uch\_str}}$ is 1.438 K/W at 10 cm^3/min and 0.264 K/W at 100-cm^3/min volumetric flow rates, which means the heat conduction increased significantly toward the ambient. Based on the simulation results, the partial thermal resistances of the microchannel structure are the following: 1.58 K/W at 10 cm^3/min and 0.31 K/W at 100 cm^3/min. The difference between the analytical model and the simulation is caused by neglecting the spreading thermal resistance. Spreading resistance occurs when a small heat source is in contact with a larger cuboid; in this case, with the interposer layer; and as a consequence, the heat does not distribute uniformly through the interposer, and consequently, it does not transfer the heat efficiently to the walls of the microchannels for convective cooling. In our calculations, we supposed that toward the lid there is no heat flow, so the heat extracted by natural convection from the lid surface was neglected.

The second scenario provides an excellent opportunity to observe the impact of an additional heat flow path on the assembly. Table III and Fig. 8 show the maximum temperatures in the center of the chiplets as well as a reduction in junction temperatures compared with the first scenario. This is because the heat flow is distributed between the two main heat paths.

Fig. 9 presents the temperature distribution for the fourth scenario, when the CPU is divided into four cores, each with its own dissipation. In comparison with the second scenario, the maximum temperature is higher, since 70% of the power of the CPU is now applied to one core, which has 25% of the total area. It is also visible, that the highest temperature is at the bottom left corner of the CPU in Fig. 9, since heat cannot be removed as effectively at the corner of the chiplet as in the middle due to the decreased effective cross-sectional area toward the bumps at the bottom and the lid on the top.

Similar to Table III and Fig. 8, Table IV and Fig. 10 show the junction temperatures in the middle of the chiplets if only microchannels provide the cooling (third scenario), or an HSF is also present beside the microchannels (fourth scenario). The

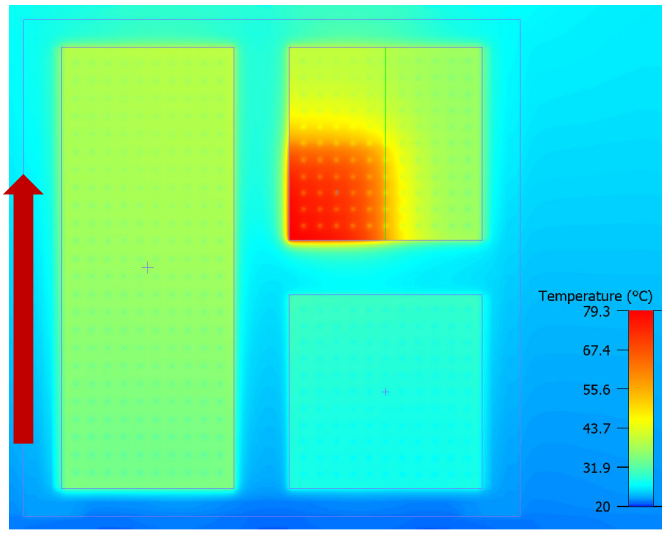

Fig. 9. Temperature distribution along the dies in the fourth scenario. (Red arrow indicates the direction of the fluid flow.)

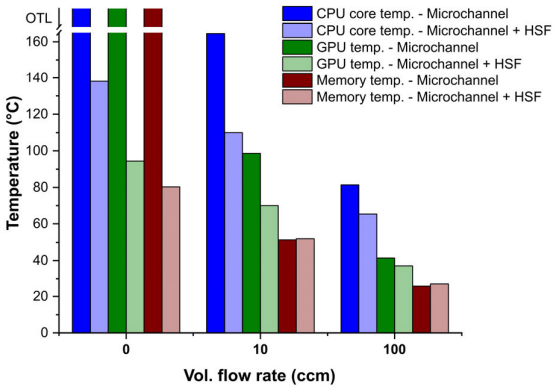

Fig. 10. Junction temperatures in the middle of the chiplets/Core No. 1 with the given flow rate for the third scenario compared with the fourth scenario.

TABLE IV

JUNCTION TEMPERATURES IN THE MIDDLE OF THE CHIPLETS/CORE NO. 1 WITH THE GIVEN FLOW RATE FOR THE THIRD SCENARIO (MICROCHANNELS ONLY) COMPARED WITH THE FOURTH SCENARIO (MICROCHANNELS AND HSF)

	Microchannel			Microchannel + HSF		
Vol. flow rate [ccm]	CPU core temp. [°C]	GPU temp. [°C]	Memory temp. [°C]	CPU core temp. [°C]	GPU temp. [°C]	Memory temp. [°C]
0	OTL	OTL	OTL	138.1	94.4	80.3
10	164.5	98.6	51.4	110	70	52
100	81.4	41.4	26	65.4	37.1	27.2

temperature monitor point in the CPU is placed at the center of Core No. 1. In comparison with Table III and Fig. 8, it can be seen that the temperatures of the GPU and the memory module are almost the same, but the temperatures of the Core No. 1 are significantly higher (about 20 °C–30 °C than in the first and second scenarios).

VI. CONCLUSION

In this article, a novel approach for thermal management for heterogeneously integrated 2.5-D packaging is presented based on forming microchannel structure inside the interposer.

Our results have proved that our previously presented chip-level, microscale embedded cooling solution can also be effective in these kinds of integrated structures, even if it is not fabricated in the backside of the dies itself, but in the silicon interposer layer. The role of the secondary heat flow path was found to be more important in the case of the chiplets fabricated with different manufacturing technologies and large total package power dissipation. This highlights the importance of the novel cooling solution presented in this article.

REFERENCES

[1] J. Van Houdt, "3D memories and ferroelectrics," in *Proc. IEEE Int. Memory Workshop (IMW)*, May 2017, pp. 1–3, doi: 10.1109/IMW.2017.7939066.

[2] S. Razdan, P. D. Dobbelaere, J. Xue, A. Prasad, and V. Patel, "Advanced 2.5D and 3D packaging technologies for next generation silicon photonics in high performance networking applications," in *Proc. IEEE 72nd Electron. Compon. Technol. Conf. (ECTC)*, May 2022, pp. 428–435, doi: 10.1109/ECTC51906.2022.00075.

[3] A. Goel, A. Rawat, and B. Rawat, "Benchmarking of analog/RF performance of Fin-FET, NW-FET, and NS-FET in the ultimate scaling limit," *IEEE Trans. Electron Devices*, vol. 69, no. 3, pp. 1298–1305, Mar. 2022, doi: 10.1109/TED.2021.3140158.

[4] V. B. Sreenivasulu and V. Narendar, "Design insights of nanosheet FET and CMOS circuit applications at 5-nm technology node," *IEEE Trans. Electron Devices*, vol. 69, no. 8, pp. 4115–4122, Aug. 2022, doi: 10.1109/TED.2022.3181575.

[5] P. S. Parkinson et al., "Novel techniques for scaling deep trench DRAM capacitor technology to 0.11 μm and beyond," in *Proc. Int. Symp. VLSI Technol., Syst. Appl.*, 2003, pp. 21–24, doi: 10.1109/VTSA.2003.1252541.

[6] J. H. Lau, *Fan-Out Wafer-Level Packaging*. Singapore: Springer, 2018.

[7] D. Stow, Y. Xie, T. Siddiqua, and G. H. Loh, "Cost-effective design of scalable high-performance systems using active and passive interposers," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Design (ICCAD)*, Nov. 2017, pp. 728–735.

[8] J. Kim et al., "Architecture, chip, and package co-design flow for 2.5D IC design enabling heterogeneous IP reuse," in *Proc. 56th ACM/IEEE Design Autom. Conf. (DAC)*, Jun. 2019, pp. 1–6.

[9] P. Coudrain et al., "Active interposer technology for chiplet-based advanced 3D system architectures," in *Proc. IEEE 69th Electron. Compon. Technol. Conf. (ECTC)*, May 2019, pp. 569–578, doi: 10.1109/ECTC.2019.00092.

[10] J. H. Lau, "Chiplet heterogeneous integration," in *Semiconductor Advanced Packaging*. Singapore: Springer, pp. 413–439, 2021.

[11] G. Shan, Y. Zheng, C. Xing, D. Chen, G. Li, and Y. Yang, "Architecture of computing system based on chiplet," *Micromachines*, vol. 13, no. 2, p. 205, 2022, doi: 10.3390/mi13020205.

[12] S. S. Salvi and A. Jain, "A review of recent research on heat transfer in three-dimensional integrated circuits (3-D ICs)," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 11, no. 5, pp. 802–821, May 2021, doi: 10.1109/TCPMT.2021.3064030.

[13] Ç. Köroglu and E. Pop, "High thermal conductivity insulators for thermal management in 3D integrated circuits," *IEEE Electron Device Lett.*, vol. 44, no. 3, pp. 496–499, Mar. 2023, doi: 10.1109/LED.2023.3240676.

[14] Y. Hu and Y. Joshi, "Single-phase microfluidic cooling of 2.5D-SICs for heterogeneous integration," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 10, no. 9, pp. 1499–1506, Sep. 2020, doi: 10.1109/TCPMT.2020.2994584.

[15] S. K. Rajan, A. Kaul, T. E. Sarvey, G. S. May, and M. S. Bakir, "Monolithic microfluidic cooling of a heterogeneous 2.5-D FPGA with low-profile 3-D printed manifolds," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 11, no. 6, pp. 974–982, Jun. 2021, doi: 10.1109/TCPMT.2021.3082013.

[16] C. H. Hoang et al., "Design and thermal analysis of a 3-D printed impingement pin fin cold plate for heterogeneous integration application," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 12, no. 7, pp. 1091–1099, Jul. 2022, doi: 10.1109/TCPMT.2022.3185401.
[17] W. Hua, L. Zhang, and X. Zhang, "Research on passive cooling of electronic chips based on PCM: A review," *J. Mol. Liquids*, vol. 340, Oct. 2021, Art. no. 117183, doi: 10.1016/j.molliq.2021.117183.
[18] C. H. Hoang et al., "A review of recent developments in pumped two-phase cooling technologies for electronic devices," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 11, no. 10, pp. 1565–1582, Oct. 2021, doi: 10.1109/TCPMT.2021.3117572.
[19] A. Bar-Cohen et al., "The ICECool fundamentals effort on evaporative cooling of microelectronics," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 11, no. 10, pp. 1546–1564, Oct. 2021, doi: 10.1109/TCPMT.2021.3111114.
[20] J. Wu et al., "Design and test of a low junction-to-case thermal resistance packaging method," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 11, no. 12, pp. 2130–2139, Dec. 2021, doi: 10.1109/TCPMT.2021.3127954.
[21] B. Dang, M. S. Bakir, D. Chandra Sekar, C. R. King, and J. D. Meindl, "Integrated microfluidic cooling and interconnects for 2D and 3D chips," *IEEE Trans. Adv. Packag.*, vol. 33, no. 1, pp. 79–87, Feb. 2010, doi: 10.1109/TADVP.2009.2035999.
[22] A. Koyuncuoğlu, R. Jafari, T. Okutucu-Özyurt, and H. Külah, "Heat transfer and pressure drop experiments on CMOS compatible microchannel heat sinks for monolithic chip cooling applications," *Int. J. Thermal Sci.*, vol. 56, pp. 77–85, Jun. 2012, doi: 10.1016/j.ijthermalsci.2012.01.006.
[23] G. Rózsás, G. Bognár, G. Takács, P. G. Szabó, and B. Plesz, "Process and measurement of electroplated back-contact integrated microchannel cooling devices for CPV cells," in *Proc. 26th Int. Workshop Thermal Investigations ICs Syst. (THERMINIC)*, Sep. 2020, pp. 1–6.
[24] X. Li and Y. Xuan, "Embedded cooling of high-heat-flux hotspots using self-adaptive microchannel/pin-fin hybrid heat sink," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 12, no. 10, pp. 1627–1635, Oct. 2022, doi: 10.1109/TCPMT.2022.3211009.
[25] C. Peng et al., "Investigation on the reliability of die-attach structures for double-sided cooling power module," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 11, no. 5, pp. 793–801, May 2021, doi: 10.1109/TCPMT.2021.3065704.
[26] K. Matsumoto, H. Mori, and Y. Orii, "Thermal performance evaluation of dual-side cooling for a three-dimensional (3D) chip stack: Additional cooling from the laminate (substrate) side," in *Proc. Int. Conf. Electron. Packag. (ICEP)*, Apr. 2016, pp. 163–168, doi: 10.1109/ICEP.2016.7486804.
[27] G. Takács, G. Bognár, E. Bándy, G. Rózsás, and P. G. Szabó, "Fabrication and characterisation of microscale heat sinks," *Microelectron. Rel.*, vol. 79, pp. 480–487, Dec. 2017, doi: 10.1016/j.microrel.2017.05.028.
[28] G. Takács, P. G. Szaó, and G. Bognár, "Enhanced thermal characterisation method of microscale heatsink structures," *Microelectron. Rel.*, vol. 67, pp. 21–28, Dec. 2016, doi: 10.1016/j.microrel.2016.09.019.
[29] A. Y. Çengel and R. H. Turner, *A Fundamentals of Thermal-Fluid Sciences*, 2nd ed. London, U.K.: McGraw-Hill, 2004.
[30] G. L. Morini, "Single-phase convective heat transfer in microchannels: A review of experimental results," *Int. J. Thermal Sci.*, vol. 43, no. 7, pp. 631–651, Jul. 2004, doi: 10.1016/j.ijthermalsci.2004.01.003.
[31] *JESD51 1-14: Methodology for the Thermal Measurement of Component Packages (Single Semiconductor Device)*, JEDEC Standard JESD51, Arlington, VA, USA, JEDEC Solid State Technology Association, 2010.
[32] Gy. Bognár, G. Takács, L. Pohl, and P. G. Szabó, "Thermal modeling of integrated microscale heatsink structures," *Microsyst. Technol.*, vol. 24, pp. 433–444, Feb. 2017.
[33] *Heterogeneous Integration Roadmap 2021 Edition*, IEEE Electron. Packag. Soc., Piscataway, NJ, USA, , ch. 20. [Online]. Available: https://eps.ieee.org/images/files/HIR_2021/ch20_thermal1.pdf
[34] *High Effective Thermal Conductivity Test Board for Leaded Surface Mount Packages*, JEDEC Standard N°51–7, Arlington, VA, USA, JEDEC Solid-State Technology Association, 1999. Accessed: Feb. 23, 2023. [Online]. Available: https://www.jedec.org/system/files/docs/jesd51-7.PDF
[35] S. Chennupaty, "Intel's high performance graphics solutions in thin and light mobile form factors," in *Proc. IEEE Hot Chips, A Symp. High Perform. Chips*, Aug. 2018, pp. 19–21.
[36] *Product Page of Siemens' Simcenter Flotherm CFD Simulation Tool*. Accessed: Feb. 23, 2023. [Online]. Available: https://www.plm.automation.siemens.com/global/en/products/simcenter/flotherm.html

György Bognár received the M.Sc. degree in electronic engineering from the Faculty of Electrical Engineering and Computer Science, Budapest University of Technology and Economy (BME), Budapest, Hungary, in 2002, respectively, the M.Edu. degree in electrical engineering from the Faculty of Economic and Social Sciences, BME, in 2008, the M.A. degree in talent development from Károly Eszterházy University, Eger, Hungary, in 2010, and the Ph.D. degree in electronic engineering from the Faculty of Electrical Engineering and Computer Science, BME, in 2010.

He joined the Department of Electron Devices, BME, as an Assistant Lecturer, in 2004, where he has been an Associate Professor, since 2010. From 2014 to 2018, he was a Supervisor of the National Fundamental Research Project (Integrated Thermal Management in System-on-Package Devices) funded by the Hungarian Scientific Research Fund (OTKA). Since 2019, he has been the Head of the Micro- and Nanoelectronics Section, Scientific Association for Infocommunications, Budapest. Since 2020, he has been a Supervisor of the National Basic Research Project (Novel Solutions for Thermal Issues in Chiplet-Based System-on-Package Devices) funded by the National Research, Development, and Innovation Office (NKFIH). He worked on several national and international projects as a Researcher. Since 2007, he has been dealing with the realization technology, hydrodynamic and thermal modeling of microscale cooling structures, and microchannel heatsinks. His actual research fields are thermal management of system-in-package (SiP) devices, thermal simulation, modeling, and technology of heterogeneously integrated 2.5-D and 3-D packages with embedded thermal management solutions. His additional research topics included developing and thermal-modeling contactless temperature measuring systems. Other interests are the design of frequency divider circuits and new architecture injection-locked-oscillator-based resonant clock distribution networks.

Gábor Takács received the M.Sc. and Ph.D. degrees in electrical engineering from the Budapest University of Technology and Economics, Budapest, Hungary, in 2013 and 2018, respectively.

He joined the Department of Electron Devices, BME, in 2015, as an Assistant Lecturer, where he was promoted to a Senior Lecturer, in 2018. Since 2013, he has been dealing with the characterization of hydrodynamic and thermal performance of mini- and microscale cooling structures. Other interest is the analog CMOS integrated circuit design. He is currently a Professor in the Joint International Master at the Smart Systems Integrated Solution Erasmus Mundus Program. Since 2019, he has been a member of the Micro- and Nanoelectronics Section, Scientific Association for Infocommunications. As of 2020, he has been a Work Package Leader of the National Fundamental Research Project No. 135224 (Novel Solutions for Thermal Issues in Chiplet-Based System-on-Package Devices) funded by the National Research, Development, and Innovation Office (NKFIH). He has been working as a Technical Expert in the European Union's Horizon 2020 Funded Project DigiFed No. 872088.

Péter G. Szabó was born in Budapest, Hungary, in 1983. He received the M.Sc. and Ph.D. degrees in electrical engineering from the Budapest University of Technology and Economics (BME), Budapest, in 2007 and 2012, respectively.

He has been working with the Department of Electron Devices, BME, for 12 years, where he is currently an Associate Professor. He also worked as a Contract Researcher for a number of Small and Medium-Sized Enterprises (SMEs) and large enterprises. In 2019, he was admitted to the Micro- and Nanoelectronics Section, Scientific Association for Infocommunications. As of 2020, he has been a Work Package Leader of the National Fundamental Research Project No. 135224 (Novel Solutions for Thermal Issues in Chiplet-Based System-on-Package devices) funded by the National Research, Development, and Innovation Office (NKFIH). He has been working as a Technical Expert in the European Union's Horizon 2020 Funded Project DigiFed No. 872088. He is an expert in numerous design suites [MEMS and printed circuit board (PCB)] and numerical simulation tools [computational fluid dynamics (CFD) and Finite Element Analysis (FEA)]. His main research areas are related to electronics cooling both in micro and macroscale. His other research fields include modeling, design and characterization of MEMS devices and electrothermal systems.